


RESEARCH

Open Access



RNA-sequencing across three matched tissues reveals shared and tissue-specific gene expression and pathway signatures of COPD

Jarrett D. Morrow^{1*} , Robert P. Chase¹, Margaret M. Parker¹, Kimberly Glass¹, Minseok Seo¹, Miguel Divo², Caroline A. Owen², Peter Castaldi¹, Dawn L. DeMeo^{1,2}, Edwin K. Silverman^{1,2} and Craig P. Hersh^{1,2}

Abstract

Background: Multiple gene expression studies have been performed separately in peripheral blood, lung, and airway tissues to study COPD. We performed RNA-sequencing gene expression profiling of large-airway epithelium, alveolar macrophage and peripheral blood samples from the same subset of COPD cases and controls from the COPDGene study who underwent bronchoscopy at a single center. Using statistical and gene set enrichment approaches, we sought to improve the understanding of COPD by studying gene sets and pathways across these tissues, beyond the individual genomic determinants.

Methods: We performed differential expression analysis using RNA-seq data obtained from 63 samples from 21 COPD cases and controls (includes four non-smokers) via the R package DESeq2. We tested associations between gene expression and variables related to lung function, smoking history, and CT scan measures of emphysema and airway disease. We examined the correlation of differential gene expression across the tissues and phenotypes, hypothesizing that this would reveal preserved and private gene expression signatures. We performed gene set enrichment analyses using curated databases and findings from prior COPD studies to provide biological and disease relevance.

Results: The known smoking-related genes *CYP1B1* and *AHRR* were among the top differential expression results for smoking status in the large-airway epithelium data. We observed a significant overlap of genes primarily across large-airway and macrophage results for smoking and airway disease phenotypes. We did not observe specific genes differentially expressed in all three tissues for any of the phenotypes. However, we did observe hemostasis and immune signaling pathways in the overlaps across all three tissues for emphysema, and amyloid and telomere-related pathways for smoking. In peripheral blood, the emphysema results were enriched for B cell related genes previously identified in lung tissue studies.

Conclusions: Our integrative analyses across COPD-relevant tissues and prior studies revealed shared and tissue-specific disease biology. These replicated and novel findings in the airway and peripheral blood have highlighted candidate genes and pathways for COPD pathogenesis.

Keywords: COPD, Emphysema, Genomics, RNA-seq, Transcriptomics

* Correspondence: jarrett.morrow@channing.harvard.edu

¹Channing Division of Network Medicine, Brigham and Women's Hospital, 181 Longwood Avenue, Boston, MA 02115, USA

Full list of author information is available at the end of the article



Background

Chronic obstructive pulmonary disease (COPD) is characterized by progressive airflow obstruction accompanied by chronic inflammation. It is a major and growing cause of morbidity and mortality worldwide [1]. Although environmental exposures such as cigarette smoking are risk factors, a genetic component to susceptibility has been observed [2–5]. Genomic regions influencing COPD susceptibility have been identified at multiple loci through genome-wide association studies [6–12]. Airway inflammation and remodeling and emphysematous destruction in the lung contribute to disease severity and progression [13, 14], with macrophage activity having an important role [15, 16]. The recapitulation of these gene expression signals in peripheral blood remains elusive. However, gene expression in blood has been used as proxy in identification of COPD subtypes [17], and peripheral blood gene expression underlines the systemic effects of COPD inflammation [18–20].

Several published COPD studies have performed microarray gene expression profiling [21]. Specifically, studies in the airway epithelium have focused on expression changes related to smoking [22, 23] and COPD status [24, 25], including targeted RNA-seq profiling [26]. Studies of gene expression in peripheral blood have also focused on COPD [19, 27, 28] and smoking [29, 30], including RNA-seq profiling [31]. Given the putative role macrophages have in inflammatory lung disease [32], gene expression profiling of these cells has been performed in the context of COPD [33] and smoking [34]. In addition to the airway studies, there have also been several COPD and emphysema gene expression studies involving resected lung tissue [35–39], including RNA-seq profiling in a cohort of males [40] and RNA-seq profiling of early COPD and emphysema in males [41].

Despite the volume of this previous work, the expression signatures for alveolar macrophages, bronchial epithelium, and peripheral blood have not previously been studied within the same population at the same time. However, gene expression in nasal and bronchial brushing samples from the same subjects has been compared [42]. Another study of nasal and bronchial gene expression was performed in independent cohorts [43]. A study of lung tissue, small airway, and peripheral blood gene expression, with tissue samples obtained from separate cohorts, involved both emphysema and lung function phenotypes [44]. Overlapping gene expression signatures have been studied in alveolar macrophages and peripheral monocytes isolated from separate cohorts [45]. Gene expression signatures have been explored across many tissues in the Genotype-Tissue Expression (GTEx) project [46], leveraging network methods to identify tissue-specific gene and transcription factor

regulation [47–49] and examining the overall blood-lung gene expression overlap [50].

