European Review for Medical and Pharmacological Sciences 2013; 17: 2784-2790

Screening and further analyzing differentially expressed genes in acute idiopathic pulmonary fibrosis with DNA microarray

F. MIN, F. GAO[1], Z. LIU[2]

Department of Obstetrics, First People's Hospital Affiliated to Shanghai Jiaotong University, Shanghai, China
[1]Department of Respiratory Medicine, Shanghai JianGong Hospital, Shanghai, China
[2]Department of Respiratory Medicine, First People's Hospital Affiliated to Shanghai Jiaotong University, Shanghai, China
Fang Min and *Fengying Gao* are the co-first authors to this article

Abstract. – BACKGROUND: Acute idiopathic pulmonary fibrosis (IPF) is a serious and progressive form of lung disease, and millions of people suffer from this disease in the world. To provide clues for getting a better understanding of the mechanism of this disease, we identified and further analyzed the differential expressed genes in IPF.

METHOD: In this study, we downloaded the gene expression microarray (GSE10667) from Gene Expression Omnibus (GEO) database. The dataset contained a total of 23 samples, including 15 normal controls and 8 diseases samples (IPF). Then, we identified the differentially expressed genes between normal and disease samples with packages in R language. Consequently, the PPI network was also constructed for the products of these DEGs, and modules in the network were analyzed by Cytoscape's plug-in Mcode and Bingo. Furthermore, enrichment analysis was performed by DAVID to illustrate the altered pathways in IPF. The drug compounds for PLK1 were screened in DrugBank.

RESULTS: Atotal of 349 genes were identified as differentially expressed genes between normal and disease samples, and we constructed a protein-protein interaction network which included 200 pairs of proteins. Then three modules were identified in our network. Function of these modules were predicted to be related to protein kinase binding, extracellular matrix structural and structural constituent of cytoskeleton, respectively. Finally, we focused on module A including 18 DEGs.

CONCLUSIONS: PLK1 (Polo like kinge-1) in this module was predicted as a marker gene in IPF, which was related to cell cycle pathway. Several compounds were found which may be the potential drug for IPF.

Key Words:
Acute idiopathic pulmonary fibrosis, Differentially expressed genes, PPI network, Enrichment analysis.

Introduction

Acute respiratory distress syndrome (ARDS) is a serious lung syndrome which occurs in the produce of various non-cardiac diseases, such as infection, shock, injures and burns. It is characterized by diffuse pulmonary interstitial edema and alveolar edema, which were caused by damages of lung capillary endothelial cells and alveolar epithelial cells[1,2].

Acute idiopathic pulmonary fibrosis is also a main inducer accounting for ARDS. Idiopathic pulmonary fibrosis (IPF) is a chronic lung disease characterized by fibrosis of the supporting framework of the lungs[3]. Nowadays, IPF affects about 5 million people in the global. The development of this disease is typically progressive. Initially, the main symptoms is dry coughing and breathless. Then, significant disability of lung function will appear. At the end-stage, copulmonale often occurs, and in that case, many patients will die because of classic signs of right heart failure[4]. Moreover, according to asurvey, the median survival of patient with IPF is less than five years from the time of diagnosis[3].

Recently, many new molecular therapeutic targets and biology processes have been linked to IPF, such amesenchymal cell proliferation, excess extracellular matrix deposition, hypercoagulation status, epithelial injury, dysregulated repair and epithelial cell activation[5,6]. Matrix metalloproteinases (MMPs) are also reported to be associated with the IPF. The aberrant regulation of extracellular matrix (ECM) is mainly caused by the imbalance between the MMPs and their inhibitors, tissue inhibitor of metalloproteinases (TIMP), which can disrupt the degradation of pulmonary ECM, thereby, leading to excess ECM deposition in alveolar interstitium[7,8].

Corresponding Author: Zhenwei Liu, PhD, MD; e-mail: zhenweiliu69@163.com

Despite extensive investigation, the cause of IPF remains unclear. Medical therapy is ineffective in the treatment of IPF. However, researches of screening the DEGs in IPF using bioinformatics methods remain rare. Moreover, genome-wide approaches have reshaped the landscape of the problem. Meta-analysis of multi-study data has allowed the identification of DEGs that may have biomarker potential.

