RESEARCH ARTICLE

Chronic obstructive pulmonary disease candidate gene prioritization based on metabolic networks and functional information

Xinyan Wang[1☯], Wan Li[2☯], Yihua Zhang[2☯], Yuyan Feng[2], Xilei Zhao[2], Yuehan He[2], Jun Zhang[3]*, Lina Chen[2]*

1 Department of Respiratory, the Second Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang, China, 2 College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Heilongjiang, China, 3 Department of pharmacy, Heilongjiang Province Land Reclamation Headquarters General Hospital, Harbin, Heilongjiang, China

☯ These authors contributed equally to this work.
* chenlina@ems.hrbmu.edu.cn (LC); zhangjun13902003@163.com (JZ)

Abstract

Chronic obstructive pulmonary disease (COPD) is a multi-factor disease, in which metabolic disturbances played important roles. In this paper, functional information was integrated into a COPD-related metabolic network to assess similarity between genes. Then a gene prioritization method was applied to the COPD-related metabolic network to prioritize COPD candidate genes. The gene prioritization method was superior to ToppGene and ToppNet in both literature validation and functional enrichment analysis. Top-ranked genes prioritized from the metabolic perspective with functional information could promote the better understanding about the molecular mechanism of this disease. Top 100 genes might be potential markers for diagnostic and effective therapies.

Introduction

Chronic obstructive pulmonary disease (COPD) is the third leading cause of morbidity and mortality worldwide [1]. As a complex disease, COPD is caused by many factors, including smoking, advanced age, medications, systemic inflammation and especially metabolic disturbances [2]. For example, disturbances in glucose metabolism are more common in COPD patients than in COPD free individuals [3]. Schols found that COPD patients had an elevated energy metabolism [4]. Cathepsin S and cystatin C plasma levels were significantly higher in the COPD group than in the healthy group, and might serve as potential biomarkers for COPD [5].

Molecular changes occurring in the process of metabolism-related complex diseases could be represented in terms of metabolic networks [6], which have been used in many researches from various aspects. Shang et al. identified disease-related metabolites from a global metabolic network based on the assumption that the metabolites related to the same disease tend to be

modularized in metabolic networks. Good performance and robustness were achieved for different disease classes, especially for respiratory diseases [7]. Oberhardt and colleagues integrated gene expression data with the metabolic network in Pseudomonas aeruginosa-infected chronic cystic fibrosis lung and demonstrated how the tradeoffs between growth and other important cellular processes shifted during disease progression [8]. Integrating other information into metabolic networks could help to better reveal disease mechanisms. Blais et al. manually curated metabolic networks to capture metabolic features. Then they integrated high-throughput transcriptomics data to predict biomarker changes in response to 76 environmental and pharmaceutical compounds for hepatocytes, which were validated with literature-based evidence and new experimental data [9]. Since genes with similar functions tend to be associated with similar diseases and vice versa [10–13], further investigation into COPD-related metabolic networks integrated with functional information is needed for better understanding of its mechanism.

Thus, in this paper, a gene prioritization method was applied to a COPD-related metabolic network, in which functional similarity was used to assess similarity between genes. Candidate genes in the COPD-related metabolic network were prioritized considering disease risks transferred between genes.

Materials and methods

Data

A human metabolic network was constructed by integrating interaction relationships from multiple databases, including the Human Metabolome Database (HMDB, http://www.hmdb.ca/) [14], HumanCYC (https://humancyc.org/) [15], BioGRID (https://thebiogrid.org/) [16], Reactome (http://www.reactome.org/) [17], Edinburgh Human Metabolic Network (EHMN) [18] and Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg/) [19]. Protein IDs from these databases were converted to their corresponding gene official symbols. The integrated human metabolic network contained 5776 genes and 589199 interaction relationships between them.

29 COPD disease genes were obtained from Online Mendelian Inheritance in Man (OMIM, https://www.omim.org/) [20], the Disease Ontology (DO, http://disease-ontology.org/) [21], Phenotype-Genotype Integrator (PheGenI) (https://www.ncbi.nlm.nih.gov/gap/phegeni) [22], DISEASES (http://diseases.jensenlab.org/) [23] and Menche's research [24].

Then, a COPD-related metabolic network was built using COPD disease genes and their direct interactors from the integrated human metabolic network. The COPD-related metabolic network was comprised of 6601 interactions (edges) between 1361 genes (nodes), 10 of which were COPD disease genes, and others were candidate genes.

Gene annotation information was collected from Gene Ontology (GO, http://www.geneontology.org/) [25]. All annotation terms for human genes in three ontologies, i.e. biological processes, molecular functions and cellular components, were extracted.

Calculation of network weights

Network weights for the COPD-related metabolic network included two aspects: node (gene) weights and edge (interaction) weights.

The gene weight w_g for gene g was calculated as the fraction of GO terms annotated by g in all GO terms annotated by human genes:

$$w_g = \frac{|T_g|}{|T_{all}|}$$

where T_g represents GO terms annotated by g and T_{all} represents all GO terms annotated by human genes. $|X|$ is the number of elements in the set X.

The interaction weight $w_{(g,h)}$ was the functional similarity of two interacting genes g and h, as we defined in [26]:

$$w_{(g,h)} = \sum_{t \in T_g \cap T_h} \frac{1}{|G_t|}$$

where T_g and T_h are GO terms annotated by gene g and h, respectively. G_t is the set of genes annotated to a GO term t.

Prioritization of candidate genes

The prioritization of candidate genes was performed based on disease risk scores of each gene obtained from an iteration process considering disease risks transferred between genes:

$$D^{(i+1)} = (1 - \beta)QD^{(i)} + \beta D^{(0)}$$

where $D^{(i)}$ is the vector of risk scores of all genes at step i, $\beta \in (0,1)$ is a parameter to measure the importance between genes and interactions. After assessing the performance using $\beta = 0.1$, $0.2, \cdots, 0.9$, $\beta = 0.1$ was chosen as the optimal parameter.

Q is the disease risk transition probability matrix, which is composed of transition probabilities. The transition probability $q(g|h)$ of disease risk going from gene h to gene g was defined as

$$q(g|h) = \frac{w_{(h,g)}}{\displaystyle\sum_{l \in neighbor(h)} w_{(l,g)}}$$

where $w_{(h,g)}$ is the interaction weight between interacting genes h and g, $neighbor(h)$ is the set of genes that interact with gene h.

$D^{(0)}$ is the vector of initial disease risk scores for all genes, which was composed of scores $d_g^{(0)}$ for gene g in the COPD-related metabolic network:

$$d_g^{(0)} = \frac{w_g}{\displaystyle\sum_{m \in \text{COPD-related metabolic network}} w_m}$$

The iteration process was carried out until the difference between $D^{(i)}$ and $D^{(i+1)}$ was less than a threshold, 10^{-9}. Candidate genes were prioritized based on their corresponding risk scores.

