Molecules and Cells

Identification of Putative Regulatory Alterations Leading to Changes in Gene Expression in Chronic Obstructive Pulmonary Disease

Dong-Yeop Kim[1], Woo Jin Kim[2], Jung-Hyun Kim[2], Seok-Ho Hong[2], and Sun Shim Choi[1,*]

[1]Division of Biomedical Convergence, College of Biomedical Science, Institute of Bioscience & Biotechnology, [2]Department of Internal Medicine, School of Medicine, Kangwon National University, Chuncheon 24341, Korea
*Correspondence: schoi@kangwon.ac.kr
http://dx.doi.org/10.14348/molcells.2019.2442
www.molcells.org

Various genetic and environmental factors are known to be associated with chronic obstructive pulmonary disease (COPD). We identified COPD-related differentially expressed genes (DEGs) using 189 samples accompanying either adenocarcinoma (AC) or squamous cell carcinoma (SC), comprising 91 normal and 98 COPD samples. DEGs were obtained from the intersection of two DEG sets separately identified for AC and SC to exclude the influence of different cancer backgrounds co-occurring with COPD. We also measured patient samples named group 'I', which were unable to be determined as normal or COPD based on alterations in gene expression. The Gene Ontology (GO) analysis revealed significant alterations in the expression of genes categorized with the 'cell adhesion', 'inflammatory response', and 'mitochondrial functions', i.e., well-known functions related to COPD, in samples from patients with COPD. Multi-omics data were subsequently integrated to decipher the upstream regulatory changes linked to the gene expression alterations in COPD. COPD-associated expression quantitative trait loci (eQTLs) were located at the upstream regulatory regions of 96 DEGs. Additionally, 45 previously identified COPD-related miRNAs were predicted to target 66 of the DEGs. The eQTLs and miRNAs might affect the expression of 'respiratory electron transport chain' genes and 'cell proliferation' genes, respectively, while both eQTLs and miRNAs might affect the expression of 'apoptosis' genes. We think that our present study will contribute to our understanding of the molecular etiology of COPD accompanying lung cancer.

Keywords: COPD, differentially expressed genes, lung cancer, regulatory alterations, *STX8*

INTRODUCTION

Chronic obstructive pulmonary disease (COPD), a complex age-related disease, has two components: chronic bronchitis, characterized by productive cough, and emphysema, demonstrated by destruction of the lung parenchyma (Agustí et al., 2012; Mannino and Buist, 2007; Rabe et al., 2007; Vestbo et al., 2013). COPD patients also suffer from shortness of breath due to chronic airway obstruction and inflammation (Fabbri et al., 2003; Vestbo et al., 2013). COPD is generally diagnosed by chronic and irreparable impairment of lung airflow (Aaron et al., 2007; Barnett, 2005; Calverley et al., 2007; Vestbo et al., 2013). The most prominent cause of COPD is cigarette smoking (CS), although this factor is not the only cause of the disease, and not all smokers have the disease (Buist et al., 2008; Kim et al., 2017; Lundback et al., 2003; Mannino et al., 2003; Stone et al., 1983; Swanney et al., 2008).

Several recent studies have shown that genetic factors

Received 30 November, 2018; revised 23 January, 2019; accepted 12 February, 2019; published online 26 March, 2019

eISSN: 0219-1032

contribute to COPD (Anderson and Bozinovski, 2003; Foreman et al., 2012; Jeong et al., 2018; Sakao et al., 2003; Sandford et al., 1998). For instance, loss-of-function mutation of α1-antitrypsin is a well-known genetic risk factor, although it is responsible for only 1-5% of COPD patients (Smith and Harrison, 1997; Stoller and Aboussouan, 2005). In addition, with the recent success of genome-wide association studies (GWAS), the list of COPD-associated genes is expanding rapidly, including FAM13A, HHIP, IREB2, RAB4B, EGLN2, MIA, CYP2A6 (Hardin and Silverman, 2014; Hobbs et al., 2017; Kim and Lee, 2015). Note that many COPD-associated single nucleotide polymorphisms (SNPs) are located in noncoding regions such as intergenic and intronic regions (Artigas et al., 2011; Repapi et al., 2010) rather than in protein-coding regions, which is also the case for other disease-associated SNPs. These noncoding but disease-associated SNPs may contribute to the altered regulation of gene expression, splicing, and epigenetic modifications.

Interestingly, CS can perturb gene expression by affecting various epigenetic markers including DNA methylation and chromatin modification (Belinsky et al., 2002; Kim et al., 2001; Lee and Pausova, 2013). A variety of epigenetic machineries for regulating downstream gene expression are known to be altered in COPD (Lawless et al., 2009; Schamberger et al., 2014). All these studies consistently indicate that both genetic and epigenetic alterations are important in the etiology of COPD. However, the mechanism by which these genes and their genetic mutations mediate the pathogenesis of the disease remains to be elucidated.

Meanwhile, an understanding of the perturbations in gene expression in various diseases will potentially contribute to the identification of molecular targets to develop new therapeutic drugs or prognostic modalities. Transcriptome studies using either microarrays or RNA-Seq approaches have been employed for those purposes in COPD as well (Chen et al., 2008; Kim et al., 2015a; Rangasamy et al., 2009; Steiling et al., 2013; Wang et al., 2008). For instance, according to Wang et al. (2008), extracellular matrix (ECM) and apoptosis genes are upregulated, whereas genes involved in anti-inflammatory functions are down-regulated in a microarray of 48 human lung samples, including normal tissues and samples from patients with various stages of COPD ranging from GOLD (Global Initiative for Chronic Obstructive Lung Diseases) stage 0 to GOLD stage 3. In addition, several other genes involved in inflammatory responses, including cytokines and chemokines, and genes involved in oxidative stress responses are associated with COPD progression (Chen et al., 2008; Kim et al., 2015a; Rangasamy et al., 2009; Steiling et al., 2013).

However, alterations in gene expression are notably heterogeneous among studies with different designs using different cell types (Novak et al., 2002), which is not surprising because 12 different cell types constitute the whole lung tissue (Wang et al., 2008). Moreover, most subjects with COPD were also diagnosed with lung cancers, which was the reason for undergoing lung resection, and thus the DEG results cannot be as easily interpreted as the genes that are altered by COPD alone (Wang et al., 2008). As shown in the study by Spira et al. (2007), gene expressions in normal airway epithelial cells derived from patients with lung cancers differ in a cancer-specific manner.

In the present study, we thus attempt to avoid complexities in estimating DEGs associated with COPD (i.e., COPD-DEGs) driven by different cancer backgrounds co-occurring with COPD or by likely misclassified patient samples named group 'I', basically by revisiting the previous study published by Kim et al. (2015a). In addition, we integrate the COPD-DEGs with various multi-omics data, revealing novel insights relevant to COPD, which may ultimately contribute to improving our understanding of the molecular etiology of COPD or to identifying molecular targets for the development of a novel diagnostic or prognostic strategy.

MATERIALS AND METHODS

Data download

For the present study, we obtained the RNA-Seq data produced from COPD patients with lung cancers from Kim et al. (2015a) (https://www.ncbi.nlm.nih.gov/geo/, GSE57148). Refer to Kim et al. (2015a) for a detailed procedure describing the process to align and map the raw FASTQ files to the reference genome (GRCh37/hg19) and to generate FPKM values for all samples using Tophat and Cufflinks (Trapnell et al., 2009; 2010; 2012).

DEG analysis

FPKM values downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/, GSE57148) were converted into $\log_2$ (FPKM + 1) values, and quantile normalization was carried out for the expression values using R. Subsequently, after assessing the normality using the Shapiro-Wilk test, we applied two different *t*-tests to identify DEGs, depending on whether the FPKMs exhibited a normal distribution: a two-sample *t*-test was applied to the data with a normal distribution, whereas the non-parametric Wilcoxon Rank-Sum test was used for the data that did not display normal distribution. We referenced the method used by several previous studies (Lee et al. 2015; Ocampo-Candiani et al., 2018; Park et al., 2016; Zhang et al., 2017). Two cutoff values were selected to identify DEGs; $Q < 0.01$ and an absolute fold change (FC) estimated by (COPD mean FPKM/normal mean FPKM) ≥ 1.5 for all samples, and $Q < 0.01$ and $|FC| \geq 2$ for the samples excluding group 'I'. In addition, we further classified the 189 samples into three different groups - normal ('N'), intermediate ('I'), and COPD ('C') - using *k*-means clustering (Hartigan and Wong, 1979).