The foundation of this study is the integration of RNA-seq profiling across three COPD-relevant tissues from the same COPDGene (Genetic Epidemiology of COPD) study subjects, mitigating variation typically seen when studying tissue samples from different subjects. Gene expression in the airway epithelium, alveolar macrophages and whole blood samples were tested for association with measures of lung function, airway disease, emphysema severity and cigarette smoke exposure. Given data across three tissues and 11 phenotype variables, we believed a comprehensive hypothesis could not be the goal. Instead, highlighting private and overlapping gene signatures when present was the more effective approach. Using statistical methods and a gene set enrichment framework, we sought to detect expression signatures across the tissues, highlighting systemic and tissue-specific signatures of lung disease and damage. By integrating these findings with previous COPD lung tissue studies and a recent COPD Genome-wide Association Study (GWAS), we sought to place our results in the context of lung disease biology and shed light on the functional role of genes previously identified at genome-wide significant COPD GWAS loci. Similar integration approaches have been previously applied in COPD [43, 44, 51]. Systems biology has the potential to reveal the molecular architecture of complex traits and disease [52] in part by examining broad biological information rather than individual genomic determinants. We hypothesized that this systems biology study would inform blood biomarker identification, motivate hypotheses regarding the systemic functions of lung disease, and potentially identify novel genes and pathways for COPD and emphysema, as targets for functional, translational and diagnostic studies.

Methods

Study subjects and bronchoscopy procedure

Subjects were enrolled in the COPDGene study [53] and participated in the five-year follow-up phase. COPDGene is a longitudinal cohort study that includes non-Hispanic Whites and African Americans enrolled at 21 centers across the United States. The subjects include a small number of non-smokers and more than 10,000 current and former cigarette smokers with a minimum 10 pack-years smoking history. Cases have airflow obstruction ($FEV_1/FVC < 0.7$) and control subjects had normal spirometry ($FEV_1\%$ predicted $\geq 80\%$ and $FEV_1/FVC \geq 0.7$). Subjects returned for the follow-up visit approximately 5 years after enrollment. At this second phase visit, the subjects completed questionnaires and underwent pre- and post-bronchodilator spirometry, volumetric computed tomography (CT) of the chest,

and had blood drawn for a complete blood cell count and biomarker studies. Emphysema severity was quantified via image analysis of chest CT data as the percentage of lung voxels below -950 HU [54].

A single physician performed flexible bronchoscopy on all subjects in the substudy, using intravenous sedation and topical anesthesia. Bronchoalveolar lavage (BAL) was performed in two lung segments with 60 ml of normal saline in each segment, and the BAL fluid obtained from each subject was pooled. Large airway brushings were performed in the right mainstem bronchus and placed in RLT buffer (Qiagen). Whole blood was collected in a PaxGene RNA tube on the day of bronchoscopy. Study subjects provided separate written informed consent for the bronchoscopy study, which was approved by the institutional review board at Partners Healthcare.

Differential gene expression

We performed differential gene expression (DGE) analysis in each tissue individually using the R/Bioconductor package DESeq2 [55], testing associations between transcript expression levels and lung function, emphysema, smoking and airway disease phenotypes. The base statistical model included the covariates age, sex, race, pack-years of smoking, a categorical variable for smoking status, and RIN (RNA Integrity Number). For analysis of emphysema variables, BMI was included as a covariate. We controlled for cell distribution in peripheral blood using the covariates: white blood cell (WBC) count, and the percentages of neutrophils, lymphocytes, monocytes and eosinophils. A summary of all models is provided in Table 1. Latent effects were addressed using

surrogate variables as covariates. These were obtained using the function svaseq in the R/Bioconductor package sva [56]. Prior to svaseq processing, coarse filtering was performed by excluding transcripts with an average count per sample of less than one. Only the surrogate variables lacking a statistically significant association with the phenotype variable of interest were included as covariates. For the differential expression analysis, adjustment for multiple testing controlled for false discovery rate (FDR). In this FDR calculation, the method DESeq2 excludes transcripts with mean normalized counts across all samples that are below a set threshold. By default, this threshold maximizes the number of significant results found at a user-specified FDR. The FDR was chosen to be 10% for this study, as the smaller sample size dictates a value greater than 5%. Results lacking an adjusted p -value (NA/not available) are not statistically significant, as they represent genes with mean normalized counts below the threshold. For the continuous variables, the log₂ fold change is per unit of change of that variable.