In this study, a total of 349 genes were identified as differentially expressed genes (DEGs) between normal and disease samples based on GSE10667, and then the PPI network for the products of these DEGs was constructed. Consequently, Cytoscape and its plugin Mcode were used to make the visualization of network and identify the modules in the network. Furthermore, GO (Gene Ontology) enrichment analysis was performed by DAVID to demonstrate the function of the genes in our modules. Finally, PLK1 (Polo like kinase-1) was chosen as a marker gene, and cell cycle pathway was found to be enriched in IPF. Moreover, we also screened the drug compounds in DrugBank for PLK1.

Materials and Methods

Microarray Data

The transcription profile of GSE10667[9,10] (a total of 23 samples, including 8 disease samples collected from patients with ARDS and 15 normal controls) was obtained from NCBI GEO database (http://www.ncbi.nlm.nih.gov/geo/) which is based on Agilent-014850 Whole Human Genome Microarray 4x44K G4112F.

Differentially Expressed Analysis

The format of dataset was transformed, and then missing value estimation was performed for these dataset for accurate estimation of missing data. Furthermore, these inputted missing data were normalized. Finally, the limma package[11] in R language with multiple testing correction (BH: Benjamini Hochberg method) was applied to identify genes which were significantly differentially expressed between disease samples and normal control[12]. We defined p-value < 0.05, FDR (False Discovery Rate) < -0.05 and |logFC| > 1 to be statistically significant.

Construction of PPI Network by String

The protein-protein interaction (PPI) network for the products of DEGs was constructed by String[13]. String is a database containing known and predicted relationship of protein interactions. The interactions include direct (physical) and indirect (functional) associations. Here we used it to predict interaction between products of these DEGs.

Analysis of Modules in Interaction Network

Gene products in same module often have same or similar function, they work together to carry out their biological function. Therefore, Cytoscape and its plugin Mcode were used to make the visualization of network and identify the modules in the network [parameters: Degree cutoff ≥ 2 (degrees of each nodes in module were larger than 2 at least), K-core ≥ 2 (subgraphs of each node in module were more than 2 at least)]. Then Bingo was used to determine which Gene Ontology (GO) terms were significantly overrepresented in the genes in our module (adj p < 0.01).

GO Enrichment Analysis

The online biological tool, DAVID (The Database for Annotation, Visualization and Integrated Discovery)[14] was used to perform GO term enrichment analysis for each group of genes in our PPI network.

Drug Discovery

Using PLK1 as the target genes, we screened the drug compounds in DrugBank[15]. DrugBank is a database, which has been widely used to facilitate in silico drug target discovery, drug design, drug docking or screening, drug metabolism prediction, drug interaction prediction and general pharmaceutical education[16].

Results

Differential Expression Analysis

In our criterion (p value < 0.05, FDR < 0.05, |logFC| > 1), a total of 349 DEGs were identified by limma package in R language. Among them, 87 genes were down-regulated and other 262 genes were up-regulated.

Construction of PPI Network by String

The products of these 349 DEGs were analyzed by the online tools STRING to predict the interaction relationship between them. As a result, a PPI network was constructed which in-

cludes 290 pairs of proteins (Figure 1). Among them, interaction relationships of 9 pairs were experimental verified and the others were predicted in STRING according to the sequencing, structure and characteristic of the proteins.

Analysis of Modules in Interaction Network

We identified the modules in our PPI network. Three modules were identified, which included 18, 5 and 4 genes, respectively (Figure 2). Then the function of each module was annotated based on GO (Table I). Interestingly, the genes in these three modules were mainly associated with protein kinase binding, extracellular matrix structural constituent and structural constituent of cytoskeleton, respectively.

Function and Pathway Annotation

To demonstrate the altered biology pathway in our modules, pathway enrichment analysis was performed by using DAVID. We defined p-value < 0.05 to be statistically significant. For genes in module B and C, no significantly enrichment was found. Otherwise genes in module A were identified to be enriched in cell cycle and p53 signaling pathway (Table II). Among them, the cell cycle pathway is mostly significant, and PLK1 was predicted to participate in the pathway.

Drug Discovery

Since PLK1 was predicted to be an important gene in this study, we also screened the drug compounds in DrugBank for it. Several compounds have been found, including nocodazole and some vitamins. These compounds are proved to be act on PLK1 by previous studies, and the related papers were listed in Table III.