Subgrouping of the COPD samples

The 189 patient samples in the GSE57148, comprising 91 samples labeled 'normal' and 98 samples labeled 'COPD', were further divided into COPD coupled with adenocarcinoma (AC-COPD; 58 normal versus 36 COPD) and COPD coupled with squamous cell carcinoma (SC-COPD; 33 normal versus 62 COPD). Detailed clinical information on the sex, age, FEV1/FVC ratio, and cancer types of the samples was retrieved from Kim et al. (2015a) and is provided as Supplementary Table S1 in the paper; all samples were collected from male patients.

Obtaining COPD-related miRNAs

A total of 45 differentially expressed miRNAs (DE-miRNAs) detected in COPD were obtained from Kim and Lee (2017). Information about miRNAs and their target mRNAs was obtained from miRTarBase (ver. 7.0, http://mirtarbase.mbc.nctu.edu.tw)(Chou et al., 2017), where a total of 2,599 miRNAs and their target genes are deposited. Forty-three DE-miRNAs, missing 2 miRNAs, were linked to a total of 4,786 mRNAs.

Obtaining COPD-related genes and eQTLs

A total of 910 previously known COPD-associated genes were obtained from the gene-disease association (GDA) database of DisGeNET (ver. 5.0, http://www.disgenet.org/)(Piñero et al., 2016). Lung expression quantitative trait locus (eQTL) data were obtained from GTEx Portal (ver. 7, https://www.gtexportal.org/)(Lonsdale et al., 2013), where a total of 25,283 eQTL-gene interactions with 24,653 SNPs and 25,243 genes were found. Among them, we used 11,297 eQTL genes that were shown to have significant interactions with gene expression at an FDR-adjusted *p*-value (Q) < 0.05.

Data analysis

All statistical tests and their related diagrams were implemented with R (ver. 3.5.0, https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf) and Bioconductor packages (ver. 3.7)(Huber et al., 2015). *k*-means clustering was also performed with the same R package by setting $k = 3$. HaploReg (ver. 4.1, https://pubs.broadinstitute.org/mammals/haploreg)(Ward and Kellis, 2015) was used to annotate SNPs in the context of regulatory regions such as transcription factor binding sites (TFBSs), enhancers, promoters, and DNaseI hypersensitivity sites (DHSs). Gene Ontology (GO) analysis was conducted using the DAVID tool (ver. 6.8; https://david.ncifcrf.gov/)(Huang et al., 2008). A gene set enrichment analysis (GSEA) was performed using the GSEA package (ver. 3.0)(Subramanian et al., 2005); we used 'biological process' (bp) of curated gene sets c5, which contain gene sets collected from the GO database (MSigDB, ver. 6.2)(Subramanian et al., 2005). The 'ReactomeFI' Cytoscape plugin was used to investigate the functional interactions among the DEGs we identified. In addition, one of the 'ReactomeFI' pull-down menu items, 'cluster FI Network' (Wu et al., 2010), was used to identify GO terms for the clustered modules. Other batch jobs were performed with custom-built Python scripts (ver. 3.6.0).

Quantitative real-time PCR (qRT-PCR) analysis

Total RNA was extracted from two types of mouse COPD models, i.e., five mice from the elastase-induced model (Suki et al., 2017) and three mice from the smoking-induced model (Cavarra et al., 2001); please refer to previous studies to find detailed protocols for constructing mouse COPD model systems (Cavarra et al., 2001; Huh et al., 2011; Kim et al., 2015b; Suki et al., 2017). Total RNA was isolated using the RNeasy Mini Kit (Qiagen, Germany) according to the manufacturer`s instructions, and the concentration was quantified with an Epoch Microplate Spectrophotometer (Biotek Instruments, Inc.). The quality of total RNAs was then tested by measuring the ratio of absorbance at 260 nm and 280 nm. Reverse transcription was conducted using the TOPscript™ RT DryMIX kit (Enzynomics); for PCR, 1 μl of synthesized cDNA was used with AccuPower PCR pre-MIX (Bioneer). Quantitative real-time PCR (qPCR) was performed to quantify the mRNA expression of candidate genes using the ABI Step One Plus System Instrument (Applied Biosystems). The qPCRs were performed according to the manufacturer's instructions with TOPreal™ qPCR 2X PreMIX SYBR Green with high ROX (Enzynomics) in a Microamp fast 96-well reaction plate. Levels of relative gene expression were normalized to GAPDH expression. The primer sequences are provided in Supplementary Table S2.

RESULTS

Overall schematic of the workflow

We collected RNA-Seq data from a total of 189 samples from the Gene Expression Omnibus (GEO) database (see Methods) to revisit the previous findings (Kim et al., 2015a) regarding COPD-associated gene expression signatures. As described in the original paper, the samples were all derived from either adenocarcinoma (AC) or squamous cell carcinoma (SC) of the lungs, regardless of whether they were labeled normal or COPD. Therefore, the terms 'normal' and 'COPD' here refer to lung cancer patients without COPD and lung cancer patients with COPD, respectively. We thus decided to classify COPD patients into two groups by the lung cancer types from which they suffered simultaneously, i.e., COPD patients with adenocarcinoma (AC-COPD) and COPD patients with squamous cell carcinoma (SC-COPD)(Fig. 1). Accordingly, the samples from AC and SC patients without COPD were named 'AC-normal' and 'SC-normal', respectively.

Consistent with the study by Kim et al. (2015), patients with COPD tended to be older, to have more pack-years of CS, and to have a significantly lower FEV1/FVC ratio than individuals without COPD (Supplementary Fig. S1). However, a significantly greater proportion of patients with COPD presented SC than AC; in contrast, a greater proportion of normal samples presented AC than SC (Supplementary Fig. S1). The biased distribution of patients with COPD between the AC and SC groups might skew the results of the DEG estimation; specifically, the DEGs between normal controls and patients with COPD might be skewed by DEGs between AC and SC. Therefore, in the present work, we decided to analyze COPD-associated gene expression signatures separately for AC and SC; the overall schematic of the present work is depicted in Fig. 1.

Existence of an intermediate group of patients with ambiguous expression patterns

We separately identified DEGs for the two COPD groups classified by the two cancer types, AC and SC, using cutoff values of $Q < 0.01$ and $|\text{fold change (FC)}| \geq 1.5$ (see Methods). As a result, 150 DEGs were identified between 58 AC-normal and 36 AC-COPD samples (named AC-DEGs), and 58 DEGs were identified between 33 SC-normal and 62 SC-COPD samples (named SC-DEGs).

