European Review for Medical and Pharmacological Sciences 2014; 18: 3557-3563

Bioinformatics analysis of molecular mechanisms of chronic obstructive pulmonary disease

W. CHEN, Y.-Q. HONG, Z.-L. MENG

Department of Respiratory Medicine, Huai'an First People's hospital, Nanjing Medical University, Huaian, China

Abstract. – OBJECTIVE: This study was designed to explore the molecular mechanisms of Chronic Obstructive Pulmonary Disease (COPD) with DNA Microarray.

MATERIALS AND METHODS: The gene expression profile GSE475 was downloaded from Gene Expression Omnibus (GEO) database. There were 7 tissue samples of human diaphragm muscle available, including 4 normal samples and 3 samples from COPD patients. The differentially expressed genes (DEGs) were identified by LIMMA package in R language and were further analyzed using bioinformatics methods. Firstly, DEGs were classified into different COG clusters by BLAST. Then, the protein-protein interaction (PPI) network was constructed by STRING and pathways of DEGs were analyzed by FuncAssociate. Finally, the DEGs enriched diseases were obtained by EASE.

RESULTS: We selected 524 DEGs including 118 down-regulated DEGs and 406 up-regulated DEGs. The most significant pathway was JAK/STAT signaling pathway and the DEGs of IL6 and SOCS3 were directly participated in this pathway. Furthermore, the DEGs of SOCS3, IL4, IL18R1, IL1R1, and IL6 were participated in the disease of pulmonary fibrosis.

CONCLUSIONS: Our findings suggest that IL6 and SOCS3 play important roles in COPD and have the potential to serve as therapeutic targets of COPD.

Key Words:

Chronic obstructive pulmonary disease, Differentially expressed gene, Disease enrichment analysis, Interaction network, Pathway.

Introduction

Chronic obstructive pulmonary disease (COPD) is a chronic inflammatory disorder characterized by chronic obstruction of expiratory flow affecting peripheral airways and infiltration of inflammatory cells into the airways[1]. COPD is projected to become the third commonest cause of death and rank fifth commonest cause of disability by 2020 in worldwide according to a study published by the World Bank/World Health Organization[2,3].

COPD is a complex inflammatory disease that involves multiple inflammatory mediators[4]. There are several chemotactic signals that have the potential for inflammatory cells recruitment in COPD, including monocyte chemoattractant protein-1 (MCP-1), leukotriene B4 (LTB4), interleukin-8 (IL-8) and related CXC chemokines, including GRO- (growth-related oncoprotein) and ENA-78 (epithelial neutrophil activating protein) which are increased in COPD airways[5-7]. The concentrations of IL-6, IL-8, TNF-a and LTB4 in sputum are increased during an exacerbation[8-10] and patients who have frequent exacerbations have higher levels of IL-6 and lower concentrations of SLPI (Secretory Leukocyte Protease Inhibitor), even when COPD is stable[9,11]. However, the underlying molecular mechanism of COPD is not entirely clear and need to be fully elucidated.

In order to further explore the molecular mechanisms of COPD, we downloaded the gene expression profile of GSE475 from Gene Expression Omnibus (GEO) database, including 4 normal samples and 3 samples from COPD patients. Compared with normal samples, we identified the differentially expressed genes (DEGs). The screened DEGs were further analyzed using bioinformatics methods. We constructed protein-protein interaction (PPI) networks to investigate the critical DEGs. And we identified significant pathways of DEGs. Furthermore, the enrichment analysis of DEGs related diseases was conducted. We anticipate our work will improve the understanding to the molecular mechanisms of COPD and provide novel sight for the development of therapeutic strategy.

Materials and Methods

Affymetrix Microarray Data

The gene expression profile GSE475 was downloaded from Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/). Total 7 tissue

Corresponding Author: Yongqing Hong, MD; e-mail: yongqinghongr@hotmail.com

samples of human diaphragm muscle were available for further analysis, including 4 normal samples and 3 samples from COPD patients. The platform information was GPL96 [HG-U133A] Affymetrix Human Genome U133A Array. The annotation information of all probe sets was provided by Affymetrix Company where we downloaded the raw file for further analysis.