To further examine the functional relevance between the top-ranked genes and COPD, literature validation was performed for top 100 genes of the gene prioritization in literature of PubMed (http://www.ncbi.nlm.nih.gov/pubmed). Then, functional enrichment analysis was applied for top 100 genes using the Functional Annotation Tool in the Database for Annotation, Visualization and Integrated Discovery (DAVID, http://david.abcc.ncifcrf.gov/) v6.8 [27, 28]. GO functions and KEGG pathways with corrected P value (Benjamini) less than 0.05 were significant.

Evaluation and comparison of the performance

Leave-one-out cross-validation (LOOCV) was carried out to assess the performance of the gene prioritization method. For all COPD disease genes, one gene was removed as a test gene at each time, and was added to candidate genes. The gene prioritization process was used to

prioritize all the candidate genes. This process was repeated by setting every COPD disease gene to a test gene. The receiver operating characteristic (ROC) curves were plotted and the area under the curve (AUC) was computed based on the ranks of test genes. These results were compared with those of ToppGene and ToppNet using the same disease and candidate genes as our gene prioritization method did.

ToppGene and ToppNet are two tools in the ToppGene Suite (https://toppgene.cchmc.org) [29], which is an online bioinformatics tool for prioritizing candidate genes based on comprehensive factors, including GO annotation, phenotype, signaling pathway and protein interaction, from a set of genes known to be associated with the disease of interest.

Literature validation and functional enrichment analysis were also performed for top 100 genes of ToppGene and ToppNet to compare their efficiency with our gene prioritization method.

Results

Gene prioritization

COPD candidate genes were prioritized in the COPD-related metabolic network according to their risk scores in descending order. The top-ranked genes were more likely to be related to COPD. To further illustrate their correlation with COPD, literature validation and functional enrichment analysis were applied for top 100 genes (S1 Table).

In these genes, 56% (56/100) have been validated by literature. Higher proportion of validation was achieved for higher ranked genes. For example, 66% of top 50 genes and 90% of top 10 genes were validated to be associated with COPD by literature. For the first ranked gene CYP2E1, its polymorphisms were found to be over-represented in COPD patients [30]. Protein levels of SOD1 (rank: 3) were significantly higher in both tumor and non-tumor lung specimens of COPD patients than in lung cancer patients with no COPD. This result indicated that SOD1 could participate in antioxidant defense of the lungs in COPD patients [31]. Genetic variations in enzyme-coding genes CYP2C9 (rank: 4) and CYP1B1 (rank: 8) have shown potentially risk of tobacco-related diseases, including COPD. Their corresponding enzymes metabolize polycyclic aromatic hydrocarbons found in tobacco smoke and generate disease-causing metabolites [32]. The SNP rs2682825 in the gene NOS1 (rank: 5) was revealed to be associated with qualitative COPD phenotypes [33].

Top 100 genes were significantly enriched in 143 GO functions, 86 (60.140%) of which were annotated by COPD disease genes and regarded as COPD-related functions (Some are illustrated in Fig 1). "Heme binding" is a process through which the enzyme heme oxygenase-1 catalyzed the oxidative degradation of heme to play a protective role as an antioxidant in the lung [34]. The promoter polymorphism of the gene coding for the enzyme has been shown to be associated with the severity and prognosis of COPD patients [35]. Fathy et al. found that "angiogenesis" was significantly decreased among COPD patients compared with controls after evaluating angiogenesis by counting microvessels highlighted using anti-CD34 antibody as a measure of microvascular density [36]. Busch et al. found five differentially methylated CpG probes significantly associated with COPD among African-Americans. The top differentially methylated CpG site was mapped to the gene MAML1, which affected NOTCH-dependent "angiogenesis" in lungs [37]. Recent researches indicated that Gram-negative bacteria-derived vesicles in "extracellular regions" could evoke neutrophilic pulmonary inflammation, a key pathology of COPD [38]. Levels of several damage-associated molecular patterns were also increased in lung fluids, the lung "extracellular region", of COPD patients [39].

34 KEGG pathways were significantly enriched by top 100 genes. 32 (94.118%) were COPD-related pathways that were annotated by COPD disease genes (Some are illustrated in
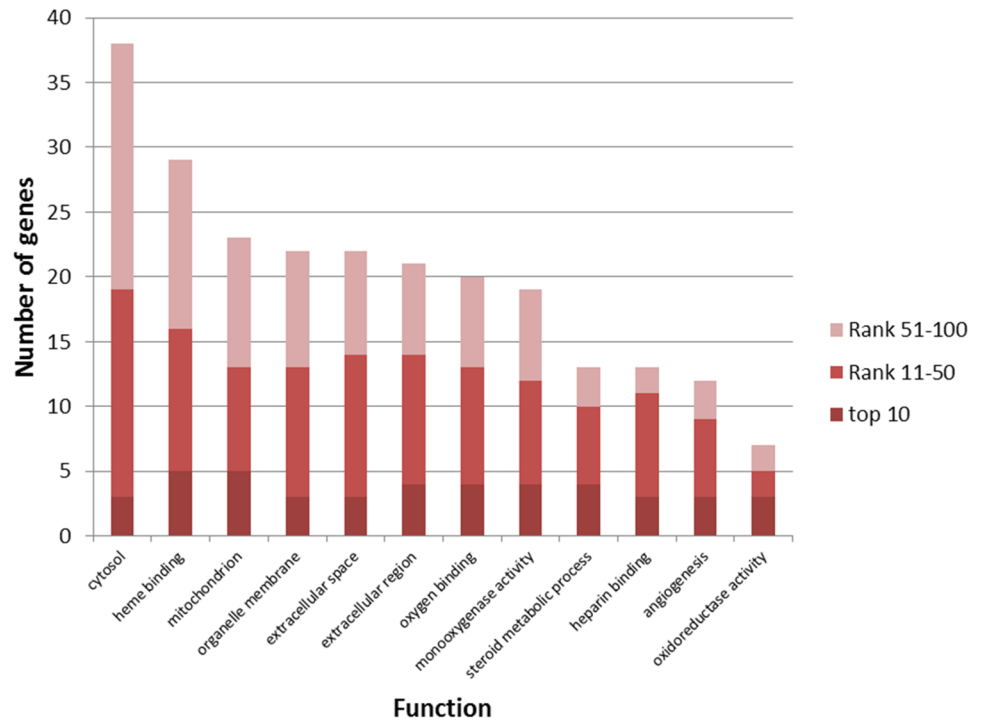
Fig 1. Some of COPD-related GO functions significantly enriched by top 100 genes. GO functions (horizontal axis) were significantly enriched by top 100 genes (the number in the vertical axis) using DAVID (Benjamini corrected P value<0.05).

