Analysis of protein-protein interaction network in chronic obstructive pulmonary disease

Y.P. Yuan[1]*, Y.H. Shi[2]* and W.C. Gu[1]

[1]Department of Respiration, Shanghai Pudong New Area People's Hospital, Shanghai, China
[2]Department of Geriatrics, Gongli Hospital, Shanghai, China

*These authors contributed equally to this study.
Corresponding author: W.C. Gu
E-mail: wenchaogu@hotmail.com

ABSTRACT. Chronic obstructive pulmonary disease (COPD) is a growing cause of morbidity and mortality throughout the world. The purpose of our study was to uncover biomarkers and explore its pathogenic mechanisms at the molecular level. The gene expression profiles of COPD samples and normal controls were downloaded from Gene Expression Omnibus. Matlab was used for data preprocessing and SAM4.0 was applied to determine the differentially expressed genes (DEGs). Furthermore, a protein-protein interaction (PPI) network was constructed by mapping the DEGs into PPI data, and functional analysis of the network was conducted with BiNGO. A total of 348 DEGs and 765 interactive genes were identified. The hub genes were mainly involved in metabolic processes and ribosome biogenesis. Several genes related to COPD in the PPI network were found, including CAMK1D, ALB, KIT, and DDX3Y. In conclusion, CAMK1D, ALB, KIT, and DDX3Y were chosen as candidate genes, which have the potential

to be biomarkers or candidate target molecules to apply in clinical diagnosis and treatment of COPD.

Key words: Chronic obstructive pulmonary disease; Biomarker; Protein-protein interaction network

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is the occurrence of emphysema or chronic bronchitis, and it remains a major public health problem (Rabe et al., 2007). COPD is mainly caused by tobacco smoking, which induces an abnormal inflammatory response in the lung (Pauwels et al., 2001). Global prevalence of COPD has been about 3.29 million patients in 2010, accounting for 4.8% of the world population (Vos et al., 2013). Similarly, COPD is the fourth leading cause of death in the USA. Recent research showed that 59 people were diagnosed with COPD in a population of 51,804 in England (Simpson et al., 2010). Yet, prevention and diagnosis of COPD remain relatively unverified by the public as well as public health organizations (Roberts et al., 2011).

In recent years, COPD has been extensively explored from cellular and molecular perspectives (Barnes et al., 2003). Although several molecular mechanisms associated with this disease have been proposed (Di Stefano et al., 2004), the pathogenic mechanisms of COPD have not been fully investigated. Notably, the factual need to identify new biomarkers in COPD is highlighted (Cazzola and Novelli, 2010). As the functional dysregulation of a protein-protein interaction (PPI) is the underlying cause of COPD, signal pathways represent a new venue in disease treatment. Various cellular pathways have been implicated in COPD. For instance, it is reported that the nuclear factor erythroid 2-related factor 2 (NFE2L2) pathway is involved in the defense mechanism against oxidants of lung and associated with oxidant-antioxidant imbalance, which contributes to the development of COPD (Sandford et al., 2012). Other researchers showed that the Stat3 pathway also relates to COPD. Stat3 upregulation is observed in lung tissue with COPD from both smokers and non-smokers, indicating that it serves as biomarker for COPD diagnosis and prognosis in mice and humans (Qu et al., 2009).

Meanwhile, the implication of several chemokines and their receptors in COPD is supported by several studies, which are potentially used in anti-inflammatory therapies (Donnelly and Barnes, 2006). Others have revealed that emerging anti-cytokine therapies can be useful in COPD treatment (Marin et al., 2011). For instance, simvastatin, which is an anti-inflammatory drug, exhibits effectiveness in treating COPD via inhibition of the mevalonic acid cascade in alveolar macrophages.

In this study, we undertook a molecular approach to predict a PPI network in COPD. A number of essential hub genes in the COPD PPI network were identified, and the pathogenic mechanisms of COPD were analyzed. We anticipate that this study could reveal some new discoveries in the management of COPD.

MATERIAL AND METHODS

Affymetrix microarray data

We obtained the gene expression profile of GSE29133 (Fujino et al., 2012) on COPD

patients with normal controls from the GEO (Gene Expression Omnibus) database, which was based on the GPL570 Platform (Affymetrix Human Genome U133 Plus 2.0 Array). A total of 6 samples of alveolar epithelial type II were available for further analysis, including 3 samples from COPD patients and 3 samples from control groups.

