Bioinformatic analysis of next-generation sequencing data to identify dysregulated genes in fibroblasts of idiopathic pulmonary fibrosis

CHAU-CHYUN SHEU[1-3], WEI-AN CHANG[1,2], MING-JU TSAI[1-3], SSU-HUI LIAO[1], INN-WEN CHONG[2,3] and PO-LIN KUO[1]

[1]Graduate Institute of Clinical Medicine, College of Medicine, Kaohsiung Medical University; [2]Division of Pulmonary and Critical Care Medicine, Kaohsiung Medical University Hospital; [3]Department of Internal Medicine, School of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung 807, Taiwan, R.O.C.

Received October 7, 2018; Accepted January 29, 2019

DOI: 10.3892/ijmm.2019.4086

Abstract. Idiopathic pulmonary fibrosis (IPF) is a lethal fibrotic lung disease with an increasing global burden. It is hypothesized that fibroblasts have a number of functions that may affect the development and progression of IPF. However, the present understanding of cellular and molecular mechanisms associated with fibroblasts in IPF remains limited. The present study aimed to identify the dysregulated genes in IPF fibroblasts, elucidate their functions and explore potential microRNA (miRNA)-mRNA interactions. mRNA and miRNA expression profiles were obtained from IPF fibroblasts and normal lung fibroblasts using a next-generation sequencing platform, and bioinformatic analyses were performed in a step-wise manner. A total of 42 dysregulated genes (>2 fold-change of expression) were identified, of which 5 were verified in the Gene Expression Omnibus (GEO) database analysis, including the upregulation of neurotrimin (*NTM*), paired box 8 (*PAX8*) and mesoderm development LRP chaperone, and the downregulation of ITPR interacting domain containing 2 and Inka box actin regulator 2 (*INKA2*). Previous data indicated that *PAX8* and *INKA2* serve roles in cell growth, proliferation and survival. Gene Ontology analysis indicated that the most significant function of these 42 dysregulated genes was associated with the composition and function of the extracellular matrix (ECM). A total of 60 dysregulated miRNAs were also identified, and 1,908 targets were predicted by the miRmap database. The integrated analysis of mRNA and miRNA expression data, combined with GEO verification, finally identified *Homo sapiens* (hsa)-miR-1254-*INKA2* and hsa-miR-766-3p-*INKA2* as the potential miRNA-mRNA interactions in IPF fibroblasts. In summary, the results of the present study suggest that dysregulation of *PAX8*, hsa-miR-1254-*INKA2* and hsa-miR-766-3p-*INKA2* may promote the proliferation and survival of IPF fibroblasts. In the functional analysis of the dysregulated genes, a marked association between fibroblasts and the ECM was identified. These data improve the current understanding of fibroblasts as key cells in the pathogenesis of IPF. As a screening study using bioinformatics approaches, the results of the present study require additional validation.

Introduction

Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive, fibrotic lung disease. IPF predominantly affects elderly men. Patients with IPF usually present with exertional dyspnea, dry cough and inspiratory bibasilar Velcro-like crackles on lung auscultation. At present, high-resolution computed tomography is the most important diagnostic tool for IPF, which typically presents with usual interstitial pneumonia pattern (1,2). IPF is a lethal lung disease. The prognosis of IPF is poorer compared with a number of types of cancer (3), with an estimated median survival of 3-5 years if untreated (2). The incidence of IPF, hospitalization rates and mortality due to IPF have increased during previous years, suggesting an increasing global burden of this disease (4-6).

The pathogenesis of IPF remains largely unknown. Data from observational studies suggest that cigarette smoking, air pollution (7), environmental exposure (8), chronic viral infection (9) and esophageal reflux (10) predispose individuals to IPF (4). There is also increasing evidence for genetic predisposition to IPF: Recent genome-wide association studies identified that several genetic variants, primarily involved in epithelial cell-cell adhesion and integrity, the innate immune response, host defense and DNA repair, are associated with increased risk of IPF (11-13). These together indicate that gene-environmental interactions serve important roles in the

Correspondence to: Professor Po-Lin Kuo, Graduate Institute of Clinical Medicine, College of Medicine, Kaohsiung Medical University, 100 Shih-Chuan 1st Road, Kaohsiung 807, Taiwan, R.O.C.
E-mail: kuopolin@seed.net.tw

Key words: bioinformatics, fibroblasts, idiopathic pulmonary fibrosis, Inka box actin regulator 2, next-generation sequencing, paired box 8

pathogenesis of IPF. In genetically predisposed individuals, repetitive environmental exposures lead to an aberrant injury-remodeling process (14). Activated lung epithelial cells produce pro-fibrotic growth factors, chemokines and other mediators, resulting in abnormal wound healing characterized by mesenchymal transition of epithelial cells, activation and differentiation of fibroblasts and myofibroblasts. These fibroblasts and myofibroblasts secrete excessive amounts of extracellular matrix proteins including fibrillary collagens, vimentin and fibronectin (15), contributing to the destruction of lung architecture by progressive scarring.

It is hypothesized that fibroblasts have a number of functions that may affect the development and progression of IPF (16). Although previous studies have described the differences between IPF and normal lung fibroblasts (17-21), the knowledge concerning the cellular and molecular mechanisms associated with fibroblasts in IPF remains limited, and their specific roles requires additional investigation (16,22).

For an improved understanding of the roles fibroblasts serve in the pathogenesis of IPF, the present study used a next generation sequencing (NGS) platform and a step-wise bioinformatic approach to analyze the expression levels of mRNAs and microRNAs (miRNAs) and their interactions in IPF fibroblasts compared with normal lung fibroblasts. The aims of the present study were to identify the top differentially expressed genes and subsequently elucidate the function of these dysregulated genes, and to explore potential miRNA-mRNA interactions in IPF fibroblasts.

Materials and methods

Study design. The study design is summarized in Fig. 1. Firstly, cell cultures of the IPF fibroblasts and healthy lung fibroblasts were generated. Then, total RNA of the fibroblasts were extracted and sent for RNA and small RNA sequencing using the NGS platform. The differentially expressed genes (>2 fold-change) were analyzed with Ingenuity® Pathway Analysis (IPA) and the Database for Annotation, Visualization and Integrated Discovery (DAVID) for pathway analysis and functional interpretation. These dysregulated genes were additionally verified in representative Gene Expression Omnibus (GEO) datasets. The differentially expressed miRNAs (>2 fold-change) were analyzed with miRmap (mirmap.ezlab.org) (23) for target prediction. Then, Venn diagrams were used to determine genes with potential miRNA-mRNA interactions. These potential miRNA-mRNA interactions were confirmed using a second miRNA prediction database, TargetScan (www.targetscan.org) (24). Finally, a literature search for the functions of these dysregulated genes was performed, and a hypothesis was generated.

Culture of primary cells. Diseased human lung fibroblasts (DHLF)-IPF (cat. no. CC-7231) and normal human lung fibroblasts (NHLF; cat. no. CC-2512) were purchased from Lonza Walkersville Inc. (Walkersville, MD, USA). The cells were grown in FGW™-2 Fibroblast Growth Medium-2 Bulletkit™ (Lonza Walkersville Inc.; cat. no. CC-3132), containing 0.5 ml human fibroblast growth factor-basic, 0.5 ml insulin, 10 ml fetal bovine serum and 0.5 ml GA-1000. To avoid losing any original characteristics over multiple generations, the DHLF-IPF and NHLF cells used for the NGS analysis were harvested from the 1st generation of cells following cultivation from the primary cells.

NGS. The expression profiles of miRNAs and mRNAs were examined using NGS. Protocol was described in our previous studies (25-27). In brief, total RNA from the DHLF-IPF and NHLF cells were extracted with TRIzol® reagent (Thermo Fisher Scientific, Inc., Waltham, MA, USA) at Welgene Biotech Co., Ltd. (Taipei, Taiwan). Purified RNA was quantified at an optical density of 260 nm using an ND-1000 spectrophotometer (NanoDrop Technologies; Thermo Fisher Scientific, Inc.) and the RNA quality was examined using a Bioanalyzer 2100 (Agilent Technologies, Inc., Santa Clara, CA, USA) with RNA 6000 LabChip® kit (Agilent Technologies, Inc.). Next, the samples were prepared for library construction and sequencing using an Illumina preparation kit (Illumina, Inc., San Diego, CA, USA). For small RNA sequencing, the harvested cDNA constructs containing 18-40-nucleotide RNA fragments (140-155 nucleotides in length with all adapters) were selected. The libraries were then sequenced on an Illumina instrument (75 single-end cycles). Following trimming and removal of reads with low quality scores using Trimmomatics (28), the qualified reads were analyzed with miRDeep2 (29) and the UCSC Genome Browser (genome.ucsc.edu/) (30). The miRNAs with low levels (<1 normalized read per million) in either DHLF-IPF or NHLF cells were excluded.

For transcriptome sequencing, the library was constructed with a SureSelect Strand Specific RNA Library Preparation kit (Agilent Technologies, Inc.), and sequenced on a Solexa platform (150 paired-end cycles), using the TruSeq SBS kit (Illumina, Inc.). Similar with small RNA sequencing data, Trimmomatics (version 0.36) was implemented to trim or remove the reads with low quality score. Qualified reads were analyzed using HISAT2 (version 2.1.0) (31). The genes with low expression levels [<0.3 fragment per kilobase of transcript per million mapped reads (FPKM)] in either DHLF-IPF or NHLF cells were excluded. The 2-tailed P-values were calculated [2(Cufflinks version 2.2.1)] with non-grouped samples used 'blind mode', in which all samples were treated as replicates of a single global 'condition' and used to build one model for statistical analysis (32,33). Genes with $P<0.05$ [($-\log_{10}$) P-value >1.3] and >2 fold-changes were considered to be significantly dysregulated genes.

miRmap and TargetScan database analysis. miRmap is an open-source software library, which provides comprehensive miRNA-target prediction (23). By calculating the complementary ability of miRNA-mRNA interactions, the putative target genes of each miRNA may be identified. By combining thermodynamic, evolutionary, probabilistic and sequence-based features, the predictor also estimates mRNA-repression strength for ranking potential candidate targets. The prediction produces a list of putative target genes with an miRmap score, a predictive reference value. In the present study, an miRmap score ≥99 was used as the criterion for selecting putative miRNA targets.

