



# Integrative Analysis of Transcriptome-Wide Association Study and mRNA Expression Profiles Identifies Candidate Genes Associated With Idiopathic Pulmonary Fibrosis

Weiming Gong<sup>1†</sup>, Ping Guo<sup>1†</sup>, Lu Liu<sup>1</sup>, Qingbo Guan<sup>2,3,4\*</sup> and Zhongshang Yuan<sup>1\*</sup>

<sup>1</sup> Department of Biostatistics, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan, China,

<sup>2</sup> Department of Endocrinology, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, China,

<sup>3</sup> Shandong Clinical Medical Center of Endocrinology and Metabolism, Jinan, China, <sup>4</sup> Shandong Institute of Endocrine and Metabolic Diseases, Jinan, China

## OPEN ACCESS

### Edited by:

Sheng Yang,  
Nanjing Medical University, China

### Reviewed by:

Xiaohui Niu,  
Huazhong Agricultural University,  
China

Haitao Yang,  
Hebei Medical University, China

### \*Correspondence:

Qingbo Guan  
doctorguanqingbo@163.com  
Zhongshang Yuan  
yuanzhongshang@sdu.edu.cn

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Computational Genomics,  
a section of the journal  
Frontiers in Genetics

**Received:** 09 September 2020

**Accepted:** 17 November 2020

**Published:** 10 December 2020

### Citation:

Gong W, Guo P, Liu L, Guan Q  
and Yuan Z (2020) Integrative Analysis  
of Transcriptome-Wide Association  
Study and mRNA Expression Profiles  
Identifies Candidate Genes  
Associated With Idiopathic Pulmonary  
Fibrosis. *Front. Genet.* 11:604324.  
doi: 10.3389/fgene.2020.604324

Idiopathic pulmonary fibrosis (IPF) is a type of scarring lung disease characterized by a chronic, progressive, and irreversible decline in lung function. The genetic basis of IPF remains elusive. A transcriptome-wide association study (TWAS) of IPF was performed by FUSION using gene expression weights of three tissues combined with a large-scale genome-wide association study (GWAS) dataset, totally involving 2,668 IPF cases and 8,591 controls. Significant genes identified by TWAS were then subjected to gene ontology (GO) and pathway enrichment analysis. The overlapped GO terms and pathways between enrichment analysis of TWAS significant genes and differentially expressed genes (DEGs) from the genome-wide mRNA expression profiling of IPF were also identified. For TWAS significant genes, protein–protein interaction (PPI) network and clustering modules analyses were further conducted using STRING and Cytoscape. Overall, TWAS identified a group of candidate genes for IPF under the Bonferroni corrected  $P$  value threshold ( $0.05/14929 = 3.35 \times 10^{-6}$ ), such as *DSP* ( $P_{TWAS} = 1.35 \times 10^{-29}$  for lung tissue), *MUC5B* ( $P_{TWAS} = 1.09 \times 10^{-28}$  for lung tissue), and *TOLLIP* ( $P_{TWAS} = 1.41 \times 10^{-15}$  for whole blood). Pathway enrichment analysis identified multiple candidate pathways, such as herpes simplex infection ( $P$  value =  $7.93 \times 10^{-5}$ ) and antigen processing and presentation ( $P$  value =  $6.55 \times 10^{-5}$ ). 38 common GO terms and 8 KEGG pathways shared by enrichment analysis of TWAS significant genes and DEGs were identified. In the PPI network, 14 genes (*DYNLL1*, *DYNC1LI1*, *DYNLL2*, *HLA-DRB5*, *HLA-DPB1*, *HLA-DQB2*, *HLA-DQA2*, *HLA-DQB1*, *HLA-DRB1*, *POLR2L*, *CENPP*, *CENPK*, *NUP133*, and *NUP107*) were simultaneously detected by hub gene and module analysis. In conclusion, through integrative analysis of TWAS and mRNA expression profiles, we identified multiple novel candidate genes, GO terms and pathways for IPF, which contributes to the understanding of the genetic mechanism of IPF.

**Keywords:** idiopathic pulmonary fibrosis, transcriptome-wide association study, gene expression profiling, pathway enrichment, protein–protein interaction network

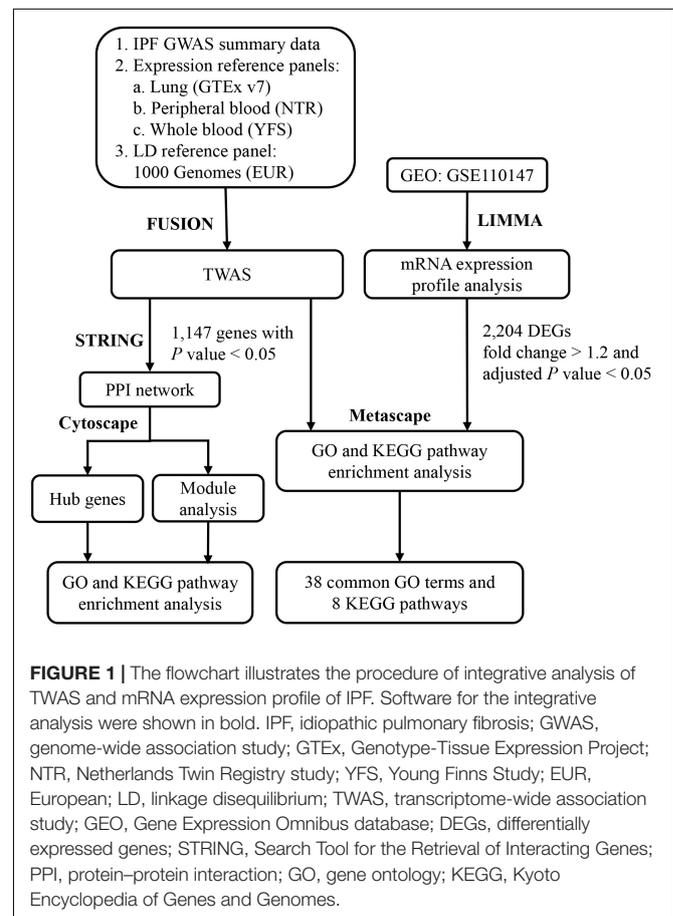
## INTRODUCTION

Idiopathic pulmonary fibrosis (IPF) is a chronic interstitial lung disease characterized by the formation of scar tissue and a progressive, and irreversible decline in lung function (Raghu et al., 2011; Lederer and Martinez, 2018), with a median survival time from diagnosis of 2–4 years (Ley et al., 2011). The incidence of IPF is increasing worldwide and has been estimated to be 3–9 cases per 100,000 people per year in Europe and North America, and fewer than four cases per 100,000 people per year in East Asia and South America (Hutchinson et al., 2015). IPF has been confirmed to be related to varieties of environmental and genetic factors. Potential risk factors for IPF include aging, male sex, smoking, certain occupational exposures (Baumgartner et al., 2000), gastroesophageal reflux (Tobin et al., 1998; Bedard Methot et al., 2019), herpesvirus infection (Tang et al., 2003), air pollution (Sack et al., 2017), and obstructive sleep apnea (Kim et al., 2017). Genome-wide association studies (GWAS) on IPF (Mushiroda et al., 2008; Fingerlin et al., 2013, 2016; Noth et al., 2013; Allen et al., 2017, 2020) have identified common genetic variants related to IPF, highlighting the significance of several IPF susceptibility factors, such as telomere maintenance, host defense, cell-cell adhesion. Rare genetic variants regarding surfactant dysfunction and telomere biology have also been identified in studies of familial pulmonary fibrosis (Nogee et al., 2001; Armanios et al., 2007; Cogan et al., 2015; Stuart et al., 2015).

Genome-wide association study have significantly succeeded in identifying IPF-related susceptibility genetic loci. However, a great number of genetic variations identified reside in non-coding regions, which are generally difficult to characterize biologically. Indeed, one common sense for GWAS is that most disease-associated genetic variants are located in non-coding regions, resulting in the hypothesis that the underlying biological mechanism of disease may be closely related to gene expression regulation. Furthermore, several expression quantitative trait loci (eQTLs) studies have illustrated that the information on expression regulation may play a pivotal role in disease development (Albert and Kruglyak, 2015). Transcriptome-wide association study (TWAS) is widely utilized in integrating GWAS with eQTL studies for investigating the causal genes associated with complex traits or diseases (Gamazon et al., 2015; Gusev et al., 2016; Yuan et al., 2020). Therefore, TWAS analysis may help us to identify novel genes associated with IPF. On the other hand, the genome-wide mRNA expression profiling of IPF provides the opportunity to identify differentially expressed genes (DEGs). Furthermore, omics integrative analysis can combine different types of omics data and provides more comprehensive insights than that offered by any single type of omics data (Liu et al., 2013). These integrative analyses are implemented and expected to rebuild meaningful biological networks by integrating information from different types of data, thus have the potential to provide a more novel and reliable understanding with respect to the underlying biological mechanisms. Statistically, complementary information can be better captured and exploited by such data integration analyses (Yu and Zeng, 2018). Indeed, in high-throughput genomic

studies, one common sense is that the analysis from single dataset often lack of reproducibility and integrative analysis can efficiently investigate and make full use of multiple datasets in a cost-efficient manner to enhance reproducibility (Yu and Zeng, 2018). This motivated us to perform a comprehensive integrative analysis of TWAS and mRNA expression profile of IPF, which may provide the better understanding of the molecular mechanisms of IPF.