PPI data

PPIs imply physical contact between proteins and are important for analyzing biological functions of proteins (De Las Rivas and Fontanillo, 2010). Moreover, they are vital for every process in living cells. PPI data were downloaded from HPRD (Human Protein Reference Database, http://www.hprd.org).

Differentially expressed gene (DEG) analysis

We used SAM4.0 (Tusher et al., 2001) to analyze the gene expression profile. By choosing "Two-class paired", we performed gene set-enrichment analysis of the DEGs. With absolute value of fold-change greater than 4 as the strict cutoff, the genes being differentially expressed between COPD and normal controls were screened out.

PPI network construction

To identify the latent meaning of a PPI relationship, the DEGs were mapped into the PPI data that had been obtained from HPRD. The human PPI network (Mathivanan et al., 2006; Amanchy et al., 2007) was then constructed. A total of 9465 proteins and 37,080 interactions were obtained for further investigation. The topology of a network is defined as the arrangement of nodes and links or interconnections between those nodes (Lee and Kalb, 2008). With network topology analysis of the PPI, we determined that our network was in line with the features of the biology network.

Functional annotation of PPI network

BiNGO (http://www.psb.ugent.be/cbd/papers/BiNGO/Home.html) is a Java-based tool to determine which Gene Ontology (GO) terms are significantly overrepresented in a set of genes (Maere et al., 2005). With BiNGO, a list of genes or interactive subgraphs of biological networks was visualized in Cytoscape.

RESULTS

Identification of DEGs between COPD and normal controls

For microarray data set GSE29133, a total of 348 DEGs with the absolute fold-change greater than 4 were selected, including 163 upregulated genes and 185 downregulated genes.

PPI network construction

The functions of DEGs in COPD samples were systemically analyzed with mapping

the DEGs to PPI data. As shown in Figure 1, the PPI network was formed, and a total of 348 DEGs and 765 interactive genes were obtained (Figure 1). From Figure 1, we could see that the DEGs of CamKI-like protein kinase (CAMK1D), DEAD box polypeptide 3 (DDX3Y), albumin (ALB), and tyrosine kinase receptors (KIT) were hub nodes of the network. Nearly 20 genes with high degrees (10 to 100) are shown in Figure 2.
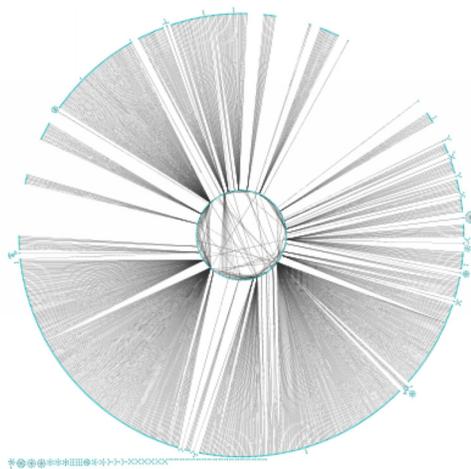

Figure 1. Networks of DEGs and interactive genes. The black regions indicate the most highly connected nodes, suggesting that CAMK1D, DDX3Y, KIT, and ALB are hub genes in the interaction network and therefore more probably play an important role in the functionality of the network.
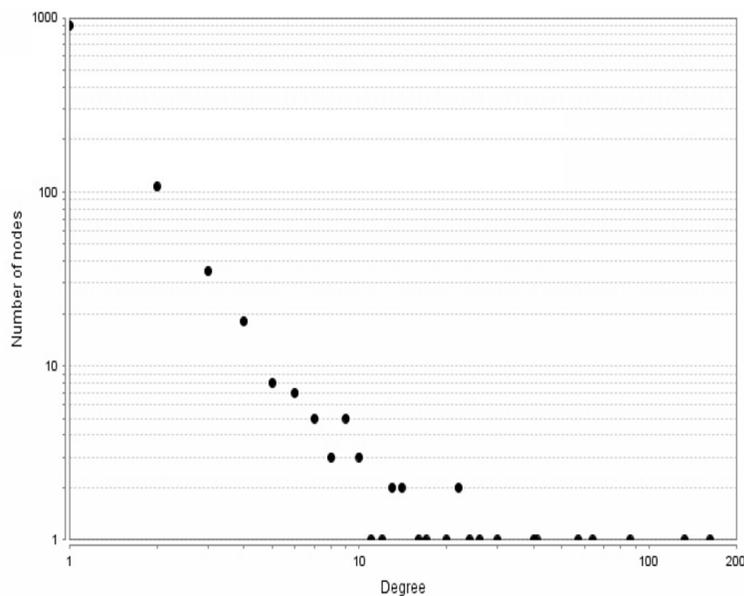
Figure 2. Degree distribution for the expression correlation and the association network. The distribution conforms to the power law distribution.