TargetScan is an online database for predicting biological targets of miRNAs (24). It searches for the presence of conserved 8, 7 and 6 mer sites, which match the seed region of

each miRNA. Predictions are ranked based on the predicted targeting efficacy or the probability of conserved targeting.

IPA. IPA (Qiagen Inc., Valencia, CA, USA) (34) is a web-based software application for the analysis, integration and interpretation of data derived from 'omics experiments, including RNA sequencing, small RNA sequencing, microarrays, metabolomics and proteomics. IPA enables the rapid understanding and visualization of data, and provides a range of data in addition to pathway analysis, including the identification of key regulators and activity to explain expression patterns, prediction of downstream effects on biological and disease processes, and provision of targeted data on genes, proteins, chemicals and drugs. The dysregulated genes from the IPF fibroblasts were uploaded to IPA (version 2.3) to identify the top canonical pathways, upstream regulators, and molecular and cellular functions.

DAVID database analysis. DAVID is a powerful gene functional classification tool that integrates multiple functional annotation databases, including Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes databases (https://david.ncifcrf.gov/) (35,36). By calculating the similarity of global annotation profiles with an agglomeration algorithm method, a list of notable genes may be classified into clusters of associated 'Biological Process, 'Cellular Components', and 'Molecular Functions'. It also provides an Expression Analysis Systematic Explorer (EASE) score, a modified one-tailed Fisher's Exact P-value, for analysis. The reference score represents how specifically the user genes are involved in the category. An EASE score=0.1 was selected as the default, and =1 to extend clustering range in the analysis in the present study.

GEO database analysis. The GEO is a web database that collects submitted high-throughput gene-expression data of microarrays, chips or NGS (www.ncbi.nlm.nih.gov/geo) (37). Microarrays of accession numbers GSE24206, which included whole lung tissue samples from patients with early and advanced IPF (38) and GSE44723, which included samples of cultured lung fibroblasts from patients with IPF (39), were used in the present study. GSE44723 assessed gene expression in fibroblasts obtained from 4 normal controls, 4 patients with rapidly progressing IPF and 6 patients with slowly progressing IPF. GSE24206 assessed gene expression in whole lung biopsy or explant samples from 6 healthy controls, 8 patients with early IPF and 9 patients with advanced IPF. The databases used the platform of Affymetrix Human Genome U133 Plus 2.0 Array. Raw data extracted from the GEO were replotted using GraphPad Prism 7 software (GraphPad Software, Inc., La Jolla, CA, USA).

Statistical analysis. Statistical analyses were performed using SAS 9.4 software (SAS Institute Inc., Cary, NC, USA). To identify the differentially expressed genes, the Wilcoxon rank-sum test was used to compare gene expressions between two groups (patients with IPF vs. controls), and the Kruskal-Wallis test followed by Benjamini-Hochberg multiple-testing corrections was used to compare gene expressions across three groups. $P<0.05$ was considered to indicate a statistically significant difference. Dysregulated genes identified from the IPF and normal fibroblasts NGS analysis were considered verified if they met the following criteria: Exhibiting concordant up-/downregulation in the GSE24206 and GSE44723 datasets, and demonstrating significantly different expression between patients with IPF and controls or across three groups in at least one dataset.

Results

Gene expression profiles and miRNA changes in IPF fibroblasts. Samples from DHLF-IPF and NHLF cells were sequenced for mRNA and small RNA using an NGS platform. The volcano plot of dysregulated genes in fibroblasts of patients with IPF (DHLF-IPF cells) compared with normal controls (NHLF cells) is presented in Fig. 2A. The red nodes represent the significantly upregulated genes, and the green nodes represent significantly downregulated genes. The density plot of the deep sequencing results between the DHLF-IPF and NHLF cells is presented in Fig. 2B. The gene expression of DHLF-IPF cells demonstrated increased FPKM values and increased density compared with NHLF cells. Gene expression analysis revealed 42 differentially expressed genes with fold-change >2, including 23 upregulated and 19 downregulated genes (Table I). miRNA expression analysis revealed 60 miRNAs with fold-change >2, including 43 upregulated and 17 downregulated miRNAs (Table II). According to miRmap analysis, 1,908 targets were predicted from these dysregulated miRNAs.

IPA and DAVID analysis of dysregulated genes in IPF fibroblasts. The pathways and functional annotation of the 42 dysregulated genes in IPF fibroblasts were first analyzed using IPA. The results are summarized in Table III. The top canonical pathways were cAMP-mediated signaling, GP6 signaling pathway, cardiac-adrenergic signaling, hepatic fibrosis/hepatic stellate cell activation and calcium transport I. The top upstream regulators were fibrillin-1, bromodomain-containing protein 4, transcription factor GATA-6, TSIX and DNA-binding protein RFX5. Fig. 4 demonstrates the network of these top 5 upstream regulators and dysregulated genes. The top molecular and cellular functions included cell cycle, cell morphology, cellular development, lipid metabolism and molecular transport.

The functional annotation of the 42 dysregulated genes was subsequently analyzed using DAVID. Notably, results within the 'Biological Process', 'Cellular Component' and 'Molecular Function' categories all indicated that the most significant function of these dysregulated genes was associated with the composition and function of the extracellular matrix (ECM) (Fig. 3).

GEO database analysis of dysregulated genes in IPF fibroblasts. To additionally verify the 42 dysregulated genes in clinical samples from patients with IPF, the GEO database was searched and two representative microarray datasets were selected: GSE44723 (IPF cultured lung fibroblasts) and GSE24206 (early and advanced IPF whole lung tissues). Dysregulated genes were considered verified if they exhibited concordant up-/downregulation in the two GSE44723 and

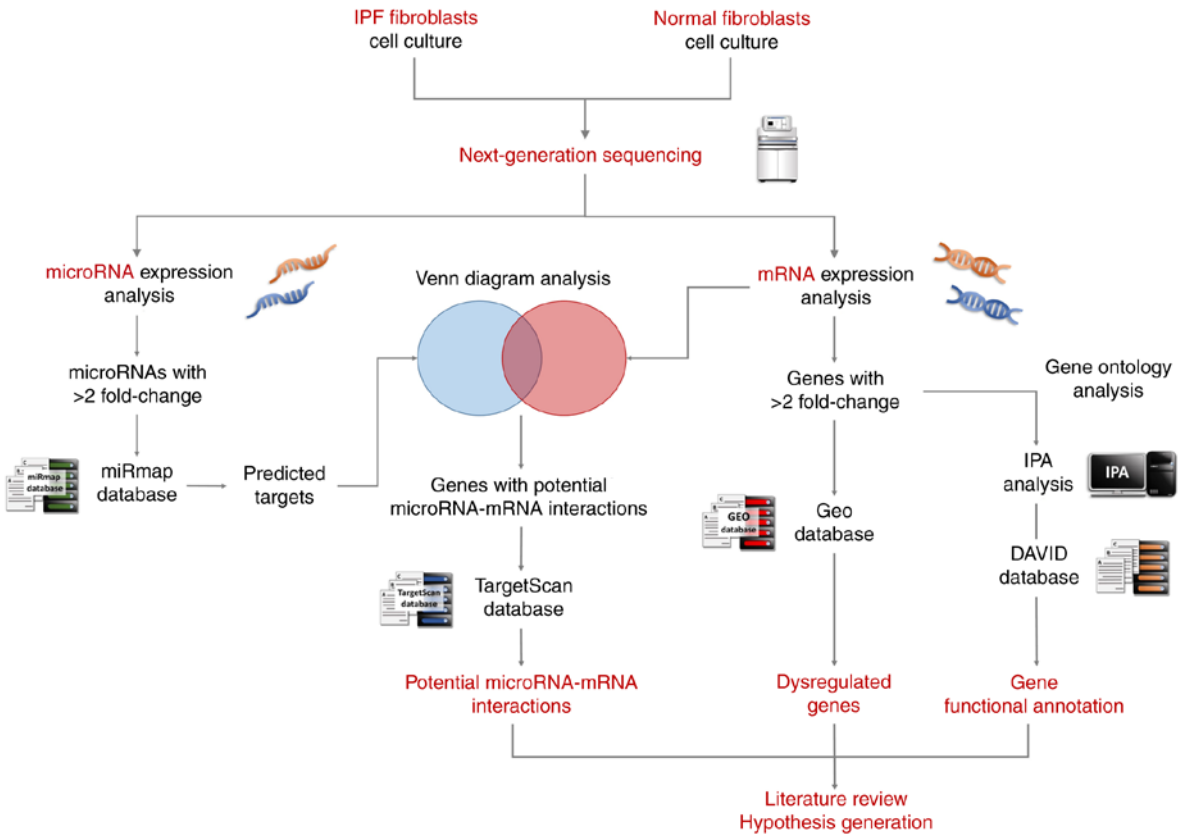

Figure 1. Schematic illustration of the study design. Following culture of the IPF fibroblasts and healthy lung fibroblasts, the total RNA were extracted and deep sequenced using a next-generation sequencing platform. The differentially expressed genes (>2 fold-change) were identified, and analyzed with IPA and DAVID for pathway analysis and functional interpretation. The GEO IPF databases were analyzed to confirm dysregulated genes identified in the present study. Conversely, the differentially expressed miRNAs (>2 fold-change) were analyzed with miRmap for target prediction. Then, genes with potential miRNA-mRNA interactions were determined by Venn diagram analysis. These miRNA-mRNA interactions were verified by a second miRNA prediction database, TargetScan. Finally, a literature search was performed and a hypothesis was generated. IPF, idiopathic pulmonary fibrosis; IPA, Ingenuity® Pathway Analysis; DAVID, Database for Annotation, Visualization, and Integrated Discovery; GEO, Gene Expression Omnibus.