In the present study, we leveraged expression imputation from a large-scale IPF GWAS dataset to perform a TWAS analysis in peripheral blood, whole blood and lung tissue. The TWAS significant genes and DEGs identified by mRNA expression profiling of IPF were then subjected to gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis. Through this analysis, the common GO terms and KEGG pathways were identified. Furthermore, for significant genes identified by TWAS for IPF, STRING and Cytoscape software were applied to implement protein–protein interaction (PPI) network and clustering modules analyses. Our results may provide novel insights into the understanding of the molecular mechanisms underlying the development of IPF. The detailed procedure of integrative analysis was displayed in **Figure 1**.



## MATERIALS AND METHODS

### GWAS of IPF

The current and largest-scale GWAS summary data of IPF were used (Allen et al., 2020). Briefly, it included 2,668 IPF cases and 8,591 controls of European ancestry from a meta-analysis of three case-control studies and restricted to unrelated individuals of European ancestry. Genotype data were imputed using the Haplotype Reference Consortium r1.1 panel. Stringent quality control was performed for the genotyped data. In each separate study, a genome-wide analysis of IPF susceptibility was conducted using SNPTEST v2.5.2, adjusting for the first 10 principal components to account for fine-scale population structure. Only biallelic autosomal variants that had a minor allele count  $\geq 10$ , were in Hardy-Weinberg Equilibrium ( $P > 1 \times 10^{-6}$ ), and well-imputed (imputation quality  $R^2 > 0.5$ ) in at least two studies were included. Detailed description related to study participants, genotyping, imputation, association analysis, and quality control can be found in the previous IPF GWAS study (Allen et al., 2020).

### TWAS of IPF

FUSION software was applied here for tissue-related TWAS analysis (Gusev et al., 2016). Briefly, FUSION leveraged a set of reference individuals to measure both gene expression and SNPs, and then to impute the *cis* genetic component of expression into a much larger set of individuals using their SNP genotype data. The imputed expression data can be viewed as a linear model of genotypes with weights based on the correlation between SNPs and gene expression in the reference data while accounting for linkage disequilibrium among SNPs. FUSION uses pre-computed gene expression weights together with disease GWAS summary statistics to evaluate the association between the expression levels of genes and target diseases (Gusev et al., 2016). The genetic values of expression were computed as one probe set at a time using SNP genotyping data located 500 kb on both sides of the gene boundary. The pre-computed expression reference weights of different tissues were downloaded from the FUSION websites<sup>1</sup>. For IPF TWAS, we used three expression reference panels, including lung, peripheral blood and whole blood, and a TWAS *P* value was obtained for each gene. Gene expression weights of lung were driven from the Genotype-Tissue Expression Project (GTEx v7;  $n = 383$ ) (GTEx Consortium et al., 2017). Gene expression weights of peripheral blood and whole blood reference panels were driven from the Netherlands Twin Registry study (NTR) ( $n = 1,247$ ) (Boomsma et al., 2006; Wright et al., 2014) and Young Finns Study (YFS) ( $n = 1,264$ ) (Raitakari et al., 2008), respectively.

### Gene Expression Profile Associated With IPF

The IPF gene expression profile data of lung tissue were obtained from the Gene Expression Omnibus database (access number: GSE110147) (Cecchini et al., 2018). Briefly, fresh frozen lung

samples were obtained from the organs of 22 patients with IPF; normal lung tissue ( $n = 11$ ) was obtained from the tissue flanking lung cancer resections. RNA was extracted and hybridized on Affymetrix microarrays. Individual-level gene expression data were included in the mRNA expression profile analysis implemented by LIMMA package (Ritchie et al., 2015). The DEGs between IPF patients and controls were identified at fold change  $> 1.2$  and adjusted *P* value  $< 0.05$ . Detailed description of sample characteristics, experimental design, statistical analysis, and quality control can be found in the previous study (Cecchini et al., 2018).

### Gene Set Enrichment Analysis

The IPF-related genes identified by TWAS and mRNA expression profiling were, respectively, subjected to GO and KEGG pathway enrichment analysis implemented by Metascape (Zhou et al., 2019)<sup>2</sup>. Note that in the enrichment analysis for IPF-related genes identified by TWAS, we included all the genes with a TWAS *P* value less than 0.05, rather than those under the Bonferroni corrected *P* value threshold ( $0.05/14929 = 3.35 \times 10^{-6}$ ), to increase the ability to identify more biological processes relevant to IPF and to make the results more stable by including more input genes (Reimand et al., 2019). A *P* value was calculated by Metascape for each GO term and pathway. Terms with a *P* value  $< 0.01$ , a minimum count of 3, and an enrichment factor  $> 1.5$  were collected and grouped into clusters based on their membership similarities. Kappa scores were used as the similarity metric when performing hierarchical clustering on the enriched terms, and sub-trees with a similarity of  $> 0.3$  were considered a cluster. The most statistically significant term within a cluster was chosen to represent the cluster. Finally, the Metascape analysis of TWAS was compared with that of mRNA expression profiles of lung tissue to identify the common GO terms and pathways shared by enrichment analysis for IPF-related genes from TWAS and for DEGs from mRNA expression profiling of IPF. Note that the common GO terms were obtained by overlapping the original GO enrichment results before grouping into clusters.

### Protein-Protein Interaction Network, Hub Genes, and Module Analysis

The PPI network of TWAS significant genes was constructed by the online Search Tool for the Retrieval of Interacting Genes (Szklarczyk et al., 2019) (STRING; 2017 release) database to evaluate the interactive relationships among the genes. Interactions with a combined score  $> 0.9$  were defined as statistically significant. Cytoscape software (Shannon et al., 2003) (version 3.5.1) was applied to visualize the integrated regulatory networks. The cytoHubba plugin and Molecular Complex Detection (MCODE) plugin in Cytoscape were used to identify hub genes and screen modules of the PPI network. All parameters of the plugin were set at their default values. Again, GO and KEGG enrichment of hub genes and genes in modules were also analyzed by Metascape.

<sup>1</sup><http://gusevlab.org/projects/fusion/>

<sup>2</sup><http://metascape.org>

## RESULTS

### TWAS Analysis Results

Totally, 14,929 genes were analyzed by TWAS in this study. Overall, TWAS identified 29 genes under the Bonferroni corrected  $P$  value threshold ( $0.05/14929 = 3.35 \times 10^{-6}$ ) and 1,147 genes with  $P$  value  $< 0.05$  (**Supplementary Table S1**), such as *DSP* ( $P_{TWAS} = 1.35 \times 10^{-29}$  for lung tissue), *MUC5B* ( $P_{TWAS} = 1.09 \times 10^{-28}$  for lung tissue), *TOLLIP* ( $P_{TWAS} = 1.41 \times 10^{-15}$  for whole blood), *MAPT* ( $P_{TWAS} = 9.60 \times 10^{-15}$  for lung tissue), and *DEPTOR* ( $P_{TWAS} = 8.58 \times 10^{-9}$  for lung tissue). The top 30 genes identified by TWAS are summarized in **Table 1**.

### Gene Set Enrichment Analysis

A total of 1,147 genes with a TWAS  $P$  value  $< 0.05$  were included in the GO enrichment analysis. Metascape detected 76 GO terms under  $P$  value  $< 0.01$  (**Figure 2** and **Supplementary Table S2**), such as antigen processing and presentation of peptide or polysaccharide antigen via major histocompatibility complex (MHC) class II ( $P_{TWAS} = 1.43 \times 10^{-8}$ ), vacuolar

part ( $P_{TWAS} = 5.98 \times 10^{-6}$ ), ncRNA metabolic process ( $P_{TWAS} = 1.61 \times 10^{-5}$ ), snRNA transcription by RNA polymerase II ( $P_{TWAS} = 3.03 \times 10^{-5}$ ), and SWI/SNF complex ( $P_{TWAS} = 5.81 \times 10^{-5}$ ). For KEGG pathway enrichment analysis of the genes identified by TWAS, Metascape detected 30 candidate pathways for IPF under  $P$  value  $< 0.01$  (**Figure 3** and **Supplementary Table S3**), such as *Staphylococcus aureus* infection ( $P_{TWAS} = 3.69 \times 10^{-7}$ ), allograft rejection ( $P_{TWAS} = 4.45 \times 10^{-6}$ ), asthma ( $P_{TWAS} = 7.65 \times 10^{-6}$ ), type I diabetes mellitus ( $P_{TWAS} = 1.32 \times 10^{-5}$ ), inflammatory bowel disease (IBD) ( $P_{TWAS} = 7.37 \times 10^{-5}$ ), and herpes simplex infection ( $P_{TWAS} = 7.93 \times 10^{-5}$ ).