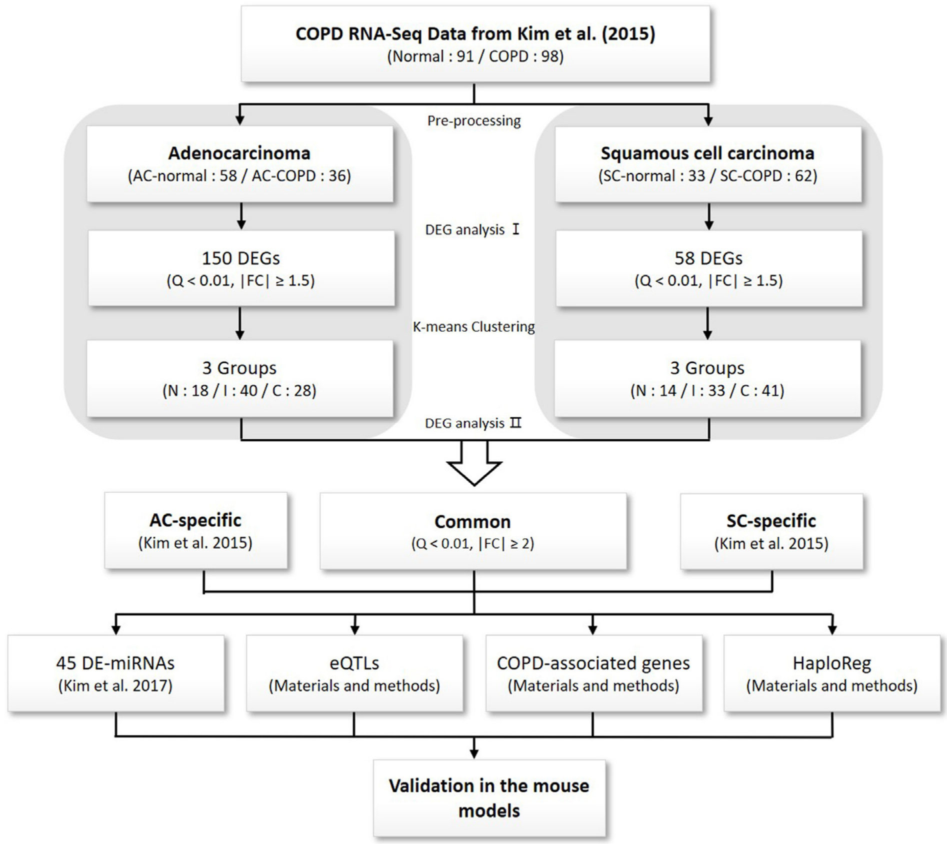

Fig. 1. Overall schematic of the workflow. AC: adenocarcinoma; SC: squamous cell carcinoma; Q: FDR-adjusted *p*-value; FC: fold change; N: normal group; I: intermediate group; C: COPD group.

We conducted heatmap analysis accompanied by unsupervised hierarchical clustering for AC and SC to investigate whether these DEGs demarcated the normal and COPD samples (Figs. 2A and 2D). The classification of samples by the DEGs was not perfect in either cancer types (i.e., some normal samples are clustered in an intermingled way with some COPD samples, and vice versa), leading to the recognition of a third group of patient samples, separate from normal and COPD.

We thus performed *k*-means clustering of samples by setting a parameter $k = 3$ for AC and SC and integrated the resulting sample clusters by principal component analysis (PCA). The third group, the intermediate group, which was classified as neither normal nor COPD, is represented by gray dots in the PCA graph (Fig. 2B and 2E). For AC and SC, the samples were thus reclassified into three groups, i.e., group 'N' (normal), group 'I' (intermediate), and group 'C' (COPD); see Methods.

Interestingly, the median FEV1/FVC ratio of group 'I' was approximately equivalent to group 'N', although the ratio ranged widely from less than 50% (clinical COPD) to greater than 80% (clinically normal) for both AC and SC (Figs. 2C and 2F). Based on this result, group 'I' may contain misclassified patient samples, although they had been clinically diagnosed with or without COPD.

Defining reliable DEGs for deciphering the molecular etiology of COPD NC-DEGs

After establishing a third group of patients, group 'I', we decided to re-identify the three sets of DEGs to determine how the gene expression patterns of the patients in group 'I' e differed. We defined NC-DEGs, NI-DEGs, IC-DEGs by comparing gene expression between groups 'N' and 'C', 'N' and 'I', and 'I' and 'C', respectively. Several statistical cutoff values for defining the three sets of DEGs were tested, from which we found that very few NI- and IC-DEGs remained when $Q < 0.01$ and $|FC| \geq 2$ were applied. However, 237 NC-DEGs still remained as DEGs at those stringent cutoffs (Supplementary Table S3). The fact that very few DEGs appeared between groups 'N' and 'I', and groups 'I' and 'C' suggests that estimating COPD-DEGs by including group 'I' in the original labels as either normal or COPD renders the DEGs to be less reliable markers for demarcating samples. In other words, NC-DEGs, rather than NI-DEGs and IC-DEGs, are more confident gene sets to investigate the molecular etiology of COPD, which is why we chose to use NC-DEGs for a reliable identification of COPD-DEGs.

Next, we tried to obtain common COPD-DEGs between AC-COPD and SC-COPD groups by collecting common NC-DEGs identified based on the intersection between the NC-DEG sets of AC and SC. As shown in Fig. 3, 237 common

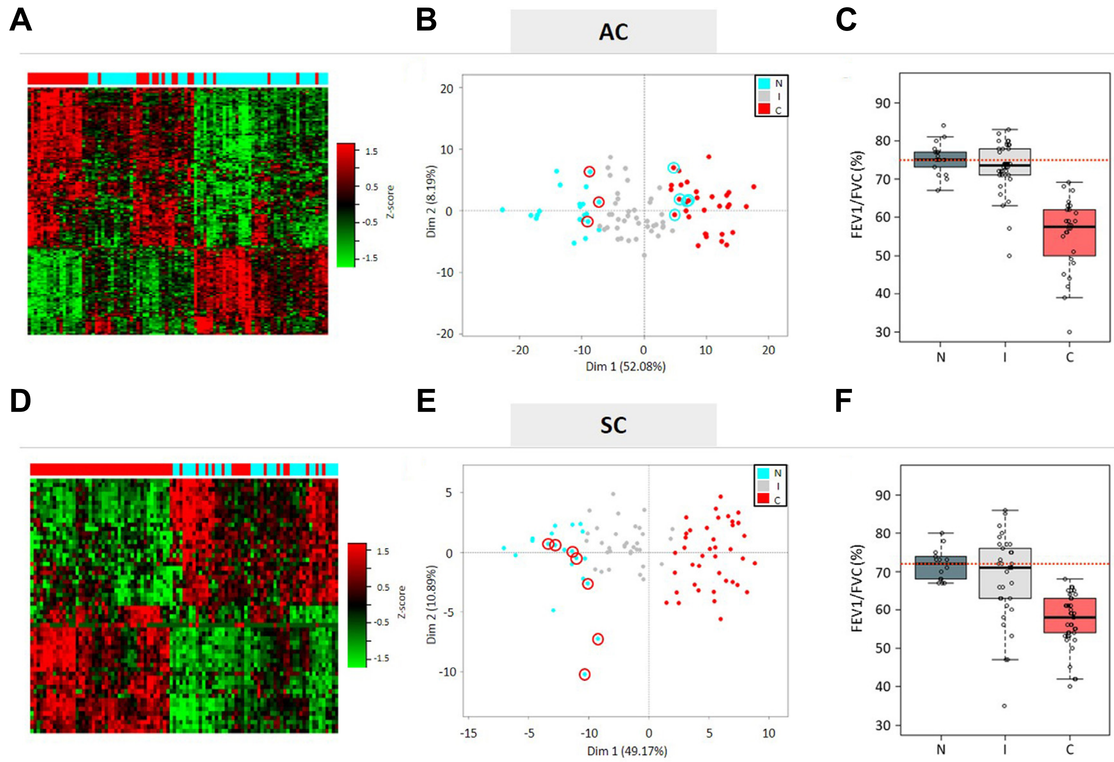

Fig. 2. Heatmap and PCA analysis of DEGs and a comparison of clinical characteristics of N, I and C samples in the AC and SC groups. DEGs were isolated separately for AC and SC, as displayed. Left panels (A and D) show the results of the heatmap analysis accompanied by the unsupervised clustering of samples using the 150 AC-DEGs and 58 SC-DEGs. Dendrograms showing distances between the samples, generally displayed at the top of heatmaps, are omitted, and colored bars are used instead to represent normal samples (colored cyan blue) and COPD samples (colored red). The middle panels (B and E) show the results of the PCA accompanied by *k*-means clustering ($k = 3$). Each dot represents one patient sample; normal, intermediate, and COPD samples are colored cyan, gray, and red, respectively. Some outlier samples that were clinically diagnosed as normal but showed similar gene expression to COPD, or vice versa, as indicated with colored cyan blue and red circles, respectively. These outliers were excluded from further analyses. The right panels (C and F) show boxplots of the FEV1/FVC ratios for N, I and C samples. The red dotted lines represent the median value of N. Significance was tested using Student's *t*-test ($*p < 0.05$; $**p < 0.01$; $***p < 0.001$).