Function analysis of the PPI network

BiNGO provided functional analysis of the network, and the enrichment results are displayed in Figure 3. To determine the function of proteins, several GO categories were enriched among the genes in the PPI network, such as macromolecule metabolic process (GO: 43170), primary metabolic process (GO: 44238), cellular macromolecule metabolic process (GO: 44260), metabolic process (GO: 8152), ribosome biogenesis (GO: 42254), and cellular metabolic process (GO: 44237). A total of 46 GO terms were identified and the top 10 GO terms with a highly significant correlation are listed in Table 1.
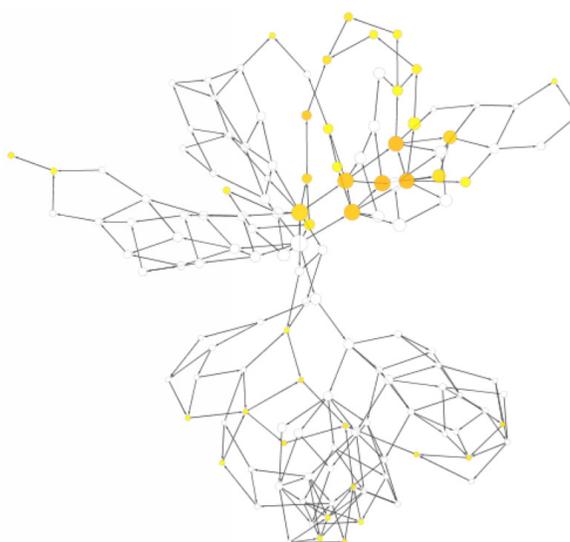
Figure 3. Network-enrichment results of BiNGO. The sub-network is colored according to the analysis of GO terms, performed with BiNGO. Nodes from the expression correlation are colored. Dark yellow nodes represent the most enriched clusters and bright yellow nodes exhibited the most enriched clusters.

Table 1. Top 10 significantly enriched GO terms in COPD.

Term	Description	FDR	Genes (partially)
GO:43170	Macromolecule metabolic process	6.63E-05	SKP1, MCM5, RPL34
GO:44238	Primary metabolic process	6.63E-05	PRL13, PRL30, PWP2, MRPL4
GO:44260	Cellular macromolecule metabolic process	6.63E-05	ATG10, LDB3, MCM5, CDC45
GO:8152	Metabolic process	1.67E-04	SAT1, MMS19, RPL13, MRPL4, SKP1, CDS1
GO:42254	Ribosome biogenesis	1.85E-04	RPL30, RPL34, TSR1, NOB1
GO:44237	Cellular metabolic process	1.85E-04	RPL13, SKP1, SPI1, MCM5
GO:22613	Ribonucleoprotein complex biogenesis	7.96E-04	PWP2, RPL30, TSR1, NOB1
GO:44267	Cellular protein metabolic process	7.96E-04	ATG10, RPL39, RPL13, RPL30
GO:9987	Cellular process	1.00E-03	LDB3, RPL13, PDX1, RTF1
GO:19538	Protein metabolic process	2.66E-03	ATG10, RPL13, MRPS16, PTK2

DISCUSSION

Network analysis can be used to extract new biological information of disease (Diez et al., 2010). In this study, we investigated the molecular mechanism of COPD in the field

of bioinformatics. We found that the expression of 348 genes was altered in COPD samples, including 163 upregulated genes and 185 downregulated genes. The DEGs were mainly involved in several metabolic processes and ribosome biogenesis. In addition, a PPI network was constructed. We identified the PPI network, which comprised 348 DEGs and 765 interactive genes, and the hub genes involved in regulation, signal transduction and biosynthesis of cellular components.

To explore the pathogenic mechanisms of COPD in the PPI network, we used a bioinformatic method in disease treatment (Zhang et al., 2012). As expected, several significantly enriched GO terms in COPD were discovered. The GO-enrichment analysis in our study suggested that the DEGs were involved in metabolic processes, ribosome biogenesis and biosynthesis of cellular components.

