

Integrated Bioinformatic Approach to Identify Potential Biomarkers against Idiopathic Pulmonary Fibrosis

Md. Asad Ullah^{1,3*}, Bishajit Sarkar^{1,3}, Yusha Araf^{2,3}, Md. Nazmul Islam Prottoy^{1,3}, Ananna Saha¹, Tanjila Jahan¹, Aisha Siddiqua Boby¹, Md. Shariful Islam¹

¹Department of Biotechnology and Genetic Engineering, Faculty of Biological Sciences, Jahangirnagar University, Dhaka, Bangladesh

²Department of Genetic Engineering and Biotechnology, Faculty of Life Sciences, Shahjalal University of Science and Technology, Sylhet, Bangladesh

³Swift Integrity Computational Lab, Dhaka, Bangladesh

Correspondence: *ullah1194@gmail.com

Abstract

Idiopathic Pulmonary Fibrosis (IPF) is a chronic and progressive lung disease that leads to gradual decline in lung function. The molecular mechanism and risk factors of this disease are still obscure. Poorly understood etiology of this disease is the major obstacle in the identification of potential biomarkers and drug targets. In this study, microarray gene expression data of normal and IPF patient has been utilized for the statistical analysis of differentially expressed genes (DEGs) with a view to identifying potential molecular signatures using network-based system. Then their functional enrichment analysis revealed their predominant involvement in transcription, protein acetylation, extracellular matrix organization, apoptic process, inflammatory response etc. Protein-Protein Interaction (PPI) network revealed (UBC, PTEN, SOS1, PTK2, FGFR1, YAP1, FOXO1, RACK1, BMP4 and CD44) as hub proteins in IPF. Subsequent regulatory network analysis suggested (E2F1, STAT3, PPARG, MEF2A, FOXC1, GATA3, YY1, GATA2, NFKB1, and FOXL1) as the best regulatory transcriptional signatures and (hsa-mir-155-5p, hsa-mir-16-5p, hsa-mir-17-5p, hsa-mir-19a-3p, hsa-mir-192-5p, hsa-mir-92a-3p, hsa-mir-26b-5p, hsa-mir-335-5p, hsa-mir-124-3p, and hsa-let-7b-5p) as the best post-transcriptional signatures. This study represents proteome and RNA signatures of IPF which might be useful to uphold the present efforts in the discovery of potential biomarkers and treatments of this disease.

Keywords: Biomarkers, Hub proteins, Idiopathic pulmonary fibrosis, Molecular signatures, Treatment

1. Introduction

Idiopathic Pulmonary Fibrosis (IPF) is a progressive and degenerative interstitial pneumonia that results in the abnormal scarring of the lungs of affected individual (Selman et al., 2001). Historically IPF was thought to be a chronic inflammatory process but there is now growing evidence that IPF response is mediated by abnormally activated alveolar epithelial cells (AECs) which secrete mediators that drive the growth of fibroblast and myofibroblast foci. These foci in turn secrete excessive extracellular matrix which then form abnormal scarring resulting in the alteration of lung architecture (King et al., 2001; Fernandez and Eickelberg, 2012). Previously IPF was accused to lead to the steady and predictable decline in lung function but now there are enough evidences that IPF also leads to the heart failure, pulmonary embolism and acute respiratory deterioration (Collard et al., 2007). The clinical epidemiology of IPF remains poorly understood since the first identification of this disease. However, environmental exposures, genetic factors and tobacco smoking are considered as the potential risk factors of IPF. There are now two approved drugs i.e., Pirfenidone and Nintedanib but unfortunately surgical therapy and lung transplantation are merely accessible to every people (Ley and Collard, 2013; Hughes et al., 2016). Older age with a median of 66 years is the major clinical feature of IPF and in addition, older age has also been shown to exhibit poorer prognosis (Ley et al., 2011). Report suggests that, IPF prevalence in USA varies between 43 and 67 cases per 100,000 population. Moreover, the occurrence and prevalence of IPF increases with age and the incidence in males is higher than females (Nalysnyk et al., 2012).

Microarray technology now allows the easy identification of variance in gene expression which in turn contributes in the identification of potential biomarkers of specific interest (Giltneane and Rimm, 2004; Kerr et al., 2000). Analysis of differential gene expression signatures is crucial in the identification of biomarkers and understanding of lung pathologies in IPF (DePianto et al., 2015; Li et al., 2014). However, although these studies provide fruitful findings but the prediction of actual mechanism of biological condition using the data of differentially expressed genes (DEGs) is often difficult and may come with erroneous interpretation sometime (Crow et al., 2019).

In this study, we have employed integrated network-based strategies to predict potential biomarkers which may contribute in the identification of new drug target, and early diagnosis of IPF. Thereafter we have also tried to discuss the specific roles of identified signatures in IPF (**Figure 1**).

2. Materials and Methods

2.1. Data Retrieval and Identification of Differentially Expressed Genes

We retrieved GSE24206 microarray data from NCBI-GEO (National Center for Biotechnology Information-Gene Expression Omnibus) database (Meltzer et al., 2011). The dataset comprises the expression profile of lung tissues from different parts of lungs of 5 normal donors and 11 IPF patients. After retrieval, the data was analyzed using GEAP (Gene Expression Analysis Platform) to differentiate the upregulated and downregulated genes (Nunes et al., 2018). Log₂ transformation was applied and differentially expressed genes (DEGs) were sorted with adjusted P value<0.01 filter since the lower value corresponds to more accurate prediction.

2.2. Functional Enrichment Analysis of DEGs

Both upregulated and downregulated gene sets were analyzed by DAVID (Database for Annotation, Visualization, and Integrated Discovery) (version 6.8) for gene over-representation to elucidate gene ontology (GO) terms and pathways involved with DEGs (Sherman and Lempicki, 2009). P values were adjusted using the Hochberg and Benjamini test and gene count >2 were set as the cut-off point during the analysis.

2.3. Construction of Protein-Protein Interaction Network and Identification of Hub Proteins

STRING database was utilized for the reconstruction of protein-protein interaction (PPI) network with NetworkAnalyst (Szklarczyk et al., 2017; Xia et al., 2015). Topological and expression analysis of the DEGs were performed using NetworkAnalyst. Hub proteins in the generic PPI network with top 10 most connected nodes were identified with cytoHubba plugin using bottleneck interaction matrix on Cytoscape (version 3.7.2) (Chin et al., 2014; Shannon et al., 2003). Top 2 modules in the network was also analyzed using MCODE plugin (Saito et al., 2012).

2.4. Identification of Regulatory Molecules

DEGs were searched against JASPAR with the help of NetworkAnalyst to construct transcription factor (TF)-DEGs interaction network (Sandelin et al., 2004). Micro RNA (miRNA)-DEGs interaction network was constructed searching the DEGs against TarBase (Sethupathy et al., 2006). Top 10 interacting TFs and miRNAs were selected and analyzed.