NC-DEGs (the bottom panel), including 131 upregulated and 106 downregulated genes, differentiated COPD from normal samples without ambiguity for both AC (the top left panel) and SC (the top right panel) using a heatmap accompanied by unsupervised hierarchical clustering. Notably, samples from group 'I' were excluded from this analysis of both AC and SC.

Common DEGs for both AC and SC reveal well-known COPD-related genes

The intersection between NC-DEGs for AC and SC produced two additional groups of gene sets, 'AC-specific' (316 genes) and 'SC-specific' (62 genes), as well as the common NC-DEGs (Fig. 4). Moreover, according to the direction of the changes in expression (up- or downregulation), these three categories of NC-DEGs were divided into six categories: 'Common-up' (131 genes), 'Common-down' (106 genes), 'AC-specific-up' (140 genes), 'AC-specific-down' (176 genes), 'SC-specific-up' (29 genes), and 'SC-specific-down' (33 genes)(Fig. 4A).

We performed a GO analysis of each of the 6 categories of DEGs, confirming several previously well-known COPD-related genes in the 'common' category. For instance, GO functions such as 'inflammatory response', 'cell adhesion', and 'ECM' were significantly enriched in the 'Common-up' category, whereas mitochondria-related functions were significantly enriched in the 'Common-down' category (Fig. 4A). Interestingly, all these functional GO terms were completely consistent with the findings from recent reviews aiming to decipher the molecular pathology of COPD (Chen et al., 2008; Wang et al., 2008). Similar GO functional terms appeared in both AC-specific and SC-specific categories, although the numbers of genes associated with specific functional terms differed, suggesting that genes involved in molecular pathways associated with COPD etiology were commonly altered as well. Notably, however, inflammatory and apoptosis genes were prominent functional terms for the AC-specific category, but not the SC-specific category, whereas cell proliferation genes showed the opposite trend (Fig. 4A). We postulate that this difference in functional GO

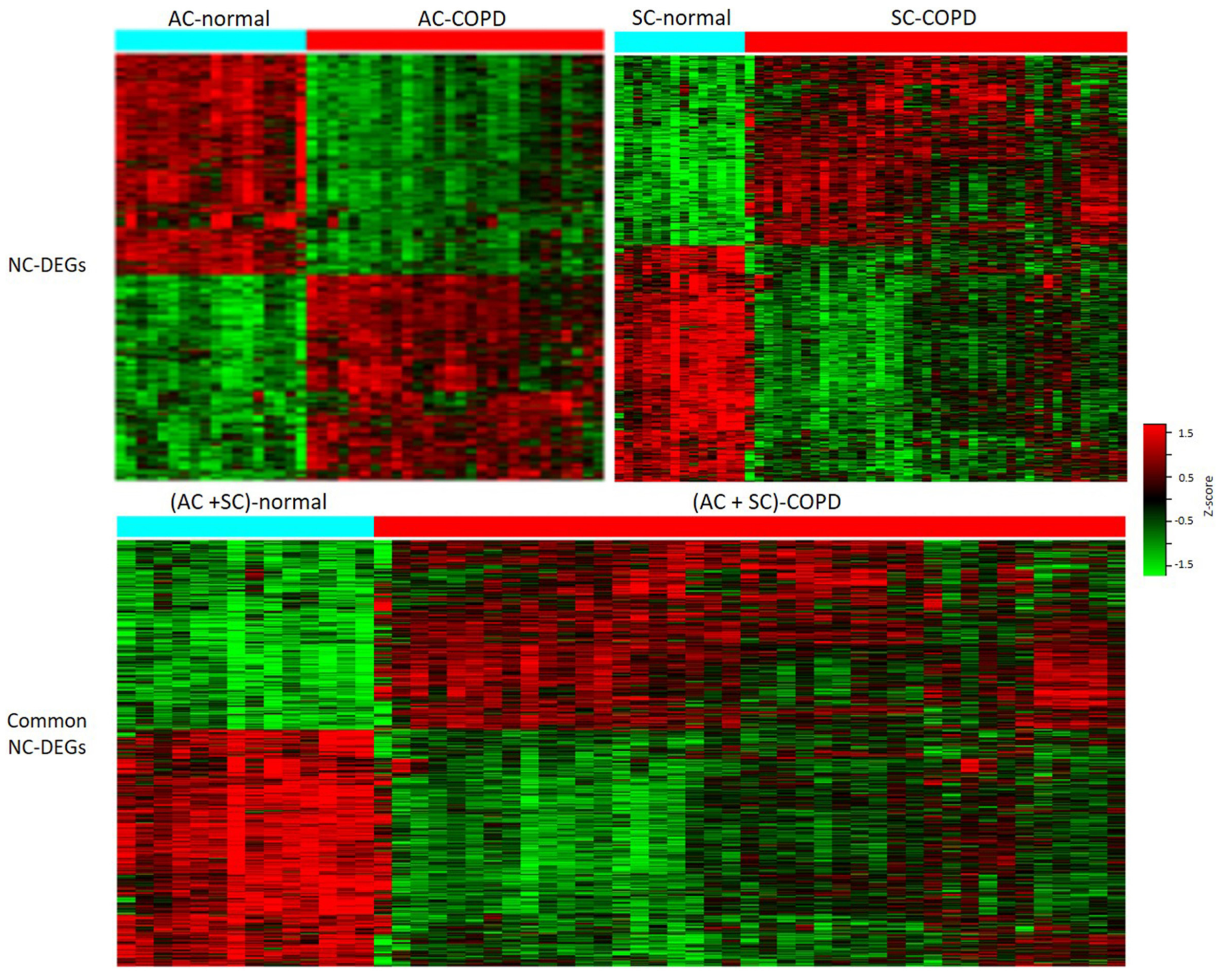

Fig. 3. Heatmaps created with NC-DEGs. NC-DEGs were isolated for AC- and SC-COPD samples, as described in the main text. Heatmap analyses coupled with unsupervised clustering, as described in Fig. 2, were conducted for AC and SC, respectively (top panels). By examining the intersection of the DEG results from the AC- and SC-COPD groups, we identified commonly altered NC-DEGs named common NC-DEGs (see the main text). Heatmap analyses coupled with unsupervised clustering were performed on the common NC-DEGs and all (AC plus SC) samples (bottom panels).

terms exists because the etiology of COPD in patients with AC and SC is not identical due to different genetic and epigenetic conditions associated with each cancer type. Notably, the GSEA generally provided the same interpretation as the GO analysis (data not shown).

We then created a functional interaction network with the ReactomeFI Cytoscape plugin (see the Methods) using these 615 NC-DEGs assigned into the six different categories (Fig. 4B), confirming that these genes, regardless of the categories to which they are assigned, are associated with COPD. Most of the input DEGs were inter-connected in a large network and the functional terms of each clustered subgroup in the network were consistent with the functions mentioned above, suggesting that these DEGs may participate in the pathogenic pathway leading to COPD. The integration of an additional source of functional information about these DEGs, i.e., previously known COPD-associated genes, including *EGFR*, *PTGS2*, *IL6*, *CXCR1* and *CXCR2*, obtained from DisGeNET (see the Methods), into this Cytoscape-generated diagram, as depicted by the gene symbols inside the bigger circles, provided strong support for the hypothesis that these DEGs potentially represent very reliable gene sets for deciphering the molecular etiology of COPD.