Data Preprocessing and Analysis of DEGs

Firstly, the probe-level data in CEL files were converted into expression measures. For each sample, the expression values of all probes for a given gene were reduced to a single value by taking the average expression value. And then, the missing parts of data were imputed[12] and the complete data were standardized[13]. The Limma (Linear Models for Microarray Data) package in R language was used to identify DEGs between the normal and COPD diaphragm muscle samples[14]. The p-value < 0.05 and |logFC| >1 were used as the cut-off criteria.

Functional Enrichment Analysis of DEGs

The comparative analysis between the sequences of DEGs and the database of clusters of orthologous groups (COG, http: //www.ncbi.nlm.nih.gov/COG) of proteins was conducted by using BLAST (Basic Local Alignment Search Tool)[15,16], and the E-value<1e-04 was used as the comparing similarity threshold. Then, the functional annotations and COG classifications of DEGs were obtained.

PPI Network Construction

Many activities of life were achieved by the combination and dissociation of proteins. The signal transduction networks of physiological activities and responses to the external and internal environment of cells were formed by the PPI[17]. Further researches about PPI are the prerequisites for recognizing and understanding all kinds of life phenomena[18].

Therefore, the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) software, a database of known and predicted protein interactions, was used to find the interactive products of DEGs for constructing the PPI networks[19]. Hub proteins were identified as those interacted with most partners.

Pathway Annotation of DEGs

DEGs in PPI networks were input into the FuncAssociate software to perform pathway annotation based on the functional enrichment analysis[20]. And the false discovery rate (FDR) < 0.05 was used as the cut-off criterion.

Analysis of DEGs Related Diseases

In this study, the enrichment analysis of DEGs related diseases was conducted by the EASE (Expressing Analysis Systematic Explorer)[21] The DEGs related diseases were picked out by the Fisher algorithm in the EASE, and $p < 0.05$ was used as the cut-off criterion.

Results

Identification of DEGs

For dataset GSE475, a total of 524 DEGs were identified at $p < 0.05$ and |logFC|>1 between the 4 normal and 3 COPD diaphragm muscle samples. And there were 118 down-regulated DEGs and 406 up-regulated DEGs.

Functional Enrichment Analysis of DEGs

The GO (Gene Ontology) functional nodes classifications of DEGs were classified based on the similarity between the sequences of DEGs and the sequences at the GO nodes recorded in the GO. As shown in Table I, we obtained 17 functional classifications with various regulatory effects, such as regulation of transcription and regulation of RNA metabolic process.

PPI Networks Construction

The STRING software was used to excavate the interactive products of DEGs for constructing the PPI networks. The constructed PPI networks were shown in Figure 1. For the PPI network, IL-6 and SOCS3 (suppressor of cytokine signaling 3) were the two hub genes with most interactions.

Pathway Annotation of DEGs

As shown in Table II, we screened 7 significant pathways and the most significant pathway was the JAK/STAT (Janus activated kinase/signal transducers and activators of transcription) signaling pathway (hsa04630). Total 11 DEGs were significantly enriched in this pathway. The IL6 and SOCS3 were directly participated in this pathway, which were shown in Figure 2.

Analysis of DEGs Related Diseases

The enrichment analysis of DEGs related diseases was conducted by the EASE. As shown in

Table I. The COG list of DEGs.