3. Result

3.1. Transcriptome Signatures

Publicly available GEO data from NCBI was utilized in this study in order to identify differentially expressed genes. A total of 660 DEGs were identified with 276 downregulated genes and 384 upregulated genes within the selected parameters after statistical analysis (**Figure 2**).

Thereafter the selected genes were analyzed to understand the functional enrichment reflecting significant GO terms and enriched pathway. Top 5 GO terms were retained for both upregulated and downregulated genes (**Table 1**). The DEGs were then subjected to analyze their involvement in biological pathway. Selected genes showed sign of their involvement in KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway (**Figure 3**).

3.2. Proteome Signatures

DEGs were used to construct generic PPI network which generated almost scale free topological densely connected network (**Figure 4**). Then the generic PPI network was utilized in the construction of hub protein network with top 10 most connected nodes (**Figure 5**). UBC, PTEN, SOS1, PTK2, FGFR1, YAP1, FOXO1, RACK1, BMP4 and CD44 were identified as most interacted hub proteins. After that, top 2 modules in the existing network was also analyzed (**Figure 6**). Then the functional enrichment of best 2 modules were also analyzed. Module 1 was mostly involved in protein predominantly in protein acetylation function (P value: 6.86×10^{-6}) and module 2 was reported to be involved in G-protein coupled receptor (GPCR) signaling (P value 2.8×10^{-4}).

3.3. Regulatory Signatures

To identify the regulatory signatures, transcriptional regulatory elements (**Figure 8**) and post-transcriptional regulatory elements (**Figure 9**) were identified from TF/miRNA-DEGs networks.

Top 10 transcription factors (E2F1, STAT3, PPARG, MEF2A, FOXC1, GATA3, YY1, GATA2, NFKB1, and FOXL1) and miRNAs (hsa-mir-155-5p, hsa-mir-16-5p, hsa-mir-17-5p, hsa-mir-19a-3p, hsa-mir-192-5p, hsa-mir-92a-3p, hsa-mir-26b-5p, hsa-mir-335-5p, hsa-mir-124-3p, and hsa-let-7b-5p) were selected from the networks as the best signatures.

4. Discussion

In order to identify potential biomarkers or drug target in IPF gene expression pattern was analyzed from array data. 660 differentially expressed genes were identified after statistical analysis. Again, their involvement in different biological processes, cellular compartmentation and molecular functions were also analyzed. The selected DEGs indicated their predominant involvement in transcription, extracellular matrix organization, cell adhesion, apoptotic process and binding activity (**Table 1**). IPF is assumed to be occurred by the manifestation of multiple pathways. A range of cascades including apoptotic pathway, inflammatory cytokines, extracellular matrix regulatory factors, vascular endothelial remodeling have been implicated in animal models of fibrosis (Maher et al., 2007).

Moreover, in order to identify hub proteins and regulatory biomolecules their PPI network and TF/miRNA-DEGs network were also analyzed. Protein-protein interaction is central to understand the functional relationships between different proteins and a protein-protein interaction map provides the functional organization of the proteome of specific interest (Xenarios and Eisenberg, 2001; Stelzl et al., 2005). We utilized BottleNeck method to construct the hub protein interactome. BottleNeck is more accurate in predicting the protein interactome

than degree matrix (Yu et al., 2007). Among the top 10 hub proteins, UBC (Ubiquitin-C) mediates proteasome dependent proteolysis, transcriptional regulation, apoptosis and many other functions (Radici et al., 2013). Elevated level of UBC has been reported in patients with Chronic Obstructive Pulmonary Disease (COPD) and ubiquitination and proteolysis have been shown to play a crucial role in COPD (Létuvé et al., 2010; Debigaré et al., 2010). PTEN (Phosphatase and Tensin Homolog) is a cell membrane phosphatase. Decreased PTEN activity has been demonstrated in the cell membrane of IPF fibroblast (Xia et al., 2010). Forkhead Box (FOXO) 1 along with FOXO3 have been shown to have favorable inhibitory effect on fibrosis activation and reduction in extracellular matrix production. These two proteins have already been suggested as potential drug target in fibrosis (Xin et al., 2018; Al-Tamari et al., 2018). The transmembrane glycoprotein CD44 comprises several isoforms and plays crucial role in cell-cell adhesion. A different form of CD44 expression pattern was observed in plexiform lesions of idiopathic pulmonary arterial hypertension (IPAH) (Ohta-Ogo et al., 2012). Yet another study evident different expression patterns of CD44 including more isoforms in fibrotic lung samples (Kasper et al., 1995). Recent study with Yes-Associated Protein 1 (YAP1), a key regulator of Hippo pathway, revealed its overexpression leading to cell proliferation, migration and collagen production in IPF whereas knockdown of YAP1 resulted in reduced fibroblast aggregation with amelioration of the fibrosis condition both *in vivo* and *in vitro* (Chen et al., 2019). FGFR1 (Fibroblast Growth Factor 1) plays key role in the development of squamous cell lung cancer (Sekine et al., 2014). Upregulation of RACK1 (Receptor of Activated C Kinase 1) has been shown to induce hepatic fibrosis in mice (Jia et al., 2013).

Moreover, two modules (**Figure 6**) were selected from the generic PPI network which was predicted to majorly be involved in protein acetylation and GPCR signaling. Abnormal pattern of

histone acetylation has been observed in IPF lung tissue which was evident by low level of COX-2 production in laboratory experiment (Coward et al., 2009). Moreover, abnormal histone deacetylation has also been reported in lung tissues of IPAH patient (Nozik-Grayck et al 2016). Transcriptional signatures i.e., TFs and Post-transcriptional signatures i.e., miRNAs provide potential sources in biomarker identification and drug targeting (Islam et al., 2018). DEGs were mapped to construct TF/miRNA-DEGs interaction network in order to find regulatory biomolecule signatures. Among the top 10 selected TFs (**Figure 7**), GATA3 Binding Protein (GATA3) overexpression in mouse model has been shown to enhance the development of IPF (Kimura et al., 2006). Forkhead Box C1 (FOXC1) encoding gene has been found to be differentially methylated in laboratory experiment with IPF subject (Yang et al., 2014). MEF2A (Myocyte Enhancer Factor 2A) has been reported to influence the growth and proliferation of lung fibroblast (Han et al., 2015). PPARG (Peroxisome Proliferator-Activated Receptor Gamma) is assumed to control the gene expression in lung fibrosis and it has already been suggested as a potential drug target in IPF (Dumoulin et al., 2010). GATA2 (GATA2 Binding Protein) deficiency is a frequent feature in pulmonary diseases. Association of GATA2 deficiency has been evident in both fibrosis and Pulmonary Alveolar Proteinosis (PAP) (Svobodova et al., 2015; Ballerie et al., 2016). STAT3 (Signal Transduce and Activator of Transcription 3) plays multiple crucial roles in maintaining lung homeostasis and it is assumed to be one major target in lung fibrosis (Prêle et al., 2012). YY1 (Yin Yang 1) is a transcriptional repressor which has been suggested as novel regulator of pulmonary fibrosis. Overexpression of YY1 has been reported in both human IPF and murine model and it has been claimed that YY1 promotes fibrogenesis at least in part by increasing collagen or α -SMA secretion (Lin et al., 2011).