https://doi.org/10.1371/journal.pone.0184299.g001

Fig 2). "Metabolic pathways" was enriched by the most genes. COPD has been linked to the dysregulation of many "metabolic pathways", such as "Steroid hormone biosynthesis". These metabolic pathways might be useful targets for novel COPD therapies [40, 41]. "Steroid hormone biosynthesis" was associated with COPD since steroid hormones are involved in lung development, pulmonary inflammation, and lung cancer. Signaling and exposure of estrogen, a group of steroid hormones, played a role in pulmonary disorders, including COPD [42]. Inflammation caused by COPD could be reduced by enhancing the anti-inflammatory effects of steroids [43]. "Metabolism of xenobiotics by Cytochrome P450" was significantly regulated by a set of genes in regulating inflammatory airway diseases, such as COPD [44]. HoffMann et al. found that "retinol metabolism" (Fig 3) was the most significantly differentially regulated pathway between pulmonary hypertension patients with COPD and idiopathic pulmonary fibrosis. They also pointed out that genes related to "retinol metabolism" might play an important role in differentiating processes involved in vascular remodeling of pulmonary hypertension caused by COPD and other lung diseases [45]. Signaling pathways were also involved in COPD. The "PI3K/Akt signaling pathway" is required for urokinase plasminogen activator receptor-mediated Epithelial-mesenchymal transition in human small airway epithelial cells, which played a crucial role in small airway fibrosis of COPD patients [46].

Functional enrichment analysis demonstrated the correlation of COPD and top 100 genes, most of which have been validated by literature. Other genes without literature validation could also be enriched in these COPD-related functions or pathways. For example, CYP51A1 (rank: 12), UGT1A7 (rank: 37) and FN1 (rank: 38) were annotated to "heme binding" function, "Metabolism of xenobiotics by cytochrome P450" pathways and "PI3K-Akt signaling pathway", respectively (Fig 4).
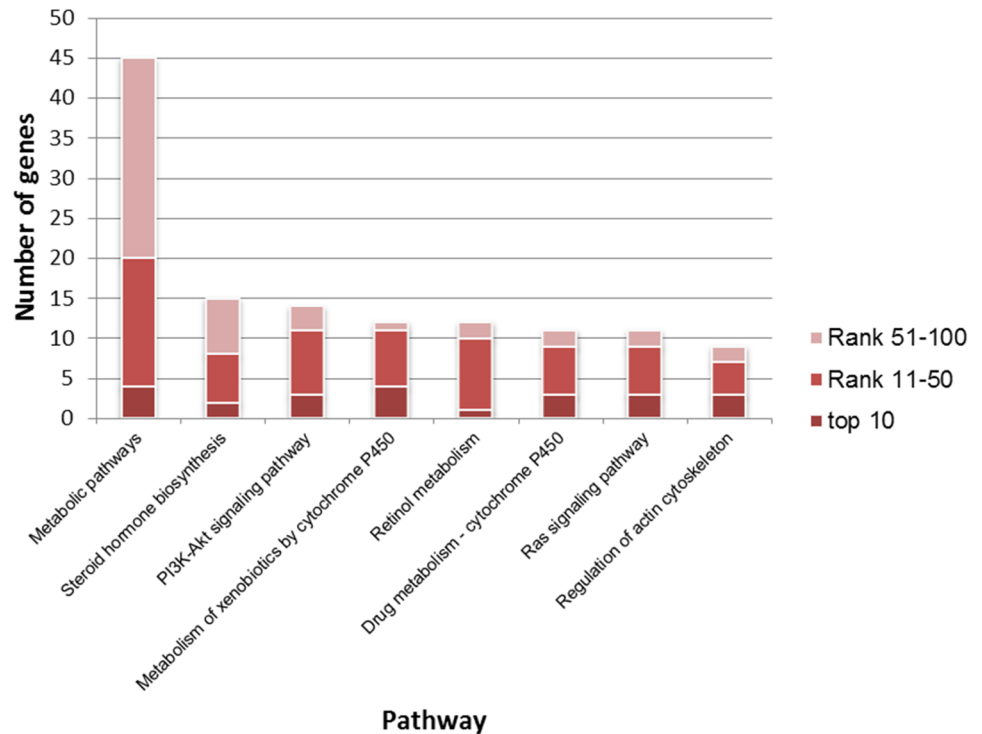
Fig 2. Some of COPD-related KEGG pathways significantly enriched by top 100 genes. KEGG pathways (horizontal axis) were significantly enriched by top 100 genes (the number in the vertical axis) using DAVID (Benjamini corrected P value<0.05).

https://doi.org/10.1371/journal.pone.0184299.g002

Performance evaluation and comparison

The gene prioritization performance was assessed using LOOCV, AUC of which was compared with that of ToppGene and ToppNet (Fig 5). It was showed that AUC of our gene prioritization method was 0.949, which was higher than that of both ToppGene (0.912) and ToppNet (0.854).

Then the performance of the three methods was compared on their literature validation (Fig 6). For ToppGene, 37% of its top 100, 42% of its top 50, and 70% of its top 10 genes were validated, while for ToppNet, 34% of top 100, 36% of top 50, and 50% of top 10 were validated to be involved in COPD. All of these proportions were less than those of our gene prioritization method (56%, 66% and 90%).

The performance of the three methods was also compared on COPD-related function or pathway proportion of top 100 genes (Table 1). Results of functional enrichment analysis showed that 100 and 75 functions, as well as 20 and 19 pathways were significantly enriched by top 100 genes of ToppGene and ToppNet, respectively. In these functions and pathways, 55% (55/100) and 56% (42/75) of functions, and 75% (15/20) and 47.368% (9/19) pathways were COPD-related. The numbers and proportions were both less than those of our gene prioritization method. That is, functions and pathways significantly enriched by top genes of our gene prioritization method were more associated with COPD, which indicated that our top genes were more likely to be related to COPD.