Term	Count	E-value
GO:0006357~regulation of transcription from RNA polymerase II promoter	41	7.57E-10
GO:0051252~regulation of RNA metabolic process	67	4.96E-08
GO:0006355~regulation of transcription, DNA-dependent	66	4.97E-08
GO:0010033~response to organic substance	37	7.22E-08
GO:0048545~response to steroid hormone stimulus	18	1.25E-07
GO:0009719~response to endogenous stimulus	26	1.77E-07
GO:0009725~response to hormone stimulus	24	4.35E-07
GO:0051254~positive regulation of RNA metabolic process	27	1.29E-06
GO:0045893~positive regulation of transcription, DNA-dependent	26	3.60E-06
GO:0051098~regulation of binding	14	6.16E-06
GO:0007167~enzyme linked receptor protein signaling pathway	21	7.17E-06
GO:0051101~regulation of DNA binding	12	1.70E-05
GO:0016481~negative regulation of transcription	24	1.85E-05
GO:0031667~response to nutrient levels	15	2.12E-05
GO:0045944~positive regulation of transcription from RNA polymerase II promoter	21	2.34E-05
GO:0010604~positive regulation of macromolecule metabolic process	35	2.78E-05
GO:0010629~negative regulation of gene expression	25	2.78E-05

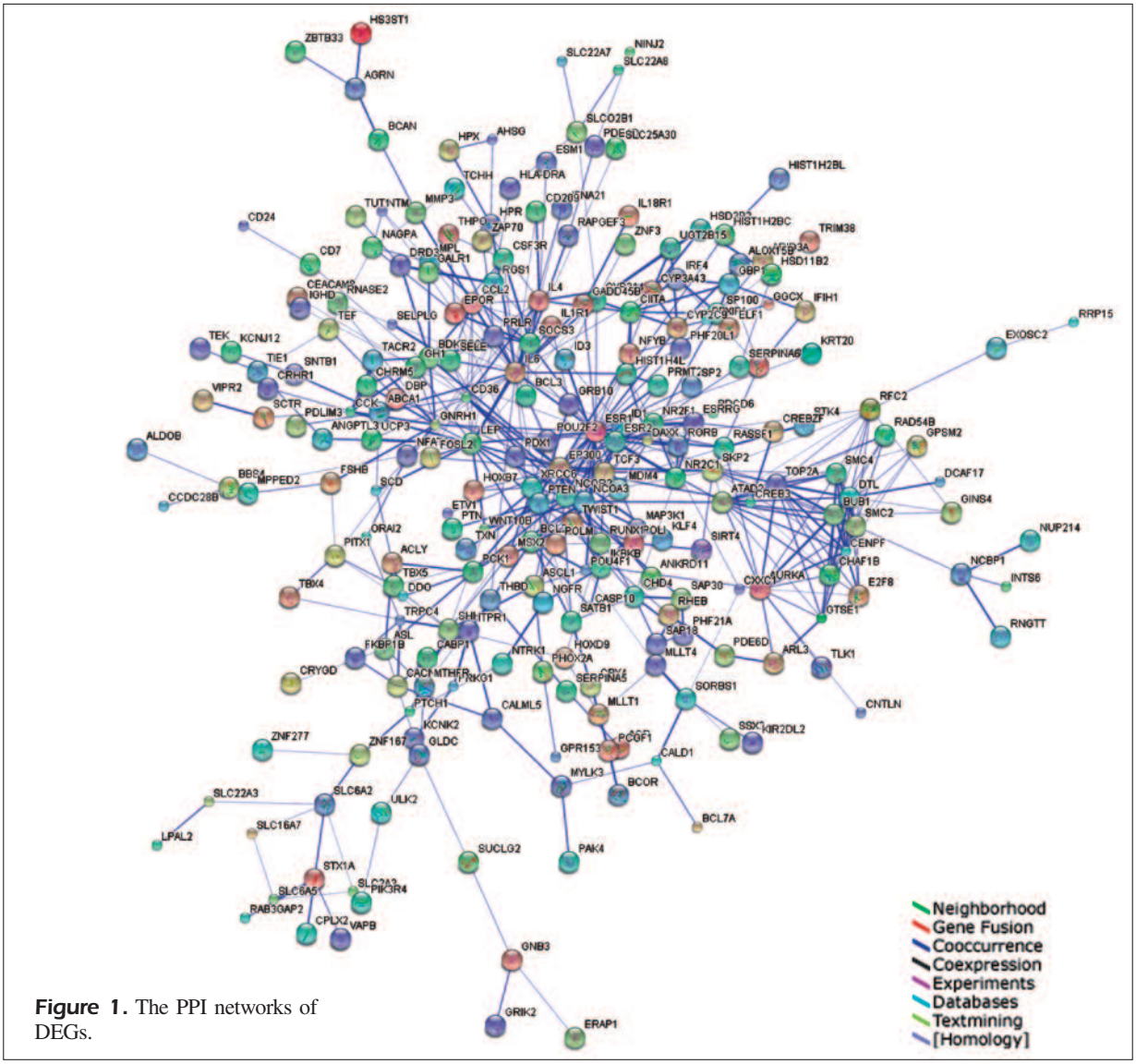

Figure 1. The PPI networks of DEGs.

Table II. Significantly expressed pathways in the interactive network.

Pathway	Count	*p* value	DEGs
hsa04630: JAK-STAT signaling pathway	11	0.004798	LEP, IL4, GH1, IFNA21, IL6, EP300, PRLR, SOCS3, CSF3R, EPOR, MPL
hsa04080: Neuroactive ligand-receptor interaction	14	0.010287	DRD3, TACR2, GRIK2, BDKRB2, VIPR2, SCTR, GRM5, CRHR1, LEP, GH1, CHRM5, GALR1, PRLR, FSHB
hsa04720: Long-term potentiation	6	0.025939	GRM5, EP300, CALML5, RAPGEF3, CACNA1C, ITPR1
hsa00140: Steroid hormone biosynthesis	5	0.026305	CYP3A43, CYP3A4, HSD3B2, HSD11B2, UGT2B15
hsa04060: Cytokine-cytokine receptor interaction	13	0.028086	IL4, IFNA21, IL18R1, IL1R1, IL6, CCL2, LEP, GH1, PRLR, CSF3R, EPOR, NGFR, MPL
hsa04622: RIG-I-like receptor signaling pathway	6	0.030567	CASP10, IFNA21, IFIH1, MAP3K1, DDX3Y, IKBKB
hsa00830: Retinol metabolism	5	0.043885	CYP3A43, CYP3A4, CYP2C9, UGT2B15, RDH5