Among the top 10 selected miRNAs, hsa-mir-155-5p, hsa-mir-16-5p and hsa-mir-26b-5p have been predicted to be involved in multiple pathways giving rise to interstitial lung diseases (ILD) and suggested as non-invasive biomarker (Mishra et al., 2018), hsa-mir-17-5p and hsa-mir-19a-3p have been shown to be involved in the prognosis of small cell lung cancer (Hayashita et al., 2005; Mancuso et al., 2016). Again, downregulation of hsa-mir-335-5p expression has been reported in COPD patients with PiZZ (Glu342Lys) inherited alpha1-antitrypsin deficiency (AATD) (Esquinas et al., 2017). In yet other laboratory experiments, hsa-mir-124-3p has been shown to participate in pulmonary vascular remodeling in association with a long non-coding RNA (lncRNA) called MALAT1 (Metastasis Associated Lung Adenocarcinoma Transcript 1), and hsa-let-7b-5p has been shown to be downregulated in patients with cystic fibrosis (CF) (Wang et al., 2019; Ideozu et al., 2019).

Finally, the identified hub proteins and regulatory biomolecules may provide potentials sources of non-invasive biomarkers which could be used in targeting drugs against IPF and early diagnosis of this disease. *In vitro* and *in vivo* experimental evidences also suggested their potential roles and connections with IPF and other related lung diseases. This study suggests (UBC, PTEN, SOS1, PTK2, FGFR1, YAP1, FOXO1, RACK1, BMP4 and CD44) as the best proteome signatures, (E2F1, STAT3, PPARG, MEF2A, FOXC1, GATA3, YY1, GATA2, NFKB1, and FOXL1) as the best transcription regulatory signatures, and (hsa-mir-155-5p, hsa-mir-16-5p, hsa-mir-17-5p, hsa-mir-19a-3p, hsa-mir-192-5p, hsa-mir-92a-3p, hsa-mir-26b-5p, hsa-mir-335-5p, hsa-mir-124-3p, and hsa-let-7b-5p) as the best post-transcription regulatory signatures in IPF within the selected strategies employed in the exploration of biomarkers.

5. Conclusion

The underlying mechanism of IPF is poorly understood and this is the major hindrance to develop effective drug and early diagnosis method. In this study network-based approach has been utilized to analyze microarray data in order to identify potential transcriptome, proteome and regulatory signatures. Different hub genes involved in apoptic process, transcription, inflammatory responses, extracellular matrix organization have been identified to play key role. Thereafter multiple hub proteins, transcription factors and micro RNAs playing crucial roles in IPF and other related lung diseases were also identified and these findings were supported by other available laboratory experimental data. We suggest further exploration and possible laboratory experiments with the identified signatures to confirm their superiority in the diagnosis and drug targeting against IPF. Hopefully, this study will raise research interest among researchers and contribute in the discovery of non-invasive and effective biomarkers and potential drug targets.

Acknowledgement

Authors are thankful to the members of Swift Integrity Computational Lab, Dhaka, Bangladesh.

Conflict of Interest

Authors declare that there is no conflict of interest regarding the publication of this manuscript.

Funding Statement

Authors received no funding from external sources.

Data Availability Statement

All the data generated and analyzed are summarized in this manuscript.

References

- Al□Tamari, H.M., Dabral, S., Schmall, A., Sarvari, P., Ruppert, C., Paik, J., DePinho, R.A., Grimminger, F., Eickelberg, O., Guenther, A. and Seeger, W., 2018. FoxO3 an important player

in fibrogenesis and therapeutic target for idiopathic pulmonary fibrosis. *EMBO molecular medicine*, 10(2), pp.276-293.

Ballerie, A., Nimubona, S., Meunier, C., Gutierrez, F.L., Desrues, B., Delaval, P. and Jouneau, S., 2016.

Association of pulmonary alveolar proteinosis and fibrosis: patient with GATA2 deficiency. *European Respiratory Journal*, 48(5), pp.1510-1514.

Chen, Y., Zhao, X., Sun, J., Su, W., Zhang, L., Li, Y., Liu, Y., Zhang, L., Lu, Y., Shan, H. and Liang,

H., 2019. YAP1/Twist promotes fibroblast activation and lung fibrosis that conferred by miR-15a loss in IPF. *Cell Death & Differentiation*, p.1.

Chin, C.H., Chen, S.H., Wu, H.H., Ho, C.W., Ko, M.T. and Lin, C.Y., 2014. cytoHubba: identifying hub

objects and sub-networks from complex interactome. *BMC systems biology*, 8(4), p.S11.

Collard, H.R., Moore, B.B., Flaherty, K.R., Brown, K.K., Kaner, R.J., King Jr, T.E., Lasky, J.A., Loyd,

J.E., Noth, I., Olman, M.A. and Raghu, G., 2007. Acute exacerbations of idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine*, 176(7), pp.636-643.

Coward, W.R., Watts, K., Feghali-Bostwick, C.A., Knox, A. and Pang, L., 2009. Defective histone

acetylation is responsible for the diminished expression of cyclooxygenase 2 in idiopathic pulmonary fibrosis. *Molecular and cellular biology*, 29(15), pp.4325-4339.

Crow, M., Lim, N., Ballouz, S., Pavlidis, P. and Gillis, J., 2019. Predictability of human differential

gene expression. *Proceedings of the National Academy of Sciences*, 116(13), pp.6491-6500.

Debigaré, R., Côté, C.H. and Maltais, F., 2010. Ubiquitination and proteolysis in limb and respiratory

muscles of patients with chronic obstructive pulmonary disease. *Proceedings of the American Thoracic Society*, 7(1), pp.84-90.

DePianto, D.J., Chandriani, S., Abbas, A.R., Jia, G., N'Diaye, E.N., Caplazi, P., Kauder, S.E., Biswas, S., Karnik, S.K., Ha, C. and Modrusan, Z., 2015. Heterogeneous gene expression signatures correspond to distinct lung pathologies and biomarkers of disease severity in idiopathic pulmonary fibrosis. *Thorax*, 70(1), pp.48-56.

Dumoulin, E., Martel, M., Cloutier, D. and Cantin, A.M., 2010. Nrf2 Versus Peroxisome Proliferator-activated Receptor-gamma (PPAR γ) As Therapeutic Targets For Idiopathic Pulmonary Fibrosis. In A19. *PRECLINICAL TESTING IN MOUSE MODELS OF PULMONARY FIBROSIS* (pp. A1055-A1055). American Thoracic Society.

