Identification of novel candidate genes in East Asian COPD patients by the functional summary-based imputation and the unified test for molecular signatures: a transcriptome-wide association study

Ye Tian[1], Shufang Shan[2], Qixue Bao[1], Siquan Zhou[1], Xia Jiang[1], Mengqiao Wang[1], Shu Yin[3], Jingyuan Xiong[1], Guo Cheng[2]

[1]Healthy Food Evaluation Research Center, West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu, Sichuan 610041, China;

[2]Department of Pediatrics, Laboratory of Molecular Translational Medicine, Center for Translational Medicine, Key Laboratory of Birth Defects and Related Diseases of Women and Children (Sichuan University), Ministry of Education, West China Second University Hospital, Sichuan University, Chengdu, Sichuan 610041, China;

[3]School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China.

To the Editor: Chronic obstructive pulmonary disease (COPD), a clinical syndrome encompassing a spectrum of chronic, progressive, and debilitating respiratory disorders, is increasingly recognized as a multisystemic heterogeneous disease.[1] Because the characteristics of East Asian COPD patients differ from those of Westerners, it would be practical to investigate whether variations in COPD genetic performance exist between East Asian and Western populations.[2]

A genome-wide association study (GWAS) identifies single-nucleotide polymorphisms for complex human diseases and traits, while a transcriptome-wide association study (TWAS) can highlight novel genes beyond the initial GWAS. Compared with GWAS, TWAS has a lower multiple comparison burden with results manifested in the form of genes, demonstrating a more direct and higher relevance for subsequent functional annotation. Twelve novel COPD loci were previously reported based on European GWAS data. However, no TWAS has been conducted on COPD patients in East Asian with cross-tissue analysis, which hampers the substantial sharing of local expression regulation across tissues.

To report gene-trait associations in East Asian COPD patients, we mined the summary data from the largest meta-GWAS of COPD in BioBank Japan (BBJ), which included 162,653 controls and 4017 COPD patients. Functional Summary-based Imputation (FUSION, http://gusevlab.org/projects/fusion/) was used in single-tissue TWAS to analyze data from 48 tissues in GTEx v8 (https://www.genome.gov/Funded-Programs-Projects/Genotype-Tissue-Expression-Project) models.[3] In cross-tissue analysis, we used the Unified Test for Molecular Signatures (UTMOST, https://github.com/Joker-Jerome/UTMOST)

to estimate expression quantitative trait loci (eQTLs) in all tissues to increase the chance of identifying multitissue eQTLs,[4] and the same candidate genes identified from FUSION and UTMOST were used to infer a link between COPD and its complications. While several other methods can be used in TWAS, FUSION and UTMOST are suitable and reliable for TWAS summary statistics. A predictive model of COPD transcription profiles from the Gene Expression Omnibus (GEO) database was used to validate the TWAS results. Differential expression analysis was performed using the online analysis tool GEO2R (https://www.ncbi.nlm.nih.gov/geo/geo2r/). And the expression profiles of COPD patients and normal individuals were compared to identify the differentially expressed genes (DEGs). For genes identified by FUSION, UTMOST, and DEG analysis, we also performed Gene Ontology (GO, http://geneontology.org/) functional analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG, https://www.genome.jp/kegg/) pathway enrichment analysis and DisGeNET analysis (https://www.disgenet.org/).