Statistically significant miRNA-DEG pairs and eQTL-DEG pairs

We next attempted to further characterize the 237 common NC-DEGs by integrating them with other multi-omics data. Previously, Kim and Lee (2017) reported a total of 45 miRNAs to be differentially expressed in COPD patients (i.e., DE-miRNAs) compared with normal controls. It is hypothesized that the alteration of miRNA expression could lead to alterations in target gene expression, considering that miRNAs are well-established regulators of gene expression. Consequently, we searched the 237 common NC-DEGs for genes predicted to be targeted by these 45 DE-miRNAs by examining

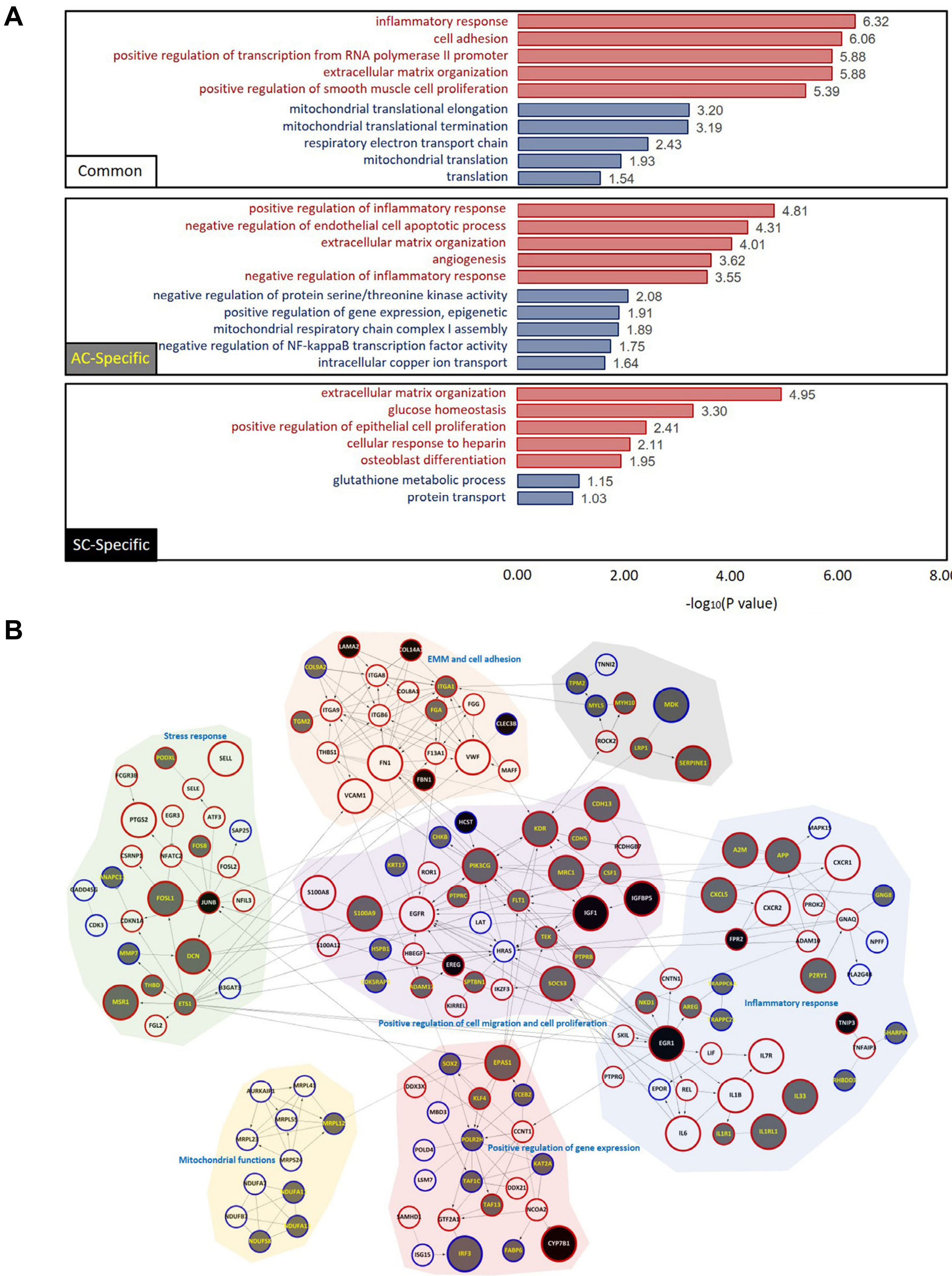

Fig. 4. GO and network analysis of the common NC-DEGs. (A) GO analysis of up- or downregulated 'AC-specific', 'SC-specific', and 'common' DEGs, as described in the main text. Red and blue represent GO functional terms for up- and downregulated genes, respectively. The top five GO terms selected based on the -$\log_{10}$(p-value) are depicted in bar graphs and are presented on the right side of the bars. (B) A functional interaction network was created with the Reactome FI plugin of the Cytoscape tool (see the Methods). Each gene in each circle is linked by functional interaction; red outlines: upregulated, blue outlines: downregulated, black symbols inside blank balls: common DEGs, yellow symbols inside gray balls: AC-specific DEGs, white symbols inside black balls: SC-specific DEGs. The larger sized circle indicates known COPD-associated genes.

the intersection of these DEGs with previously known miRNA-target mRNA pairs (see the Methods). As a result, 43 of the 45 miRNAs were associated with a total of 4,786 mRNA targets. By overlapping the 237 DEGs with the 4,786 miRNA targets, 66 NC-DEGs were identified (Tables 1 and Supplementary S4). Subsequently, a random permutation experiment was conducted to determine the extent of the significant enrichment of the 66 miRNA-DEG pairs among the 237 DEGs within the total of 4,786 genes. We confirmed that the 66 genes linked to the DE-miRNAs were impossible to obtain by chance, indicating that the 66 DEGs were significantly enriched through an interaction with DE-miRNAs (Fig. 5A).

Another mechanism of gene expression alteration is genetic mutations that occur in *cis*-acting regulatory elements, such as promoters and enhancers. The genetic changes associated with gene expression have been studied by the identification of eQTLs. We acquired this eQTL information from the GTEx database (see Methods) to investigate whether the NC-DEGs are affected by any eQTL. Ninety-six common NC-DEGs mapped onto the locations of eQTL SNPs within their regulatory regions (Tables 1 and Supplementary S5). A similar random permutation experiment to the analysis of the DEG-miRNA pairs was applied. During each

Table 1. Results of the GO analysis

Category (n)	Gene Ontology	*p*-value
miRNA specific (46)	positive regulation of smooth muscle cell proliferation	4.26E-07
	positive regulation of transcription from RNA polymerase II promoter	2.61E-05
	positive regulation of cell proliferation	1.64E-04
	cellular response to tumor necrosis factor	2.79E-03
	cell adhesion	6.11E-03
Both (20)	positive regulation of ERK1 and ERK2 cascade	9.54E-04
	negative regulation of apoptotic process	1.49E-03
	positive regulation of fibroblast proliferation	1.68E-03
	positive regulation of MAP kinase activity	2.00E-03
	leukocyte migration	8.26E-03
eQTL specific (76)	respiratory electron transport chain	2.00E-03
	lymphocyte homeostasis	1.66E-02
	positive regulation of intrinsic apoptotic signaling pathway in response to DNA damage	1.98E-02
	inflammatory response	3.74E-02
	protein phosphorylation	6.53E-02

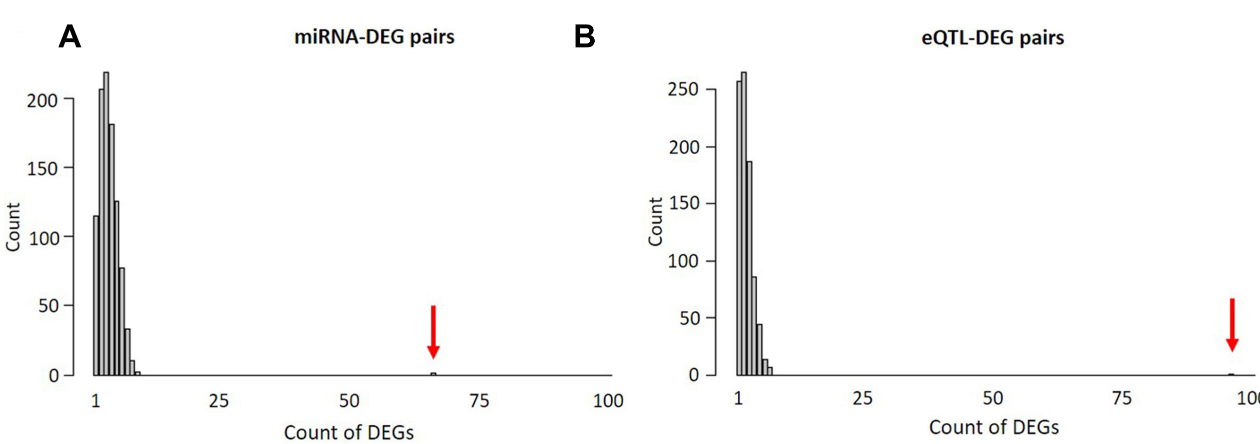

Fig. 5. Permutation test of miRNA-DEG pairs and eQTL-DEG pairs. A random permutation experiment was conducted to the extent of the significance of the number of miRNA-DEGs, 66 genes (A), and the number of eQTL-DEGs, 96 genes (B), by comparing the values with random chance. A. During each random iteration that was repeated at least 1,000 times, the number of DEGs among the 237 randomly selected genes from the total of 4,786 genes determined to be DE-miRNA target genes was estimated and plotted against the observed number, i.e., 66. B. During each random iteration that was repeated at 1,000 times, the number of DEGs among the 237 randomly selected genes from the total of 11,297 genes with lung eQTLs retrieved from GTEx database (see the Methods) was estimated and plotted against the observed number, i.e., 96.