Esquinas, C., Janciauskiene, S., Gonzalo, R., de Xaxars, G.M., Olejnicka, B., Belmonte, I., Barrecheguren, M., Rodriguez, E., Nunez, A., Rodriguez-Frias, F. and Miravittles, M., 2017. Gene and miRNA expression profiles in PBMCs from patients with severe and mild emphysema and PiZZ alpha1-antitrypsin deficiency. *International journal of chronic obstructive pulmonary disease*, 12, p.3381.

Fernandez, I.E. and Eickelberg, O., 2012. New cellular and molecular mechanisms of lung injury and fibrosis in idiopathic pulmonary fibrosis. *The Lancet*, 380(9842), pp.680-688.

Giltane, J.M. and Rimm, D.L., 2004. Technology insight: Identification of biomarkers with tissue microarray technology. *Nature Reviews Clinical Oncology*, 1(2), p.104.

Han, J.H., Hwang, A.R., Nam, D.H., Kim, S., Choi, H.C., Lim, J.H. and Woo, C.H., 2015. ERK5 regulates basic fibroblast growth factor-induced type 1 plasminogen activator inhibitor expression and cell proliferation in lung fibroblasts. *Life sciences*, 135, pp.1-8.

Hayashita, Y., Osada, H., Tatematsu, Y., Yamada, H., Yanagisawa, K., Tomida, S., Yatabe, Y., Kawahara, K., Sekido, Y. and Takahashi, T., 2005. A polycistronic microRNA cluster, miR-17-

92, is overexpressed in human lung cancers and enhances cell proliferation. *Cancer research*, 65(21), pp.9628-9632.

Hughes, G., Toellner, H., Morris, H., Leonard, C. and Chaudhuri, N., 2016. Real world experiences: pirfenidone and nintedanib are effective and well tolerated treatments for idiopathic pulmonary fibrosis. *Journal of clinical medicine*, 5(9), p.78.

Ideozu, J.E., Zhang, X., Rangaraj, V., McColley, S. and Levy, H., 2019. Microarray profiling identifies extracellular circulating miRNAs dysregulated in cystic fibrosis. *Scientific reports*, 9(1), pp.1-11.

Islam, T., Rahman, R., Gov, E., Turanli, B., Gulfidan, G., Haque, A., Arga, K.Y. and Haque Mollah, N., 2018. Drug targeting and biomarkers in head and neck cancers: insights from systems biology analyses. *OmicS: a journal of integrative biology*, 22(6), pp.422-436.

Jia, D., Duan, F., Peng, P., Sun, L., Liu, X., Wang, L., Wu, W., Ruan, Y. and Gu, J., 2013. Up-regulation of RACK1 by TGF- β 1 promotes hepatic fibrosis in mice. *PloS one*, 8(3), p.e60115.

Kasper, M., Günthert, U., Dall, P., Kayser, K., Schuh, D., Haroske, G. and Müller, M., 1995. Distinct expression patterns of CD44 isoforms during human lung development and in pulmonary fibrosis. *American journal of respiratory cell and molecular biology*, 13(6), pp.648-656.

Kerr, M.K., Martin, M. and Churchill, G.A., 2000. Analysis of variance for gene expression microarray data. *Journal of computational biology*, 7(6), pp.819-837.

Kimura, T., Ishii, Y., Yoh, K., Morishima, Y., Iizuka, T., Kiwamoto, T., Matsuno, Y., Homma, S., Nomura, A., Sakamoto, T. and Takahashi, S., 2006. Overexpression of the transcription factor GATA-3 enhances the development of pulmonary fibrosis. *The American journal of pathology*, 169(1), pp.96-104.

- King Jr, T.E., Pardo, A. and Selman, M., 2011. Idiopathic pulmonary fibrosis. *The Lancet*, 378(9807), pp.1949-1961.
- Létuvé, S., Kozhich, A., Humbles, A., Brewah, Y., Dombret, M.C., Grandsaigne, M., Adle, H., Kolbeck, R., Aubier, M., Coyle, A.J. and Pretolani, M., 2010. Lung chitinolytic activity and chitotriosidase are elevated in chronic obstructive pulmonary disease and contribute to lung inflammation. *The American journal of pathology*, 176(2), pp.638-649.
- Ley, B. and Collard, H.R., 2013. Epidemiology of idiopathic pulmonary fibrosis. *Clinical epidemiology*, 5, p.483.
- Ley, B., Collard, H.R. and King Jr, T.E., 2011. Clinical course and prediction of survival in idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine*, 183(4), pp.431-440.
- Li, P., Li, J., Chen, T., Wang, H., Chu, H., Chang, J., Zang, W., Wang, Y., Ma, Y., Du, Y. and Zhao, G., 2014. Expression analysis of serum microRNAs in idiopathic pulmonary fibrosis. *International journal of molecular medicine*, 33(6), pp.1554-1562.
- Lin, X., Sime, P.J., Xu, H., Williams, M.A., LaRussa, L., Georas, S.N. and Guo, J., 2011. Yin yang 1 is a novel regulator of pulmonary fibrosis. *American journal of respiratory and critical care medicine*, 183(12), pp.1689-1697.
- Maher, T.M., Wells, A.U. and Laurent, G.J., 2007. Idiopathic pulmonary fibrosis: multiple causes and multiple mechanisms?. *European Respiratory Journal*, 30(5), pp.835-839.

- Mancuso, G., Bovio, E., Rena, O., Rrapaj, E., Mercalli, F., Veggiani, C., Paganotti, A., Andorno, S. and Boldorini, R., 2016. Prognostic impact of a 3-MicroRNA signature in cytological samples of small cell lung cancer. *Cancer cytopathology*, 124(9), pp.621-629.
- Meltzer, E.B., Barry, W.T., D'Amico, T.A., Davis, R.D., Lin, S.S., Onaitis, M.W., Morrison, L.D., Sporn, T.A., Steele, M.P. and Noble, P.W., 2011. Bayesian probit regression model for the diagnosis of pulmonary fibrosis: proof-of-principle. *BMC medical genomics*, 4(1), p.70.
- Mishra, S., Shah, M.I., Sarkar, M. and Rout, C., 2018. Integrated analysis of non-coding RNAs for the identification of promising biomarkers in interstitial lung diseases. *Gene Reports*, 11, pp.87-93.
- Nalysnyk, L., Cid-Ruzafa, J., Rotella, P. and Esser, D., 2012. Incidence and prevalence of idiopathic pulmonary fibrosis: review of the literature. *European Respiratory Review*, 21(126), pp.355-361.
- Nozik-Grayck, E., Woods, C., Stearman, R.S., Venkataraman, S., Ferguson, B.S., Swain, K., Bowler, R.P., Geraci, M.W., Ihida-Stansbury, K., Stenmark, K.R. and McKinsey, T.A., 2016. Histone deacetylation contributes to low extracellular superoxide dismutase expression in human idiopathic pulmonary arterial hypertension. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 311(1), pp.L124-L134.
- Nunes, I.J.G., 2018. Gene expression analysis platform (GEAP): uma plataforma flexível e intuitiva para análise de transcriptoma.
- Ohta-Ogo, K., Hao, H., Ishibashi-Ueda, H., Hirota, S., Nakamura, K., Ohe, T. and Ito, H., 2012. CD44 expression in plexiform lesions of idiopathic pulmonary arterial hypertension. *Pathology international*, 62(4), pp.219-225.