A total of 2,204 DEGs were identified by mRNA expression profiling analysis of IPF (**Supplementary Table S4**), and were then conducted GO and KEGG pathway enrichment analysis (**Supplementary Tables S5, S6**). 38 common GO terms were identified by enrichment analysis of IPF-related genes identified by TWAS and DEGs (**Supplementary Table S7**), including antigen processing and presentation of peptide or polysaccharide antigen via MHC class II (GO:0002504,  $P_{TWAS} = 1.43 \times 10^{-8}$ ,  $P_{mRNA} = 8.87 \times 10^{-3}$ ), lytic vacuole (GO:0000323,  $P_{TWAS} = 6.25 \times 10^{-6}$ ,  $P_{mRNA} = 5.91 \times 10^{-3}$ ), ncRNA metabolic process (GO:0034660,  $P_{TWAS} = 1.61 \times 10^{-5}$ ,  $P_{mRNA} = 4.43 \times 10^{-7}$ ), microtubule organizing center (GO:0005815,  $P_{TWAS} = 1.24 \times 10^{-4}$ ,  $P_{mRNA} = 2.46 \times 10^{-23}$ ), and autophagy (GO:0002504,  $P_{TWAS} = 1.10 \times 10^{-3}$ ,  $P_{mRNA} = 2.11 \times 10^{-4}$ ). **Table 2** summarizes the top 20 common GO terms detected by enrichment analysis of IPF-related genes identified by TWAS and DEGs. We also detected 8 common KEGG pathways (**Table 3**), such as staphylococcus aureus infection ( $P_{TWAS} = 3.69 \times 10^{-7}$ ,  $P_{mRNA} = 9.17 \times 10^{-3}$ ), herpes simplex infection ( $P_{TWAS} = 7.93 \times 10^{-5}$ ,  $P_{mRNA} = 8.76 \times 10^{-4}$ ), HTLV-I infection ( $P_{TWAS} = 8.84 \times 10^{-5}$ ,  $P_{mRNA} = 5.92 \times 10^{-4}$ ), phagosome ( $P_{TWAS} = 8.99 \times 10^{-5}$ ,  $P_{mRNA} = 1.93 \times 10^{-4}$ ), and systemic lupus erythematosus ( $P_{TWAS} = 2.25 \times 10^{-3}$ ,  $P_{mRNA} = 3.12 \times 10^{-3}$ ).

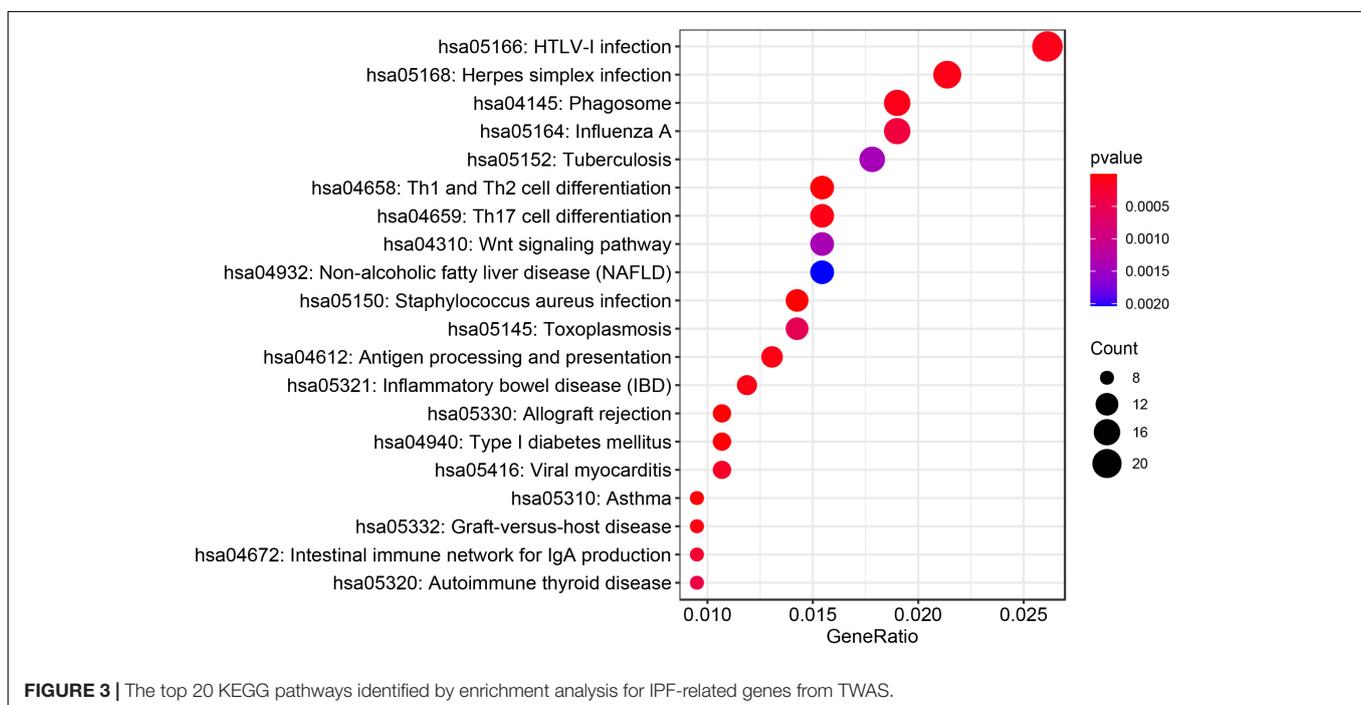
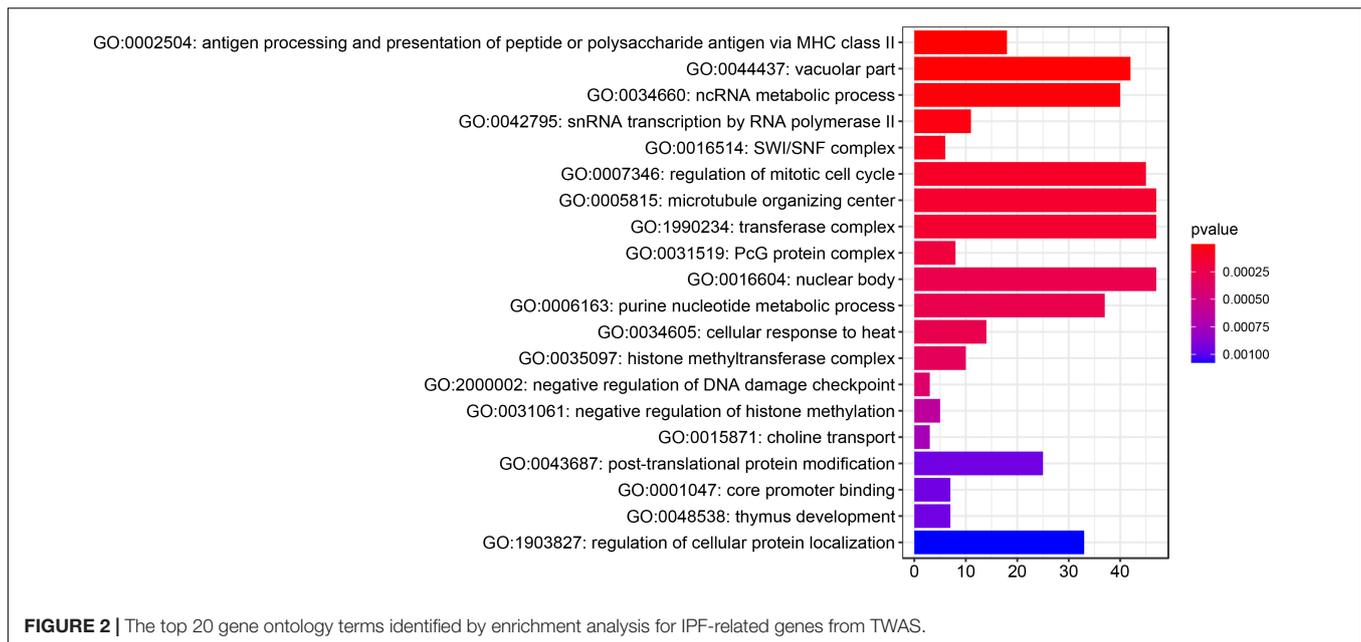
### PPI Network, Hub Gene and Module Analysis

To evaluate the association of IPF-related genes identified by TWAS, a PPI network was constructed by STRING and visualized by Cytoscape, containing 329 nodes and 893 edges (**Supplementary Figure S1**). The top 20 hub genes were identified by cytoHubba plugin that uses 12 different algorithms (**Supplementary Table S8**). Then, from the genes that can be detected by more than five algorithms, 16 hub genes with the highest degree of connectivity were selected to build the hub gene PPI network (**Figure 4A**). The enrichment analysis showed that the IPF-related processes hub genes were enriched in antigen processing and presentation of exogenous peptide antigen via MHC class II, chromosome, centromeric region, and cytoplasmic dynein complex.