Results

Differential gene expression

RNA-seq data were available in one or more of the three tissues for 39 subjects, encompassing 94 total samples (Additional file 1: Figure S1). This study focused on the 21 subjects having data available in all three tissues (see Additional file 2: Methods). We performed RNA-seq profiling followed by pathway and enrichment analysis (Fig. 1) on the 63 samples from these 21 subjects (6 COPD cases and 15 controls; Additional file 3: Table S1). We observed greater clustering by tissue than by individual. Emphysema

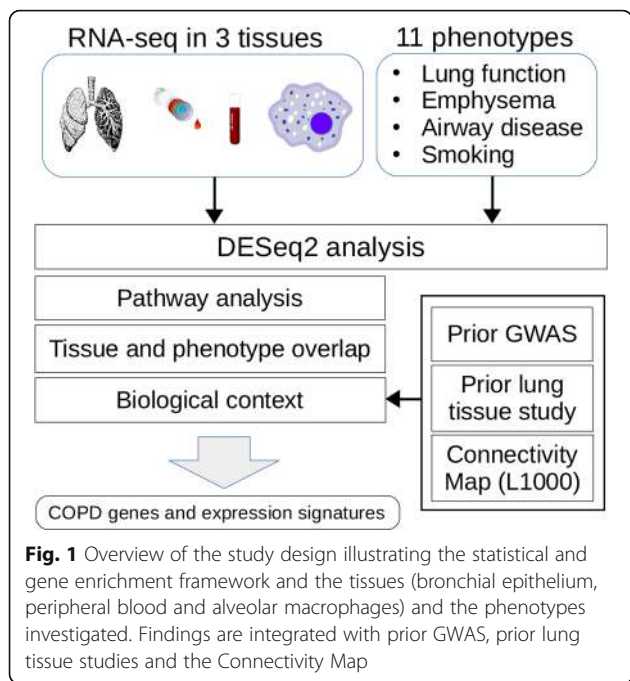
Table 1 Models for expression association with outcomes of interest

Phenotype category	Variable of interest	Model
Lung function	COPD case-control	EXP ~ COPD + age + sex + race + pack-years + smoking + RIN + SVs
	FEV1% predicted	EXP ~ FEV1.PP + age + sex + race + pack-years + smoking + RIN + SVs
	FEV1/FVC	EXP ~ FEV1FVC + age + sex + race + pack-years + smoking + RIN + SVs
Smoking	pack-years	EXP ~ pack-years + age + sex + race + RIN + SVs
	Smoking status	EXP ~ smoking + age + sex + race + RIN + SVs
Emphysema	pctEmph	EXP ~ pctEmph + age + sex + race + BMI + pack-years + smoking + RIN + SVs
	perc15	EXP ~ perc15 + age + sex + race + BMI + pack-years + smoking + RIN + SVs
	adj_density	EXP ~ adj_density + age + sex + race + BMI + pack-years + smoking + RIN + SVs
Airway disease	Pi10	EXP ~ Pi10 + age + sex + race + pack-years + smoking + RIN + SVs
	AWT	EXP ~ AWT + age + sex + race + pack-years + smoking + RIN + SVs
	WallAreaPct	EXP ~ WallAreaPct + age + sex + race + pack-years + smoking + RIN + SVs

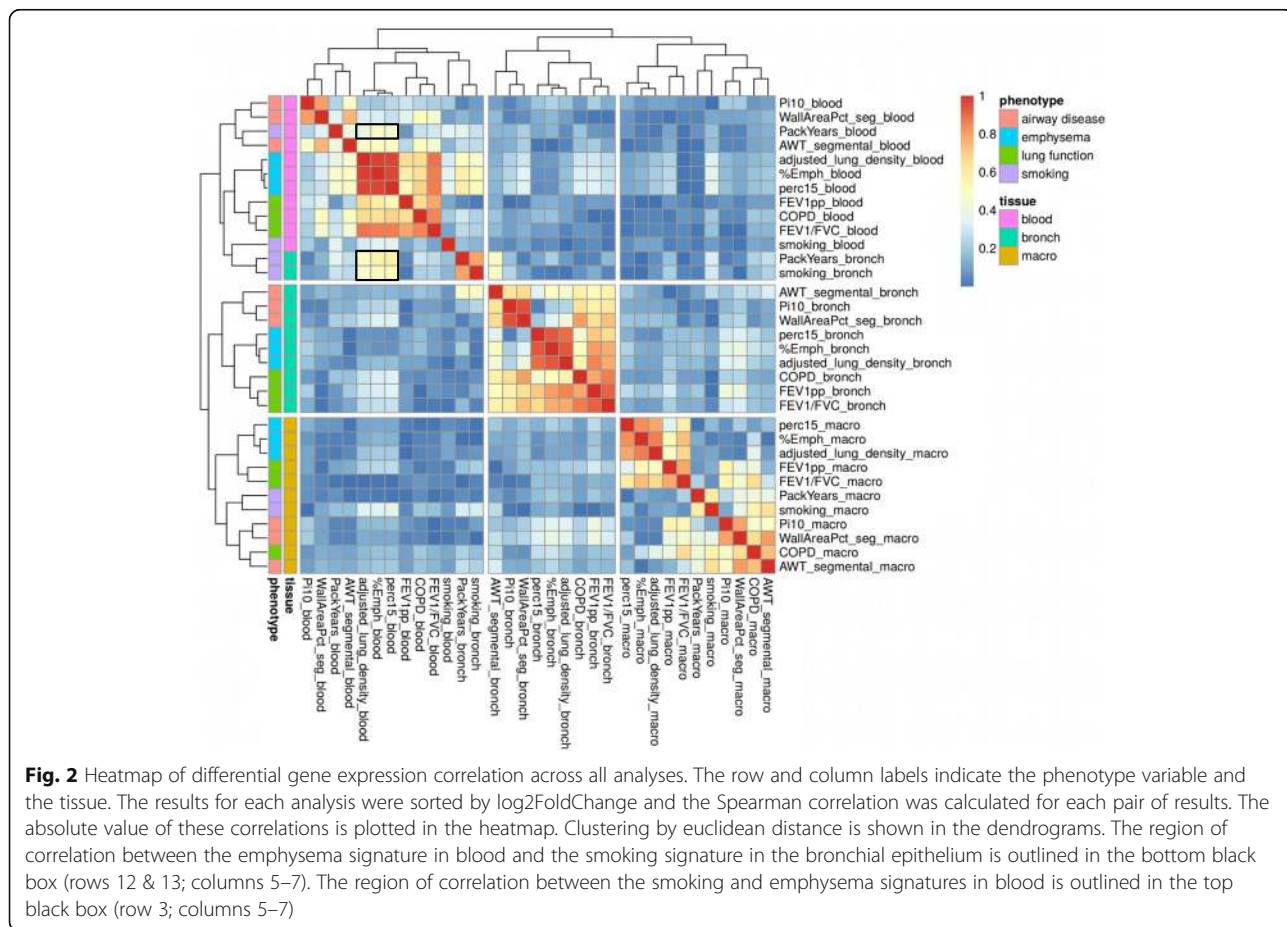
Models for peripheral blood also included: WBC, and the percentages of neutrophils, lymphocytes, monocytes and eosinophils