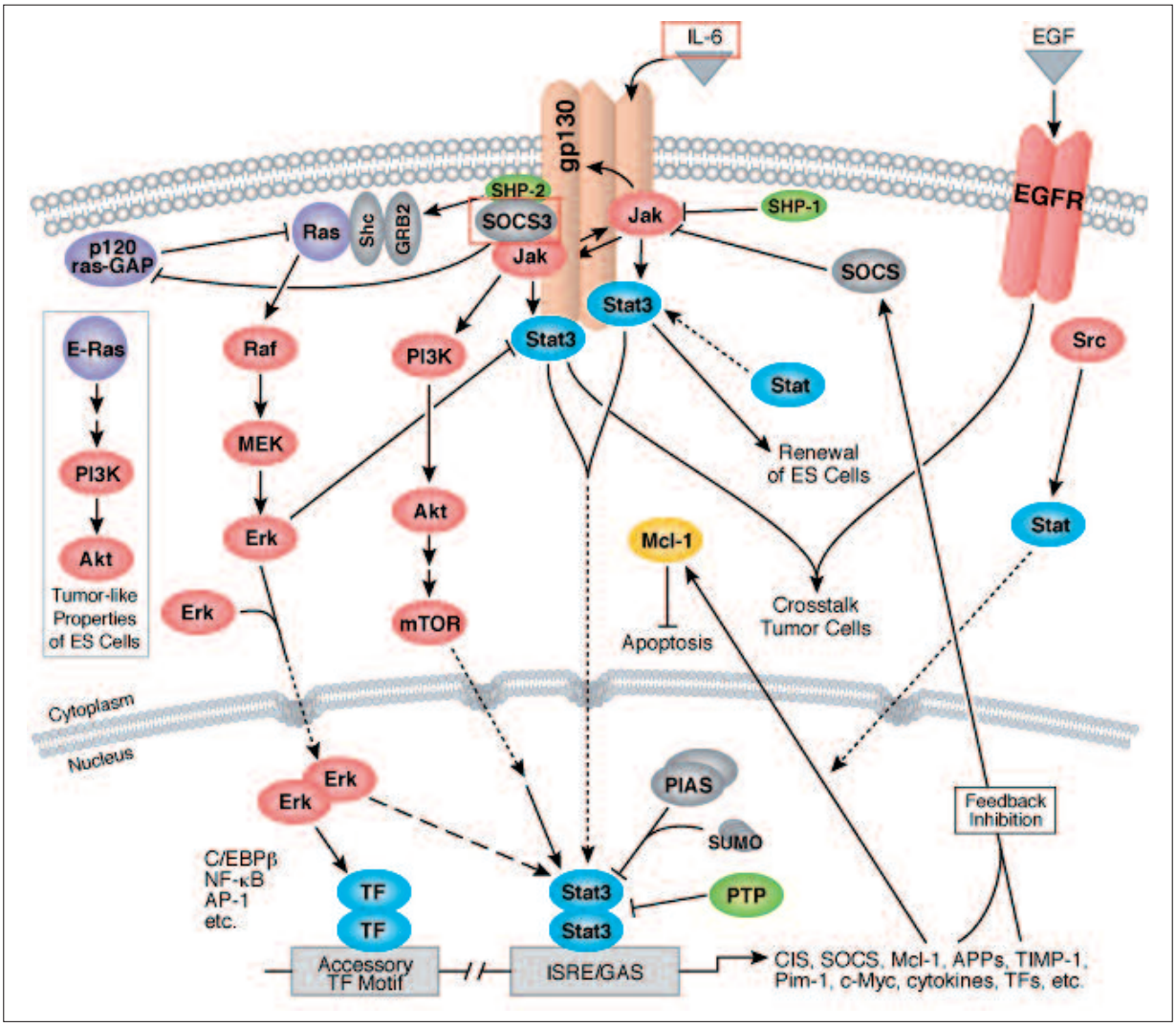

Figure 2. The JAK-STAT signaling pathway (the screened DEGs were in the red boxes).

Table III, we identified 19 relevant diseases which DEGs significantly enriched, such as breast cancer involved 23 DEGs, hypertension involved 16 DEGs and atherosclerosis involved 13 DEGs. Furthermore, the DEGs of IL4, IL18R1, IL1R1 and IL6 were participated in the pulmonary fibrosis that has close relationship with COPD.

Discussion

COPD is a severe respiratory disease caused by the abnormal inflammatory response of the lung to noxious particles or gases[22,23]. In this study, we used bioinformatics methods to investigate the molecular mechanisms of COPD. The results showed that the DEGs of IL6 and SOCS3 were directly participated in the JAK/STAT signaling pathway which is the most significant pathway. And DEGs of SOCS3, IL4, IL18R1, IL1R1, and IL6 were participated in the disease of pulmonary fibrosis.

In our study, it is showed that IL-6 and SOCS3 directly participated in the JAK/STAT signaling pathway and related to the disease of pulmonary fibrosis. IL-6 is an interleukin that implicated in the promotion of inflammation, cell proliferation and differentiation, and it has been found that patients with COPD had higher levels of CRP (C Reactive Protein), IL-6, fibrinogen and adiponectin[24,25]. The increased levels of IL6 are present in the airways and blood samples of patients with COPD, and increased IL6 plasma levels are associated with the risk of developing COPD during follow-up[26,27]. SOCS3 is an IL-6 responsive gene and is a specific inhibitor of the IL-6/Stat3 signaling pathway[28]. Furthermore, it has been suggested that tobacco smoke could inhibit the expression of SOCS3, and SOCS3 may represent a potential biomarker for understanding the efficacy and a novel anti-inflammatory mechanism of FP/SAL (Fluticasone propionate/Salmeterol) combination therapy in the treatment of COPD[29,30]. Therefore, the results of our work are consisted with previous studies.