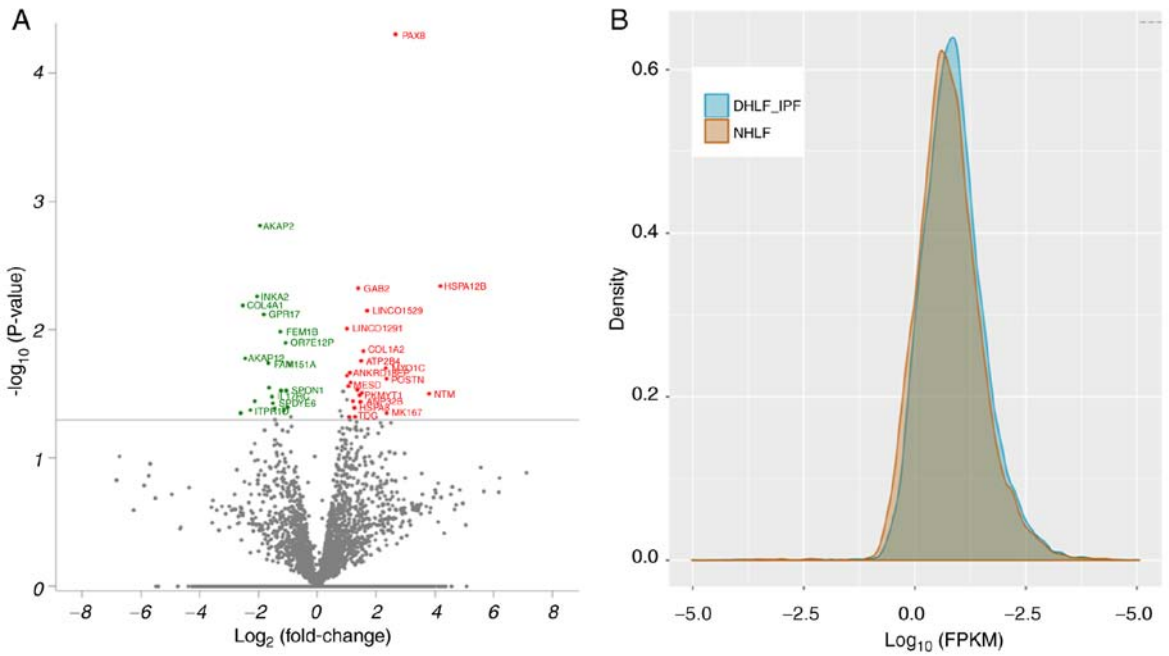

Figure 2. Differential gene expression patterns between IPF and normal fibroblasts. (A) The volcano plot of -$\log_{10}$ (P-value) vs. $\log_2$ (fold-change) demonstrated differentially expressed genes in DHLF-IPF vs. NHLF. Genes with -$\log_{10}$ (P-value) >1.3 and >2 fold-changes are plotted in red (upregulated genes) or green (downregulated genes). (B) The frequency distribution of FPKM between DHLF-IPF cells and NHLF cells was compared and presented in the density plot. IPF, idiopathic pulmonary fibrosis; FPKM, fragments per kilobase of transcript per million mapped reads; DHLF, diseased human lung fibroblasts; NHLF, normal human lung fibroblasts.

Table I. Dysregulated gene in IPF fibroblasts compared with normal lung fibroblasts.

A, Upregulated

Gene	Full gene name	DHLF-IPF FPKM	NHLF FPKM	Fold-change	P-value
HSPA12B	Heat shock protein family A (Hsp70) member 12B	59.62	3.28	18.18	0.005
NTM	Neurotrimin	117.08	8.40	13.94	0.032
PAX8	Paired box 8	86.90	13.69	6.35	0.000
MKI67	Marker of proliferation Ki-67	16.75	3.25	5.15	0.045
POSTN	Periostin	42.91	8.35	5.14	0.024
MYO1C	Myosin IC	21.87	4.34	5.04	0.020
LINC01529	Long intergenic non-protein coding RNA 1529	41.54	12.83	3.24	0.007
COL1A2	Collagen type I alpha 2 chain	69.47	23.42	2.97	0.015
CCNL1	Cyclin L1	38.61	13.41	2.88	0.031
ATP2B4	ATPase plasma membrane Ca^{2+} transporting 4	29.61	10.64	2.78	0.017
ANP32B	Acidic nuclear phosphoprotein 32 family member B	52.75	19.24	2.74	0.037
PKMYT1	Protein kinase, membrane associated tyrosine/threonine 1	571.51	212.09	2.69	0.032
GAB2	GRB2-associated binding protein 2	23.34	8.87	2.63	0.005
GPM6B	Glycoprotein M6B	39.57	15.38	2.57	0.029
TDG	Thymine DNA glycosylase	32.79	13.38	2.45	0.048
HSPA8	Heat shock protein family A (Hsp70) member 8	94.93	39.54	2.40	0.041
CD248	CD248 molecule	25.01	10.79	2.32	0.036
LINC00847	Long intergenic non-protein coding RNA 847	91.64	41.90	2.19	0.026
ANKRD18EP	Ankyrin repeat domain 18E, pseudogene	98.99	45.65	2.17	0.022
CNN3	Calponin 3	70.11	33.29	2.11	0.047
MESD	Mesoderm development LRP chaperone	93.66	44.85	2.09	0.028
IQGAP1	IQ motif containing GTPase activating protein 1	45.40	22.51	2.02	0.023
LINC01291	Long intergenic non-protein coding RNA 1291	161.02	80.15	2.01	0.010

B, Downregulated

TYW1	tRNA-YW synthesizing protein 1 homolog	24.22	146.89	-6.06	0.044
COL4A1	Collagen type IV alpha 1 chain	4.11	23.87	-5.81	0.006
AKAP12	A-kinase anchoring protein 12	4.24	23.09	-5.45	0.017
ITPRID2	ITPR interacting domain containing 2	4.96	23.91	-4.82	0.042
PRKAR1B	Protein kinase cAMP-dependent type I regulatory subunit beta	30.41	131.62	-4.33	0.036
INKA2	Inka box actin regulator 2	11.38	47.14	-4.14	0.006
AKAP2	A-kinase anchoring protein 2	12.30	47.75	-3.88	0.002
GPR17	G protein-coupled receptor 17	14.88	52.29	-3.51	0.008
FAM151A	Family with sequence similarity 151 member A	11.47	36.15	-3.15	0.018
UXT-AS1	UXT antisense RNA 1	48.81	151.15	-3.10	0.028
IL17RC	Interleukin 17 receptor C	8.18	23.55	-2.88	0.033
SPDYE6	Speedy/RINGO cell cycle regulator family member E6	10.45	30.07	-2.88	0.037

Table I. Continued.

B, Downregulated

Gene	Full gene name	DHLF-IPF FPKM	NHLF FPKM	Fold-change	P-value
MED31	Mediator complex subunit 31	25.55	69.73	-2.73	0.041
FEM1B	Fem-1 homolog B	14.77	35.13	-2.38	0.010
NCL	Nucleolin	19.51	45.91	-2.35	0.030
PTX3	Pentraxin 3	20.35	44.55	-2.19	0.042
OR7E12P	Olfactory receptor family 7 subfamily E member 12 pseudogene	126.50	268.42	-2.12	0.013
SPON1	Spondin 1	16.93	34.91	-2.06	0.030
COL4A2	Collagen type IV alpha 2 chain	12.65	25.46	-2.01	0.040

IPF, idiopathic pulmonary fibrosis; FPKM, fragments per kilobase of transcript per million mapped reads; DHLF, diseased human lung fibroblast; NHLF, normal human lung fibroblast.

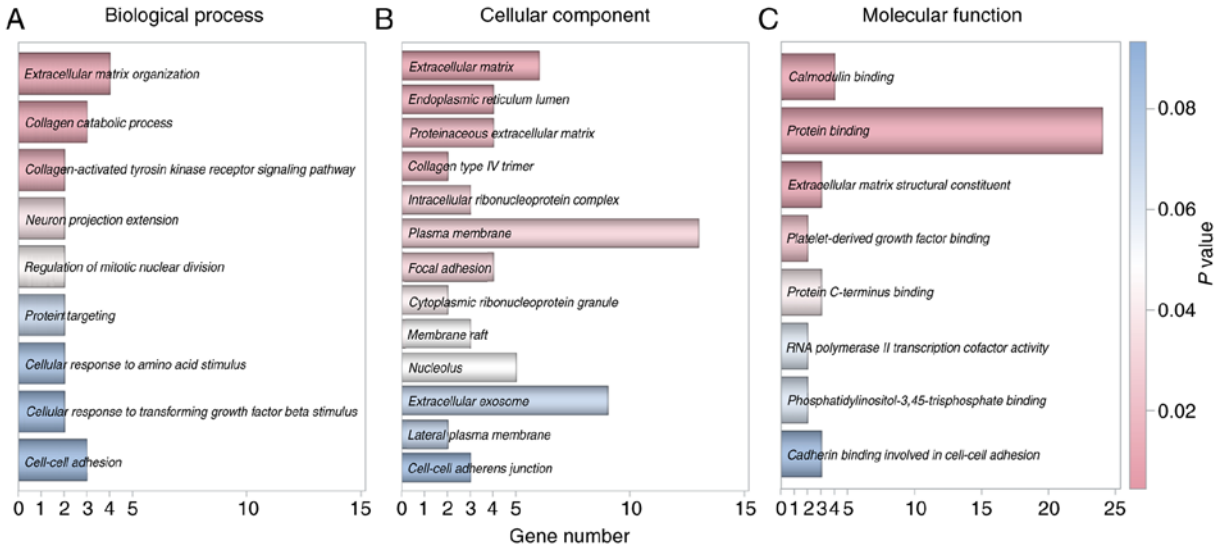

Figure 3. Gene Ontology analysis of dysregulated genes identified in idiopathic pulmonary fibrosis fibroblasts. Functional annotation of the 42 dysregulated genes was determined using Database for Annotation, Visualization, and Integrated Discovery. Results within the (A) Biological Process, (B) Cellular Component and (C) Molecular Function categories all indicated that the most significant function of these genes was associated with the composition and function of extracellular matrix.