Discussion

IPF is a serious lung disease, and until today, there is no effective way to treat it. However researching the underlying mechanism of IPF through bioinformatics may hint a new thought to the therapy of IPF.

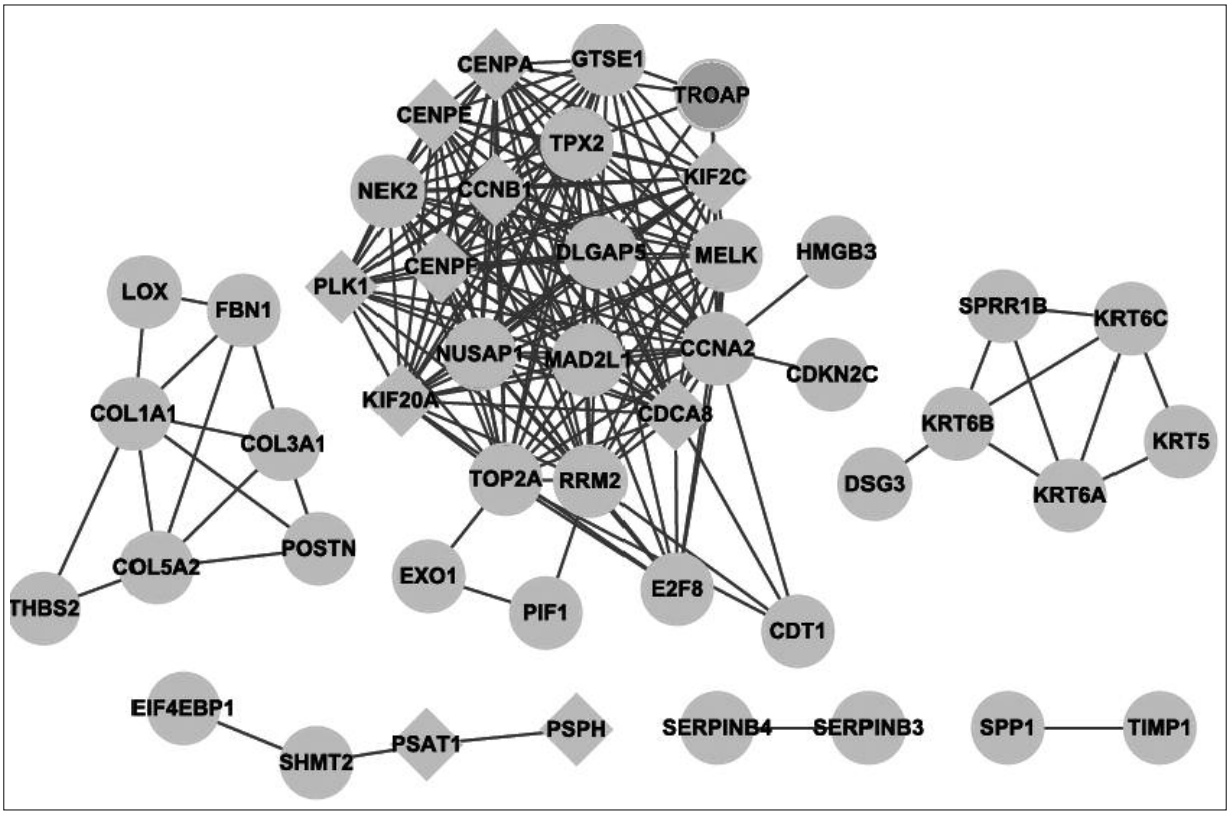

Figure 1. The PPI network for products of DEGs Diamond represents confirmed interactionswhich were proved by references in String database; Round represents unconfirmed interactions which were predicted by String according to the sequence, structure and characteristic of the proteins.