- Prêle, C.M., Yao, E., O'Donoghue, R.J., Mutsaers, S.E. and Knight, D.A., 2012. STAT3: a central mediator of pulmonary fibrosis?. *Proceedings of the American Thoracic Society*, 9(3), pp.177-182.
- Radici, L., Bianchi, M., Crinelli, R. and Magnani, M., 2013. Ubiquitin C gene: structure, function, and transcriptional regulation. *Advances in Bioscience and Biotechnology*, 4(12), p.1057.
- Saito, R., Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L., Lotia, S., Pico, A.R., Bader, G.D. and Ideker, T., 2012. A travel guide to Cytoscape plugins. *Nature methods*, 9(11), p.1069.
- Sandelin, A., Alkema, W., Engström, P., Wasserman, W.W. and Lenhard, B., 2004. JASPAR: an open access database for eukaryotic transcription factor binding profiles. *Nucleic acids research*, 32(suppl_1), pp.D91-D94.
- Sekine, Y., Hata, A., Koh, E. and Hiroshima, K., 2014. Lung carcinogenesis from chronic obstructive pulmonary disease: characteristics of lung cancer from COPD and contribution of signal transducers and lung stem cells in the inflammatory microenvironment. *General thoracic and cardiovascular surgery*, 62(7), pp.415-421.
- Selman, M., King, T.E. and Pardo, A., 2001. Idiopathic pulmonary fibrosis: prevailing and evolving hypotheses about its pathogenesis and implications for therapy. *Annals of internal medicine*, 134(2), pp.136-151.
- Sethupathy, P., Corda, B. and Hatzigeorgiou, A.G., 2006. TarBase: A comprehensive database of experimentally supported animal microRNA targets. *Rna*, 12(2), pp.192-197.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11), pp.2498-2504.

Sherman, B.T. and Lempicki, R.A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*, 4(1), pp.44-57.

Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F.H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S. and Timm, J., 2005. A human protein-protein interaction network: a resource for annotating the proteome. *Cell*, 122(6), pp.957-968.

Svobodova, T., Mejstrikova, E., Salzer, U., Sukova, M., Hubacek, P., Matej, R., Vasakova, M., Hornofova, L., Dvorakova, M., Fronkova, E. and Votava, F., 2015. Diffuse parenchymal lung disease as first clinical manifestation of GATA-2 deficiency in childhood. *BMC pulmonary medicine*, 15(1), p.8.

Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N.T., Roth, A., Bork, P. and Jensen, L.J., 2016. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic acids research*, p.gkw937.

Wang D, Xu H, Wu B, Jiang S, Pan H, Wang R, Chen J. Long non-coding RNA MALAT1 sponges miR-124-3p. 1/KLF5 to promote pulmonary vascular remodeling and cell cycle progression of pulmonary artery hypertension. *International journal of molecular medicine*. 2019 Sep 1;44(3):871-84.

Xenarios, I. and Eisenberg, D., 2001. Protein interaction databases. *Current Opinion in Biotechnology*, 12(4), pp.334-339.

- Xia, H., Khalil, W., Kahm, J., Jessurun, J., Kleidon, J. and Henke, C.A., 2010. Pathologic caveolin-1 regulation of PTEN in idiopathic pulmonary fibrosis. *The American journal of pathology*, 176(6), pp.2626-2637.
- Xia, J., Gill, E.E. and Hancock, R.E., 2015. NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. *Nature protocols*, 10(6), p.823.
- Xin, Z., Ma, Z., Hu, W., Jiang, S., Yang, Z., Li, T., Chen, F., Jia, G. and Yang, Y., 2018. FOXO1/3: Potential suppressors of fibrosis. *Ageing research reviews*, 41, pp.42-52.
- Yang, I.V., Pedersen, B.S., Rabinovich, E., Hennessy, C.E., Davidson, E.J., Murphy, E., Guardela, B.J., Tedrow, J.R., Zhang, Y., Singh, M.K. and Correll, M., 2014. Relationship of DNA methylation and gene expression in idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine*, 190(11), pp.1263-1272.
- Yu, H., Kim, P.M., Sprecher, E., Trifonov, V. and Gerstein, M., 2007. The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS computational biology*, 3(4), p.e59.