Hub genes are highly related to disease diagnosis and treatment (Chang et al., 2009). Four hub genes, CAMK1D, ALB, DDX3Y, and KIT, were identified by PPI network analysis. The interactions of hub genes with other genes related to COPD may influence the occurrence of disease. CAMK1D is enzymatically active and may play a role in the chemokine signal transduction pathway that regulates granulocyte function (Channels et al., 2006). In our study, CAMK1D was upregulated in COPD patients compared with healthy controls (fold-change = 7.06). The interactive genes of CAMK1D are SKP1, MCM5 and RPL34, which are involved in metabolic processes. Ludwig et al. (2012) took samples from patients with COPD and other diseases, and found that CAMK1D was a significant in-frame clone for distinguishing cancer patients' sera from control sera.

The ALBs are a family of globular proteins, and the main protein of human plasma is ALB (Farrugia, 2010). In our study, ALB was downregulated in COPD patients compared with healthy controls (fold-change = 0.22). The interactive genes of ALB are LDB3 and ATG10, which are involved in metabolic and cellular processes. A report showed that when considering ALB in the prediction of survival status, the survival of patients hospitalized with acute exacerbations of COPD may be enhanced (Gunen et al., 2005).

DEAD box proteins are implicated in a number of cellular processes, and DDX3Y is reported to have various cellular functions (Lee et al., 2008). To our knowledge, there are scarce studies reporting DDX3Y expression in COPD. In our study, DDX3Y was upregulated in COPD patients compared with healthy controls (fold-change = 16.56). The interactive genes of DDX3Y are RPL30, RPL13 and PWP2, which are involved in metabolic process and ribosome biogenesis.

KIT has a critical role in the regulation of normal cellular processes and the progression of many types of human cancer (Zwick et al., 2001). In our study, KIT was upregulated in COPD patients compared with healthy controls (fold-change = 5.44). Other studies have revealed that KIT is expressed at a high percentage in lung cancer, and that the KIT receptor acts as a key molecular target for the treatment of SCLC (Blackhall et al., 2003).

The most significant advantage of this study was in analyzing pathogenic mechanisms of COPD at the molecular level. The hub genes can be candidate genes in treating COPD. Moreover, the constructed PPI network has some biological implications in the management of COPD, and is related to rehabilitating COPD patients. Systematic identification of causal disease genes and biomarkers is used to measure complex disease, which may provide a deep understanding of the mechanisms underlying COPD and help develop efficient therapies (Liu et al., 2012).

CONCLUSIONS

In conclusion, we constructed a PPI network and screened out several hub genes in the network. Functional analyses revealed that four hub genes CAMK1D, ALB, KIT, and DDX3Y are associated with COPD, which may have particular potential application for exploring pathogenic mechanisms. These genes are mainly involved in metabolic processes and ribosome biogenesis. However, a large-scale experimental study is needed to confirm these findings, since our study consisted of small number of samples.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

REFERENCES

Amanchy R, Periaswamy B, Mathivanan S, Reddy R, et al. (2007). A curated compendium of phosphorylation motifs. *Nat. Biotechnol.* 25: 285-286.

Barnes PJ, Shapiro SD and Pauwels RA (2003). Chronic obstructive pulmonary disease: molecular and cellular mechanisms. *Eur. Respir. J.* 22: 672-688.

Blackhall FH, Pintilie M, Michael M, Leighl N, et al. (2003). Expression and prognostic significance of kit, protein kinase B, and mitogen-activated protein kinase in patients with small cell lung cancer. *Clin. Cancer Res.* 9: 2241-2247.

Cazzola M and Novelli G (2010). Biomarkers in COPD. *Pulm. Pharmacol. Ther.* 23: 493-500.

Chang W, Ma L, Lin L, Gu L, et al. (2009). Identification of novel hub genes associated with liver metastasis of gastric cancer. *Int. J. Cancer* 125: 2844-2853.

Channels I, Transporters V, Signalling PD and Bountra C (2006). CAMK1D: Human Ca^{2+}/calmodulin-dependent protein kinase I-delta PDB Code: 2JC6.

De Las Rivas J and Fontanillo C (2010). Protein-protein interactions essentials: key concepts to building and analyzing interactome networks. *PLoS Comput. Biol.* 6: e1000807.