In addition, module analysis conducted by MCODE plugin in Cytoscape identified several modules in the PPI network. Then, the top four significant modules were selected for subsequent analysis. A significant module, which gained the highest MCODE

**TABLE 1** | Top 30 genes identified by TWAS for IPF.

| Gene                 | Chromosome | $P_{TWAS}$             | Tissue           |
|----------------------|------------|------------------------|------------------|
| <i>DSP</i>           | 6          | $1.35 \times 10^{-29}$ | Lung             |
| <i>MUC5B</i>         | 11         | $1.09 \times 10^{-28}$ | Lung             |
| <i>TOLLIP</i>        | 11         | $1.41 \times 10^{-15}$ | Whole blood      |
| <i>DND1P1</i>        | 17         | $8.12 \times 10^{-15}$ | Lung             |
| <i>CRHR1-IT1</i>     | 17         | $9.20 \times 10^{-15}$ | Lung             |
| <i>MAPT</i>          | 17         | $9.60 \times 10^{-15}$ | Lung             |
| <i>RP11-259G18.2</i> | 17         | $1.04 \times 10^{-14}$ | Lung             |
| <i>RP11-707Q23.5</i> | 17         | $1.19 \times 10^{-14}$ | Lung             |
| <i>RP11-259G18.3</i> | 17         | $1.22 \times 10^{-14}$ | Lung             |
| <i>LRRC37A4P</i>     | 17         | $1.59 \times 10^{-14}$ | Lung             |
| <i>KANSL1-AS1</i>    | 17         | $2.75 \times 10^{-14}$ | Lung             |
| <i>RP11-259G18.1</i> | 17         | $3.81 \times 10^{-14}$ | Lung             |
| <i>KIAA1267</i>      | 17         | $2.19 \times 10^{-13}$ | Peripheral blood |
| <i>LRRC37A2</i>      | 17         | $2.91 \times 10^{-13}$ | Lung             |
| <i>WNT3</i>          | 17         | $1.21 \times 10^{-12}$ | Lung             |
| <i>DND1</i>          | 17         | $2.17 \times 10^{-12}$ | Peripheral blood |
| <i>PLEKHM1</i>       | 17         | $5.00 \times 10^{-11}$ | Peripheral blood |
| <i>FAM215B</i>       | 17         | $7.66 \times 10^{-11}$ | Lung             |
| <i>PLEKHM1</i>       | 17         | $3.08 \times 10^{-10}$ | Lung             |
| <i>RP11-158M2.5</i>  | 15         | $1.41 \times 10^{-9}$  | Lung             |
| <i>FAM13A</i>        | 4          | $2.12 \times 10^{-9}$  | Lung             |
| <i>LRRC37A</i>       | 17         | $6.40 \times 10^{-9}$  | Lung             |
| <i>DEPTOR</i>        | 8          | $8.58 \times 10^{-9}$  | Lung             |
| <i>RP11-760H22.2</i> | 8          | $1.06 \times 10^{-8}$  | Lung             |
| <i>BAHD1</i>         | 15         | $1.16 \times 10^{-8}$  | Lung             |
| <i>BRSK2</i>         | 11         | $1.49 \times 10^{-8}$  | Lung             |
| <i>RP11-798G7.5</i>  | 17         | $2.97 \times 10^{-8}$  | Lung             |
| <i>GCHFR</i>         | 15         | $1.94 \times 10^{-7}$  | Whole blood      |
| <i>RP11-64K12.8</i>  | 15         | $2.37 \times 10^{-7}$  | Lung             |
| <i>ZNF514</i>        | 2          | $4.41 \times 10^{-6}$  | Whole blood      |



score, contained 24 nodes and 150 edges. Subsequent functional enrichment analysis indicated that the genes in this module were primarily enriched in antigen processing and presentation of exogenous peptide antigen via MHC class II, chromosome, centromeric region, cytoplasmic dynein complex, and type I interferon signaling pathway (Figure 4B). Fourteen genes (*DYNLL1*, *DYNC1L1*, *DYNLL2*, *HLA-DRB5*, *HLA-DPB1*, *HLA-DQB2*, *HLA-DQA2*, *HLA-DQB1*, *HLA-DRB1*, *POLR2L*, *CENPP*, *CENPK*, *NUP133*, and *NUP107*) were simultaneously detected by both hub gene and module analysis.

## DISCUSSION

Although the biological basis of IPF has been investigated in the past years, the cellular and molecular mechanisms of IPF are very complicated and remain unclear. In the present study, we performed the first large-scale integrative analysis of TWAS and mRNA expression profiles for IPF, which successfully detected some plausible genes as well as pathways, and can potentially provide novel insights to better understand the molecular mechanisms underlying the development of IPF.

**TABLE 2** | Top 20 overlapped gene ontology terms identified by enrichment analysis for IPF-related genes from TWAS and for differentially expressed genes from mRNA expression profiling of IPF.

| ID         | Category | Description   | $P_{TWAS}$            | $P_{mRNA}$             |
|------------|----------|---|-----------------------|------------------------|
| GO:0002504 | BP       | Antigen processing and presentation of peptide or polysaccharide antigen via MHC class II | $1.43 \times 10^{-8}$ | $8.87 \times 10^{-3}$  |
| GO:0000323 | CC       | Lytic vacuole   | $6.25 \times 10^{-6}$ | $5.91 \times 10^{-3}$  |
| GO:0034660 | BP       | ncRNA metabolic process   | $1.61 \times 10^{-5}$ | $4.43 \times 10^{-7}$  |
| GO:0042795 | BP       | snRNA transcription by RNA polymerase II  | $3.03 \times 10^{-5}$ | $8.88 \times 10^{-3}$  |
| GO:0006338 | BP       | Chromatin remodeling  | $2.47 \times 10^{-4}$ | $1.37 \times 10^{-5}$  |
| GO:0007346 | BP       | Regulation of mitotic cell cycle  | $9.76 \times 10^{-5}$ | $1.43 \times 10^{-12}$ |
| GO:0005815 | CC       | Microtubule organizing center   | $1.24 \times 10^{-4}$ | $2.46 \times 10^{-23}$ |
| GO:1990234 | CC       | Transferase complex   | $1.32 \times 10^{-4}$ | $9.74 \times 10^{-6}$  |
| GO:0016604 | CC       | Nuclear body  | $2.50 \times 10^{-4}$ | $2.37 \times 10^{-16}$ |
| GO:0006163 | BP       | Purine nucleotide metabolic process   | $2.56 \times 10^{-4}$ | $2.17 \times 10^{-6}$  |
| GO:0034605 | BP       | Cellular response to heat   | $2.68 \times 10^{-4}$ | $4.12 \times 10^{-3}$  |
| GO:0044417 | BP       | Translocation of molecules into host  | $2.95 \times 10^{-3}$ | $5.04 \times 10^{-4}$  |
| GO:0043687 | BP       | Post-translational protein modification   | $9.48 \times 10^{-4}$ | $8.44 \times 10^{-6}$  |
| GO:1903827 | BP       | Regulation of cellular protein localization   | $1.08 \times 10^{-3}$ | $2.58 \times 10^{-6}$  |
| GO:0006914 | BP       | Autophagy   | $1.10 \times 10^{-3}$ | $2.11 \times 10^{-4}$  |
| GO:0098687 | CC       | Chromosomal region  | $1.31 \times 10^{-3}$ | $7.26 \times 10^{-15}$ |
| GO:0072594 | BP       | Establishment of protein localization to organelle  | $1.55 \times 10^{-3}$ | $6.00 \times 10^{-5}$  |
| GO:0008134 | MF       | Transcription factor binding  | $1.55 \times 10^{-3}$ | $4.44 \times 10^{-4}$  |
| GO:0032984 | BP       | Protein-containing complex disassembly  | $3.84 \times 10^{-3}$ | $7.07 \times 10^{-3}$  |
| GO:0000139 | CC       | Golgi membrane  | $1.87 \times 10^{-3}$ | $7.40 \times 10^{-6}$  |

Biological Processes (BP); Cellular Components (CC); Molecular Functions (MF).