Here, candidate genes were retained if they were statistically significant in two of the three analyses (FUSION, UTMOST, and DEG analysis). We anticipate that these genes will make novel contributions to understanding the genetic mechanism of COPD in East Asian populations and provide further clues for multisystemic comprehension of COPD and its complications at the transcriptional level. FUSION identified 35 genes across all tissues, and 17 were significantly expressed in multiple tissues [Supplementary Table 1, http://links.lww.com/CM9/B347 and Figure 1A]. _CIB2_ was identified in 28 tissues, including visceral adipose tissue (Omentum), adrenal gland, tibial artery, lung, and whole blood. Eighteen genes were expressed only in individual tissues.
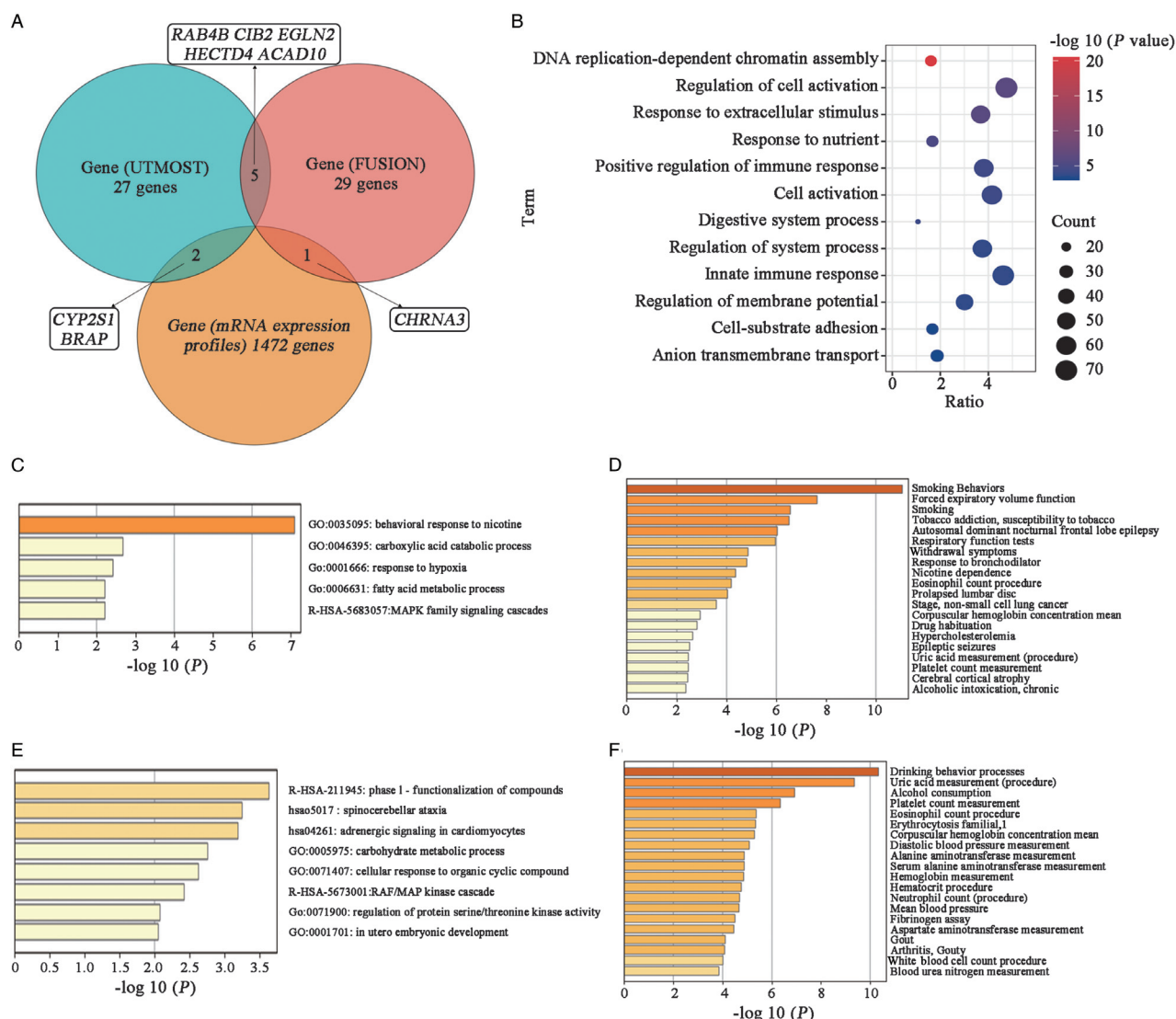
Figure 1: (A) Venn plot for FUSION, UTMOST, and DEG analysis results. (B) The 12 major GO terms were identified in DEG analysis. (C) Pathways highlighted by FUSION in GO functional and KEGG pathway enrichment analyses. (D) The summary of the enrichment analysis in DisGeNET for the COPD-associated genes by FUSION. (E) Pathways highlighted by UTMOST in GO functional and KEGG pathway enrichment analyses. (F) The summary of the enrichment analysis in DisGeNET for the COPD-associated genes by UTMOST. COPD: Chronic obstructive pulmonary disease; FUSION: Functional summary-based imputation; GO: Gene ontology; KEGG: Kyoto encyclopedia of genes and genomes; MAP: Mitogen-activated protein; MAPK: Mitogen-activated protein kinase; mRNA: Messenger RNA; UTMOST: Unified test for molecular signatures.