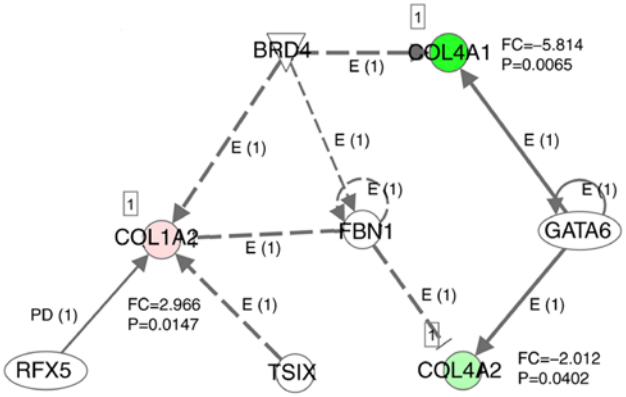

Figure 4. Network of top 5 upstream regulators and dysregulated genes in idiopathic pulmonary fibrosis fibroblasts. The top 5 upstream regulators, including FBN1, BRD4, GATA6, TSIX and RFX5, were generated using Ingenuity® Pathway Analysis based on the significance of overlap between the dysregulated genes and the genes that were regulated by the upstream regulators. The dysregulated genes targeted by these top upstream regulators were *COL1A2*, *COL4A1* and *COL4A2*, all belonging to the fibrillar collagen family. Numbers in parentheses are the number of reports supporting interactions. E, expression; FC, expression fold-change; PD, protein-DNA binding; CLO1A2, collagen alpha-2(I) chain; COL4A1, collagen alpha-1(IV) chain; COL4A2, collagen alpha-2(IV) chain; FBN1, fibrillin-1; BRD4, bromodomain-containing protein 4; GATA6, transcription factor GATA-6; RFX5, DNA-binding protein RFX5.

Table II. Dysregulated miRNAs in IPF fibroblasts vs. normal lung fibroblasts.

miRNA	DHLF-IPF Seq (norm)	NHLF Seq (norm)	Fold-change	Up/Down
hsa-miR-10b-5p	3,357.09	317.7	10.57	Up
hsa-miR-412-5p	50.6	5.86	8.63	Up
hsa-miR-329-3p	6.77	1.69	4.01	Up
hsa-miR-4661-5p	4.51	1.17	3.85	Up
hsa-miR-4521	17.7	4.82	3.67	Up
hsa-miR-3130-3p	3.8	1.17	3.25	Up
hsa-miR-369-3p	26.73	8.34	3.21	Up
hsa-miR-615-3p	11.17	3.52	3.17	Up
hsa-miR-766-3p	5.58	1.82	3.07	Up
hsa-miR-548am-5p	5.11	1.69	3.02	Up
hsa-miR-5000-3p	5.46	1.82	3.00	Up
hsa-miR-619-5p	3.44	1.17	2.94	Up
hsa-miR-486-5p	424.28	145.3	2.92	Up
hsa-miR-493-5p	369.05	132.01	2.80	Up
hsa-miR-3613-5p	10.45	3.78	2.76	Up
hsa-miR-625-3p	5.7	2.08	2.74	Up
hsa-miR-33b-5p	4.75	1.82	2.61	Up
hsa-miR-382-3p	14.49	5.6	2.59	Up
hsa-miR-379-3p	33.38	13.55	2.46	Up
hsa-miR-199b-5p	471.8	192.86	2.45	Up
hsa-miR-127-3p	292.67	120.67	2.43	Up
hsa-miR-1254	2.79	1.17	2.38	Up
hsa-miR-543	31.95	13.42	2.38	Up
hsa-miR-549a	15.92	6.91	2.30	Up
hsa-miR-193a-3p	5.7	2.48	2.30	Up
hsa-miR-665	24.11	10.56	2.28	Up
hsa-miR-376c-3p	20.55	9.12	2.25	Up
hsa-miR-487a-3p	32.9	14.99	2.19	Up
hsa-miR-582-3p	4.87	2.22	2.19	Up
hsa-miR-134-5p	174.73	80.14	2.18	Up
hsa-miR-1185-5p	3.68	1.69	2.18	Up
hsa-miR-493-3p	159.05	73.89	2.15	Up
hsa-miR-136-3p	424.16	198.72	2.13	Up
hsa-miR-136-5p	12.35	5.86	2.11	Up
hsa-miR-1185-1-3p	49.06	23.33	2.10	Up
hsa-miR-654-3p	1,401.49	669.93	2.09	Up
hsa-miR-185-3p	6.18	3	2.06	Up
hsa-miR-548ao-3p	2.14	1.04	2.06	Up
hsa-miR-668-3p	3.21	1.56	2.06	Up
hsa-miR-410-3p	865.2	424.42	2.04	Up
hsa-miR-409-3p	3,074.39	1,512.26	2.03	Up
hsa-miR-494-3p	6.3	3.13	2.01	Up
hsa-miR-1197	7.84	3.91	2.01	Up
hsa-miR-182-5p	47.63	95.52	-2.01	Down
hsa-miR-422a	1.66	3.39	-2.04	Down
hsa-miR-3200-3p	3.8	7.95	-2.09	Down
hsa-miR-335-5p	49.53	103.99	-2.10	Down
hsa-miR-138-1-3p	87.18	189.86	-2.18	Down
hsa-miR-365a-5p	1.43	3.13	-2.19	Down
hsa-miR-5699-5p	1.07	2.35	-2.20	Down
hsa-miR-664a-3p	2.38	5.86	-2.46	Down
hsa-miR-1538	2.26	5.6	-2.48	Down

Table II. Continued.

miRNA	DHLF-IPF Seq (norm)	NHLF Seq (norm)	Fold-change	Up/Down
hsa-miR-4662a-5p	2.97	7.56	-2.55	Down
hsa-miR-335-3p	341.14	920.39	-2.70	Down
hsa-miR-29c-5p	1.19	3.39	-2.85	Down
hsa-miR-9-5p	1.54	4.69	-3.05	Down
hsa-miR-4461	2.61	7.95	-3.05	Down
hsa-miR-3662	1.43	4.82	-3.37	Down
hsa-miR-4767	1.31	4.82	-3.68	Down
hsa-miR-204-5p	45.37	170.97	-3.77	Down

IPF, idiopathic pulmonary fibrosis; DHLF, diseased human lung fibroblast; NHLF, normal human lung fibroblast; Up, upregulation; Down, downregulation; hsa, *Homo sapiens*; miR, microRNA.

Table III. Summary of Ingenuity® Pathway Analysis of 42 dysregulated genes in idiopathic pulmonary fibrosis fibroblasts.

Categories	P-value	Overlap (%)	No. of molecules
Top canonical pathways			
cAMP-mediated signaling	0.0007	4/225 (1.8)	-
GP6 signaling pathway	0.0018	3/131 (2.3)	-
Cardiac-adrenergic signaling	0.0021	3/140 (2.1)	-
Hepatic fibrosis/hepatic stellate cell activation	0.0045	3/183 (1.6)	-
Calcium transport I	0.0181	1/10 (10.0)	-
Top upstream regulators			
FBN1	0.0002	-	-
BRD4	0.0021	-	-
GATA6	0.0030	-	-
TSIX	0.0037	-	-
RFX5	0.0037	-	-
Top molecular and cellular function			
Cell cycle	0.0013-0.0181	-	6
Cell morphology	0.0018-0.0395	-	4
Cellular development	0.0018-0.0106	-	4
Lipid metabolism	0.0018-0.0199	-	2
Molecular transport	0.0018-0.0055	-	2

IPA, Ingenuity® Pathway Analysis; IPF, idiopathic pulmonary fibrosis; FBN1, fibrillin-1; BRD4, bromodomain-containing protein 4; GATA6, transcription factor GATA-6; RFX5, DNA-binding protein RFX5.

GSE24206 datasets, with significantly different expression between patients with IPF and controls in at least one dataset. The results are summarized in Table IV. Based on the GEO analysis, neurotrimin (*NTM*), paired box 8 (*PAX8*) and mesoderm development LRP chaperone (*MESD*) were verified as the upregulated genes, and ITPR interacting domain containing 2 (*ITPRID2*) and Inka box actin regulator 2 (*INKA2*) as the downregulated genes in IPF. The detailed expression levels of these 5 genes in each group were plotted and compared in Fig. 5.