random iteration repeated 1,000 times, we estimated the number of DEGs coupled with eQTLs of the 237 randomly selected genes among the total of 11,297 lung eQTLs (Fig. 5B). The 96 eQTL-DEG pairs we observed were significantly enriched and not expected by random chance (Fig. 5B). Consistently, 91% (87/96) of those SNPs were located in regulatory regions, such as TFBSs, enhancers, promoters, or DHSs (Supplementary Table S5).

Relationships between GO functional classes and regulatory alterations

We then examined whether a general pattern occurred in the mechanism of regulatory alterations, mediated either by miRNAs or by eQTLs, in functional categories into which the common NC-DEGs were assigned. Two groups of DEGs, i.e., DEGs targeted by the 45 DE-miRNAs and DEGs mapped to eQTLs, were overlapped, resulting in three categories of DEGs: genes whose expression was altered only by DE-miRNAs (designated 'miRNA-specific'), genes whose expression was altered only by regulatory mutations (designated 'eQTL-specific'), and whose expression was altered by both mechanisms (designated 'both'). As shown in Table 1, only 20 NC-DEGs, including *EGFR*, *FASTK*, and *HIPK3*, were subject to regulation by both DE-miRNAs and eQTLs, and the functional GO terms of these NC-DEGs were 'apoptosis' or 'cell proliferation'. Forty-six genes, including *IL6*, *THBS1* and *HIPK2*, were 'miRNA-specific' and were grouped into 'cell proliferation' or 'cell adhesion'. The remaining 76 genes, including *ETFB*, *MMRN1* and *MRPL41*, were 'eQTL-specific' and were involved in mitochondrial functions, such as 'respiratory electron transport chain' or 'inflammatory response'. Interestingly, some genes, such as those associated with the 'respiratory electron transport chain', which are genes that are likely related to oxidative stress, were generally altered by regulatory genetic mutations rather than miRNA-mediated regulatory pathways.

Validating the alterations in the expression of selected genes in the mouse COPD models

Supplementary Table S5 summarizes our integration of COPD-related multi-omics data, from which we confirmed several novel genes, such as *STRA13* and *FGG*, as well as well-known COPD-related genes, such as *MIF* and *SOD2*. We decided to validate some of these DEGs using qRT-PCR. Since no human COPD samples were available, we tried to validate the expression of these genes in two different mouse COPD models, elastase-induced COPD (EI-COPD) and smoking-induced COPD (Sm-COPD) (see Methods). Genes were selected by referencing Supplementary Table S5 and previous literature. For instance, the *MIF* gene was selected because it contained an eQTL SNP, to which several TFBSs were mapped and linked to the DE-miRNA, and studies have implicated this gene in COPD (Russell et al., 2016; She et al., 2012). The mouse models represented models of emphysema (i.e., one of the key symptoms of COPD) rather than exact models of human COPD. Nevertheless, the expression of six of the seven selected genes was confirmed to be altered in these mouse models (Fig. 6). *Mif* and *Fastk* were specifically downregulated in EI-COPD, whereas *Ndufa7*, *Stra13 and Fgg* were specifically downregulated in

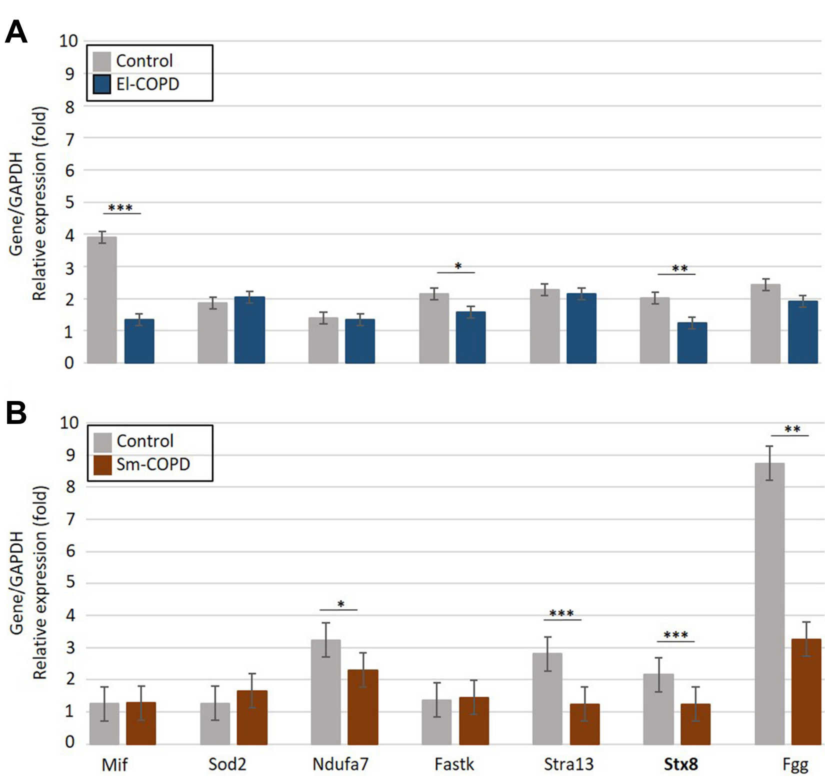

Fig. 6. Functional validation of selected DEGs using two COPD mouse model systems. Seven genes were chosen to be validated by RT-qPCR in the two mouse COPD model systems (see Methods). A. RT-qPCR results from a mouse model of elastase-induced COPD; five EI-COPD mice and five C57BL/6J mice were compared. B. RT-qPCR results from a mouse model of smoking-induced COPD; three Sm-COPD mice and three C57BL/6J mice were compared. Significance was tested by Student's *t*-test (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$).

Sm-COPD. Interestingly, the expression of the *Stx8* gene was downregulated in both models.

DISCUSSION

Here, we identified genes that were significantly altered by COPD using samples previously published by Kim et al. (2015a). We not only confirmed the previous finding, showing the perturbation in the gene expression of inflammatory genes and mitochondrial genes, but also revealed a novel aspect regarding the changes in putative regulatory regions that might affect changes in COPD-DEG expression.

A problem confronted by Kim et al. in their study (Kim et al., 2015) and by us in the present study was that all COPD samples were accompanied by lung cancers, either AC or SC. However, a substantial difference between the study by Kim et al. (2015) and our present study is the strategy used to identify DEGs; we attempted to remove the bias driven by lung cancers in patients with COPD by detecting COPD-DEGs. Another difference was the recognition of an 'intermediate' group named group 'I', i.e., patients who cannot be classified as COPD or non-COPD by gene expression patterns alone. We observed few differences in gene expression levels between the 'N' and 'I' groups and between the 'I' and 'C' groups, and the FEV1/FVC ratio of group 'I' was scattered from low to high, indicating that COPD and non-COPD samples were likely intermingled within group 'I'. In other words, the clinical diagnosis of patients in group 'I' as either COPD or normal might have been inaccurate, potentially leading to the identification of less reliable COPD-DEGs if group 'I' samples were included in the datasets. We postulate that the exclusion of samples from group 'I' and the collection of NC-DEGs helped us unambiguously identify reliable COPD-DEGs. In fact, the subsequent analyses performed after DEG identification, including the GO analysis and functional interaction network analysis, confirmed the previous findings from COPD-driven transcriptome analyses, as the expression of inflammatory genes and apoptosis genes was upregulated, and the expression of oxidation-reduction genes was downregulated. A third difference exists with regard to our effort to integrate DEGs with various omics datasets to investigate a mechanistic question underlying the expression alteration occurring in COPD. The integration of multi-omics data, including GDA, eQTLs, and miRNAs, provided us an opportunity to link changes in upstream regulatory regions and changes in downstream gene expression, through which we identified some common pathways involved in the development of COPD, regardless of co-occurring cancer types. Additionally, different functional classes of genes were altered by different lung cancer backgrounds or by the upregulation or downregulation of genes.