The JAK/STAT signaling pathway activated by cytokines, hormones and growth factors plays important role in regulating expression of genes involved in cellular activation, differentiation and

Table III. The diseases relevant to the DEGs in the interactive network.

Disease	Count	*p* value	DEGs
Bone mass	5	0.001193	GH1, IL1R1, IL6, ESR1, ESR2
Breast cancer	23	0.001195	CYP3A4, IL6, XRCC6, ESR1, AURKA, ESR2, MMP3, PTEN, HPR, LEP, CASP10, GH1, MTHFR, EP300, PRLR, NCOA3, MAP3K1, RASSF1, TXN, BUB1, PTCH1, GNB3, UGT2B15
Hypertension	16	0.003851	CYP3A4, IL6, DRD3, SLC6A2, CYP2C9, ESR1, ESR2, BDKRB2, MMP3, FKBP1B, LEP, CD36, HSD11B2, ERAP1, GNB3, CACNA1C
Anorexia nervosa	5	0.004388	UCP3, DRD3, SLC6A2, ESR1, ESR2
Uterine cancer	3	0.004403	ESR1, AURKA, ESR2
Endometriosis	8	0.004417	IL4, IL6, ESR1, ESR2, MMP3, PDCD6, PTEN, AHSG
Blood pressure, arterial	8	0.005424	MTHFR, SLC6A2, CYP2C9, ESR1, BDKRB2, GNB3, ESR2, SELE
Bone density	11	0.006942	IL4, GGCX, IL1R1, IL6, CCL2, GNRH1, DBP, ESR1, PDE4D, ESR2, AHSG
Myocardial infarction	12	0.007076	IL4, MTHFR, IL6, CCL2, THBD, CYP2C9, ABCA1, BDKRB2, GNB3, MMP3, SELE, SELPLG
Depressive disorder, major	7	0.008061	CRHR1, CCL2, SLC6A2, CYP2C9, ESR1, GNB3, ESR2
Cardiovascular disease	7	0.01584	MTHFR, THBD, CYP2C9, ESR1, GNB3, SELE, HPR
Atherosclerosis, coronary	13	0.016062	IL18R1, CCL2, ESR1, ESR2, BDKRB2, ABCA1, MMP3, HPR, MTHFR, THBD, GNB3, SELE, SELPLG
Body mass	6	0.018237	LEP, HSD3B2, MTHFR, IL6, UCP3, GNB3
Atherosclerosis, carotid	4	0.01863	IL6, CCL2, ABCA1, AHSG
Heart disease, ischemic	6	0.024328	MTHFR, IL6, ABCA1, MMP3, SELE, SELPLG
Myocardial infarction	7	0.025747	MTHFR, IL6, THBD, ESR1, BDKRB2, MMP3, THPO
Blood pressure	4	0.045991	CYP2C9, BDKRB2, ESR2, SELE
Pulmonary fibrosis	4	0.045991	IL4, IL18R1, IL1R1, IL6
Osteoporosis	4	0.045991	MTHFR, ESR1, ESR2, AHSG

survival[31,32]. Research has shown that SOCS could blind to the cytokine receptors and sites of JAK tyrosine-phosphorylation by its SH2 structure to suppress cytokine signaling, and the IL-6/JAK/Stat3 pathway induces expression of SOCS3, inhibits activation of JAK-1 and, hence, blocks IL-6 signaling in a classic feed-back loop[28,33]. Therefore, JAK/STAT signaling pathway is very critical for IL-6 and SOCS3.

Conclusions

IL6 and SOCS3 play important roles in COPD and have the potential to serve as therapeutic targets of COPD. This work could contribute to understanding the molecular mechanisms of COPD. Since the identified DEGs in COPD were based on gene chips from a small sample size, the molecular mechanism and target therapy of RCC need to be further explored and researched.

Acknowledgements

We would like to thank all respondents of the study and all the people who give the help for this work.

Conflict of Interest

The Authors declare that there are no conflicts of interest.