Identification of dysregulated genes with potential miRNA-mRNA interactions in IPF fibroblasts. Venn diagram analysis of the 1,908 target genes of 60 dysregulated miRNAs, predicted from miRmap, and the 42 dysregulated genes revealed 5 dysregulated genes with potential miRNA-mRNA interactions in IPF fibroblasts (Fig. 6). A total of 2 miRNA-mRNA interactions, *Homo sapiens* (hsa)-miR-185-3p-heat shock protein family A member 12B (*HSPA12B*) and hsa-miR185-3p-GRB associated binding protein 2 (*GAB2*), were excluded due to a lack of biological plausibility (the miRNA and mRNA were upregulated). The remaining 3 genes with miRNA-mRNA interactions were subsequently verified in a second miRNA predicting database, TargetScan. The results identified hsa-miR-185-3p-TRNA-YW synthesizing protein 1 homolog (*TYW1*), hsa-miR-3662-glycoprotein M6B (*GPM6B*),

Table IV. Gene Expression Omnibus verification of 42 dysregulated genes in IPF fibroblasts.

A, Upregulated genes

		GSE44723		GSE24206	
Genes	Fold-change	Up-/Downregulation	P-value	Up-/Downregulation	P-value
HSPA12B	18.15	N/A	-	Up	0.835
NTM	13.94	Up	0.257	Up	0.010
PAX8	6.35	Up	0.341	Up	0.032
MKI67	5.15	Down	0.582	Up	0.017
POSTN	5.14	Down	0.688	Up	0.008
MYO1C	5.04	Down	0.810	Up	0.259
LINC01529	3.24	N/A	-	N/A	-
COL1A2	2.97	Down	0.397	Up	<0.001
CCNL1	2.88	Up	0.737	Down	<0.001
ATP2B4	2.78	Up	0.147	Up	0.159
ANP32B	2.74	Up	0.443	Down	0.159
PKMYT1	2.69	Down	0.929	Up	0.832
GAB2	2.63	Down	0.873	Down	<0.001
GPM6B	2.57	Up	0.890	Down	0.013
TDG	2.45	N/A	-	Down	0.113
HSPA8	2.40	N/A	-	Up	0.081
CD248	2.32	Down	0.255	Up	0.115
LINC00847	2.19	Down	0.053	Up	0.091
ANKRD18EP	2.17	N/A	-	N/A	-
CNN3	2.11	N/A	-	Down	0.095
MESD	2.09	Up	0.958	Up	0.002
IQGAP1	2.02	Up	0.437	Down	0.717
LINC01291	2.01	N/A	-	N/A	-

B, Downregulated genes

TYW1	-6.06	Up	0.201	Down	0.716
COL4A1	-5.81	Down	0.224	Up	0.354
AKAP12	-5.45	Up	0.410	Down	0.004
ITPRID2	-4.82	Down	0.019	Down	0.031
PRKAR1B	-4.33	Down	0.194	Up	0.767
INKA2	-4.14	Down	0.213	Down	0.001
AKAP2	-3.88	N/A	-	Down	0.007
GPR17	-3.51	N/A	-	Down	0.460
FAM151A	-3.15	N/A	-	Down	0.081
UXT-AS1	-3.10	N/A	-	N/A	-
IL17RC	-2.88	Down	0.841	Down	0.630
SPDYE6	-2.88	N/A	-	Down	0.220
MED31	-2.73	Up	0.478	Up	0.831
FEM1B	-2.38	Up	0.873	Down	0.024
NCL	-2.35	N/A	-	Down	0.166
PTX3	-2.19	Up	0.918	Down	<.001
OR7E12P	-2.12	N/A	-	Down	0.937
SPON1	-2.06	Up	0.986	Down	0.036
COL4A2	-2.01	Down	0.214	Up	0.584

Fold-change indicates the results of RNA-seq analysis from the next-generation sequencing data. GEO, Gene Expression Omnibus; IPF, idiopathic pulmonary fibrosis; N/A, not available.

Table V. Dysregulated genes with potential miRNA-mRNA interaction in IPF fibroblasts.

Gene	Gene full name	mRNA fold-change	miRNA	miRNA fold-change	miRmap score	TargetScan Total context++ score
TYW1	tRNA-YW synthesizing protein 1 homolog	-6.06	hsa-miR-185-3p	2.06	99.12	-0.37
GPM6B	glycoprotein M6B	2.44	hsa-miR-3662	-3.37	99.81	-0.09
INKA2	Inka box actin regulator 2	-4.14	hsa-miR-1254	2.38	99.28	-0.32
			hsa-miR-766-3p	3.07	99.09	-0.67

IPF, idiopathic pulmonary fibrosis; miRNA, microRNA; hsa, *Homo sapiens*; TYW1, TRNA-YW synthesizing protein 1 homolog; GPM6B, glycoprotein M6B; INKA2, Inka box actin regulator 2.

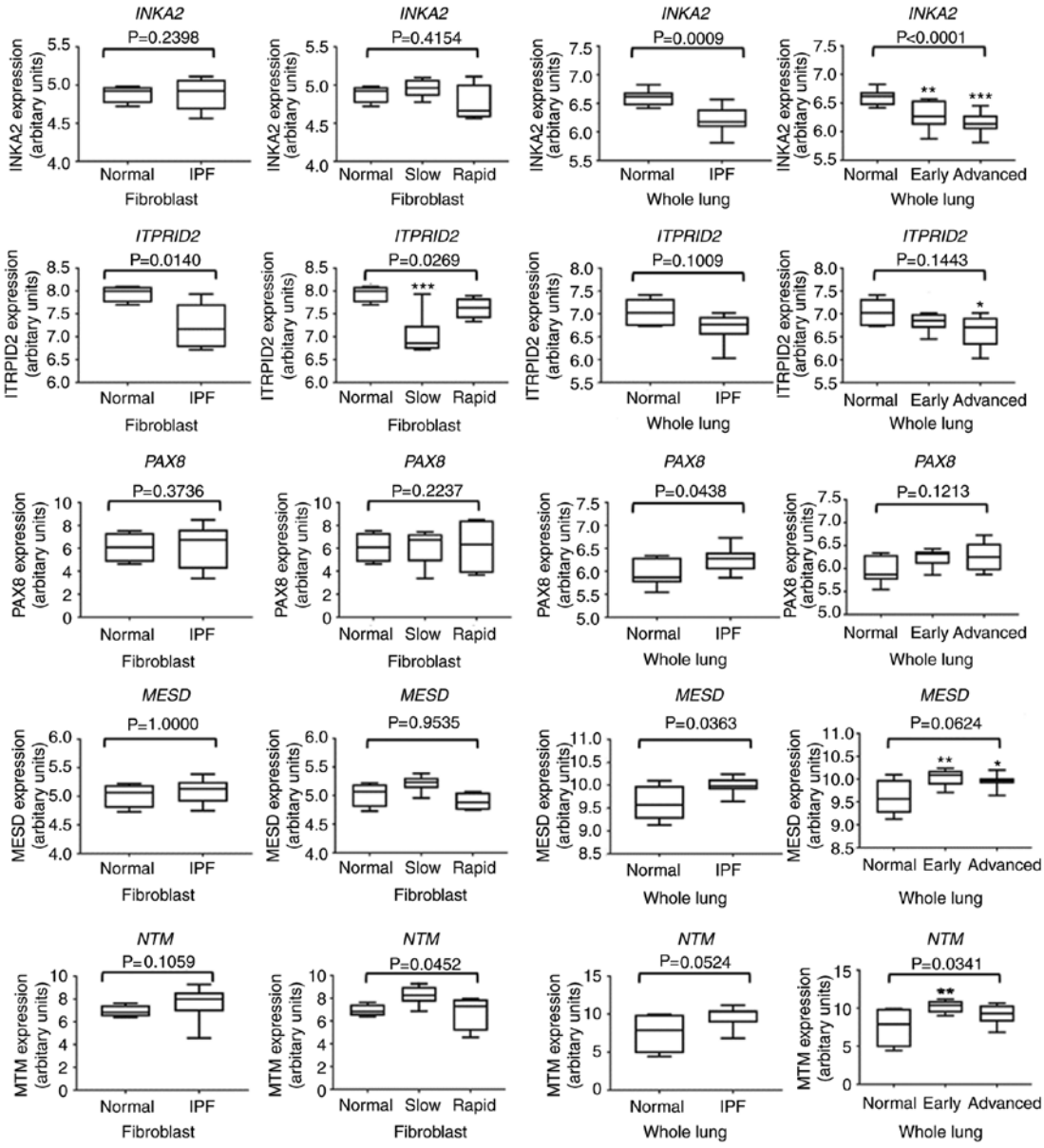

Figure 5. Gene Expression Omnibus analysis of 5 dysregulated genes in IPF fibroblasts. Representative microarray datasets GSE44723 (slowly and rapidly progressing IPF: cultured lung fibroblasts) and GSE24206 (early and advanced IPF: whole lung tissue) were used for verification of differential gene expressions identified from Next Generation Sequencing analysis. The results indicated that the expression of *INKA2* and *ITRPID2* were downregulated and *PAX8, MESD* and *NTM* were upregulated in IPF. Cultured lung fibroblasts and whole lung from healthy subjects were used as the normal controls. P-values were calculated using the Wilcoxon rank-sum test for comparisons of two groups, and the Kruskal-Wallis test for comparisons of three groups. Adjusted P-values were calculated using the Kruskal-Wallis test followed by Benjamini-Hochberg multiple-testing corrections. *Adjusted P<0.05, **adjusted P<0.01 and ***adjusted P<0.001. IPF, idiopathic pulmonary fibrosis; INKA2, Inka box actin regulator 2; ITPRID2, ITPR interacting domain containing 2; PAX8, paired box 8; MESD, mesoderm development LRP chaperone; NTM, neurotrimin.