These results demonstrated the good performance of our gene prioritization method, which was superior to both ToppGene and ToppNet.
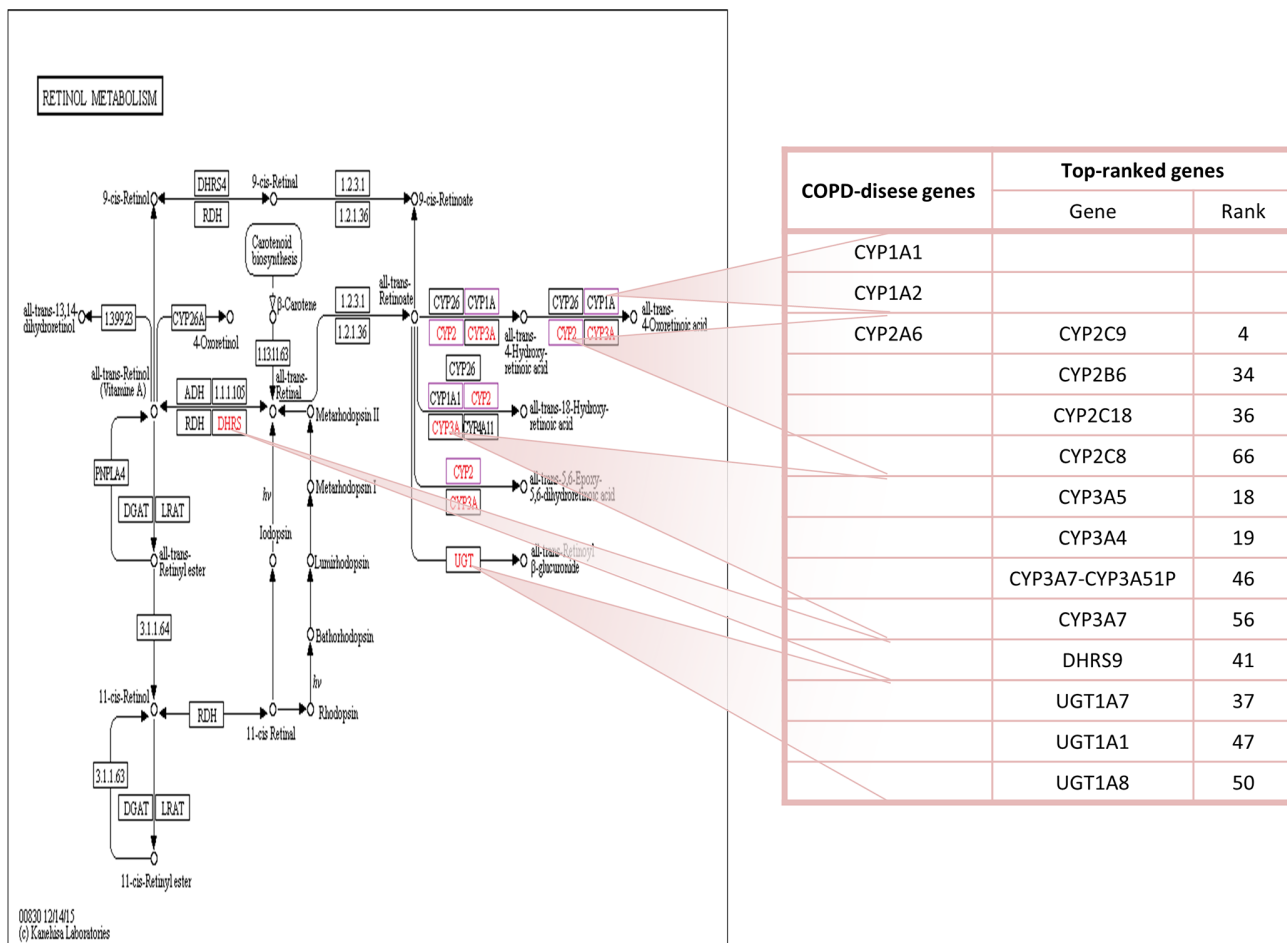
COPD-disese genes	Top-ranked genes	
	Gene	Rank
CYP1A1		
CYP1A2		
CYP2A6	CYP2C9	4
	CYP2B6	34
	CYP2C18	36
	CYP2C8	66
	CYP3A5	18
	CYP3A4	19
	CYP3A7-CYP3A51P	46
	CYP3A7	56
	DHRS9	41
	UGT1A7	37
	UGT1A1	47
	UGT1A8	50

Fig 3. The retinol metabolism pathway and its top-ranked genes. Nodes with red name contain top 100 genes, whose ranks are listed in the right table. Nodes with purple border contain COPD-disease genes.

Discussion

In this paper, a gene prioritization method was applied to a COPD-related metabolic network to prioritize COPD candidate genes according to their risk scores in descending order. Literature validation and functional enrichment analysis were assessed for top 100 genes. The performance of the gene prioritization method was better on AUC of LOOCV, literature validation and COPD-related function or pathway proportion of top 100 genes than those of ToppGene and ToppNet.

To further exhibit the performance of our gene prioritization method, a linear support vector machine classifier was applied to classify samples of an expression profile GSE57148 from Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) [47] for top 10 (the same number as COPD disease genes in the COPD-related metabolic network), top 29 (the same number as all COPD disease genes), and top 100 genes, respectively. The profile contained 98 COPD patients and 91 normal controls. To assess the performance of our top-ranked genes, the same classification process was also conducted for 10 COPD disease genes in the COPD-related metabolic network and 29 COPD disease genes from multiple databases (see Data). AUC was used to compare their classification performance. The classification performance of top 10 genes of our gene prioritization method (0.729) was slightly better than that
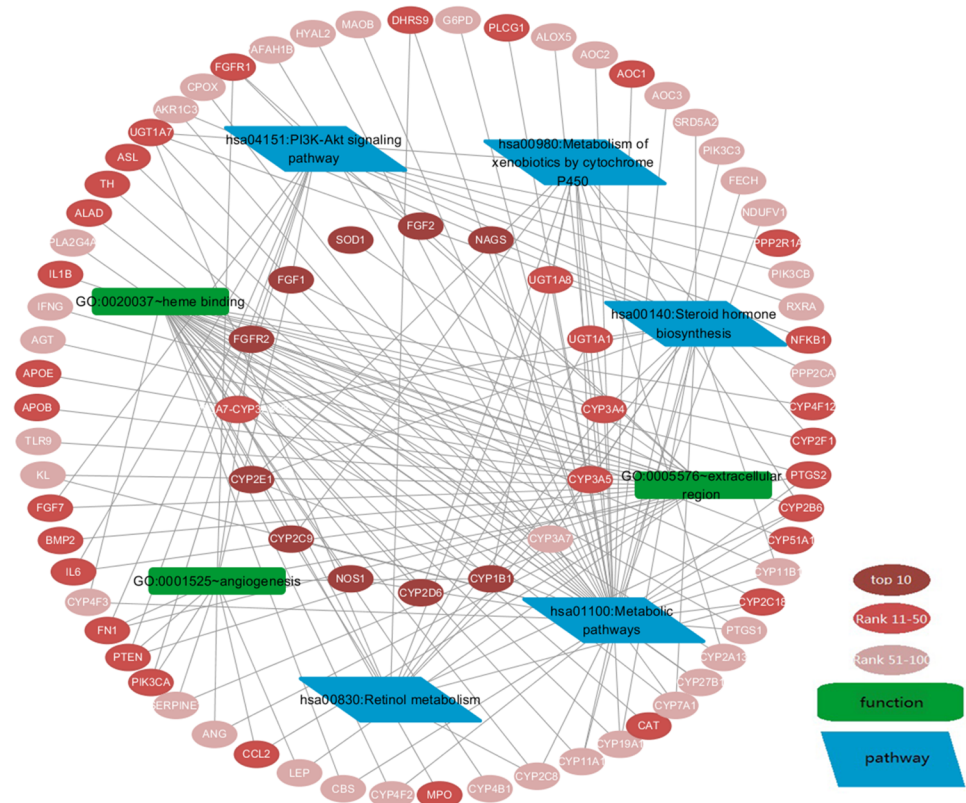
Fig 4. Top 100 genes and some of their enriched COPD-related functions or pathways. GO functions (green round rectangle) and KEGG pathways (blue parallelogram) were significantly enriched by top 100 genes (red ellipse) using DAVID (Benjamini corrected P value<0.05).