References

1) Chung KF. Cytokines in chronic obstructive pulmonary disease. Eur Respir J Suppl 2001; 34: 50s-59s.
2) Rabe KF, Hurd S, Anzueto A, Barnes PJ, Buist SA, Calverley P, Fukuchi Y, Jenkins C, Rodriguez-Roisin R, van Weel C, Zielinski J. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med 2007; 176: 532-555.
3) Lopez AD, Murray CC. The global burden of disease, 1990-2020. Nat Med 1998; 4: 1241-1243.
4) Barnes PJ, Shapiro SD, Pauwels RA. Chronic obstructive pulmonary disease: molecular and cellular mechanisms. Eur Respir J 2003; 22: 672-688.
5) Tanino M, Betsuyaku T, Takeyabu K, Tanino Y, Yamaguchi E, Miyamoto K, Nishimura M. Increased levels of interleukin-8 in BAL fluid from smokers susceptible to pulmonary emphysema. Thorax 2002; 57: 405-411.
6) Bazzoni F, Cassatella MA, Rossi F, Ceska M, Dewald B, Baggiolini M. Phagocytosing neutrophils produce and release high amounts of the neutrophil-activating peptide 1/interleukin 8. J Exp Med 1991; 173: 771-774.
7) Traves SL, Culpitt SV, Russell RE, Barnes PJ, Donnelly LE. Increased levels of the chemokines GROalpha and MCP-1 in sputum samples from patients with COPD. Thorax 2002; 57: 590-595.
8) Crooks SW, Bayley DL, Hill SL, Stockley RA. Bronchial inflammation in acute bacterial exacerbations of chronic bronchitis: the role of leukotriene B4. Eur Respir J 2000; 15: 274-280.
9) Bhowmik A, Seemungal TA, Sapsford RJ, Wedzicha JA. Relation of sputum inflammatory markers to symptoms and lung function changes in COPD exacerbations. Thorax 2000; 55: 114-120.
10) Aaron SD, Angel JB, Lunau M, Wright K, Fex C, Le Saux N, Dales RE. Granulocyte inflammatory markers and airway infection during acute exacerbation of chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2001; 163: 349-355.
11) Gompertz S, Bayley DL, Hill SL, Stockley RA. Relationship between airway inflammation and the frequency of exacerbations in patients with smoking related COPD. Thorax 2001; 56: 36-41.
12) Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botstein D, Altman RB. Missing value estimation methods for DNA microarrays. Bioinformatics 2001; 17: 520-525.
13) Fujita A, Sato JR, Rodrigues Lde O, Ferreira CE, Sogayar MC. Evaluating different methods of microarray data normalization. BMC Bioinformatics 2006; 7: 469.
14) Smyth GK. Limma: linear models for microarray data. In: Bioinformatics and Computational Biology Solutions using R and Bioconductor. Springer, New York 2005; 397-420.
15) Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990; 215: 403-410.
16) Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND, Koonin EV. The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res 2001; 29: 22-28.
17) Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao YL, Ooi CE, Godwin B, Vitols E, Vijayadamodar G, Pochart P, Machineni H, Welsh M, Kong Y, Zerhusen B, Malcolm R, Varrone Z, Collis A, Minto M, Burgess S, McDaniel L, Stimpson E, Spriggs F, Williams J, Neurath K, Ioime N, Agee M, Voss E, Furtak K, Renzulli R, Aanensen N, Carrolla S, Bickelhaupt E, Lazovatsky Y, DaSilva A, Zhong J, Stanyon CA, Finley RL, Jr., White KP, Braverman M, Jarvie T, Gold S, Leach M, Knight J, Shimkets RA, McKenna MP, Chant J, Rothberg JM. A protein interaction map of Drosophila melanogaster. Science 2003; 302: 1727-1736.
18) Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, Vidalain PO, Han JD, Chesneau A, Hao T,

Goldberg DS, Li N, Martinez M, Rual JF, Lamesch P, Xu L, Tewari M, Wong SL, Zhang LV, Berriz GF, Jacotot L, Vaglio P, Reboul J, Hirozane-Kishikawa T, Li Q, Gabel HW, Elewa A, Baumgartner B, Rose DJ, Yu H, Bosak S, Sequerra R, Fraser A, Mango SE, Saxton WM, Strome S, Van Den Heuvel S, Piano F, Vandenhaute J, Sardet C, Gerstein M, Doucette-Stamm L, Gunsalus KC, Harper JW, Cusick ME, Roth FP, Hill DE, Vidal M. A map of the interactome network of the metazoan C. elegans. Science 2004; 303: 540-543.

19) Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen LJ, von Mering C. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res 2011; 39: D561-568.

20) Berriz GF, Beaver JE, Cenik C, Tasan M, Roth FP. Next generation software for functional trend analysis. Bioinformatics 2009; 25: 3043-3044.

21) Hosack DA, Dennis G, Jr., Sherman BT, Lane HC, Lempicki RA. Identifying biological themes within lists of genes with EASE. Genome Biol 2003; 4: R70.

22) Celli BR, Cote CG, Marin JM, Casanova C, Montes de Oca M, Mendez RA, Pinto Plata V, Cabral HJ. The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. N Engl J Med 2004; 350: 1005-1012.

23) Baraldo S, Turato G, Saetta M. Pathophysiology of the small airways in chronic obstructive pulmonary disease. Respiration 2012; 84: 89-97.

24) Breyer MK, Rutten EP, Locantore NW, Watkins ML, Miller BE, Wouters EF. Dysregulated adipokine metabolism in chronic obstructive pulmonary disease. Eur J Clin Invest 2012; 42: 983-991.

25) Hirano T, Yasukawa K, Harada H, Taga T, Watanabe Y, Matsuda T, Kashiwamura S, Nakajima K, Koyama K, Iwamatsu A, Tsunasawa S, Sakiyama F, Matsui H, Takahara Y, Taniguchi T, Kishimoto T. Complementary DNA for a novel human interleukin (BSF-2) that induces B lymphocytes to produce immunoglobulin. Nature 1986; 324: 73-76.

26) van Durme YM, Lahousse L, Verhamme KM, Stolk L, Eijgelsheim M, Loth DW, Uitterlinden AG, Breteler MM, Joos GF, Hofman A, Stricker BH, Brusselle GG. Mendelian randomization study of interleukin-6 in chronic obstructive pulmonary disease. Respiration 2011; 82: 530-538.

27) Attaran D, Lari SM, Towhidi M, Marallu HG, Ayatollahi H, Khajehdaluee M, Ghanei M, Basiri R. Interleukin-6 and airflow limitation in chemical warfare patients with chronic obstructive pulmonary disease. Int J Chron Obstruct Pulmon Dis 2010; 5: 335-340.

28) Yasukawa H, Ohishi M, Mori H, Murakami M, Chinen T, Aki D, Hanada T, Takeda K, Akira S, Hoshijima M, Hirano T, Chien KR, Yoshimura A. IL-6 induces an anti-inflammatory response in the absence of SOCS3 in macrophages. Nat Immunol 2003; 4: 551-556.

29) Singh SD. Salmeterol/fluticasone propionate combination in the treatment of COPD. Expert Rev Respir Med 2007; 1: 25-33.

30) Nasreen N, Khodayari N, Sukka-Ganesh B, Peruvemba S, Mohammed KA. Fluticasone propionate and Salmeterol combination induces SOCS-3 expression in airway epithelial cells. Int Immunopharmacol 2012; 12: 217-225.

31) Darnell JE, Jr. STATs and gene regulation. Science 1997; 277: 1630-1635.

32) Tas SW, Remans PH, Reedquist KA, Tak PP. Signal transduction pathways and transcription factors as therapeutic targets in inflammatory disease: towards innovative antirheumatic therapy. Curr Pharm Des 2005; 11: 581-611.

33) Valentino L, Pierre J. JAK/STAT signal transduction: regulators and implication in hematological malignancies. Biochem Pharmacol 2006; 71: 713-721.