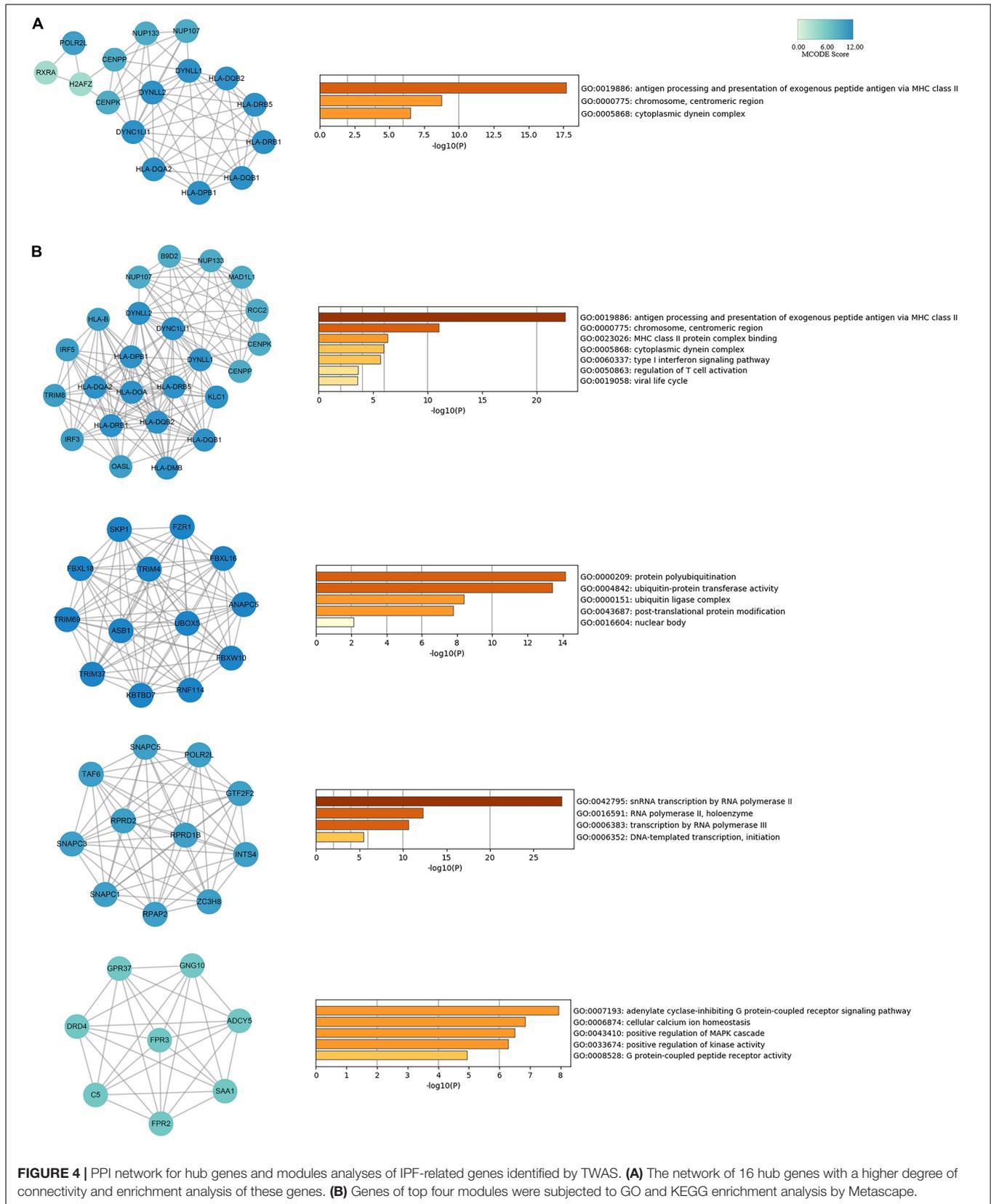
**TABLE 3** | Overlapped KEGG pathways identified by enrichment analysis for IPF-related genes from TWAS and for differentially expressed genes from mRNA expression profiling of IPF.

| ID       | Description                               | $P_{TWAS}$            | $P_{mRNA}$            |
|----------|---|-----------------------|-----------------------|
| hsa05150 | Staphylococcus aureus infection           | $3.69 \times 10^{-7}$ | $9.17 \times 10^{-3}$ |
| hsa05168 | Herpes simplex infection                  | $7.93 \times 10^{-5}$ | $8.76 \times 10^{-4}$ |
| hsa05166 | HTLV-I infection                          | $8.84 \times 10^{-5}$ | $5.92 \times 10^{-4}$ |
| hsa04145 | Phagosome                                 | $8.99 \times 10^{-5}$ | $1.93 \times 10^{-4}$ |
| hsa05164 | Influenza A                               | $3.46 \times 10^{-4}$ | $1.32 \times 10^{-3}$ |
| hsa04932 | Non-alcoholic fatty liver disease (NAFLD) | $2.04 \times 10^{-3}$ | $2.64 \times 10^{-4}$ |
| hsa05322 | Systemic lupus erythematosus              | $2.25 \times 10^{-3}$ | $3.12 \times 10^{-3}$ |
| hsa04620 | Toll-like receptor signaling pathway      | $9.97 \times 10^{-3}$ | $9.76 \times 10^{-3}$ |

Transcriptome-wide association study detected several significant IPF-related genes previously reported in GWAS, such as *DSP* and *MUC5B*. *DSP* encodes desmoplakin, which is an important component of the desmosome structure. Desmoplakin is involved in the mechanical linkage of cells, stabilization of tissue structure and the process of cell migration, proliferation, and differentiation. Accordingly, *DSP* is essential to cell-cell adhesion and epithelial barrier function (Vasioukhin et al., 2001). Previous evidence has suggested that *DSP* expression is higher in IPF lung than in the lungs of healthy control subjects and the intron 5 variant rs2076295 was found to be associated with decreased *DSP* expression, indicating that differential *DSP* expression plays an important role in IPF etiology (Mathai et al., 2016). *MUC5B* encodes for mucin 5B, which is produced by airway epithelial cells and is a major gel-forming mucin in the mucus. Mucin 5B is involved in the production of airway mucous and may have a significant role in mucociliary clearance and

airway defense (Roy et al., 2014; Evans et al., 2016). Increased *MUC5B* expression might impair mucosal defense of host and thus lead to the reduction of lung clearance of inhaled particles, dissolved chemicals, and microorganisms (Seibold et al., 2011). The *MUC5B* promoter variant rs35705950 is a common variant that accounts for a large proportion of risk for the development of familial interstitial pneumonia and IPF (Seibold et al., 2011; Todd et al., 2015; Evans et al., 2016; Zhang et al., 2019). Interestingly, a retrospective study has demonstrated improved survival of patients with this promoter variant compared with those without this variant, indicating that this variant might be a potential prognostic indicator (Peljto et al., 2013). These paradoxical findings imply that further investigation is required to clarify the biological mechanism by which this promoter variant promotes the development of IPF.

Besides, attention should be paid to genes simultaneously detected by hub gene and module analysis, such as *DYNC11L1*,



**FIGURE 4 |** PPI network for hub genes and modules analyses of IPF-related genes identified by TWAS. **(A)** The network of 16 hub genes with a higher degree of connectivity and enrichment analysis of these genes. **(B)** Genes of top four modules were subjected to GO and KEGG enrichment analysis by Metascape.

*DYNLL1*, and *DYNLL2*, which are likely to be related with IPF but not reported earlier. *DYNLL1* and *DYNLL2* are related to cell cycle spindle assembly and chromosome separation and may be involved in the change or maintenance of the spatial distribution of cytoskeletal structures (Dunsch et al., 2012). *DYNCILL1* is related to microtubule motor activity and may play a role in binding dynein to membranous organelles. These three genes belong to the cytoplasmic dynein subunit gene. Cytoplasmic dynein acts as a motor for the intracellular retrograde motility of vesicles and organelles along microtubules (Dunsch et al., 2012). Besides, human airway epithelium is characterized by the presence of ciliated cells bearing motile cilia, and specialized cell surface projections containing axonemes consisted of microtubules and dynein arms, providing ATP-driven motility (Tilley et al., 2015). In the airways, cilia function together with airway mucus, plays an important role in mediating mucociliary clearance and eliminating the inhaled particles and pathogens (Evans et al., 2016). Cilia dysfunction and clearance impairment would result in chronic airway inflammation and infection, bronchiectasis, and distal lung remodeling. Besides, the hub gene PPI network showed that the genes involved are mainly enriched in cytoplasmic dynein complex and antigen processing and presentation, suggesting the critical role of cilia function and immune response in the development of IPF.