$HECTD4$ ($P = 2.99 \times 10^{-20}$) and $C19orf54$ ($P = 2.06 \times 10^{-7}$) were expressed only in lung tissue. UTMOST identified 34 genes from all tissues ($P < 3.46 \times 10^{-6}$), and 19 were significant in multiple tissues [Supplementary Table 2, http://links.lww.com/CM9/B347 and Figure 1A]. $RAB4B$ was identified in subcutaneous and visceral adipose tissue (Omentum), transformed fibroblasts, lung, transverse colon, pituitary, and whole blood. Fifteen genes were specifically expressed in different tissues. $CYP2S1$ ($P = 3.03 \times 10^{-7}$) and $ATXN2$ ($P = 1.34 \times 10^{-6}$) were expressed only visceral adipose tissue (Omentum). Four genes ($SNRPA$, $RAB4B$, $TRAFD1$, and $EGLN2$) were significantly expressed in lung tissue.

When comparing genes identified by FUSION and UTMOST, five candidate genes ($RAB4B$, $CIB2$, $EGLN2$, $HECTD4$, and $ACAD10$) were identified in both analyses [Figure 1A]. For example, $EGLN2$ ($P_{FUSION} = 8.88 \times 10^{-7}$, $P_{UTMOST} = 1.39 \times 10^{-6}$) was

expressed in lung tissue in both analyses [Supplementary Tables 1 and 2, http://links.lww.com/CM9/B347]. When comparing TWAS identified genes with the DEG analysis, three candidate genes ($CHRNA3$, $CYP2S1$, and $BRAP$) were obtained [Figure 1A]. $CHRNA3$ was identified by both FUSION and DEG analysis. $CYP2S1$ and $BRAP$ were identified by both UTMOST and DEG analysis. GO functional and KEGG pathway enrichment analyses of the significant genes from COPD DEG analysis identified 12 common GO terms [Figure 1B].

In the GO functional and KEGG pathway enrichment analyses, pathways highlighted by FUSION included the behavioral response to nicotine, the carboxylic acid catabolic process, and the response to hypoxia [Figure 1C and Supplementary Table 3, http://links.lww. com/CM9/B347], and the summary of the enrichment analysis in DisGeNET mainly involved smoking behavior and forced expiratory volume function [Figure 1D and

Supplementary Table 3, http://links.lww.com/CM9/B347]. Biological pathways highlighted by UTMOST included phase I functionalization of compounds, spinocerebellar ataxia, and adrenergic signaling in cardiomyocytes [Figure 1E and Supplementary Table 4, http://links.lww.com/CM9/B347], and the summary of the enrichment analysis in DisGeNET mainly involved drinking behavior and uric acid measurement (procedure) [Figure 1F and Supplementary Table 4, http://links.lww.com/CM9/B347].

Although a previous TWAS on European COPD patients used FUSION and S-PrediXcan, both methods were based on tissue-specific simulations and could not simulate cross-tissue gene-trait association levels of COPD, a multisystem disease.[5] We investigated the expression of TWAS genes in different tissue contexts by comparing two approaches, FUSION and UTMOST. FUSION was used to reveal tissue-specific genetic regulatory mechanisms. UTMOST was adopted to prioritize the same associations in multiple tissues and emphasized the important associations found in tissues significantly related to COPD.