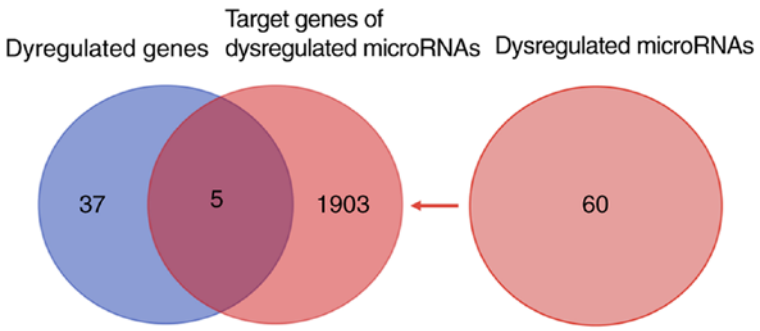

Figure 6. Venn diagram analysis of miRNA-mRNA interactions in idiopathic pulmonary fibrosis fibroblasts. RNA sequencing revealed 42 dysregulated genes (left). Small RNA sequencing revealed 60 dysregulated miRNAs, which predicted 1,908 target genes (right) based on miRmap database. Venn diagram analysis identified 5 dysregulated genes with potential miRNA-mRNA interaction. miRNA, microRNA.

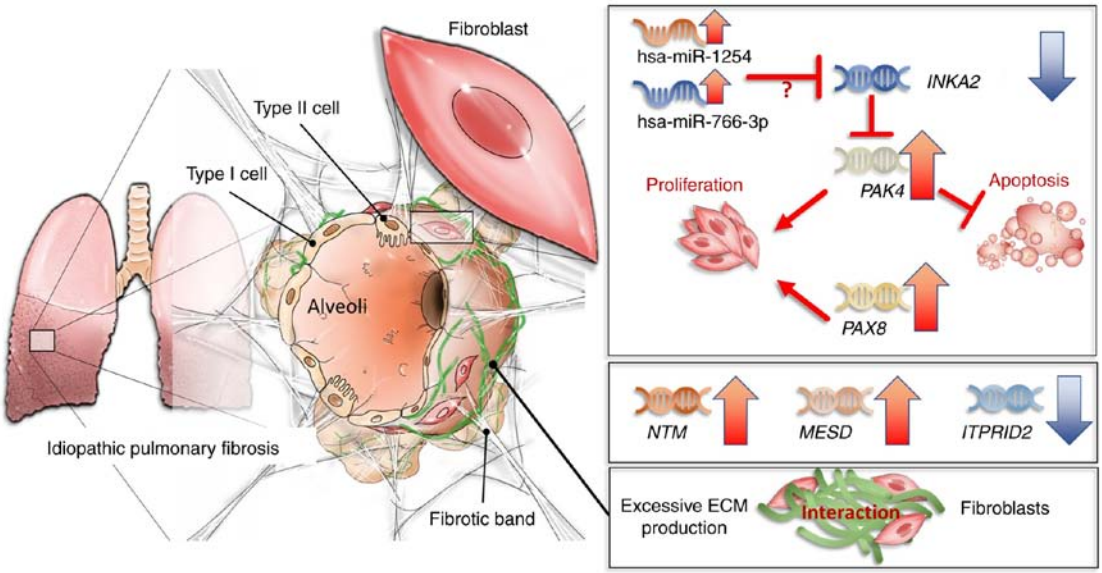

Figure 7. Proposed mechanisms demonstrating gene regulations of fibroblasts and their potential roles in IPF. Upregulation of hsa-miR-1254 and hsa-miR-766-3p may inhibit *INKA2* expression, leading to overexpression of *PAK4*, which may promote fibroblast proliferation and migration, and protect fibroblasts from apoptosis. Overexpression of *PAX8* may also have a role in fibroblast proliferation. The functions of *NTM, MESD* and *ITPRID2* remain unknown. Whether they contribute to the pro-proliferative and pro-fibrotic microenvironment requires additional study. Based on the results of the Gene Ontology analysis, the gene dysregulations in IPF fibroblasts may also have a role in the excessive production and deposition of disorganized ECM. The hsa-miR-1254-*INKA2* and hsa-miR-766-3p-*INKA2* interactions were identified based on bioinformatic analysis, and therefore requires additional experiments to confirm. IPF, idiopathic pulmonary fibrosis; hsa, *Homo sapiens*; INKA2, Inka box actin regulator 2; PAK4, serine/threonine-protein kinase PAK4; PAX8, paired box 8; MESD, mesoderm development LRP chaperone; NTM, neurotrimin; ITPRID2, ITPR interacting domain containing 2; ECM, extracellular matrix.

hsa-miR-1254-*INKA2* and hsa-miR-766-3p-*INKA2* as the potential miRNA-mRNA interactions in IPF fibroblasts (Table V).

Discussion

The pathogenic mechanisms of IPF have not yet been fully elucidated. The hallmarks of IPF are aberrant activation of lung epithelial cells, accumulation of fibroblasts and myofibroblasts, and excessive production of ECM (4,14-16). Previous data indicate that fibroblasts, as one of the key cells, serve a crucial role in the development and progression of IPF (16). An improved understanding of the gene regulation in IPF fibroblasts may assist in the development of novel therapeutics targeting this key cell. In the present study, the whole mRNA and miRNA profiles of IPF fibroblasts were obtained using an NGS platform, and bioinformatics analyses were performed in a step-wise manner. A total of 42 dysregulated genes were identified, and then bioinformatics tools IPA and DAVID were used for pathway and functional analyses. A total of 5 of the differentially expressed genes were verified in the representative GEO datasets, including *NTM*, *PAX8* and *MESD* (upregulated), and *ITPRID2* and *INKA2* (downregulated). Integrated analysis of mRNA and miRNA expression data was also performed, and hsa-miR-185-3p-*TYW1*, hsa-miR-3662-*GPM6B*, hsa-miR-1254-*INKA2* and hsa-miR-766-3p-*INKA2* were identified as the potential miRNA-mRNA interactions in IPF fibroblasts. According to the GEO verification, hsa-miR-1254-*INKA2* and hsa-miR-766-3p-*INKA2* were considered as the most likely dysregulated miRNA-mRNA

interactions in IPF fibroblasts. However, these interactions were identified based on bioinformatic analysis. Therefore, they require additional experiments to confirm the results. Hsa-miR185-3p-*HSPA12B* and hsa-miR185-3p-*GAB2* were excluded from subsequent analysis, as the miRNAs and mRNAs were dysregulated in the same manner. There is a possibility of indirect modulation, in that the upregulated hsa-miR185-3p may control one or more other targets. Which may in turn upregulate the expression levels of *HSPA12B* or *GAB2*. Nevertheless, the upregulated levels of *HSPA12B* or *GAB2* were not validated in the GEO database analysis. Whether these 2 miRNA-mRNA interactions were excluded or not did not affect the final results.

In the GO analysis, it was identified that the most important function of the identified dysregulated genes was associated with the composition and function of the ECM. Replacement of the normal lung structure with an excessive deposition of disorganized collagen and ECM is the hallmark of IPF (40). Although previous evidence suggests that fibroblasts and myofibroblasts in the fibrotic foci are the key cells leading to excessive ECM production (41), the crosstalk between epithelial cells, fibroblasts, myofibroblasts and ECM remains largely uncharacterized. The results from the present study improve the understanding of the fibroblast-ECM interaction in the pathogenesis of IPF. The development of novel therapeutic strategies targeting the fibrotic ECM may provide an opportunity to halt fibrosis and restore organ function (42). A recent study confirmed the importance of the ECM in IPF pathogenesis and treatment: Kwapiszewska *et al* (43) compared transcriptomic profiles in lung homogenates and fibroblasts obtained from patients with IPF treated with or without pirfenidone. They identified that cell migration-inducing and hyaluronan-binding protein (CEMIP) was markedly downregulated by pirfenidone treatment. They also identified that circulating CEMIP levels were significantly increased in patients with IPF compared with the healthy controls, and that pirfenidone treatment was associated with a significant decrease in CEMIP levels. CEMIP has been previously associated with ECM production, inflammation and cell proliferation (44,45). They concluded that pirfenidone exhibited effects on multiple pathways in fibroblasts and other pulmonary cells, through the regulation of the ECM structure and inflammatory reactions.

The *INKA2* gene, also known as family with sequence similarity 212 member B or chromosome 1 open reading frame 183, encodes the protein serine/threonine-protein kinase PAK4 (PAK4)-inhibitor INKA2. The PAK4-inhibitor INKA proteins, with 2 isoforms in humans, are endogenous inhibitors of PAK4 (46). The PAK proteins are a family of serine/threonine cyclin-dependent kinase inhibitor 1-activating protein kinases and have been implicated in a range of biological activities (46). PAK4 is an effector molecule for the Rho GTPase cell division control protein 42 homolog. Previous experiments have indicated that active PAK4 protects cells from apoptosis (47,48). PAK4 is associated with tumorigenesis and progression in different types of cancer through promoting cell growth, proliferation, migration and metastasis (49-55). The downregulation of *INKA2* identified in the present study may potentially upregulate the expression of *PAK4*, which in turn promotes cell proliferation and inhibits cell apoptosis in IPF fibroblasts. The NGS data suggested that the FPKM of *PAK4* was 0.2922 in IPF fibroblasts, but undetectable in normal lung fibroblasts. Based on the quality control criterion to exclude genes with FPKM <0.3 in either IPF or normal fibroblasts, *PAK4* was not able to be analyzed in the present study. The upregulation of *PAK4* in IPF fibroblasts required additional confirmation. It was also identified that the miRNAs hsa-miR-1254 and hsa-miR-766-3p, 2 putative regulators of *INKA2*, were upregulated in IPF fibroblasts. Therefore, the results suggested that the miRNA-mRNA interaction in the *INKA2* gene may serve a role in cell growth, proliferation or apoptosis inhibition in IPF fibroblasts.