smoking (ordinal variable): 0 non-smoker, 1 former smoker, 2 current smoker

Abbreviations: SVs = surrogate variables; FEV1 = forced expiratory volume in 1 s; FVC = forced vital capacity; pctEmph = % emphysema; perc15 = 15th percentile of lung density histogram; adj_density = adjusted lung density, sponge model adjustment; Pi10 = SRWA-Pi10 = square root wall area of a hypothetical airway with 10 mm internal perimeter; AWT = airway wall thickness; WallAreaPct = wall area percent; EXP = gene expression values; WBC = white blood cell count



and airway data from CT scans are available for a subset of the subjects. Subjects who never smoked were excluded from analyses, except analyses for the two models assessing the impact of smoking. Association between gene expression and 11 phenotype variables (Table 1) was tested in each tissue. We viewed the overlapping gene expression signatures of all results included in the DESeq2 FDR calculations (Additional file 3: Table S2), using a correlation heatmap (Fig. 2). Within this heatmap, we observed clustering by both tissue and phenotype, presenting as blocks of higher correlation. Specifically, the clustering by phenotype variable (mirroring our phenotype category grouping) is nested within the clustering by tissue. However, some smoking results in bronchial epithelium grouped with the peripheral blood module (top left in Fig. 2). Delving into this heatmap reveals this clustering is driven by the correlation between the emphysema signature in blood and the smoking signature in the bronchial epithelium (rows 12 & 13; columns 5–7 within black outline, Fig. 2). These signatures reside within the complete differential expression results for the specific tissue and phenotype. Another feature of note is the clustering of the smoking signature in blood with the airway-disease signature in blood, driven by the correlation between the smoking and emphysema