https://doi.org/10.1371/journal.pone.0184299.g004

of 10 COPD disease genes (0.725), while the classification performance of 29 COPD disease genes (0.837) was better than that of top 29 genes of our gene prioritization method (0.810). Top 100 genes could classify samples with high AUC (0.789). It was shown that, like COPD disease genes, top 100 genes could classify samples with good performance.

Top-ranked genes prioritized from the COPD-related metabolic network could be significantly enriched in some metabolic pathways, including "Metabolism of xenobiotics by cytochrome P450" and "Drug metabolism—cytochrome P450". In these pathways, cytochrome P450, essential enzymes for the metabolism of many medications, was involved. Besides, 23 genes in top 100 genes and 3 COPD disease genes (CYP1A1, CYP1A2 and CYP2A6) participate in components of cytochrome P450. Goblet cell-associated cytochrome P450 activity elevated leukotoxin-diol levels, which played a role in the clinical manifestations of COPD in a female-dominated disease sub-phenotype [48]. Plasma epoxyeicosatrienoic acids synthesized by cytochrome P450 enzymes and produced in lung epithelial cells might become dysfunctional in COPD because of the synergistic effect caused by smoking with cytochrome P450 polymorphisms [49].

Top-ranked genes should also be validated in GWAS data, since some of our disease genes were from the PheGenI, which merges NHGRI GWAS catalog data. We first searched public databases storing genes corresponding to GWAS results, including ClinVar [50], GWAS Central [51] and GWASdb [52]. Three genes were in top 200 genes of our prioritization method: FGF7 (Rank: 34), ACE (Rank: 115) and SLC6A4 (Rank: 123), all of which were validated in
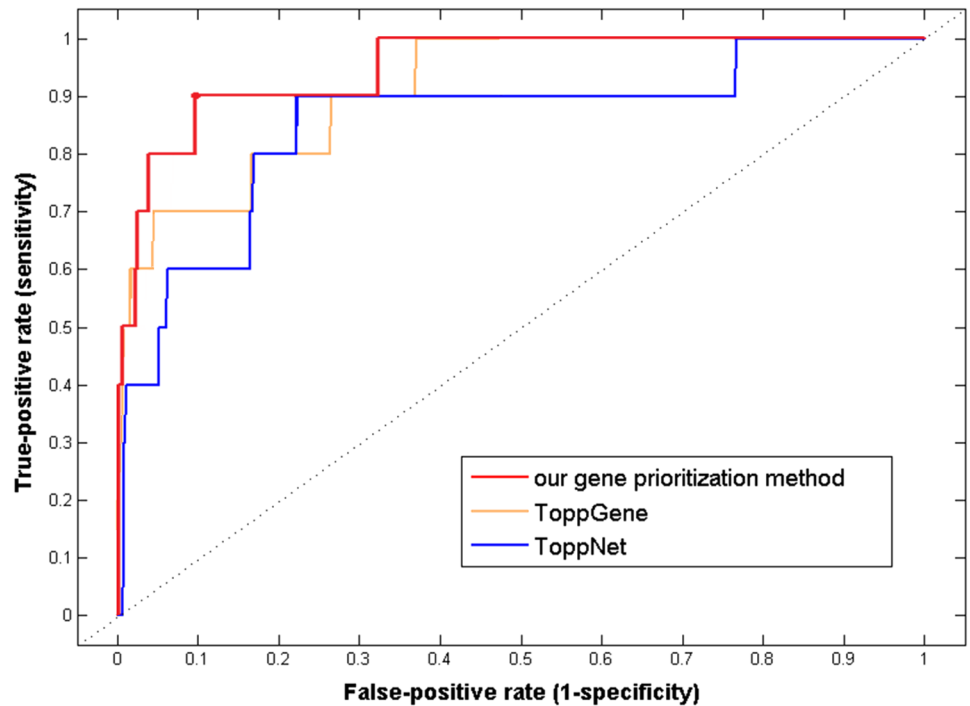
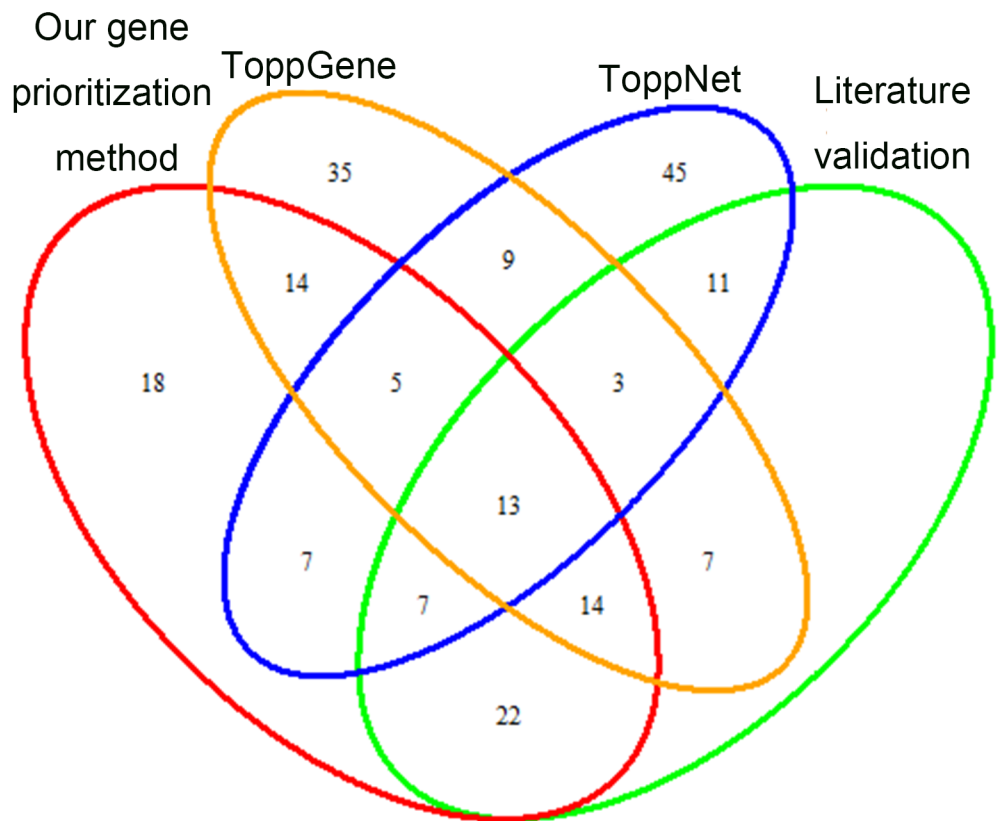**Fig 5. The ROC curves of our gene prioritization method, ToppGene and ToppNet.**

Fig 6. Overlap of top 100 genes from our gene prioritization method, ToppGene and ToppNet, and their literature validation.