Di Stefano A, Caramori G, Ricciardolo FL, Capelli A, et al. (2004). Cellular and molecular mechanisms in chronic obstructive pulmonary disease: an overview. *Clin. Exp. Allergy* 34: 1156-1167.

Diez D, Wheelock AM, Goto S, Haeggstrom JZ, et al. (2010). The use of network analyses for elucidating mechanisms in cardiovascular disease. *Mol. Biosyst.* 6: 289-304.

Donnelly LE and Barnes PJ (2006). Chemokine receptors as therapeutic targets in chronic obstructive pulmonary disease. *Trends Pharmacol. Sci.* 27: 546-553.

Farrugia A (2010). Albumin usage in clinical medicine: tradition or therapeutic? *Transfus. Med. Rev.* 24: 53-63.

Fujino N, Ota C, Takahashi T, Suzuki T, et al. (2012). Gene expression profiles of alveolar type II cells of chronic obstructive pulmonary disease: a case-control study. *BMJ Open* 2: pii: e001553.

Gunen H, Hacievliyagil SS, Kosar F, Mutlu LC, et al. (2005). Factors affecting survival of hospitalised patients with COPD. *Eur. Respir. J.* 26: 234-241.

Lee CS, Dias AP, Jedrychowski M, Patel AH, et al. (2008). Human DDX3 functions in translation and interacts with the translation initiation factor eIF3. *Nucleic Acids Res.* 36: 4708-4718.

Lee DS and Kalb JL (2008). Network Topology Analysis, SAND2008-0069. Sandia National Laboratories, Albuquerque. Accessed January 2008.

Liu X, Liu ZP, Zhao XM and Chen L (2012). Identifying disease genes and module biomarkers by differential interactions. *J. Am. Med. Inform. Assoc.* 19: 241-248.

Ludwig N, Keller A, Leidinger P, Harz C, et al. (2012). Is there a general autoantibody signature for cancer? *Eur. J. Cancer* 48: 2451-2461.

Maere S, Heymans K and Kuiper M (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21: 3448-3449.

Marin L, Colombo P, Bebawy M, Young PM, et al. (2011). Chronic obstructive pulmonary disease: patho-physiology, current methods of treatment and the potential for simvastatin in disease management. *Expert Opin. Drug Del.* 8: 1205-1220.

Mathivanan S, Periaswamy B, Gandhi TK, Kandasamy K, et al. (2006). An evaluation of human protein-protein interaction data in the public domain. *BMC Bioinformatics* 7 (Suppl 5): S19.

Pauwels RA, Buist AS, Calverley PM, Jenkins CR, et al. (2001). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop summary. *Am. J. Respir. Crit. Care Med.* 163: 1256-1276.

Qu P, Roberts J, Li Y, Albrecht M, et al. (2009). Stat3 downstream genes serve as biomarkers in human lung carcinomas and chronic obstructive pulmonary disease. *Lung Cancer* 63: 341-347.

Rabe KF, Hurd S, Anzueto A, Barnes PJ, et al. (2007). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am. J. Respir. Crit. Care Med.* 176: 532-555.

Roberts M, Mapel D, Petersen H, Blanchette C, et al. (2011). Comparative effectiveness of budesonide/formoterol and fluticasone/salmeterol for COPD management. *J. Med. Econ.* 14: 769-776.

Sandford AJ, Malhotra D, Boezen HM, Siedlinski M, et al. (2012). NFE2L2 pathway polymorphisms and lung function decline in chronic obstructive pulmonary disease. *Physiol. Genomics* 44: 754-763.

Simpson CR, Hippisley-Cox J and Sheikh A (2010). Trends in the epidemiology of chronic obstructive pulmonary disease in England: a national study of 51 804 patients. *Br. J. Gen. Pract.* 60: 277-284.

Tusher VG, Tibshirani R and Chu G (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U. S. A.* 98: 5116-5121.

Vos T, Flaxman AD, Naghavi M, Lozano R, et al. (2012). Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 380: 2163-2196.

Zhang M, Su S, Bhatnagar RK, Hassett DJ, et al. (2012). Prediction and analysis of the protein interactome in *Pseudomonas aeruginosa* to enable network-based drug target selection. *PLoS One* 7: e41202.

Zwick E, Bange J and Ullrich A (2001). Receptor tyrosine kinase signalling as a target for cancer intervention strategies. *Endocr. Relat. Cancer* 8: 161-173.