Gene ontology and KEGG pathway enrichment analysis detected several candidate biological pathways for IPF, mainly involved in immune inflammation response and infection. For instance, antigen processing and presentation of peptide or polysaccharide antigen via MHC class II. Human leukocyte antigen (HLA), encoded by the human MHC gene complex, plays a critical role in the antigen presentation of peptides and the regulation of immune response (Bodis et al., 2018). A GWAS analysis has identified two risk alleles in the HLA region (DRB1\*15:01 and DQB1\*06:02) that were related to IPF (Fingerlin et al., 2016). Besides, several studies have suggested the role of HLA region in the development of IPF (Falfan-Valencia et al., 2005; Aquino-Galvez et al., 2009; Xue et al., 2011; Zhang et al., 2012, 2015). The association between HLA and IPF may suggest the potential etiologic role of autoimmunity in IPF. Recently, a nationwide retrospective cohort study in Korea involving 38,921 IBD patients and 116,763 patients without IBD suggested that patients with IBD, especially Crohn's disease, have an increasing risk for the development of IPF (Kim et al., 2020). In addition, a 1:1 retrospective case-control study (196 IPF cases and 196 controls) has indicated that hypothyroidism, an immune-mediated process, was common among IPF patients and was found to be associated with decreased survival time as an independent predictor of mortality in IPF patients (Oldham et al., 2015). Besides, diabetes has been reported to be a risk factor of IPF (Enomoto et al., 2003; Gribbin et al., 2009). Type 1 diabetes, also known as insulin-dependent diabetes, is an organ-specific autoimmune disease. Gene variants in the HLA region have been found to be related to the susceptibility of type 1 diabetes mellitus (Nejentsev et al., 2007). Naturally, these autoimmune diseases may share genetic basis contributed by genetic variations in HLA region. Since, observational associations are prone to reverse causality and confounding, further investigation is warranted

to characterize the pathophysiologic link between this genetic variation and disease.

Another primary candidate biological pathway was shown to be infections (both viral and bacterial), which is also closely related to the antigen stimulation and immune response, such as herpes simplex and *Staphylococcus aureus* infection. Previous studies have demonstrated that virus may be involved in disease initiation. And the presence of herpes viral DNA and epithelial cell stress in the lungs of asymptomatic relatives are at risk for the development of familial IPF (Moore and Moore, 2015). A recent meta-analysis of 20 case-control studies with 1,287 participants (634 IPF cases and 653 controls) has reported that the existence of persistent or chronic viral infections significantly associate with the increasing risk of the development of IPF, but not with the aggravation of IPF (Sheng et al., 2020). Previous studies in mice models have reported that viral infection could promote the formation of lung fibrosis (Mora et al., 2006, 2007; Qiao et al., 2009). Especially, animal experiments have been applied to provide evidence of the pathogenesis of lung fibrosis regulated by gamma herpesvirus (Moore and Moore, 2015). These animal experiments have also indicated that previous infections seem to make lung epithelial cells reprogrammed during the incubation period, producing profibrotic factors, leading to the enhanced susceptibility to subsequent fibrosis damage in lung. Nevertheless, infections in susceptible hosts or the exacerbation of existing fibrosis involve active viral replication and are affected by antiviral therapy (Moore and Moore, 2015). In addition, activated leukocyte signals in IPF patients provide further support for infectious processes driving the progression of IPF. Studies have also reported that bacterial infections play a role in the progression and prognosis of IPF. A microbiome analysis of IPF bronchoscopic alveolar lavage (BAL) samples suggested that the increase in relative abundance of two operational taxonomic units (*Streptococcus* OTU1345 and *Staphylococcus* OTU1348) was positively correlated with the progression of IPF (Han et al., 2014). In another study, reduced diversity of the lung microbiome has been found to be associated with low forced vital capacity and early mortality in patients with IPF, and a mouse model demonstrated that bleomycin-induced lung fibrosis led to a decrease in the diversity and modification of microbiota (Takahashi et al., 2018). In addition, compared with the control group, bacterial load in BAL of IPF patients has been shown to be greater. The rate of decline in lung function and the mortality risk can be partly predicted by the baseline bacterial load (Molyneaux et al., 2014). *Haemophilus*, *Streptococcus*, *Neisseria*, and *Veillonella* have found to be more abundant in cases in comparison with controls. However, animal modeling implicated that infection of *Pseudomonas aeruginosa* did not aggravate bleomycin-induced fibrosis (Ashley et al., 2014), suggesting that there might be some microbial specificity in the progression of lung fibrosis or the bleomycin-induced mouse model cannot accurately reflect the alterations of the IPF disease course induced by bacterial infection in humans. To summarize, both viral and bacterial infections may play a crucial role in the progression of IPF and may be potential predictors of disease prognosis. Additional work will be warranted to investigate the biological mechanism

of infections in the progression of IPF and further explore the potential benefit of antiviral and antimicrobial therapy.

There are several limitations in our study. First, the number of genes that can be accurately imputed in the TWAS analysis is limited by the training cohort sample size, the majority of subjects were of European ancestry, and the results cannot be directly generalized to other ethnic population. Second, there may be tissue bias using lung, peripheral blood and whole blood expression reference panels. Cell-type heterogeneity within or between tissues and cross-tissue pleiotropy may introduce tissue bias. Further investigation can be implemented to address this issue if reference panels for individual cell types or states are available (Wainberg et al., 2019). Third, TWAS significant genes cannot guarantee causality, since co-regulation may lead non-causal hits. Some fine-mapping methods such as FOCUS (fine-mapping of causal gene sets) may partly address this issue, due to its ability to directly model predicted expression correlations and use them to assign genes posterior probabilities of causality (Mancuso et al., 2019). FOCUS, as a post-TWAS analysis method, can be applied on top of the genes identified by TWAS to further reduce false discoveries.

## CONCLUSION

In conclusion, we conducted a large-scale integrative analysis of TWAS and mRNA expression profiles for IPF. Our results provide novel insights into a better understanding of the genetic mechanism of IPF. Further functional biology studies are warranted to validate our findings and clarify the potential roles of identified genes and pathways in the development of IPF.

## DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: Full summary statistics for the genome-wide meta-analysis of IPF can be accessed from <https://github.com/genomicsITER/PFgenetics>. We would like to thank the Collaborative Group of genetic studies of IPF for providing us

## REFERENCES

- Albert, F. W., and Kruglyak, L. (2015). The role of regulatory variation in complex traits and disease. *Nat. Rev. Genet.* 16, 197–212. doi: 10.1038/nrg3891
- Allen, R. J., Guillen-Guio, B., Oldham, J. M., Ma, S. F., Dressen, A., Paynton, M. L., et al. (2020). Genome-wide association study of susceptibility to idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* 201, 564–574. doi: 10.1164/rccm.201905-1017OC
- Allen, R. J., Porte, J., Braybrooke, R., Flores, C., Fingerlin, T. E., Oldham, J. M., et al. (2017). Genetic variants associated with susceptibility to idiopathic pulmonary fibrosis in people of European ancestry: a genome-wide association study. *Lancet Respir. Med.* 5, 869–880. doi: 10.1016/S2213-2600(17)30387-9
- Aquino-Galvez, A., Perez-Rodriguez, M., Camarena, A., Falfan-Valencia, R., Ruiz, V., Montano, M., et al. (2009). MICA polymorphisms and decreased expression of the MICA receptor NKG2D contribute to idiopathic pulmonary fibrosis susceptibility. *Hum. Genet.* 125, 639–648. doi: 10.1007/s00439-009-0666-1
- Armanios, M. Y., Chen, J. J., Cogan, J. D., Alder, J. K., Ingersoll, R. G., Markin, C., et al. (2007). Telomerase mutations in families with idiopathic pulmonary fibrosis. *N. Engl. J. Med.* 356, 1317–1326. doi: 10.1056/NEJMoa066157

with the IPF GWAS summary data. The gene expression profile dataset of IPF analyzed for this study can be found in the GEO with accession number GSE110147 (<https://www.ncbi.nlm.nih.gov/geo/>).

## AUTHOR CONTRIBUTIONS

ZY conceived and designed the study. WG and PG performed the statistical analysis. PG wrote the manuscript. LL, QG, and ZY provided feasible advice on data analysis and drafting manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This work has been supported by grants from National Natural Science Foundation of China (81673272 and 81872712), the Natural Science Foundation of Shandong Province (ZR2019ZD02), the Key Technology Research and Development Program of Shandong Province (2017CXGC1214), and the Young Scholars Program of Shandong University (2016WLJH23).