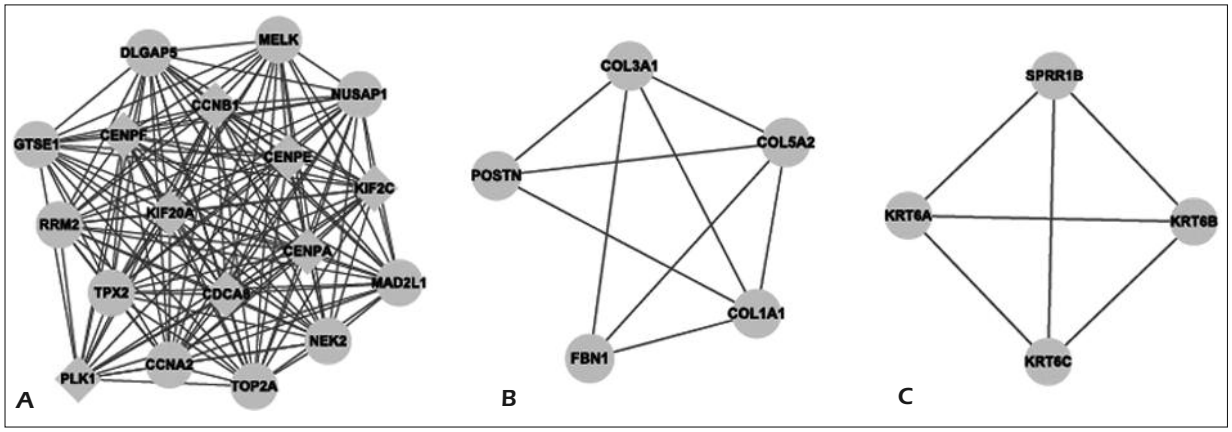

Figure 2. Three modules in PPI network.

In this study, about 300 DEGs were identified in disease samples compared to normal control. Further the PPI network for these DEGs was constructed and modules in the network were identified. Three modules were identified (Figure 2), and among them module A were focused. The function of genes in cluster A was associated with protein kinases. We also performed enrichment analysis to demonstrate the altered pathways in IPF. Finally, PLK1 was predicted to be significantly enriched in cell cycle pathway.

Polo like kinase 1 (PLK1) is a serine/threonine protein kinase, in human which is 66kDA and consists of 603 amino acids[17]. PLK1 contains two conserved polo-box domain at C-terminus, and one kinase domain at N-terminus. Polo-box

Table I. The enriched GO categories of Module A, B and C (FDR < 0.05, Module A, only list the top 10 GO categories).

Module	GO-ID	Description	FDR
Module A	19901	Protein kinase binding	1.11E-04
Module A	19900	Kinase binding	1.27E-04
Module A	19899	Enzyme binding	1.54E-03
Module A	3777	Microtubule motor activity	1.84E-03
Module A	5524	ATP binding	1.89E-03
Module A	32559	Adenyl ribonucleotide binding	1.89E-03
Module A	5515	Protein binding	1.92E-03
Module A	30554	Adenyl nucleotide binding	2.02E-03
Module A	1883	Purine nucleoside binding	2.02E-03
Module A	1882	Nucleoside binding	2.02E-03
Module B	5201	Extracellular matrix structural constituent	7.34E-08
Module B	48407	Platelet-derived growth factor binding	4.84E-05
Module B	5198	Structural molecule activity	7.96E-05
Module B	46332	SMAD binding	4.73E-04
Module B	19838	Growth factor binding	1.93E-03
Module C	5200	Structural constituent of cytoskeleton	2.63E-06
Module C	5198	Structural molecule activity	7.04E-06
Module C	30674	Protein binding, bridging	4.83E-02

Table II. The enriched pathways in Module A (p-value < 0.05).

KEGG-ID	Description	*p* value	Genes
hsa04110	Cell cycle	1.40E-04	PLK1
hsa04115	p53 signaling pathway	0.001717002	CCNB1, RRM2, GTSE1

Table III. The compounds acting on PLK1 in DrugBank.

Compound	PubMed IDs for Articles
Threonine	8127874, 18095313, 9370299, 18793607
Serine	8127874, 18095313, 9370299, 18793607
Scytonemin	15853646, 12388673
Pyridoxine	14568555, 16207494
Pyridoxal	14568555, 15381280, 16207494
Nocodazole	15522227, 18625707
Phosphothreonine	12595692
Pyridoxal	9354586, 14568555, 15381280, 16207494
Phosphoserine	12595692, 15640845
Paclitaxel	16740723, 12167714, 19806032, 17369857

domain plays an important role in both subcellular localization and regulation of kinase activity[18]. The GFP-PLK1 recombinant protein has been found to localize to centromere region during interphase, and PLK1 also plays an important role in early mitosis[19].