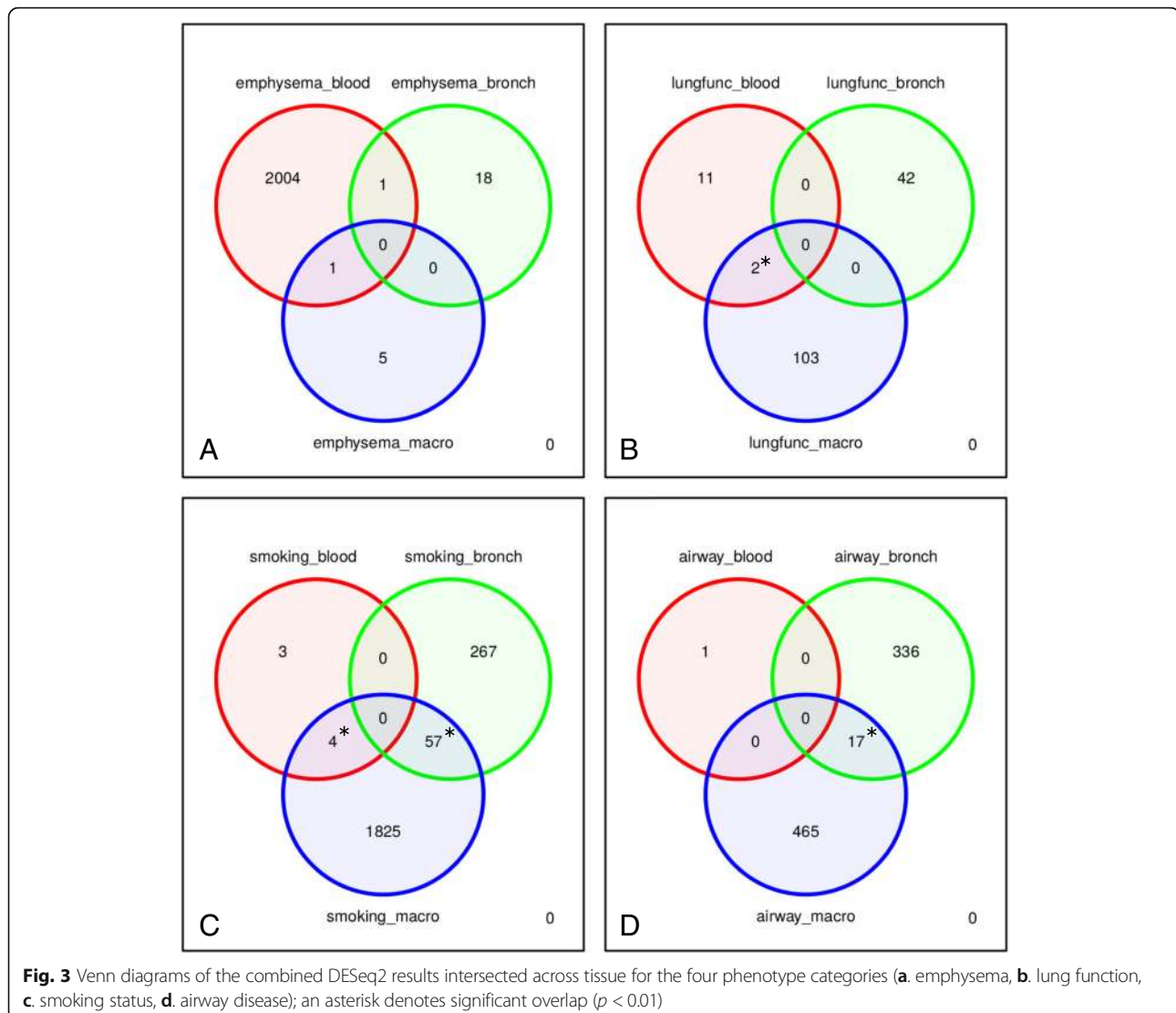
signatures in blood (row 3; columns 5–7 within black outline, Fig. 2).

The number of significant (q -value < 0.1) results across all 33 analyses (11 variables across 3 tissues) varied from zero to 1886 (Additional file 3: Table S3). Since log₂ fold change depends on the units for each variable, we did not apply a set fold change filter when determining significance. Of these 33 sets of results, 26 contained at least one significant result (Additional file 3: Tables S4–S6); the intersection pattern across the results was complex and mixed (Additional file 1: Figure S2). Next, to enhance the gene expression signal for each phenotype, we combined the significant (q -value < 0.1) genes across the phenotype variables in each of the four phenotype categories (lung function, airway disease, emphysema severity and cigarette smoke exposure), retaining the unique genes in each category. Venn diagrams highlight any cross-tissue intersections of these

combined results (Fig. 3, Additional file 3: Table S7). We performed hypergeometric tests of gene enrichment across the tissues within each phenotype category (Fig. 3 and Additional file 3: Table S8). The backgrounds in these tests were the unique common genes across each pair of tissues having an average of at least two reads in each tissue. We observed statistically significant enrichment primarily across bronchial epithelium and alveolar macrophages. Although there were statistically meaningful overlaps in the macrophage and blood sets, the number of intersecting genes was less than five. We did not observe genes differentially expressed in all three tissues for any of the four phenotype categories.