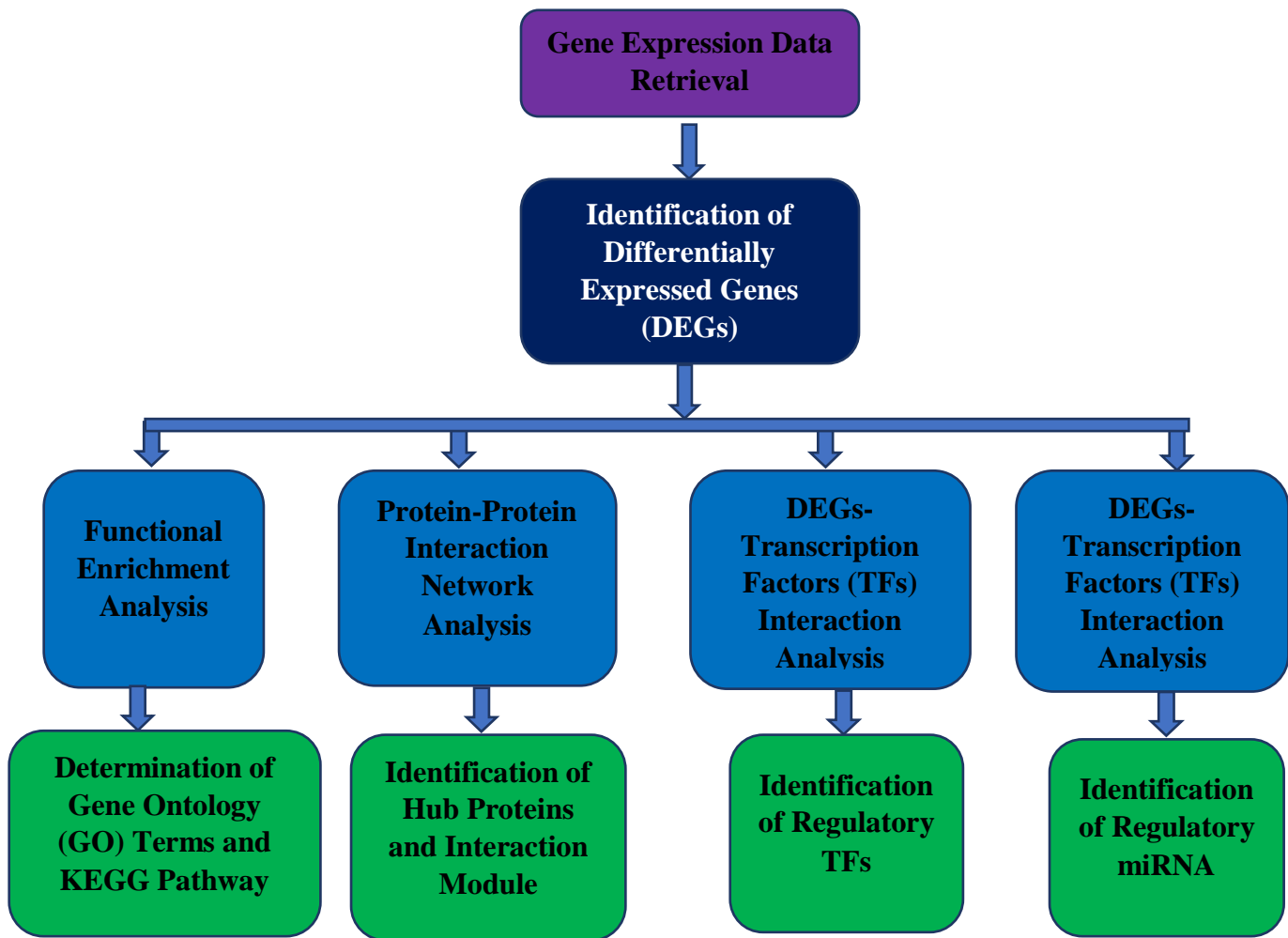


Figure 1: Flowchart of strategies employed in the overall study.

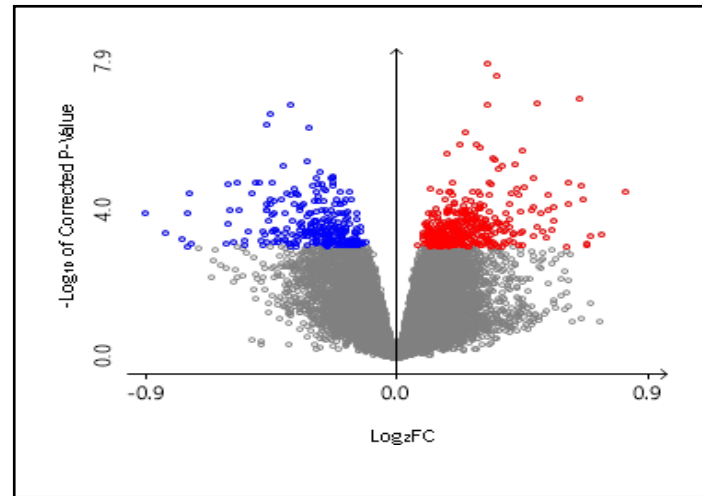


Figure 2: Volcano plot of selected differentially expressed genes (DEGs). Colored (Blue: Down regulated genes; Red: Upregulated genes) DEGs have been selected with adjusted P value > 0.01. filter.

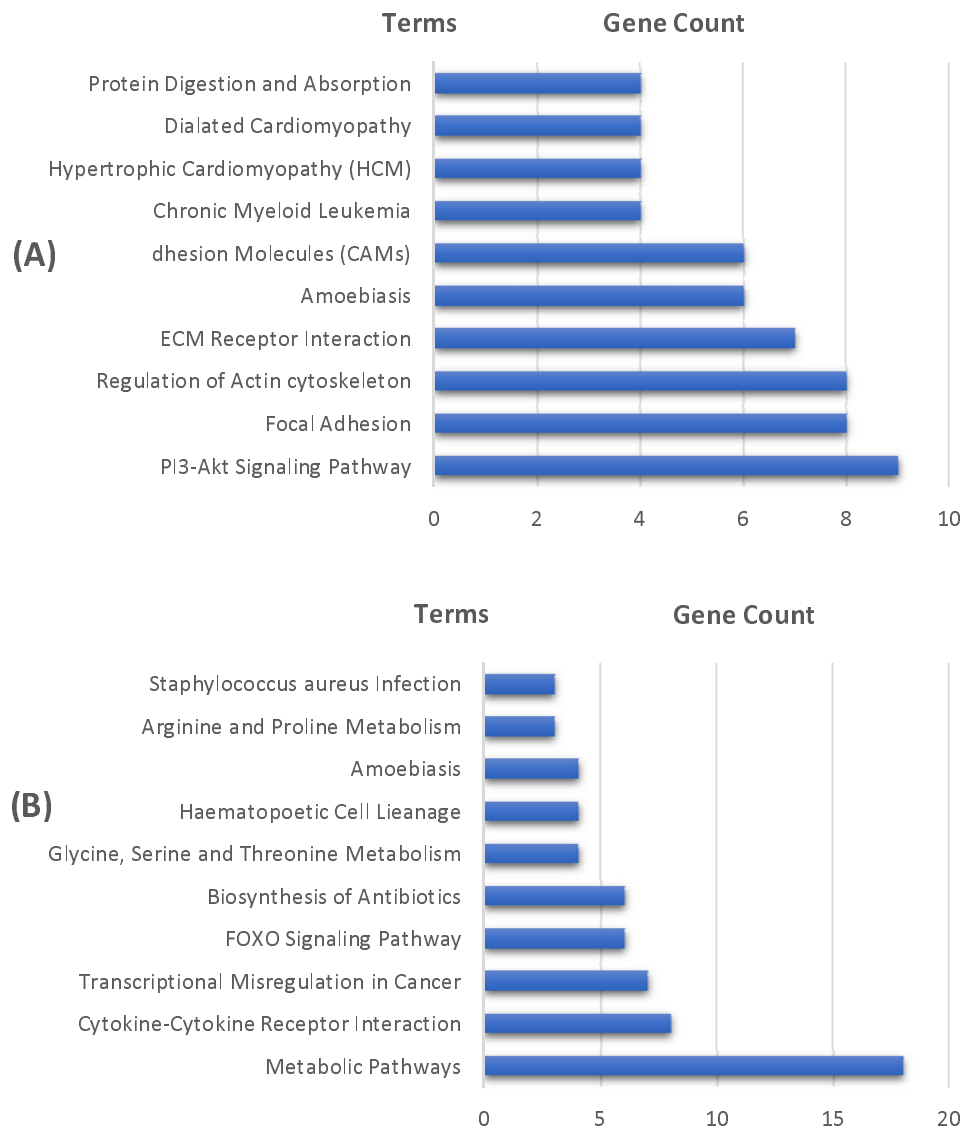


Figure 3: KEGG pathway of differentially expressed genes: (A) Upregulated genes; (B) Downregulated genes.