Table 1. The number of significantly enriched functions and pathways, and the number and proportion of COPD-related functions and pathways in these significantly enriched functions and pathways, for top 100 genes from our gene prioritization method, ToppGene and ToppNet.

	Function		Pathway	
	Significantly enriched	COPD-related (proportion)	Significantly enriched	COPD-related (proportion)
Our gene prioritization method	143	86 (60.140%)	34	32 (94.118%)
ToppGene	100	55 (55%)	20	15 (75%)
ToppNet	75	42 (56%)	19	9 (47.368%)

https://doi.org/10.1371/journal.pone.0184299.t001

literature. Additionally, GWAS were performed for genotype data we retrieved from GSE57148, a high throughput sequencing dataset from lung tissues of COPD patients versus normal controls in GEO. Disease significantly associated SNPs and SNPs in high linkage disequilibrium with them were considered (P-value$<5\times10^{-8}$). Another three genes mapped to these SNPs were in top 200 genes of our prioritization method, i.e. PTPRJ (Rank: 104), PLCG2 (Rank: 109) and ATP2B4 (Rank: 137).

The gene prioritization method we proposed here could prioritize COPD candidate genes with a good performance. However, only 6 genes of our top 200 genes could be obtained by GWAS. This implied that our top-ranked genes could be a complement to GWAS data, and our method depended on disease genes and the disease-related network based on these genes. With the intrinsic limitation of the COPD-related metabolic network, some genes involved in COPD might be filtered out. To make up this efficiency, novel information of disease genes and the metabolic network should be added and considered comprehensively.

Conclusions

COPD candidate genes were prioritized in a COPD-related metabolic network using the gene prioritization method. The correlation of COPD and top 100 genes was validated by literature and functional enrichment analysis. The performance of the gene prioritization method was better than ToppGene and ToppNet. In summary, top-ranked genes prioritized from the metabolic perspective with functional information could promote the better understanding about the molecular mechanism of this disease. Top 100 genes might act as potential markers for diagnostic and effective therapies.

Supporting information

S1 Table. Top 100 genes of our gene prioritization method and PMIDs for their correlations with COPD.
(DOC)

Author Contributions

Conceptualization: Xinyan Wang, Lina Chen.

Data curation: Xilei Zhao, Yuehan He.

Formal analysis: Wan Li, Yihua Zhang, Yuyan Feng.

Funding acquisition: Xinyan Wang, Lina Chen.

Investigation: Yihua Zhang.

Methodology: Wan Li, Yihua Zhang, Lina Chen.

Project administration: Lina Chen.

Resources: Xilei Zhao, Yuehan He.

Software: Wan Li, Yihua Zhang, Yuyan Feng.

Supervision: Jun Zhang, Lina Chen.

Validation: Wan Li, Lina Chen.

Visualization: Wan Li, Yihua Zhang.

Writing – original draft: Wan Li.

Writing – review & editing: Xinyan Wang, Wan Li, Jun Zhang, Lina Chen.

References

1. Vestbo J, Hurd SS, Agusti AG, Jones PW, Vogelmeier C, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. American journal of respiratory and critical care medicine. 2013; 187(4):347–65. Epub 2012/08/11. https://doi.org/10.1164/rccm.201204-0596PP PMID: 22878278.

2. Breyer MK, Spruit MA, Hanson CK, Franssen FM, Vanfleteren LE, Groenen MT, et al. Prevalence of metabolic syndrome in COPD patients and its consequences. PloS one. 2014; 9(6):e98013. Epub 2014/06/21. https://doi.org/10.1371/journal.pone.0098013 PMID: 24950070.

3. Mirrakhimov AE. Chronic obstructive pulmonary disease and glucose metabolism: a bitter sweet symphony. Cardiovascular diabetology. 2012; 11:132. Epub 2012/10/30. https://doi.org/10.1186/1475-2840-11-132 PMID: 23101436.

4. Schols AM. Nutritional and metabolic modulation in chronic obstructive pulmonary disease management. The European respiratory journal Supplement. 2003; 46:81s–6s. Epub 2003/11/19. PMID: 14621110.

5. Nakajima T, Nakamura H, Owen CA, Yoshida S, Tsuduki K, Chubachi S, et al. Plasma Cathepsin S and Cathepsin S/Cystatin C Ratios Are Potential Biomarkers for COPD. Disease markers. 2016; 2016:4093870. Epub 2016/12/21. https://doi.org/10.1155/2016/4093870 PMID: 27994288 other authors declare that there are no competing interests regarding the publication of this paper. This study was supported in part by a grant to the Respiratory Failure Research Group from the Ministry of Health, Labour and Welfare of Japan and KAKENHI (14770277).

6. Diez D, Agusti A, Wheelock CE. Network analysis in the investigation of chronic respiratory diseases. From basics to application. American journal of respiratory and critical care medicine. 2014; 190 (9):981–8. Epub 2014/09/26. https://doi.org/10.1164/rccm.201403-0421PP PMID: 25254605.

7. Shang D, Li C, Yao Q, Yang H, Xu Y, Han J, et al. Prioritizing candidate disease metabolites based on global functional relationships between metabolites in the context of metabolic pathways. PloS one. 2014; 9(8):e104934. Epub 2014/08/26. https://doi.org/10.1371/journal.pone.0104934 PMID: 25153931.

8. Oberhardt MA, Goldberg JB, Hogardt M, Papin JA. Metabolic network analysis of Pseudomonas aeruginosa during chronic cystic fibrosis lung infection. Journal of bacteriology. 2010; 192(20):5534–48. Epub 2010/08/17. https://doi.org/10.1128/JB.00900-10 PMID: 20709898.

9. Blais EM, Rawls KD, Dougherty BV, Li ZI, Kolling GL, Ye P, et al. Reconciled rat and human metabolic networks for comparative toxicogenomics and biomarker predictions. Nature communications. 2017; 8:14250. Epub 2017/02/09. https://doi.org/10.1038/ncomms14250 PMID: 28176778.