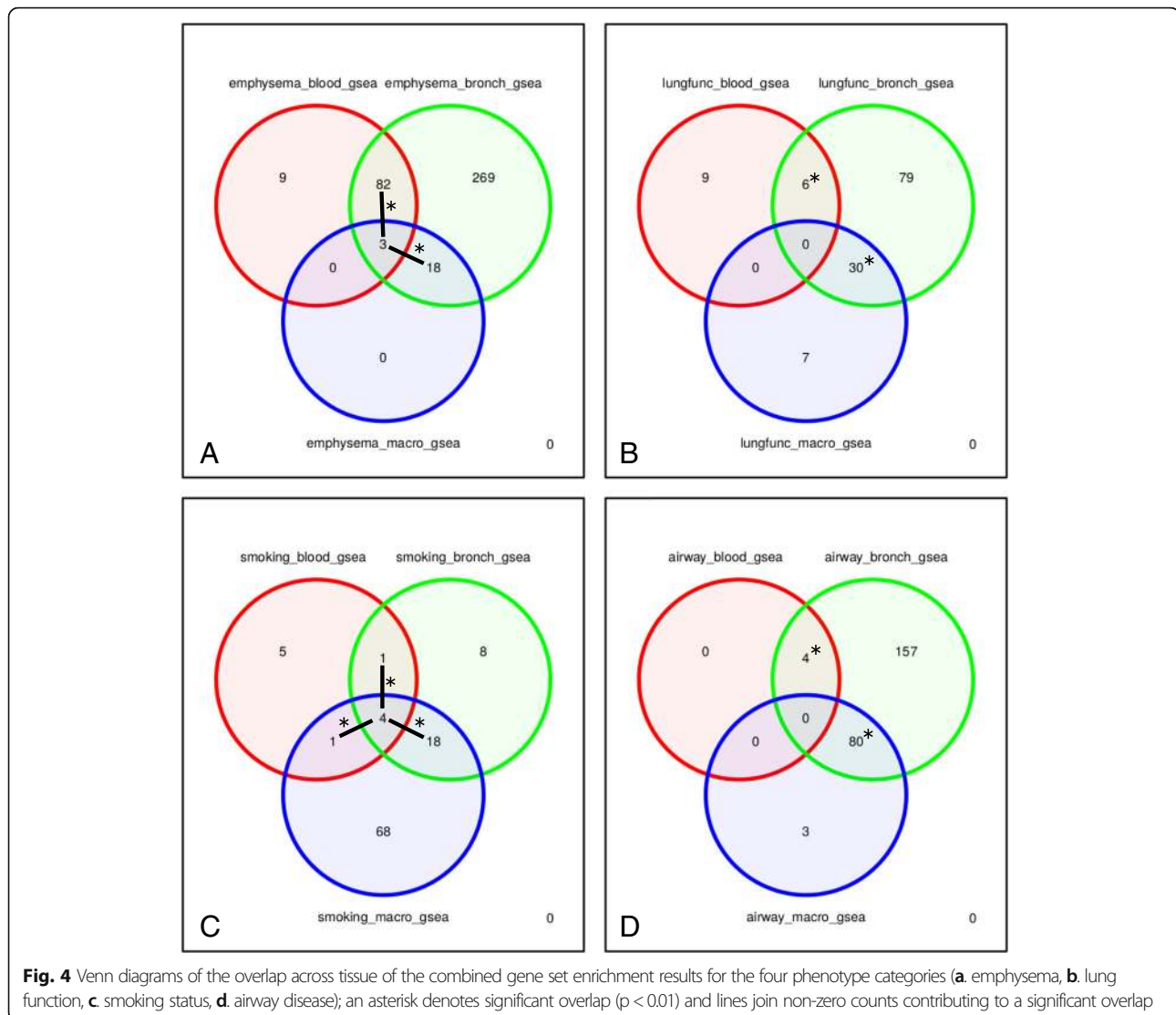
Enrichment and signature analyses

We performed pathway analyses on the results for each phenotype association using gene set enrichment, and combined the significant (q -value < 0.05) findings across



the four phenotype categories (Fig. 4, Additional file 3: Table S9). In contrast to the results for individual genes, we did observe pathway overlaps across all three tissues for emphysema and smoking. We performed hypergeometric tests of pathway enrichment across the tissues within each phenotype category (Fig. 4 and Additional file 3: Table S8). The background for these tests were the set of pathways tested. We observed statistically significant enrichment across two or more pairs of tissues for each of the phenotype categories. To provide a graphical pathway summary, we created an Enrichment Map in Cytoscape using the overlapping pathways for blood and bronchial epithelium in emphysema (Additional file 1: Figure S3). We observed network modules characterized by metabolic, cancer, and immune signaling pathways, with smaller groups containing adherens junction and focal adhesion pathways.

To further assess disease relevance, we performed gene set enrichment tests to compare the differential gene expression signatures with findings from three previous studies (see Additional file 2: Methods). The set of significant (q -value < 0.1) genes from the analysis of airway phenotypes in the bronchial epithelium was included as a disease tissue reference. We summarized these enrichment findings for both up- and down-regulated genes in a p -value heatmap (Fig. 5). From the heatmap, we observed enrichment of the bronchial epithelium airway-disease genes in macrophage results across all four phenotype categories. Within the integration with previous studies (lung tissue gene expression, lung tissue DNA methylation and COPD GWAS), we observed enrichment of lung tissue COPD and emphysema genes in our bronchial epithelium results for both lung function and emphysema. The down-regulated lung tissue genes were found enriched