The *PAX8* gene encodes protein PAX8, which is a member of the paired box family of transcription factors. PAX8 is involved in the development and differentiation of thyroid follicular cells (56,57), and is associated with the growth of cancer cells (58,59). Overexpression of *PAX8* has been observed in a number of types of cancer. It also has a critical role in the survival and proliferation of epithelial cells (60). In the present study, upregulation of *PAX8* was identified in IPF fibroblasts. These data suggested that PAX8 may serve a role in promoting the growth, survival or proliferation of fibroblasts in IPF. The upregulation of *NTM* and *MESD* (also known as mesoderm development candidate 2), and downregulation of *ITPRID2* (also known as Ki-Ras-induced actin-interacting protein or sperm specific antigen 2) were also identified in IPF fibroblasts. Neurotrimin, a neural cell adhesion molecule, may promote neurite outgrowth and adhesion via a homophilic mechanism (61,62). In addition to brain tissue, human tissue-specific expression analysis indicated a high expression of *NTM* in human lung tissue (63). MESD, a specialized chaperone for low-density lipoprotein receptor-related protein (LRP)5 and LRP6, is a universal inhibitor of Wnt/LRP signaling on the cell surface (64,65). *ITPRID2* encodes a protein that is involved in the regulation of filamentous actin and extracellular signals (66), and may be associated with structural integrity and/or signal transductions in human cancer (67). The functions of *NTM*, *MESD* and *ITPRID2* remain largely unknown. Their associations with human lung diseases, particularly IPF, require additional investigation.

The identified gene dysregulation in fibroblasts and their proposed mechanisms in IPF are summarized in Fig. 7. In summary, the present study demonstrated that *NTM*, *PAX8* and *MESD* were upregulated and *ITPRID2* and *INKA2* were downregulated in IPF fibroblasts. Dysregulation of *PAX8* and *INKA2* may participate in the regulation of proliferation and survival in IPF fibroblasts. Functional analysis of the dysregulated genes suggested a marked association between fibroblasts and ECM. The integrated analysis of miRNA and mRNA expression profiles suggested that the upregulated miRNAs hsa-miR-1254 and hsa-miR-766-3p may downregulate *INKA2* and serve important roles in IPF. These data improve the current understanding of fibroblasts as key cells in the pathogenesis of IPF. As a screening study using bioinformatics approaches, without biological or cross-platform replicates, the results of the present study require additional validation.

Acknowledgements

The authors gratefully acknowledge the assistance from the Center for Research and Development of Kaohsiung Medical University (Kaohsiung, Taiwan), and the assistance with

statistical analysis from Mr. Tse-Kuang Kai at Kaohsiung Medical University Hospital.

Funding

The present study was supported in part by research grants from the Ministry of Science and Technology (grant nos. MOST 107-2320-B-037-011-MY3 and MOST 106-2314-B-037-016-MY2), the Kaohsiung Medical University Hospital Research Foundation (grant nos. KMUHS10701, KMUHS10712, KMUH106-6R15 and KMUH107-7M06), and the Kaohsiung Medical University (grant no. KMU-DK108003).

Availability of data and materials

The data used and analyzed in this study are available from the corresponding author on reasonable request.

Authors' contributions

CCS, IWC and PLK conceived the study. CCS, WAC and MJT analyzed and interpreted the data. CCS, WAC, MJT, IWC and PLK prepared the manuscript. SHL performed cell culture and laboratory experiments. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Raghu G, Remy-Jardin M, Myers JL, Richeldi L, Ryerson CJ, Lederer DJ, Behr J, Cottin V, Danoff SK, Morell F, *et al*: Diagnosis of idiopathic pulmonary fibrosis. An official ATS/ERS/JRS/ALAT clinical practice guideline. Am J Respir Crit Care Med 198: e44-e68, 2018.
2. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, Colby TV, Cordier JF, Flaherty KR, Lasky JA, *et al*: An official ATS/ERS/JRS/ALAT statement: Idiopathic pulmonary fibrosis: Evidence-based guidelines for diagnosis and management. Am J Respir Crit Care Med 183: 788-824, 2011.
3. Vancheri C, Failla M, Crimi N and Raghu G: Idiopathic pulmonary fibrosis: A disease with similarities and links to cancer biology. Eur Respir J 35: 496-504, 2010.
4. Lederer DJ and Martinez FJ: Idiopathic pulmonary fibrosis. N Engl J Med 378: 1811-1823, 2018.
5. Hutchinson J, Fogarty A, Hubbard R and McKeever T: Global incidence and mortality of idiopathic pulmonary fibrosis: A systematic review. Eur Respir J 46: 795-806, 2015.
6. Navaratnam V, Fogarty AW, Glendening R, McKeever T and Hubbard RB: The increasing secondary care burden of idiopathic pulmonary fibrosis: Hospital admission trends in England from 1998 to 2010. Chest 143: 1078-1084, 2013.
7. Sack C, Vedal S, Sheppard L, Raghu G, Barr RG, Podolanczuk A, Doney B, Hoffman EA, Gassett A, Hinckley-Stukovsky K, *et al*: Air pollution and subclinical interstitial lung disease: The Multi-Ethnic Study of Atherosclerosis (MESA) air-lung study. Eur Respir J 50: 1700559, 2017.
8. Baumgartner KB, Samet JM, Coultas DB, Stidley CA, Hunt WC, Colby TV and Waldron JA: Occupational and environmental risk factors for idiopathic pulmonary fibrosis: A multicenter case-control study. Collaborating centers. Am J Epidemiol 152: 307-315, 2000.
9. Tang YW, Johnson JE, Browning PJ, Cruz-Gervis RA, Davis A, Graham BS, Brigham KL, Oates JA Jr, Loyd JE and Stecenko AA: Herpesvirus DNA is consistently detected in lungs of patients with idiopathic pulmonary fibrosis. J Clin Microbiol 41: 2633-2640, 2003.
10. Tobin RW, Pope CE II, Pellegrini CA, Emond MJ, Sillery J and Raghu G: Increased prevalence of gastroesophageal reflux in patients with idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 158: 1804-1808, 1998.
11. Fingerlin TE, Murphy E, Zhang W, Peljto AL, Brown KK, Steele MP, Loyd JE, Cosgrove GP, Lynch D, Groshong S, *et al*: Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. Nat Genet 45: 613-620, 2013.
12. Noth I, Zhang Y, Ma SF, Flores C, Barber M, Huang Y, Broderick SM, Wade MS, Hysi P, Scuirba J, *et al*: Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: A genome-wide association study. Lancet Respir Med 1: 309-317, 2013.
13. Peljto AL, Zhang Y, Fingerlin TE, Ma SF, Garcia JG, Richards TJ, Silveira LJ, Lindell KO, Steele MP, Loyd JE, *et al*: Association between the MUC5B promoter polymorphism and survival in patients with idiopathic pulmonary fibrosis. JAMA 309: 2232-2239, 2013.
14. Mora AL, Rojas M, Pardo A and Selman M: Emerging therapies for idiopathic pulmonary fibrosis, a progressive age-related disease. Nat Rev Drug Discov 16: 755-772, 2017.
15. Gomperts BN and Strieter RM: Fibrocytes in lung disease. J Leukoc Biol 82: 449-456, 2007.
16. Pardo A and Selman M: Lung fibroblasts, aging, and idiopathic pulmonary fibrosis. Ann Am Thorac Soc 13 (Suppl 5): S417-S421, 2016.
17. Ramos C, Montaño M, García-Alvarez J, Ruiz V, Uhal BD, Selman M and Pardo A: Fibroblasts from idiopathic pulmonary fibrosis and normal lungs differ in growth rate, apoptosis, and tissue inhibitor of metalloproteinases expression. Am J Respir Cell Mol Biol 24: 591-598, 2001.
18. Pierce EM, Carpenter K, Jakubzick C, Kunkel SL, Evanoff H, Flaherty KR, Martinez FJ, Toews GB and Hogaboam CM: Idiopathic pulmonary fibrosis fibroblasts migrate and proliferate to CC chemokine ligand 21. Eur Respir J 29: 1082-1093, 2007.
19. Kurundkar AR, Kurundkar D, Rangarajan S, Locy ML, Zhou Y, Liu RM, Zmijewski J and Thannickal VJ: The matricellular protein CCN1 enhances TGF-β1/SMAD3-dependent profibrotic signaling in fibroblasts and contributes to fibrogenic responses to lung injury. FASEB J 30: 2135-2150, 2016.
20. Conte E, Gili E, Fruciano M, Korfei M, Fagone E, Iemmolo M, Lo Furno D, Giuffrida R, Crimi N, Guenther A and Vancheri C: PI3K p110γ overexpression in idiopathic pulmonary fibrosis lung tissue and fibroblast cells: In vitro effects of its inhibition. Lab Invest 93: 566-576, 2013.
21. Vuga LJ, Ben-Yehudah A, Kovkarova-Naumovski E, Oriss T, Gibson KF, Feghali-Bostwick C and Kaminski N: WNT5A is a regulator of fibroblast proliferation and resistance to apoptosis. Am J Respir Cell Mol Biol 41: 583-589, 2009.
22. Desai O, Winkler J, Minasyan M and Herzog EL: The role of immune and inflammatory cells in idiopathic pulmonary fibrosis. Front Med 5: 43, 2018.
23. Vejnar CE and Zdobnov EM: MiRmap: Comprehensive prediction of microRNA target repression strength. Nucleic Acids Res 40: 11673-11683, 2012.
24. Agarwal V, Bell GW, Nam JW and Bartel DP: Predicting effective microRNA target sites in mammalian mRNAs. Elife 4, 2015.
25. Sheu CC, Tsai MJ, Chen FW, Chang KF, Chang WA, Chong IW, Kuo PL and Hsu YL: Identification of novel genetic regulations associated with airway epithelial homeostasis using next-generation sequencing data and bioinformatics approaches. Oncotarget 8: 82674-82688, 2017.
26. Chang WA, Tsai MJ, Jian SF, Sheu CC and Kuo PL: Systematic analysis of transcriptomic profiles of COPD airway epithelium using next-generation sequencing and bioinformatics. Int J Chron Obstruct Pulmon Dis 13: 2387-2398, 2018.
27. Tsai MJ, Chang WA, Jian SF, Chang KF, Sheu CC and Kuo PL: Possible mechanisms mediating apoptosis of bronchial epithelial cells in chronic obstructive pulmonary disease - A next-generation sequencing approach. Pathol Res Pract 214: 1489-1496, 2018.