Finally, we further validated alterations in the expression of some selected genes in the two mouse COPD model systems, in which *Stx8* in particularly emerged. Interestingly, the downregulation of S*tx8* in COPD was consistent not only with the finding in the human samples but also between the two mouse model systems (Figs. 6A and 6B), which provides the insight that some molecular etiologies may be shared between different species and even different environmental causes. *Stx8* belongs to the 'syntaxin' family, i.e., gene groups of membrane proteins participating in exocytosis, a GO term in the molecular functional category of 'chloride channel inhibitor activity'. This gene function related to chloride channel activity leads to the topic of cystic fibrosis, a well-known genetic disease caused by mutation of the chloride channel encoded by the *CFTR* gene. In future studies, it will be interesting to examine whether the downregulation of the *Stx8* gene can be a direct cause of COPD and, if so, how that effect might occur at the level of molecular mechanisms.

We think that our present study will contribute to understanding the molecular etiology of COPD coupled with lung cancers and to identifying diagnostic markers of COPD.

Note: Supplementary information is available on the Molecules and Cells website (www.molcells.org).

ACKNOWLEDGMENTS
This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2016R1D1A1B03930411).

REFERENCES

Aaron, S.D., Vandemheen, K.L., Fergusson, D., Maltais, F., Bourbeau, J., Goldstein, R., Balter, M., O'Donnell, D., McIvor, A., and Sharma, S. (2007). Tiotropium in combination with placebo, salmeterol, or fluticasone-salmeterol for treatment of chronic obstructive pulmonary disease: a randomized trial. Ann. Intern. Med. *146*, 545-555.

Agustí, A., Edwards, L.D., Rennard, S.I., MacNee, W., Tal-Singer, R., Miller, B.E., Vestbo, J., Lomas, D.A., Calverley, P.M., and Wouters, E. (2012). Persistent systemic inflammation is associated with poor clinical outcomes in COPD: a novel phenotype. PloS One *7*, e37483.

Anderson, G.P., and Bozinovski, S. (2003). Acquired somatic mutations in the molecular pathogenesis of COPD. Trends Pharmacol. Sci. *24*, 71-76.

Artigas, M.S., Loth, D.W., Wain, L.V., Gharib, S.A., Obeidat, M., Tang, W., Zhai, G., Zhao, J.H., Smith, A.V., and Huffman, J.E. (2011). Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. Nat. Genet. *43*, 1082.

Barnett, M. (2005). Chronic obstructive pulmonary disease: a phenomenological study of patients' experiences. J. Clin. Nurs. *14*, 805-812.

Belinsky, S.A., Palmisano, W.A., Gilliland, F.D., Crooks, L.A., Divine, K.K., Winters, S.A., Grimes, M.J., Harms, H.J., Tellez, C.S., Smith, T.M., et al. (2002). Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers. Cancer Res. *62*, 2370-2377.

Buist, A.S., Vollmer, W., and McBurnie, M. (2008). Worldwide burden of COPD in high-and low-income countries. Part I. The Burden of Obstructive Lung Disease (BOLD) Initiative [State of the art series. Chronic obstructive pulmonary disease in high-and low-income countries. Edited by G. Marks and M. Chan-Yeung. Number 6 in the series]. Int. J. Tuberc. Lung Dis. *12*, 703-708.

Calverley, P.M., Anderson, J.A., Celli, B., Ferguson, G.T., Jenkins, C., Jones, P.W., Yates, J.C., and Vestbo, J. (2007). Salmeterol and fluticasone propionate and survival in chronic obstructive pulmonary disease. N. Engl. J. Med. *356*, 775-789.

Cavarra, E., Bartalesi, B., Lucattelli, M., Fineschi, S., Lunghi, B., Gambelli, F., Ortiz, L.A., Martorana, P.A., and Lungarella, G. (2001). Effects of cigarette smoke in mice with different levels of α1-proteinase inhibitor and sensitivity to oxidants. Am. J. Respir. Crit. Care Med. *164*, 886-890.

Chen, Z., Kim, H.P., Ryter, S.W., and Choi, A.M. (2008). Identifying targets for COPD treatment through gene expression analyses. Int. J. Chron. Obstruct. Pulmon. Dis. *3*, 359.

Chou, C., Shrestha, S., Yang, C., Chang, N., Lin, Y., Liao, K., Huang, W., Sun, T., Tu, S., and Lee, W. (2017). miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. Nucleic Acids Res. *46*, D296-D302.

Fabbri, L.M., Romagnoli, M., Corbetta, L., Casoni, G., Busljetic, K., Turato, G., Ligabue, G., Ciaccia, A., Saetta, M., and Papi, A. (2003). Differences in airway inflammation in patients with fixed airflow obstruction due to asthma or chronic obstructive pulmonary disease. Am. J. Respir. Crit. Care Med. *167*, 418-424.

Foreman, M.G., Campos, M., and Celedón, J.C. (2012). Genes and chronic obstructive pulmonary disease. Medical Clinics *96*, 699-711.

Hardin, M., and Silverman, E.K. (2014). Chronic obstructive pulmonary disease genetics: a review of the past and a look into the future. Int. J. Chron. Obstruct. Pulmon. Dis. *1*, 33-46.

Hartigan, J.A., and Wong, M.A. (1979). Algorithm AS 136: A k-means clustering algorithm. J. R. Stat. Soc. C-Appl. *28*, 100-108.

Hobbs, B.D., De Jong, K., Lamontagne, M., Bossé, Y., Shrine, N., Artigas, M.S., Wain, L.V., Hall, I.P., Jackson, V.E., and Wyss, A.B. (2017). Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. Nat. Genet. *49*, 426.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols *4*, 44.

Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., and Girke, T. (2015). Orchestrating high-throughput genomic analysis with bioconductor. Nat. Methods *12*, 115.

Huh, J.W., Kim, S., Lee, J.H., Lee, J., Van Ta, Q., Kim, M., Oh, Y., Lee, Y., and Lee, S. (2011). Bone marrow cells repair cigarette smoke-induced emphysema in rats. Am. J. Physiol.-lung C. *301*, L255-L266.

Jeong, I., Lim, J., Oh, D.K., Kim, W.J., and Oh, Y. (2018). Gene expression profile of human lung in a relatively early stage of COPD with emphysema. Int. J. Chronic. Obstr. *13*, 2643.

Kim, W.J., and Lee, C.Y. (2017). Environmental exposures and chronic obstructive pulmonary disease. Mol. Cellular Toxicol. *13*, 251-255.

Kim, W.J., Lim, J.H., Hong, Y., Hong, S., Bang, C.Y., Lee, J.S., Oh, Y., and Kim, J.H. (2017). Altered miRNA expression in lung tissues of patients with chronic obstructive pulmonary disease. Mol. Cellular Toxicol. *13*, 207-212.

Kim, W.J., Lim, J.H., Lee, J.S., Lee, S., Kim, J.H., and Oh, Y. (2015). Comprehensive analysis of transcriptome sequencing data in the lung tissues of COPD subjects. Int. J. Genomics *2015*, 206937.

Kim, Y., Kim, J., Huh, J.W., Lee, S.W., Choi, S.J., and Oh, Y. (2015). The therapeutic effects of optimal dose of mesenchymal stem cells in a murine model of an elastase induced-emphysema. Tuberc. Respir. Dis. *78*, 239-245.