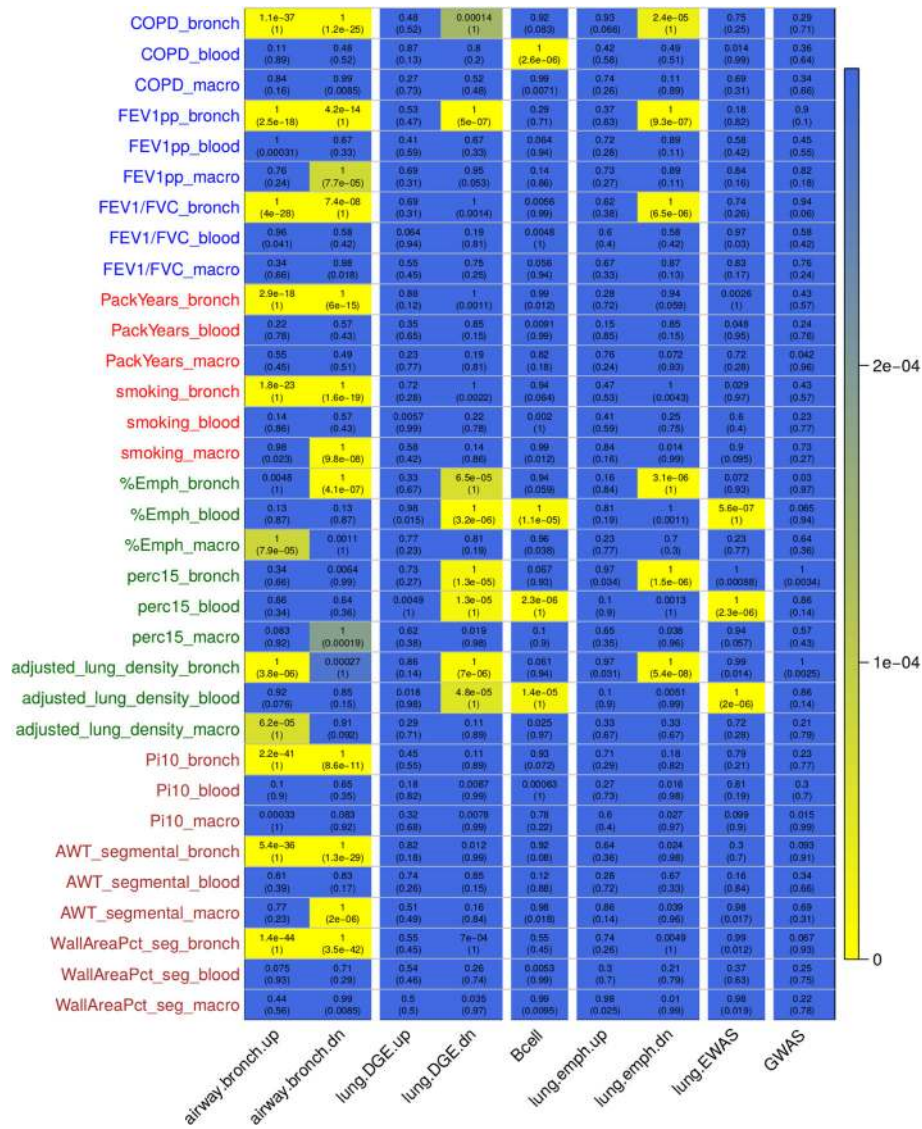


Fig. 5 Heatmap summary of *p*-values from gene set enrichment tests using a set of significant (*q*-value < 0.1) airway disease results in the bronchial epithelium, and findings from previous GWAS and lung tissue studies. The top *p*-value corresponds to enrichment test in the up-regulated genes and the bottom (*p*-value) refers to enrichment in down-regulated genes. The row labels are color-coded by phenotype category (blue = lung function, red = smoking; green = emphysema, brown = airway)

in the genes up-regulated in bronchial epithelium by disease status. We also observed enrichment of our previously published COPD-associated B cell lung tissue expression module [38] and lung tissue DNA methylation genes [57] within the emphysema results for peripheral blood. We did not find enrichment for the lung emphysema genes in these phenotype variables nor did we find enrichment of the B cell module in the bronchial epithelium or macrophage results. We extracted the significant (*q*-value < 0.1) differential expression genes intersecting the external gene sets (Additional file 3: Table S10). These enrichment experiments provided lung disease context by linking to gene expression and epigenetic signatures of COPD in lung tissue.

We queried the Connectivity Map (CMap CLUE) using our bronchial epithelium results for airway disease. The perturbagens of interest have negative scores, as these signatures demonstrate reversal of disease severity for both the A549 cell line (Additional file 3: Table S11) and HCC515 cell line (Additional file 3: Table S12). The top chemical perturbagen from the A549 results was lomerizine and the top result for HCC515 was ephedrine.

Discussion

We integrated RNA-sequencing across three matched COPD-relevant tissues using tests of association with

lung function, airway disease, emphysema severity and cigarette smoke exposure, and a gene set enrichment framework. This has revealed expression signatures across the tissues in the context of each phenotype, highlighting systemic and tissue-specific signatures and pathways of lung disease and damage. We did not observe any genes differentially expressed across all three tissues. However, we did find pathways overlapping the three tissues in emphysema and smoking. Disease relevance and biology were elucidated through integration with previous COPD lung tissue studies and a recent COPD GWAS.