## ACKNOWLEDGMENTS

We would like to thank Collaborative Group of genetic studies of IPF for providing us with the IPF GWAS summary data and Gene Expression Omnibus database for providing the IPF gene expression profile data.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.604324/full#supplementary-material>

- Ashley, S. L., Jegal, Y., Moore, T. A., van Dyk, L. F., Laouar, Y., and Moore, B. B. (2014). gamma-Herpes virus-68, but not *Pseudomonas aeruginosa* or influenza A (H1N1), exacerbates established murine lung fibrosis. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 307, L219–L230. doi: 10.1152/ajplung.00300.2013
- Baumgartner, K. B., Samet, J. M., Coultas, D. B., Stidley, C. A., Hunt, W. C., Colby, T. V., et al. (2000). Occupational and environmental risk factors for idiopathic pulmonary fibrosis: a multicenter case-control study. Collaborating Centers. *Am. J. Epidemiol.* 152, 307–315. doi: 10.1093/aje/152.4.307
- Bedard Methot, D., Leblanc, E., and Lacasse, Y. (2019). Meta-analysis of gastroesophageal reflux disease and idiopathic pulmonary fibrosis. *Chest* 155, 33–43. doi: 10.1016/j.chest.2018.07.038
- Bodis, G., Toth, V., and Schwarting, A. (2018). Role of human leukocyte antigens (HLA) in autoimmune diseases. *Rheumatol. Ther.* 5, 5–20. doi: 10.1007/s40744-018-0100-z
- Boomsma, D. I., de Geus, E. J., Vink, J. M., Stubbe, J. H., Distel, M. A., Hottenga, J. J., et al. (2006). Netherlands twin register: from twins to twin families. *Twin. Res. Hum. Genet.* 9, 849–857. doi: 10.1375/183242706779462426
- Cecchini, M. J., Hosein, K., Howlett, C. J., Joseph, M., and Mura, M. (2018). Comprehensive gene expression profiling identifies distinct and overlapping

- transcriptional profiles in non-specific interstitial pneumonia and idiopathic pulmonary fibrosis. *Respir Res.* 19:153. doi: 10.1186/s12931-018-0857-1
- Cogan, J. D., Kropski, J. A., Zhao, M., Mitchell, D. B., Rives, L., Markin, C., et al. (2015). Rare variants in RTEL1 are associated with familial interstitial pneumonia. *Am. J. Respir. Crit. Care Med.* 191, 646–655. doi: 10.1164/rccm.201408-1510OC
- Dunsch, A. K., Hammond, D., Lloyd, J., Schermelleh, L., Gruneberg, U., and Barr, F. A. (2012). Dynein light chain 1 and a spindle-associated adaptor promote dynein asymmetry and spindle orientation. *J. Cell Biol.* 198, 1039–1054. doi: 10.1083/jcb.201202112
- Enomoto, T., Usuki, J., Azuma, A., Nakagawa, T., and Kudoh, S. (2003). Diabetes mellitus may increase risk for idiopathic pulmonary fibrosis. *Chest* 123, 2007–2011. doi: 10.1378/chest.123.6.2007
- Evans, C. M., Fingerlin, T. E., Schwarz, M. I., Lynch, D., Kurche, J., Warg, L., et al. (2016). Major histocompatibility complex and alveolar epithelial apoptosis in idiopathic pulmonary fibrosis. *Hum. Genet.* 118, 235–244. doi: 10.1007/s00439-005-0035-7
- Fingerlin, T. E., Murphy, E., Zhang, W., Peljto, A. L., Brown, K. K., Steele, M. P., et al. (2013). Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat. Genet.* 45, 613–620. doi: 10.1038/ng.2609
- Fingerlin, T. E., Zhang, W., Yang, I. V., Ainsworth, H. C., Russell, P. H., Blumhagen, R. Z., et al. (2016). Genome-wide imputation study identifies novel HLA locus for pulmonary fibrosis and potential role for auto-immunity in fibrotic idiopathic interstitial pneumonia. *BMC Genet.* 17:74. doi: 10.1186/s12863-016-0377-2
- Gamazon, E. R., Wheeler, H. E., Shah, K. P., Mozaffari, S. V., Aquino-Michaels, K., Carroll, R. J., et al. (2015). A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* 47, 1091–1098. doi: 10.1038/ng.3367
- Gribbin, J., Hubbard, R., and Smith, C. (2009). Role of diabetes mellitus and gastro-oesophageal reflux in the aetiology of idiopathic pulmonary fibrosis. *Respir. Med.* 103, 927–931. doi: 10.1016/j.rmed.2008.11.001
- GTEX Consortium, Laboratory, Data Analysis & Coordinating Center (Ldacc)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEX (eGTEX) groups, Nih Common Fund, NIH/NCI; NIH/NHGRI et al. (2017). Genetic effects on gene expression across human tissues. *Nature* 550, 204–213. doi: 10.1038/nature24277
- Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B. W., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* 48, 245–252. doi: 10.1038/ng.3506
- Han, M. K., Zhou, Y., Murray, S., Tayob, N., Noth, I., Lama, V. N., et al. (2014). Lung microbiome and disease progression in idiopathic pulmonary fibrosis: an analysis of the COMET study. *Lancet Respir. Med.* 2, 548–556. doi: 10.1016/S2213-2600(14)70069-4
- Hutchinson, J., Fogarty, A., Hubbard, R., and McKeever, T. (2015). Global incidence and mortality of idiopathic pulmonary fibrosis: a systematic review. *Eur. Respir. J.* 46, 795–806. doi: 10.1183/09031936.00185114
- Kim, J., Chun, J., Lee, C., Han, K., Choi, S., Lee, J., et al. (2020). Increased risk of idiopathic pulmonary fibrosis in inflammatory bowel disease: a nationwide study. *J. Gastroenterol. Hepatol.* 35, 249–255. doi: 10.1111/jgh.14838
- Kim, J. S., Podolanczuk, A. J., Borker, P., Kawut, S. M., Raghu, G., Kaufman, J. D., et al. (2017). Obstructive sleep apnea and subclinical interstitial lung disease in the multi-ethnic study of atherosclerosis (MESA). *Ann. Am. Thorac. Soc.* 14, 1786–1795. doi: 10.1513/AnnalsATS.201701-091OC
- Lederer, D. J., and Martinez, F. J. (2018). Idiopathic pulmonary fibrosis. *N. Engl. J. Med.* 378, 1811–1823. doi: 10.1056/NEJMr1705751
- Ley, B., Collard, H. R., and King, T. E. Jr. (2011). Clinical course and prediction of survival in idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* 183, 431–440. doi: 10.1164/rccm.201006-0894CI
- Liu, Y., Devescovi, V., Chen, S., and Nardini, C. (2013). Multilevel omic data integration in cancer cell lines: advanced annotation and emergent properties. *BMC Syst. Biol.* 7:14. doi: 10.1186/1752-0509-7-14
- Mancuso, N., Freund, M. K., Johnson, R., Shi, H., Kichaev, G., Gusev, A., et al. (2019). Probabilistic fine-mapping of transcriptome-wide association studies. *Nat. Genet.* 51, 675–682. doi: 10.1038/s41588-019-0367-1
- Mathai, S. K., Pedersen, B. S., Smith, K., Russell, P., Schwarz, M. I., Brown, K. K., et al. (2016). Desmoplakin variants are associated with idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* 193, 1151–1160. doi: 10.1164/rccm.201509-1863OC
- Molyneaux, P. L., Cox, M. J., Willis-Owen, S. A., Mallia, P., Russell, K. E., Russell, A. M., et al. (2014). The role of bacteria in the pathogenesis and progression of idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* 190, 906–913. doi: 10.1164/rccm.201403-0541OC
- Moore, B. B., and Moore, T. A. (2015). Viruses in idiopathic pulmonary fibrosis. Etiology and Exacerbation. *Ann. Am. Thorac. Soc.* 12(Suppl. 2), S186–S192. doi: 10.1513/AnnalsATS.201502-088AW
- Mora, A. L., Torres-Gonzalez, E., Rojas, M., Corredor, C., Ritzenthaler, J., Xu, J. G., et al. (2006). Activation of alveolar macrophages via the alternative pathway in herpesvirus-induced lung fibrosis. *Am. J. Respir. Cell Mol. Biol.* 35, 466–473. doi: 10.1165/rcmb.2006-0121OC
- Mora, A. L., Torres-Gonzalez, E., Rojas, M., Xu, J., Ritzenthaler, J., Speck, S. H., et al. (2007). Control of virus reactivation arrests pulmonary herpesvirus-induced fibrosis in IFN-gamma receptor-deficient mice. *Am. J. Respir. Crit. Care Med.* 175, 1139–1150. doi: 10.1164/rccm.200610-1426OC
- Mushiroda, T., Wattanapokayakit, S., Takahashi, A., Nukiwa, T., Kudoh, S., Ogura, T., et al. (2008). A genome-wide association study identifies an association of a common variant in TERT with susceptibility to idiopathic pulmonary fibrosis. *J. Med. Genet.* 45, 654–656. doi: 10.1136/jmg.2008.057356
- Nejentsev, S., Howson, J. M., Walker, N. M., Szeszko, J., Field, S. F., Stevens, H. E., et al. (2007). Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. *Nature* 450, 887–892. doi: 10.1038/nature06406
- Nogee, L. M., Dunbar, A. E. III, Wert, S. E., Askin, F., Hamvas, A., and Whitsett, J. A. (2001). A mutation in the surfactant protein C gene associated with familial interstitial lung disease. *N. Engl. J. Med.* 344, 573–579. doi: 10.1056/NEJM200102223440805
- Noth, I., Zhang, Y., Ma, S. F., Flores, C., Barber, M., Huang, Y., et al. (2013). Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: a genome-wide association study. *Lancet Respir. Med.* 1, 309–317. doi: 10.1016/S2213-2600(13)70045-6
- Oldham, J. M., Kumar, D., Lee, C., Patel, S. B., Takahashi-Manns, S., Demchuk, C., et al. (2015). Thyroid disease is prevalent and predicts survival in patients with idiopathic pulmonary fibrosis. *Chest* 148, 692–700. doi: 10.1378/chest.14-2714
- Peljto, A. L., Zhang, Y. Z., Fingerlin, T. E., Ma, S. F., Garcia, J. G. N., Richards, T. J., et al. (2013). Association between the MUC5B promoter polymorphism and survival in patients with idiopathic pulmonary fibrosis. *Jama J. Am. Med. Assoc.* 309, 2232–2239. doi: 10.1001/jama.2013.5827
- Qiao, J., Zhang, M., Bi, J., Wang, X., Deng, G., He, G., et al. (2009). Pulmonary fibrosis induced by H5N1 viral infection in mice. *Respir. Res.* 10:107. doi: 10.1186/1465-9921-10-107
- Raghu, G., Collard, H. R., Egan, J. J., Martinez, F. J., Behr, J., Brown, K. K., et al. (2011). 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. doi: 10.1164/rccm.2009-040GL
- Raitakari, O. T., Juonala, M., Rönnemaa, T., Keltikangas-Järvinen, L., Räsänen, L., Pietikäinen, M., et al. (2008). Cohort profile: the cardiovascular risk in young finns study. *Int. J. Epidemiol.* 37, 1220–1226. doi: 10.1093/ije/dym225
- Reimand, J., Isserlin, R., Voisin, V., Kucera, M., Tannus-Lopes, C., Rostamianfar, A., et al. (2019). Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat. Protoc.* 14, 482–517. doi: 10.1038/s41596-018-0103-9
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43:e47. doi: 10.1093/nar/gkv007
- Roy, M. G., Livraghi-Butrico, A., Fletcher, A. A., McElwee, M. M., Evans, S. E., Boerner, R. M., et al. (2014). Muc5b is required for airway defence. *Nature* 505, 412–416. doi: 10.1038/nature12807
- Sack, C., Vedral, S., Sheppard, L., Raghu, G., Barr, R. G., Podolanczuk, A., et al. (2017). Air pollution and subclinical interstitial lung disease: the multi-ethnic