Moreover, PLK1 plays a key role in cell cycle pathway. It is recognized as an early trigger for G2/M transition. The cell cycle is series of events that take place in a cell leading to its division and duplication. In the course of evolution, cells develop and establish a serial of mechanism to ensure that the alternation of each phase of cell cycle is in order. Many regulators have been found to be related to the cell cycle. Cell cycle proteins, cyclin-dependent kinase, and cyclin-dependent kinase inhibitor are the three main categories of these regulators[20-22] PLK1 belongs to cyclin-dependent kinase.

In cell cycle pathway, PLK1 is phosphorylated and activated by CDK1, further the phosphorylated PLK1 can activate Cdc25C, a phosphatase that can dephosphorylate and activate the cyclinB/cdc2 complex[23]. These enzymes are controlled by DNA structure checkpoints, which de-

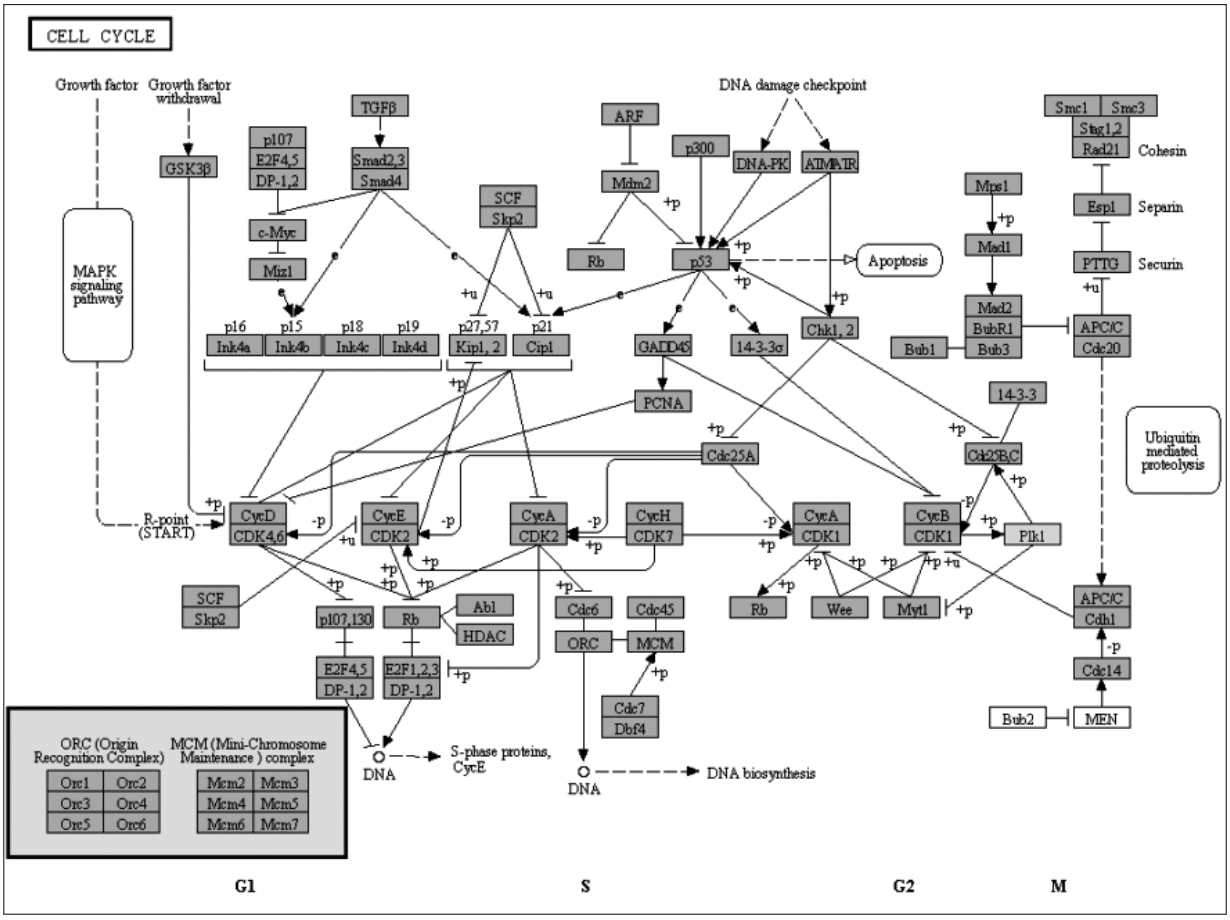

Figure 3. The pathway enrichment analysis of module A Highlight box represents PLK1, which was significantly related to cell cycle pathway.