Among FUSION- and UTMOST-identified genes, HECT domain E3 ubiquitin protein ligase 4 (encoded by HECTD4) was suggested playing a role in obesity, inflammation, and metabolic syndrome, all of which are inextricably linked with COPD. Interestingly, HECTD4 was recently reported as a candidate gene regulating fat distribution in East Asian population. As a novel target gene for COPD identified by our analyses, HECTD4 deserves more investigation, particularly among East Asians. Calcium and integrin binding family member 2 (encoded by CIB2) was found to block the translocation of sphingosine kinase 1 (encoded by SPHK1) and inhibit downstream signaling. Since SPHK1 is a cell survival regulator and its overexpression is observed in lung cancer, the association between CIB2 and COPD merits further evaluation. Ras-related protein Rab-4B (encoded by RAB4B) was significantly expressed not only in lung tissue but also in subcutaneous adipose and visceral adipose tissue (Omentum) in our analysis. Lung and adipose tissues are important sources of proinflammatory mediators in COPD patients. RAB4B deficiency in T cells promotes adipose tissue inflammation, providing new insights into the genetic underpinning of the inflammatory response in COPD. In agreement with previous GWAS in European populations and case-control study in Chinese populations, we identified EGLN2 as a potential candidate gene in East Asian COPD patients.[5] EGLN2 regulates the activity of nuclear factor kappa-B (NF-κB), an inflammatory transcription factor. In addition, a previous study showed that smoking was associated with the EGLN2 variant (hypoxic response), supporting the critical role of EGLN2 in COPD pathogenesis.

Comparing FUSION and UTMOST results with DEGs from COPD messenger RNA (mRNA) expression profiles, we obtained three candidate genes. The CHRNA3 (rs1051730) polymorphism was associated with COPD, lung cancer, and nicotine dependence. Bombesin receptor-activated protein (BRAP, encoded by BRAP) is widely expressed in human airway epithelial cells. BRAP was demonstrated to modulate NF-κB activity by enhancing the activity of histone deacetylase, suggesting that BRAP is an important regulator of immune and inflammatory responses in the human airway epithelium, providing a target for understanding the pathogenesis of COPD.

To take into account genes with minor differential expression but of considerable biological importance, GO functional and KEGG pathway enrichment analyses were performed to evaluate the function and distribution of COPD-associated genes. In these enriched pathways, behavioral response to nicotine (GO:0035095), smoking behavior (GO:C1519383), and forced expiratory volume function (GO:C0016529) were highlighted in our study, supporting the crucial role of tobacco smoking in COPD.[6]

In conclusion, we explored the BBJ GWAS dataset for COPD to identify and provide insights into candidate causal genes. Eight genes were identified by integrating FUSION, UTMOST, and DEG analysis. Previous small-scale GWAS reported several risk loci for COPD at CHRNA3 and EGLN2. In our study, six novel genes (RAB4B, CIB2, HECTD4, ACAD10, CYP2S1, and BRAP) were identified, providing transcriptional explanations and potential predictions for the genetics of COPD in East Asian populations.

Conflicts of interest

None.

References

1. Chan SMH, Selemidis S, Bozinovski S, Vlahos R. Pathobiological mechanisms underlying metabolic syndrome (MetS) in chronic obstructive pulmonary disease (COPD): clinical significance and therapeutic strategies. Pharmacol Ther 2019;198:160–188. doi: 10.1016/j.pharmthera.2019.02.013.
2. Takahashi S, Betsuyaku T. The chronic obstructive pulmonary disease comorbidity spectrum in Japan differs from that in western countries. Respir Investig 2015;53:259–270. doi: 10.1016/j.resinv.2015.05.005.
3. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BW, et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet 2016;48:245–252. doi: 10.1038/ng.3506.4.
4. Hu Y, Li M, Lu Q, Weng H, Wang J, Zekavat SM, et al. A statistical framework for cross-tissue transcriptome-wide association analysis. Nat Genet 2019;51:568–576. doi: 10.1038/s41588-019-0345-7.
5. Ding Y, Niu H, Yang H, Sun P, Chen Y, Duan M, et al. EGLN2 and RNF150 genetic variants are associated with chronic obstructive pulmonary disease risk in the Chinese population. Int J Chron Obstruct Pulmon Dis 2015;10:145–151. doi: 10.2147/COPD.S73031.
6. Halpin DMG. Precision medicine in chronic obstructive pulmonary disease. Chin Med J 2022;135:1156–1162. doi: 10.1097/CM9.0000000000002042.