28. Bolger AM, Lohse M and Usadel B: Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114-2120, 2014.
29. Friedländer MR, Mackowiak SD, Li N, Chen W and Rajewsky N: miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res 40: 37-52, 2012.
30. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM and Haussler D: The human genome browser at UCSC. Genome Res 12: 996-1006, 2002.
31. Kim D, Langmead B and Salzberg SL: HISAT: A fast spliced aligner with low memory requirements. Nat Methods 12: 357-360, 2015.
32. Galipon J, Ishii R, Suzuki Y, Tomita M and Ui-Tei K: Differential binding of three major human ADAR isoforms to coding and long non-coding transcripts. Genes 8: E68, 2017.
33. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL and Pachter L: Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7: 562-578, 2012.
34. Krämer A, Green J, Pollard J Jr and Tugendreich S: Causal analysis approaches in ingenuity pathway analysis. Bioinformatics 30: 523-530, 2014.
35. Huang da W, Sherman BT and Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4: 44-57, 2009.
36. Huang da W, Sherman BT and Lempicki RA: Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37: 1-13, 2009.
37. Edgar R, Domrachev M and Lash AE: Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30: 207-210, 2002.
38. Meltzer EB, Barry WT, D'Amico TA, Davis RD, Lin SS, Onaitis MW, Morrison LD, Sporn TA, Steele MP and Noble PW: Bayesian probit regression model for the diagnosis of pulmonary fibrosis: Proof-of-principle. BMC Med Genomics 4: 70, 2011.
39. Peng R, Sridhar S, Tyagi G, Phillips JE, Garrido R, Harris P, Burns L, Renteria L, Woods J, Chen L, *et al*: Bleomycin induces molecular changes directly relevant to idiopathic pulmonary fibrosis: A model for 'active' disease. PLoS One 8: e59348, 2013.
40. Barratt SL, Creamer A, Hayton C and Chaudhuri N: Idiopathic pulmonary fibrosis (IPF): An overview. J Clin Med 7: E201, 2018.
41. Sgalla G, Iovene B, Calvello M, Ori M, Varone F and Richeldi L: Idiopathic pulmonary fibrosis: Pathogenesis and management. Respir Res 19: 32, 2018.
42. Walraven M and Hinz B: Therapeutic approaches to control tissue repair and fibrosis: Extracellular matrix as a game changer. Matrix Biol 71-72: 205-224, 2018.
43. Kwapiszewska G, Gungl A, Wilhelm J, Marsh LM, Thekkekara Puthenparampil H, Sinn K, Didiasova M, Klepetko W, Kosanovic D, Schermuly RT, *et al*: Transcriptome profiling reveals the complexity of pirfenidone effects in idiopathic pulmonary fibrosis. Eur Respir J 52: 1800564, 2018.
44. Soroosh A, Albeiroti S, West GA, Willard B, Fiocchi C and de la Motte CA: Crohn's disease fibroblasts overproduce the novel protein KIAA1199 to create proinflammatory hyaluronan fragments. Cell Mol Gastroenterol Hepatol 2: 358-368, e4, 2016.
45. Li L, Yan LH, Manoj S, Li Y and Lu L: Central role of CEMIP in tumorigenesis and its potential as therapeutic target. J Cancer 8: 2238-2246, 2017.
46. Baskaran Y, Ang KC, Anekal PV, Chan WL, Grimes JM, Manser E and Robinson RC: An *in cellulo*-derived structure of PAK4 in complex with its inhibitor Inka1. Nat Commun 6: 8681, 2015.
47. Gnesutta N, Qu J and Minden A: The serine/threonine kinase PAK4 prevents caspase activation and protects cells from apoptosis. J Biol Chem 276: 14414-14419, 2001.
48. Gnesutta N and Minden A: Death receptor-induced activation of initiator caspase 8 is antagonized by serine/threonine kinase PAK4. Mol Cell Biol 23: 7838-7848, 2003.
49. Callow MG, Clairvoyant F, Zhu S, Schryver B, Whyte DB, Bischoff JR, Jallal B and Smeal T: Requirement for PAK4 in the anchorage-independent growth of human cancer cell lines. J Biol Chem 277: 550-558, 2002.
50. Siu MK, Chan HY, Kong DS, Wong ES, Wong OG, Ngan HY, Tam KF, Zhang H, Li Z, Chan QK, *et al*: p21-activated kinase 4 regulates ovarian cancer cell proliferation, migration, and invasion and contributes to poor prognosis in patients. Proc Natl Acad Sci USA 107: 18622-18627, 2010.
51. Wong LE, Chen N, Karantza V and Minden A: The Pak4 protein kinase is required for oncogenic transformation of MDA-MB-231 breast cancer cells. Oncogenesis 2: e50, 2013.
52. Guo Q, Su N, Zhang J, Li X, Miao Z, Wang G, Cheng M, Xu H, Cao L and Li F: PAK4 kinase-mediated SCG10 phosphorylation involved in gastric cancer metastasis. Oncogene 33: 3277-3287, 2014.
53. Wang C, Li Y, Zhang H, Liu F, Cheng Z, Wang D, Wang G, Xu H, Zhao Y, Cao L, *et al*: Oncogenic PAK4 regulates Smad2/3 axis involving gastric tumorigenesis. Oncogene 33: 3473-3484, 2014.
54. Xia M, Wei J and Tong K: MiR-224 promotes proliferation and migration of gastric cancer cells through targeting PAK4. Pharmazie 71: 460-464, 2016.
55. Zhang X, Zhang X, Li Y, Shao Y, Xiao J, Zhu G and Li F: PAK4 regulates G6PD activity by p53 degradation involving colon cancer cell growth. Cell Death Dis 8: e2820, 2017.
56. Pasca di Magliano M, Di Lauro R and Zannini M: Pax8 has a key role in thyroid cell differentiation. Proc Natl Acad Sci USA 97: 13144-13149, 2000.
57. Mansouri A, Chowdhury K and Gruss P: Follicular cells of the thyroid gland require *Pax8* gene function. Nat Genet 19: 87-90, 1998.
58. Muratovska A, Zhou C, He S, Goodyer P and Eccles MR: *Paired-Box* genes are frequently expressed in cancer and often required for cancer cell survival. Oncogene 22: 7989-7997, 2003.
59. Ghannam-Shahbari D, Jacob E, Kakun RR, Wasserman T, Korsensky L, Sternfeld O, Kagan J, Bublik DR, Aviel-Ronen S, Levanon K, *et al*: PAX8 activates a p53-p21-dependent pro-proliferative effect in high grade serous ovarian carcinoma. Oncogene 37: 2213-2224, 2018.
60. Di Palma T, Filippone MG, Pierantoni GM, Fusco A, Soddu S and Zannini M: Pax8 has a critical role in epithelial cell survival and proliferation. Cell Death Dis 4: e729, 2013.
61. Struyk AF, Canoll PD, Wolfgang MJ, Rosen CL, D'Eustachio P and Salzer JL: Cloning of neurotrimin defines a new subfamily of differentially expressed neural cell adhesion molecules. J Neurosci 15: 2141-2156, 1995.
62. Gil OD, Zanazzi G, Struyk AF and Salzer JL: Neurotrimin mediates bifunctional effects on neurite outgrowth via homophilic and heterophilic interactions. J Neurosci 18: 9312-9325, 1998.
63. Fagerberg L, Hallström BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, Habuka M, Tahmasebpoor S, Danielsson A, Edlund K, *et al*: Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Mol Cell Proteomics 13: 397-406, 2014.
64. Lin C, Lu W, Zhai L, Bethea T, Berry K, Qu Z, Waud WR and Li Y: Mesd is a general inhibitor of different Wnt ligands in Wnt/LRP signaling and inhibits PC-3 tumor growth in vivo. FEBS Lett 585: 3120-3125, 2011.
65. Lu W, Liu CC, Thottassery JV, Bu G and Li Y: Mesd is a universal inhibitor of Wnt coreceptors LRP5 and LRP6 and blocks Wnt/beta-catenin signaling in cancer cells. Biochemistry 49: 4635-4643, 2010.
66. Inokuchi J, Komiya M, Baba I, Naito S, Sasazuki T and Shirasawa S: Deregulated expression of *KRAP*, a novel gene encoding actin-interacting protein, in human colon cancer cells. J Hum Genet 49: 46-52, 2004.
67. Fujimoto T, Koyanagi M, Baba I, Nakabayashi K, Kato N, Sasazuki T and Shirasawa S: Analysis of *KRAP* expression and localization, and genes regulated by *KRAP* in a human colon cancer cell line. J Hum Genet 52: 978-984, 2007.