- study of atherosclerosis (MESA) air-lung study. *Eur. Respir. J.* 50:1700559. doi: 10.1183/13993003.00559-2017
- Seibold, M. A., Wise, A. L., Speer, M. C., Steele, M. P., Brown, K. K., Loyd, J. E., et al. (2011). A common MUC5B promoter polymorphism and pulmonary fibrosis. *N. Engl. J. Med.* 364, 1503–1512.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Sheng, G., Chen, P., Wei, Y., Yue, H., Chu, J., Zhao, J., et al. (2020). viral infection increases the risk of idiopathic pulmonary fibrosis: a meta-analysis. *Chest* 157, 1175–1187. doi: 10.1016/j.chest.2019.10.032
- Stuart, B. D., Choi, J., Zaidi, S., Xing, C., Holohan, B., Chen, R., et al. (2015). Exome sequencing links mutations in PARN and RTEL1 with familial pulmonary fibrosis and telomere shortening. *Nat. Genet.* 47, 512–517. doi: 10.1038/ng.3278
- Szkarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613. doi: 10.1093/nar/gky1131
- Takahashi, Y., Saito, A., Chiba, H., Kuronuma, K., Ikeda, K., Kobayashi, T., et al. (2018). Impaired diversity of the lung microbiome predicts progression of idiopathic pulmonary fibrosis. *Respi. Res.* 19:34. doi: 10.1186/s12931-018-0736-9
- Tang, Y. W., Johnson, J. E., Browning, P. J., Cruz-Gervis, R. A., Davis, A., Graham, B. S., et al. (2003). Herpesvirus DNA is consistently detected in lungs of patients with idiopathic pulmonary fibrosis. *J. Clin. Microbiol.* 41, 2633–2640. doi: 10.1128/Jcm.41.6.2633-2640.2003
- Tilley, A. E., Walters, M. S., Shaykhiev, R., and Crystal, R. G. (2015). Cilia dysfunction in lung disease. *Annu. Rev. Physiol.* 77, 379–406. doi: 10.1146/annurev-physiol-021014-071931
- Tobin, R. W., Pope, C. E. II, Pellegrini, C. A., Emond, M. J., Sillery, J., and Raghu, G. (1998). Increased prevalence of gastroesophageal reflux in patients with idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* 158, 1804–1808. doi: 10.1164/ajrccm.158.6.9804105
- Todd, N. W., Atamas, S. P., Luzina, I. G., and Galvin, J. R. (2015). Permanent alveolar collapse is the predominant mechanism in idiopathic pulmonary fibrosis. *Expert Rev. Respir. Med.* 9, 411–418. doi: 10.1586/17476348.2015.1067609
- Vasioukhin, V., Bowers, E., Bauer, C., Degenstein, L., and Fuchs, E. (2001). Desmoplakin is essential in epidermal sheet formation. *Nat. Cell Biol.* 3, 1076–1085. doi: 10.1038/ncb1201-1076
- Wainberg, M., Sinnott-Armstrong, N., Mancuso, N., Barbeira, A. N., Knowles, D. A., Golan, D., et al. (2019). Opportunities and challenges for transcriptome-wide association studies. *Nat. Genet.* 51, 592–599. doi: 10.1038/s41588-019-0385-z
- Wright, F. A., Sullivan, P. F., Brooks, A. I., Zou, F., Sun, W., Xia, K., et al. (2014). Heritability and genomics of gene expression in peripheral blood. *Nat. Genet.* 46, 430–437. doi: 10.1038/ng.2951
- Xue, J., Gochuico, B. R., Feghali-Bostwick, C. A., Noth, I., Nathan, S. D., Rosen, G., et al. (2011). The HLA Class II Allele DRB1\*1501 is over-represented in patients with idiopathic pulmonary fibrosis. *PLoS One* 6:e14715. doi: 10.1371/journal.pone.0014715
- Yu, X.-T., and Zeng, T. (2018). “Integrative analysis of omics big data,” in *Computational Systems Biology: Methods and Protocols*, ed. T. Huang (New York, NY: Springer New York), 109–135. doi: 10.1007/978-1-4939-7717-8\_7
- Yuan, Z., Zhu, H., Zeng, P., Yang, S., Sun, S., Yang, C., et al. (2020). Testing and controlling for horizontal pleiotropy with probabilistic Mendelian randomization in transcriptome-wide association studies. *Nat. Commun.* 11:3861. doi: 10.1038/s41467-020-17668-6
- Zhang, H. P., Zou, J., Xie, P., Gao, F., and Mu, H. J. (2015). Association of HLA and cytokine gene polymorphisms with idiopathic pulmonary fibrosis. *Kaohsiung J. Med. Sci.* 31, 613–620. doi: 10.1016/j.kjms.2015.10.007
- Zhang, J., Xu, D. J., Xu, K. F., Wu, B., Zheng, M. F., Chen, J. Y., et al. (2012). HLA-A and HLA-B gene polymorphism and idiopathic pulmonary fibrosis in a Han Chinese population. *Respir. Med.* 106, 1456–1462. doi: 10.1016/j.rmed.2012.06.015
- Zhang, Q. H., Wang, Y., Qu, D. H., Yu, J. Y., and Yang, J. L. (2019). The possible pathogenesis of idiopathic pulmonary fibrosis considering MUC5B. *Biomed. Res. Int.* 2019:9712464. doi: 10.1155/2019/9712464
- Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* 10:1523. doi: 10.1038/s41467-019-09234-6

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Gong, Guo, Liu, Guan and Yuan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.