10. Li P, Guo M, Wang C, Liu X, Zou Q. An overview of SNP interactions in genome-wide association studies. Briefings in functional genomics. 2015; 14(2):143–55. https://doi.org/10.1093/bfgp/elu036 PMID: 25241224.

11. Zou Q, Li J, Song L, Zeng X, Wang G. Similarity computation strategies in the microRNA-disease network: a survey. Briefings in functional genomics. 2016; 15(1):55–64. https://doi.org/10.1093/bfgp/elv024 PMID: 26134276.

12. Zeng X, Liao Y, Liu Y, Zou Q. Prediction and validation of disease genes using HeteSim Scores. IEEE/ACM transactions on computational biology and bioinformatics. 2016. https://doi.org/10.1109/TCBB.2016.2520947 PMID: 26890920.

13. Liu Y, Zeng X, He Z, Zou Q. Inferring microRNA-disease associations by random walk on a heterogeneous network with multiple data sources. IEEE/ACM transactions on computational biology and bioinformatics. 2016. https://doi.org/10.1109/TCBB.2016.2550432 PMID: 27076459.

14. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, et al. HMDB 3.0—The Human Metabolome Database in 2013. Nucleic acids research. 2013; 41(Database issue):D801–7. Epub 2012/11/20. https://doi.org/10.1093/nar/gks1065 PMID: 23161693.

15. Romero P, Wagg J, Green ML, Kaiser D, Krummenacker M, Karp PD. Computational prediction of human metabolic pathways from the complete human genome. Genome biology. 2005; 6(1):R2. Epub 2005/01/12. https://doi.org/10.1186/gb-2004-6-1-r2 PMID: 15642094.

16. Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. Nucleic acids research. 2006; 34(Database issue):D535–9. Epub 2005/12/31. https://doi.org/10.1093/nar/gkj109 PMID: 16381927.

17. Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R, et al. The Reactome pathway Knowledgebase. Nucleic acids research. 2016; 44(D1):D481–7. Epub 2015/12/15. https://doi.org/10.1093/nar/gkv1351 PMID: 26656494.

18. Ma H, Sorokin A, Mazein A, Selkov A, Selkov E, Demin O, et al. The Edinburgh human metabolic network reconstruction and its functional analysis. Molecular systems biology. 2007; 3:135. Epub 2007/09/21. https://doi.org/10.1038/msb4100177 PMID: 17882155.

19. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research. 2000; 28(1):27–30. Epub 1999/12/11. PMID: 10592173.

20. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. Nucleic acids research. 2015; 43(Database issue):D789–98. Epub 2014/11/28. https://doi.org/10.1093/nar/gku1205 PMID: 25428349.

21. Kibbe WA, Arze C, Felix V, Mitraka E, Bolton E, Fu G, et al. Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. Nucleic acids research. 2015; 43(Database issue):D1071–8. Epub 2014/10/29. https://doi.org/10.1093/nar/gku1011 PMID: 25348409.

22. Ramos EM, Hoffman D, Junkins HA, Maglott D, Phan L, Sherry ST, et al. Phenotype-Genotype Integrator (PheGenI): synthesizing genome-wide association study (GWAS) data with existing genomic resources. European journal of human genetics: EJHG. 2014; 22(1):144–7. Epub 2013/05/23. https://doi.org/10.1038/ejhg.2013.96 PMID: 23695286.

23. Pletscher-Frankild S, Palleja A, Tsafou K, Binder JX, Jensen LJ. DISEASES: text mining and data integration of disease-gene associations. Methods. 2015; 74:83–9. Epub 2014/12/09. https://doi.org/10.1016/j.ymeth.2014.11.020 PMID: 25484339.

24. Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, et al. Disease networks. Uncovering disease-disease relationships through the incomplete interactome. Science. 2015; 347(6224):1257601. Epub 2015/02/24. https://doi.org/10.1126/science.1257601 PMID: 25700523.

25. Gene Ontology Consortium: going forward. Nucleic acids research. 2015; 43(Database issue):D1049–56. Epub 2014/11/28. https://doi.org/10.1093/nar/gku1179 PMID: 25428369.

26. Jiang J, Li W, Liang B, Xie R, Chen B, Huang H, et al. A Novel Prioritization Method in Identifying Recurrent Venous Thromboembolism-Related Genes. PloS one. 2016; 11(4):e0153006. Epub 2016/04/07. https://doi.org/10.1371/journal.pone.0153006 PMID: 27050193.

27. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic acids research. 2009; 37(1):1–13. Epub 2008/11/27. https://doi.org/10.1093/nar/gkn923 PMID: 19033363.

28. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols. 2009; 4(1):44–57. Epub 2009/01/10. https://doi.org/10.1038/nprot.2008.211 PMID: 19131956.

29. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic acids research. 2009; 37(Web Server issue):W305–11. Epub 2009/05/26. https://doi.org/10.1093/nar/gkp427 PMID: 19465376.

30. Arif E, Vibhuti A, Alam P, Deepak D, Singh B, Athar M, et al. Association of CYP2E1 and NAT2 gene polymorphisms with chronic obstructive pulmonary disease. Clinica chimica acta; international journal of clinical chemistry. 2007; 382(1–2):37–42. Epub 2007/04/20. https://doi.org/10.1016/j.cca.2007.03.013 PMID: 17442289.

31. Mateu-Jimenez M, Sanchez-Font A, Rodriguez-Fuster A, Aguilo R, Pijuan L, Fermoselle C, et al. Redox Imbalance in Lung Cancer of Patients with Underlying Chronic Respiratory Conditions. Mol Med. 2016. Epub 2016/01/17. https://doi.org/10.2119/molmed.2015.00199 PMID: 26772773 Medicine, or other interests that might be perceived to influence the results and discussion reported in this paper.

32. Kaur-Knudsen D, Bojesen SE, Nordestgaard BG. Cytochrome P450 1B1 and 2C9 genotypes and risk of ischemic vascular disease, cancer, and chronic obstructive pulmonary disease. Current vascular pharmacology. 2012; 10(4):512–20. Epub 2012/01/26. PMID: 22272893.

33. Demeo DL, Campbell EJ, Barker AF, Brantly ML, Eden E, McElvaney NG, et al. IL10 polymorphisms are associated with airflow obstruction in severe alpha1-antitrypsin deficiency. American journal of

respiratory cell and molecular biology. 2008; 38(1):114–20. Epub 2007/08/11. https://doi.org/10.1165/rcmb.2007-0107OC PMID: 17690329.