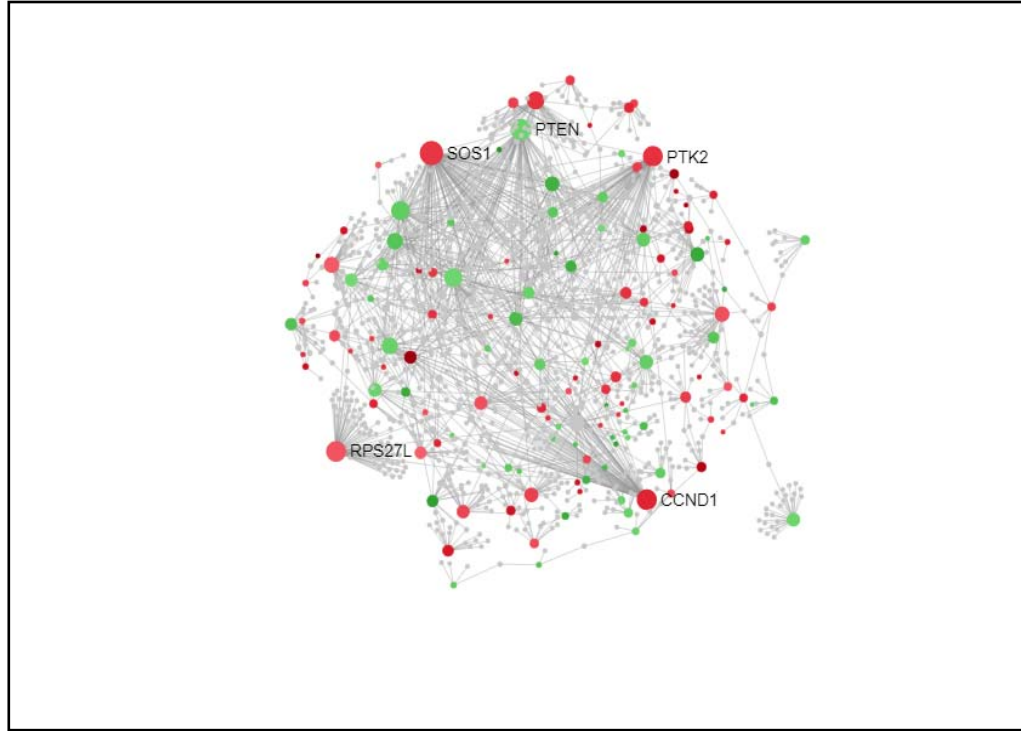


Figure 4: Protein-protein interaction (PPI) network of differentially expressed genes (DEGs). Nodes represent DEGs (Green: Upregulated genes; Red: Downregulated genes). Edges represent interaction.

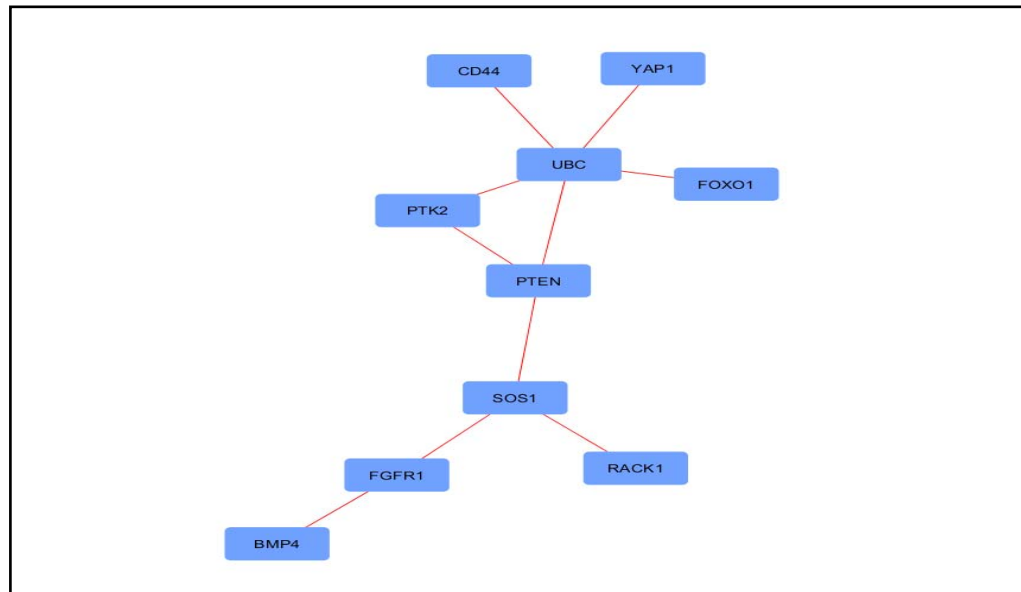


Figure 5: Hub proteins from generated protein-protein interaction (PPI) network. Nodes represent proteins and edges represent interactions.

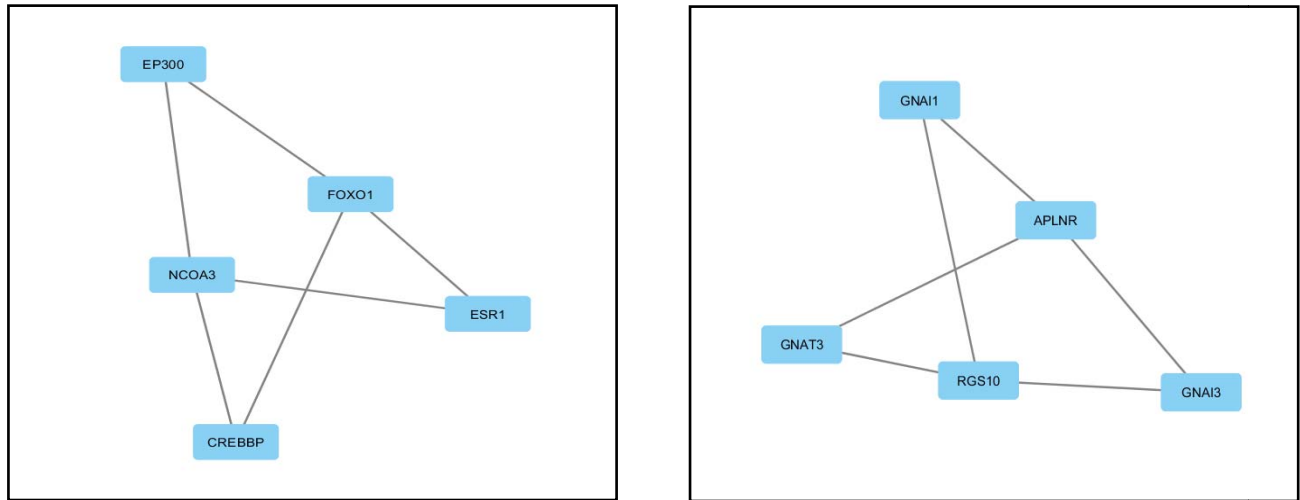


Figure 6: Best two modules obtained from protein-protein interaction (PPI) network in IPF
Nodes represent proteins and edges represent interactions