lay the onset of mitosis in the presence of unreplicated or damaged DNA. The abnormal regulation of kinase in cell cycle pathway often leads to development of many diseases, such as cancer and IPF[24,25]. More and more researchers begin to focus on the relationship between the cell cycle and IPF. For example, some researchers report that cell cycle genes encoding p53 are associated with IPF disease development and progression[26].Tumor protein 53 (p53) is a key regulator of apoptosis. It is up-regulated upon DNA damage and prevents damaged cells from becoming malignant by inducing growth arrest and cell death[27,28]. Increased levels of p53 in the lungs of IPF patients are consistent with increased apoptosis[29]. Moreover, P53-induced growth arrest is mediated by increased CDK proteins[24].

It is also been reported overexpression of FoxO3a in IPF fibroblasts increases the CDK inhibitor, p27 levels and suppresses the ability of IPF fibroblasts to proliferate on polymerized collagen[30].

Our findings consist with the previous studies, and these findings support the idea that cell cycle control plays a role in the pathology of IPF. Here, we firstly reported that PLK1 was related to IPF. Furthermore, according to our results, together with previous studies we are likely to suppose that PLK1 may be associated with IPF through CDK1.

We also screened the drug compounds in DrugBank for PLK1. Several compounds have been found to act on PLK1. Such as Nocodazole, it is an anti-neoplastic drug which is often used to interfere with microtubule polymerization. Here, we first reported it should be a potential drug for IPF. However, more experimental verifications are still needed to prove this hypothesis.

Conclusions

Molecular biology, genetics and other experimental techniques are continually updated and mature. Progress in these fields can provide us new methods and new thoughts in treatment of IPF. In this study, a total of 349 DEGs were identified, and among them PLK1 was identified as a marker gene in IPF and predicted to be involved in cell cycle pathway. What's more, we also suggest that Nocodazole may become a drug for curing the IPF. Although, more experiments are needed to verify these results, our findings still increase cause for optimism in the treatment of IPF.

Conflict of Interest

The Authors declare that they have no conflict of interests.

References

1) Bernard GR, Artigas A, Brigham KL, Carlet J, Falke K, Hudson L, Lamy M, Legall JR, Morris A, Spragg R. The American-European Consensus Conference on ARDS. Definitions, mechanisms, relevant outcomes, and clinical trial coordination. Am J Respir Crit Care Med 1994; 149: 818-824.
2) Liang K, Gan X, Deng Z. Traumatic forequarter amputation associated acute lung injury (ALI): report of one case. Eur Rev Med Pharmacol Sci 2012; 16: 974.
3) Gross TJ, Hunninghake GW. Idiopathic pulmonary fibrosis. N Engl J Med 2001; 345: 517-525.
4) Meltzer EB, Noble PW. Idiopathic pulmonary fibrosis. Orphanet J Rare Dis 2008; 3: 1-15.
5) Willis BC, Liebler JM, Luby-Phelps K, Nicholson AG, Crandall ED, Du Bois RM, Borok Z. Induction of epithelial-mesenchymal transition in alveolar epithelial cells by transforming growth factor-β1: potential role in idiopathic pulmonary fibrosis. Am J Pathol 2005; 166: 1321.
6) Zhang K, Rekhter MD, Gordon D, Phan SH. Myofibroblasts and their role in lung collagen gene expression during pulmonary fibrosis. A combined immunohistochemical and in situ hybridization study. Am J Pathol 1994; 145: 114.
7) Wynn T. Cellular and molecular mechanisms of fibrosis. J Pathol 2007; 214: 199-210.
8) Hu P, Wang J, Zhao XQ, Hu B, Lu L, Qin YH. Overexpressed C-type natriuretic peptide serves as an early compensatory response to counteract extracellular matrix remodeling in unilateral ureteral obstruction rats. Molecular Biol Rep 2012; 1-13.
9) Konishi K, Gibson KF, Lindell KO, Richards TJ, Zhang Y, Dhir R, Bisceglia M, Gilbert S, Yousem SA, Song JW. Gene expression profiles of acute exacerbations of idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 2009; 180: 167-175.
10) Rosas IO, Richards TJ, Konishi K, Zhang Y, Gibson K, Lokshin AE, Lindell KO, Cisneros J, MacDonald SD, Pardo A. MMP1 and MMP7 as potential peripheral blood biomarkers in idiopathic pulmonary fibrosis. PLoS Med 2008; 5: e93.
11) Smyth G. Limma: linear models for microarray data. In: Bioinformatics and computational biology solutions using R and Bioconductor. R Gentleman, V Carey, S Dudoit, R Irizarry, W Huber (eds), Springer, New York, 2005; 397-420.
12) Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc Series B (Methodological) 1995; 289-300.

13) Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen LJ, von Mering C. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res 2011; 39: D561-D568.

14) Da Wei Huang BTS, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2008; 4: 44-57.

15) Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res 2006; 34: D668-D672.

16) Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, Gautam B, Hassanali M. DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res 2008; 36: D901-D906.

17) Holtrich U, Wolf G, Bräuninger A, Karn T, Böhme B, Rübsamen-Waigmann H, Strebhardt K. Induction and down-regulation of PLK, a human serine/threonine kinase expressed in proliferating cells and tumors. Proc Natl Acad Sci 1994; 91: 1736-1740.

18) Cheng KY, Lowe ED, Sinclair J, Nigg EA, Johnson LN. The crystal structure of the human polo-like kinase-1 polo box domain and its phospho-peptide complex. EMBO J 2003; 22: 5757-5768.

19) Lee KS, Oh DY, Kang YH, Park JE. Self-regulated mechanism of Plk1 localization to kinetochores: lessons from the Plk1-PBIP1 interaction. Cell Div 2008; 3: 4.

20) Hunter T. Cyclins and cancer II: cyclin D and CDK inhibitors come of age. Cell 1994; 79: 573-582.

21) Evan GI, Vousden KH. Proliferation, cell cycle and apoptosis in cancer. Nature 2001; 411: 342-348.

22) Kamb A, Gruis NA, Weaver-Feldhaus J, Liu Q, Harshman K, Tavtigian SV, Stockert E, Day 3rd R, Johnson BE, Skolnick MH. A cell cycle regulator potentially involved in genesis of many tumor types. Science 1994; 264: 436-440.

23) Nigg EA. Mitotic kinases as regulators of cell division and its checkpoints. Nat Rev Mol Cell Biol 2001; 2: 21-32.

24) Korthagen NM, van Moorsel CH, Barlo NP, Kazemier KM, Ruven HJ, Grutters JC. Association between variations in cell cycle genes and idiopathic pulmonary fibrosis. PloS One 2012; 7: e30442.

25) Li CY, Huang WF, Wang QL, Wang F, Cai E, Hu B, Du JC, Wang J, Chen R, Cai XJ. Crocetin induces cytotoxicity in colon cancer cells via p53-independent mechanisms. Asian Pacific J Cancer Prev 2012; 13: 3757-3761.

26) Korthagen NM, van Moorsel CHM, Barlo NP, Kazemier KM, Ruven HJT, Grutters JC. Association between variations in cell cycle genes and idiopathic pulmonary fibrosis. PloS one 2012; 7: e30442.

27) Canman CE, Kastan MB. Role of p53 in apoptosis. Adv Pharmacol 1997; 41: 429-460.

28) Al-Khodairy FM, Kunhi M, Siddiqui YM, Arif JM, Al-Ahdal MN, Hannan MA. Defective repair of UV-induced DNA damage in cultured primary skin fibroblasts from Saudi thyroid cancer patients. Asian Pacific J Cancer Prev 2004; 5: 139-143.

29) Plataki M, Koutsopoulos AV, Darivianaki K, Delides G, Siafakas NM, Bouros D. Expression of apoptotic and antiapoptotic markers in epithelial cells in idiopathic pulmonary fibrosis. Chest 2005; 127: 266-274.

30) Nho RS, Hergert P, Kahm J, Jessurun J, Henke C. Pathological alteration of FoxO3a activity promotes idiopathic pulmonary fibrosis fibroblast proliferation on type i collagen matrix. Am J Pathol 2011; 179: 2420-2430.