Kim, D.H., Nelson, H.H., Wiencke, J.K., Zheng, S., Christiani, D.C., Wain, J.C., Mark, E.J., and Kelsey, K.T. (2001). p16(INK4a) and histology-specific methylation of CpG islands by exposure to tobacco smoke in non-small cell lung cancer. Cancer Res. *61*, 3419-3424.

Kim, W.J., and Lee, S.D. (2015). Candidate genes for COPD: current evidence and research. Int. J. Chron. Obstruct Pulmon. Dis. *10*, 2249-2255.

Lawless, M.W., O'byrne, K.J., and Gray, S.G. (2009). Oxidative stress induced lung cancer and COPD: opportunities for epigenetic therapy. J. Cell Mol. Med. *13*, 2800-2821.

Lee, H., Han, S., Rhee, H., Park, J.H., Lee, J.S., Oh, Y., Choi, S.S., Shin, S., and Kim, W.J. (2015). Differential expression of microRNAs and their target genes in non-small-cell lung cancer. Mol. Med. Rep. *11*, 2034-2040.

Lee, K.W., and Pausova, Z. (2013). Cigarette smoking and DNA methylation. Front. Genet. *4*, 132.

Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., and Young, N. (2013). The genotype-tissue expression (GTEx) project. Nat. Genet. *45*, 580.

Lundback, B., Lindberg, A., Lindstrom, M., Ronmark, E., Jonsson, A.C., Jonsson, E., Larsson, L.G., Andersson, S., Sandstrom, T., Larsson, K., et al. (2003). Not 15 but 50% of smokers develop COPD?-Report from the obstructive lung disease in northern sweden studies. Respir. Med. *97*, 115-122.

Mannino, D.M., and Buist, A.S. (2007). Global burden of COPD: risk factors, prevalence, and future trends. The Lancet *370*, 765-773.

Mannino, D.M., Ford, E.S., and Redd, S.C. (2003). Obstructive and restrictive lung disease and markers of inflammation: data from the third national health and nutrition examination. Am. J. Med. *114*, 758-762.

Novak, J.P., Sladek, R., and Hudson, T.J. (2002). Characterization of variability in large-scale gene expression data: implications for study design. Genomics *79*, 104-113.

Ocampo-Candiani, J., Salinas-Santander, M., Trevino, V., Ortiz-López, R., Ocampo-Garza, J., and Sanchez-Dominguez, C.N. (2018). Evaluation of skin expression profiles of patients with vitiligo treated with narrow-band UVB therapy by targeted RNA-seq. An. Bras. Dermatol. *93*, 843-851.

Piñero, J., Bravo, À., Queralt-Rosinach, N., Gutiérrez-Sacristán, A., Deu-Pons, J., Centeno, E., García-García, J., Sanz, F., and Furlong, L.I. (2017). DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Res. *4*, D833-D839.

Rabe, K.F., Hurd, S., Anzueto, A., Barnes, P.J., Buist, S.A., Calverley, P., Fukuchi, Y., Jenkins, C., Rodriguez-Roisin, R., and Van Weel, C. (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.

Rangasamy, T., Misra, V., Zhen, L., Tankersley, C.G., Tuder, R.M., and Biswal, S. (2009). Cigarette smoke-induced emphysema in A/J mice is associated with pulmonary oxidative stress, apoptosis of lung cells, and global alterations in gene expression. Am. J. Physiol.-lung C. *296*, L888-L900.

Repapi, E., Sayers, I., Wain, L.V., Burton, P.R., Johnson, T., Obeidat, M., Zhao, J.H., Ramasamy, A., Zhai, G., and Vitart, V. (2010). Genome-wide association study identifies five loci associated with lung function. Nat. Genet. *42*, 36.

Russell, K.E., Chung, K.F., Clarke, C.J., Durham, A.L., Mallia, P., Footitt, J., Johnston, S.L., Barnes, P.J., Hall, S.R., and Simpson, K.D. (2016). The MIF antagonist ISO-1 attenuates corticosteroid-insensitive inflammation and airways hyperresponsiveness in an ozone-induced model of COPD. PLoS One *11*, e0146102.

Sakao, S., Tatsumi, K., Hashimoto, T., Igari, H., Shino, Y., Shirasawa, H., and Kuriyama, T. (2003). Vascular endothelial growth factor and the risk of smoking-related COPD. Chest *124*, 323-327.

Sandford, A.J., Chagani, T., Weir, T.D., and Parè, P.D. (1998). α1-

antichymotrypsin mutations in patients with chronic obstructive pulmonary disease. Dis. Markers *13*, 257-260.

Schamberger, A.C., Mise, N., Meiners, S., and Eickelberg, O. (2014). Epigenetic mechanisms in COPD: implications for pathogenesis and drug discovery. Expert Opin. Drug Discov. *9*, 609-628.

She, W.B., Liu, X.S., Ni, W., Chen, S.X., and Xu, Y.J. (2012). The expression of macrophage migration inhibition factor in pulmonary tissues of smokers with or without chronic obstructive pulmonary disease. Zhonghua Nei Ke Za Zhi *51*, 863-866.

Smith, C.A., and Harrison, D.J. (1997). Association between polymorphism in gene for microsomal epoxide hydrolase and susceptibility to emphysema. The Lancet *350*, 630-633.

Spira, A., Beane, J.E., Shah, V., Steiling, K., Liu, G., Schembri, F., Gilman, S., Dumas, Y., Calner, P., and Sebastiani, P. (2007). Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. Nat. Med. *13*, 361.

Steiling, K., Van Den Berge, M., Hijazi, K., Florido, R., Campbell, J., Liu, G., Xiao, J., Zhang, X., Duclos, G., and Drizik, E. (2013). A dynamic bronchial airway gene expression signature of chronic obstructive pulmonary disease and lung function impairment. Am. J. Respir. Crit. Care Med. *187*, 933-942.

Stoller, J.K., and Aboussouan, L.S. (2005). α1-antitrypsin deficiency. The Lancet *365*, 2225-2236.

Stone, P.J., Calore, J.D., McGowan, S.E., Bernardo, J., Snider, G.L., and Franzblau, C. (1983). Functional alpha 1-protease inhibitor in the lower respiratory tract of cigarette smokers is not decreased. Science *221*, 1187-1189.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545-15550.

Suki, B., Bartolák-Suki, E., and Rocco, P.R. (2017). Elastase-induced lung emphysema models in mice. In: Borel F., Mueller C. (eds) Alpha-1 Antitrypsin Deficiency. (New York: Humana Press) pp. 67-75.

Swanney, M.P., Ruppel, G., Enright, P.L., Pedersen, O.F., Crapo, R.O., Miller, M.R., Jensen, R.L., Falaschetti, E., Schouten, J.P., Hankinson, J.L., et al. (2008). Using the lower limit of normal for the FEV1/FVC ratio reduces the misclassification of airway obstruction. Thorax *63*, 1046-1051.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. *7*, 562.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511.

Vestbo, J., Hurd, S.S., Agustí, A.G., Jones, P.W., Vogelmeier, C., Anzueto, A., Barnes, P.J., Fabbri, L.M., Martinez, F.J., and Nishimura, M. (2013). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am. J. Respir. Crit. Care Med. *187*, 347-365.

Wang, I., Stepaniants, S., Boie, Y., Mortimer, J.R., Kennedy, B., Elliott, M., Hayashi, S., Loy, L., Coulter, S., and Cervino, S. (2008). Gene expression profiling in patients with chronic obstructive pulmonary disease and lung cancer. Am. J. Respir. Crit. Care Med. *177*, 402-411.

Ward, L.D., and Kellis, M. (2015). HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Res. *44*, D877-D881.

Wu, G., Feng, X., and Stein, L. (2010). A human functional protein interaction network and its application to cancer data analysis. Genome Biol. *11*, R53.

Zhang, L., Li, S., Choi, Y., Lee, J., Gong, Z., Liu, X., Pei, Y., Jiang, A., Ye, M., and Mao, M. (2017). Systematic identification of cancer-related long noncoding RNAs and aberrant alternative splicing of quintuple-negative lung adenocarcinoma through RNA-Seq. Lung Cancer *109*, 21-27.