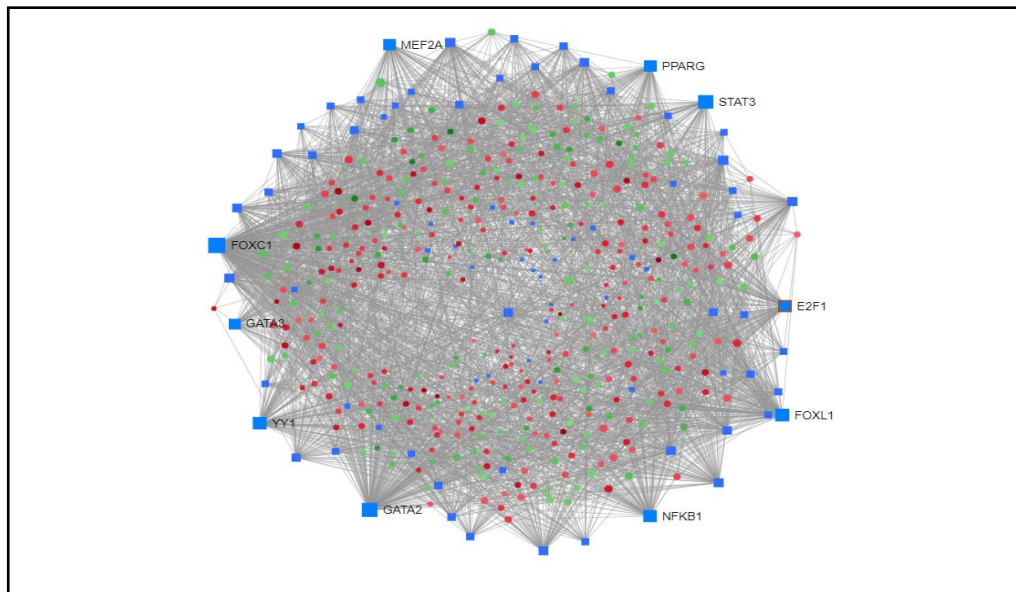


Figure 7: Interaction between transcription factor (TF) and differentially expressed genes (DEGs) network of differentially expressed genes (DEGs). Nodes represent DEGs (Green: Upregulated genes; Red: Downregulated genes; Blue: Transcription factors). Edges represent interaction.

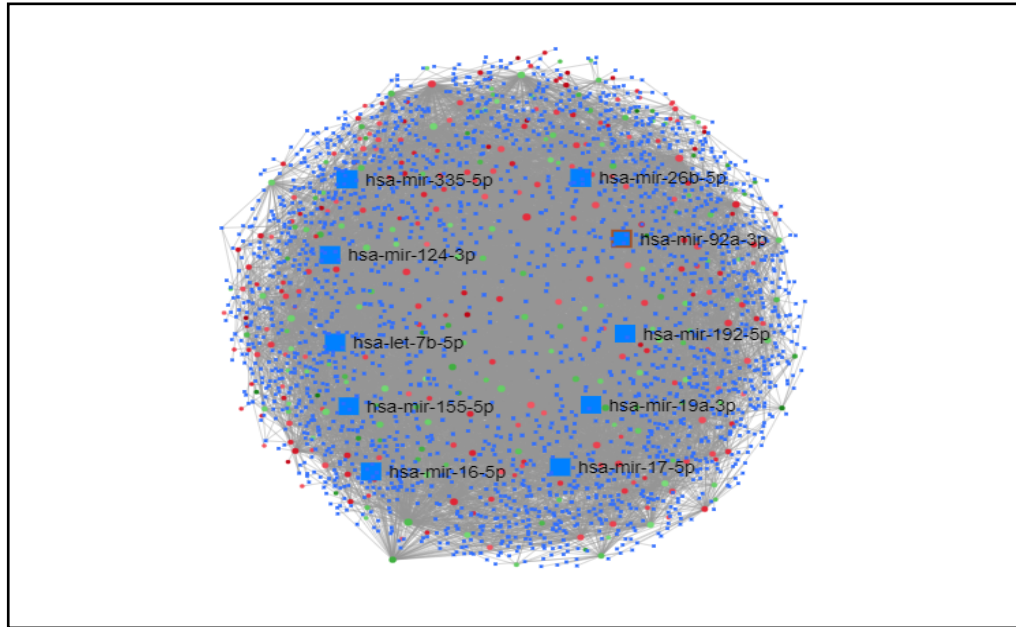


Figure 8: Interaction between miRNA and differentially expressed genes network of differentially expressed genes (DEGs). Nodes represent DEGs (Green: Upregulated genes; Red: Downregulated genes; Blue: miRNA). Edges represent interaction.

	Terms	Gene Count	Coverage (%)	P-Value
Biological Processes	Transcription, DNA-templated	36	15.70	3.0×10^{-3}
	Regulation of Transcription, DNA-Templated	26	11.40	2.8×10^{-2}
	Cell Adhesion	12	5.20	1.4×10^{-2}
	Extracellular Matrix Organization	11	4.80	7.0×10^{-5}
	Positive regulation of Apoptic process	22	4.40	6.5×10^{-3}
Cellular Components	Extracellular Region	28	12.20	2.2×10^{-2}
	Intracellular	27	11.80	3.4×10^{-3}
	Extracellular Space	25	10.90	1.4×10^{-2}
	Extracellular Matrix	14	6.10	2.8×10^{-5}
	Cell Surface	14	6.10	7.6×10^{-3}
Molecular Functions	Metal Ion Binding	39	17.00	7.2×10^{-4}
	DNA Binding	26	11.40	6.2×10^{-2}
	Nucleic Acid Binding	25	10.90	1.7×10^{-4}
	Actin Binding	9	3.90	1.1×10^{-2}
	Heparin Binding	6	2.60	3.0×10^{-2}
Downregulated Genes				
	Terms	Gene Count	Coverage (%)	P-Value
Biological Processes	Inflammatory Response	11	6.80	1.5×10^{-3}
	Positive Regulation of Cell Proliferation	10	6.20	1.8×10^{-2}
	Immune Response	9	5.60	2.8×10^{-2}
	Negative Regulation of Apoptic Process	9	5.60	4.1×10^{-2}
	Positive Regulation of Apoptice Process	8	5.00	1.4×10^{-2}
Cellular Components	Integral Component of Membrane	56	34.80	1.4×10^{-2}
	Cytoplasm	52	32.30	7.9×10^{-2}
	Plasma Membrane	48	29.80	7.3×10^{-3}
	Extracellular Region	20	12.40	6.6×10^{-2}
	Extracellular Space	19	11.80	2.7×10^{-2}
Molecular Functions	ATP Binding	20	12.40	4.0×10^{-2}
	Identical Protein Binding	13	8.10	2.2×10^{-2}
	Protein Homodimerization Activity	12	7.50	4.2×10^{-2}
	Ubiquitin Protein Ligase Activity	7	4.30	3.3×10^{-2}
	Enzyme Binding	7	4.30	5.9×10^{-2}

Table 1: Top five gene ontology (GO) terms of differentially expressed genes (DEGs) in Idiopathic Pulmonary Fibrosis.