34. Harada E, Sugishima M, Harada J, Fukuyama K, Sugase K. Distal regulation of heme binding of heme oxygenase-1 mediated by conformational fluctuations. Biochemistry. 2015; 54(2):340–8. Epub 2014/12/17. https://doi.org/10.1021/bi5009694 PMID: 25496210.

35. Zhang JQ, Fang LZ, Liu L, Fu WP, Dai LM. Effect of oral N-acetylcysteine on COPD patients with micro-satellite polymorphism in the heme oxygenase-1 gene promoter. Drug design, development and therapy. 2015; 9:6379–87. Epub 2015/12/18. https://doi.org/10.2147/DDDT.S91823 PMID: 26674585.

36. Fathy EM, Shafiek H, Morsi TS, El Sabaa B, Elnekidy A, Elhoffy M, et al. Image-enhanced broncho-scopic evaluation of bronchial mucosal microvasculature in COPD. International journal of chronic obstructive pulmonary disease. 2016; 11:2447–55. Epub 2016/10/13. https://doi.org/10.2147/COPD.S109788 PMID: 27729785.

37. Busch R, Qiu W, Lasky-Su J, Morrow J, Criner G, DeMeo D. Differential DNA methylation marks and gene comethylation of COPD in African-Americans with COPD exacerbations. Respiratory research. 2016; 17(1):143. Epub 2016/11/07. https://doi.org/10.1186/s12931-016-0459-8 PMID: 27814717.

38. Kim YS, Lee WH, Choi EJ, Choi JP, Heo YJ, Gho YS, et al. Extracellular vesicles derived from Gram-negative bacteria, such as Escherichia coli, induce emphysema mainly via IL-17A-mediated neutro-philic inflammation. J Immunol. 2015; 194(7):3361–8. Epub 2015/02/27. https://doi.org/10.4049/jimmunol.1402268 PMID: 25716999.

39. Pouwels SD, Heijink IH, ten Hacken NH, Vandenabeele P, Krysko DV, Nawijn MC, et al. DAMPs acti-vating innate and adaptive immune responses in COPD. Mucosal immunology. 2014; 7(2):215–26. Epub 2013/10/24. https://doi.org/10.1038/mi.2013.77 PMID: 24150257.

40. Azimzadeh Jamalkandi S, Mirzaie M, Jafari M, Mehrani H, Shariati P, Khodabandeh M. Signaling net-work of lipids as a comprehensive scaffold for omics data integration in sputum of COPD patients. Bio-chimica et biophysica acta. 2015; 1851(10):1383–93. Epub 2015/07/29. https://doi.org/10.1016/j.bbalip.2015.07.005 PMID: 26215076.

41. Yang L, Li J, Li Y, Tian Y, Li S, Jiang S, et al. Identification of Metabolites and Metabolic Pathways Related to Treatment with Bufei Yishen Formula in a Rat COPD Model Using HPLC Q-TOF/MS. Evi-dence-based complementary and alternative medicine: eCAM. 2015; 2015:956750. Epub 2015/07/15. https://doi.org/10.1155/2015/956750 PMID: 26170891.

42. Konings GF, Reynaert NL, Delvoux B, Verhamme FM, Bracke KR, Brusselle GG, et al. Increased levels of enzymes involved in local estradiol synthesis in chronic obstructive pulmonary disease. Molecular and cellular endocrinology. 2017; 443:23–31. Epub 2016/12/13. https://doi.org/10.1016/j.mce.2016.12.001 PMID: 27940297.

43. Hodge G, Hodge S. Steroid Resistant CD8+CD28null NKT-Like Pro-inflammatory Cytotoxic Cells in Chronic Obstructive Pulmonary Disease. Frontiers in immunology. 2016; 7:617. Epub 2017/01/10. https://doi.org/10.3389/fimmu.2016.00617 PMID: 28066427.

44. Fourtounis J, Wang IM, Mathieu MC, Claveau D, Loo T, Jackson AL, et al. Gene expression profiling fol-lowing NRF2 and KEAP1 siRNA knockdown in human lung fibroblasts identifies CCL11/Eotaxin-1 as a novel NRF2 regulated gene. Respiratory research. 2012; 13:92. Epub 2012/10/16. https://doi.org/10.1186/1465-9921-13-92 PMID: 23061798.

45. Hoffmann J, Wilhelm J, Marsh LM, Ghanim B, Klepetko W, Kovacs G, et al. Distinct differences in gene expression patterns in pulmonary arteries of patients with chronic obstructive pulmonary disease and idiopathic pulmonary fibrosis with pulmonary hypertension. American journal of respiratory and critical care medicine. 2014; 190(1):98–111. Epub 2014/06/12. https://doi.org/10.1164/rccm.201401-0037OC PMID: 24918967.

46. Wang Q, Wang Y, Zhang Y, Xiao W. The role of uPAR in epithelial-mesenchymal transition in small air-way epithelium of patients with chronic obstructive pulmonary disease. Respiratory research. 2013; 14:67. Epub 2013/06/29. https://doi.org/10.1186/1465-9921-14-67 PMID: 23806081.

47. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic acids research. 2013; 41(Database issue):D991–5. Epub 2012/11/30. https://doi.org/10.1093/nar/gks1193 PMID: 23193258.

48. Balgoma D, Yang M, Sjodin M, Snowden S, Karimi R, Levanen B, et al. Linoleic acid-derived lipid medi-ators increase in a female-dominated subphenotype of COPD. The European respiratory journal. 2016; 47(6):1645–56. Epub 2016/03/12. https://doi.org/10.1183/13993003.01080-2015 PMID: 26965288.

49. Yang L, Cheriyan J, Gutterman DD, Mayer RJ, Ament Z, Griffin JL, et al. Mechanisms of Vascular Dys-function in COPD and Effects of a Novel Soluble Epoxide Hydrolase Inhibitor in Smokers. Chest. 2017; 151(3):555–63. Epub 2016/11/26. https://doi.org/10.1016/j.chest.2016.10.058 PMID: 27884766.

50. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, et al. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic acids research. 2016; 44(D1):D862–8. Epub 2015/11/20. https://doi.org/10.1093/nar/gkv1222 PMID: 26582918.

51. Beck T, Hastings RK, Gollapudi S, Free RC, Brookes AJ. GWAS Central: a comprehensive resource for the comparison and interrogation of genome-wide association studies. European journal of human genetics: EJHG. 2014; 22(7):949–52. Epub 2013/12/05. https://doi.org/10.1038/ejhg.2013.274 PMID: 24301061.

52. Li MJ, Liu Z, Wang P, Wong MP, Nelson MR, Kocher JP, et al. GWASdb v2: an update database for human genetic variants identified by genome-wide association studies. Nucleic acids research. 2016; 44(D1):D869–76. Epub 2015/11/29. https://doi.org/10.1093/nar/gkv1317 PMID: 26615194.