Replication of airway differential gene expression

Our top two results from the differential expression analysis of smoking status in the bronchial epithelium were *CYP1A1* (cytochrome P450 family 1 subfamily A member 1) and *CYP1B1* (cytochrome P450 family 1 subfamily A member 1). These replicate previous findings in studies of smoking in the airway [23, 58] and oral mucosa [59], with *CYP1B1* also identified in the lung [60]. Significant in our analysis of smoking was *AHRR* (aryl-hydrocarbon receptor repressor), previously found to be differentially expressed by smoking status in lung tissue [60] and in the oral mucosa [59]. Both *CYP1B1* and *AHRR* were also significant in our analysis of smoking status in macrophages, and Poliska et al. also found *CYP1B1* correlated with COPD status in alveolar macrophages [45].

In our bronchial epithelium analysis of airway disease, the genes *CLDN10* (claudin 10), *TMEM2* (*CEMIP2* - cell migration inducing hyaluronidase 2) and *ALDH1A3* (aldehyde dehydrogenase 1 family member A3) were significant across the three airway-disease variables. *CLDN10* is believed to have a role in idiopathic pulmonary fibrosis (IPF) progression [61]. A gene-by-environmental tobacco smoke interaction study on the level of FEV1 identified a locus intronic to the gene *TMEM2* [62] and *TMEM2* was previously associated with lung function in the small airway [44]. Last, the gene *ALDH1A3* was found to be differentially expressed by smoking status in both the bronchial and nasal epithelium [42].

The top gene in our bronchial epithelium analysis of percent emphysema was *APOD* (apolipoprotein D), a gene found differentially expressed in a study of emphysema severity and bronchiolitis [37]. The second gene in this emphysema analysis was *CYP2A6* (cytochrome P450 family 2 subfamily A member 6) from a locus previously identified in GWAS of smoking behavior [63] and COPD [7]. These replications suggest a link to smoking-related lung disease and progression, with relevance throughout the respiratory tract.

Pathways overlap across tissues

We observed a mixed and complex overlap pattern of significant genes across all differential expression results. To better glean information from the overlaps, we focused on private and cross-tissue signatures. We combined the differential expression and pathway results across phenotype variables, based on our observations of clustering by phenotype categories in the correlation heatmap. In this context, we observed statistically significant enrichment primarily across the bronchial epithelium and alveolar macrophages. We did not observe genes differentially expressed in all three tissues for any of the four phenotype categories. However, for emphysema and smoking we did observe pathway overlaps across all three tissues. We also observed statistically significant pathway overlaps across pairs of tissues in each of the four phenotype categories. In emphysema, the pathways at the three-tissue intersection were related to hemostasis and immune signaling, both markers of systemic inflammation. The three-tissue overlap for smoking included amyloid and telomere related pathways. This is concordant with observations of amyloids as putative biomarkers of systemic inflammation and COPD [64] and the association between lung disease, lung aging, and telomere length [65].

In addition to the three-tissue intersections, the robust two-tissue pathway overlap in airway disease for the bronchial epithelium and macrophages appears to be localized with signatures of oxidative stress, highlighted by enrichment of nonsense mediated decay and metabolic pathways. The cell-cycle pathways also present in this overlap are suggestive of cellular senescence mechanisms [66, 67], particularly given the findings in emphysema for these cells [68]. A differentially expressed gene observed at this intersection and the bronchial epithelium and macrophage intersection for smoking was *SCGB1A1* (secretoglobulin family 1A member 1). This gene is expressed at high levels in the lung and encodes for CC16 (Club Cell Secretory Protein) a blood biomarker of COPD [69, 70].

Another significant pathway overlap was observed between blood and bronchial epithelium in emphysema, characterized by clusters of metabolic, cancer, and immune signaling pathways, with adherens junction and focal adhesion pathways also present. These pathways highlight signals of structural damage and systemic immune response in airway disease and emphysema [14, 71]. The significantly differentially expressed gene at this intersection was *FCNI* (ficolin 1), a gene found to be differentially expressed in peripheral blood in mild IPF [72]. In addition, functional polymorphic sites in the promoter region of *FCNI* regulate ficolin-1 expression and influence outcomes during systemic inflammation [73].

Airway signatures overlap in blood and recapitulate in lung

We observed clustering of the smoking signature in the bronchial epithelium with the smoking signature in blood in our differential expression correlation heatmap, brought about by the relationship between the emphysema signature in blood and the smoking signature in the bronchial epithelium. We also observed clustering of the smoking signature in blood with the airway-disease signature in blood, owing to the correlation between the smoking and emphysema signatures in blood. Taken together, these suggest a common and systemic marker of emphysema with a gene expression signature of smoke-induced damage [18, 44, 74].

We integrated our differential expression results with findings from previous COPD studies, along with the significant bronchial epithelium results in airway disease. We observed enrichment of the bronchial epithelium airway-disease genes in macrophage results across all four phenotype categories. This was in line with the findings when we intersected the significant genes and pathways for these tissues. Lung tissue COPD and emphysema genes were enriched in our bronchial epithelium results for both the lung function and emphysema phenotype categories, demonstrating disease relevance in lung tissue. We found that the down-regulated lung tissue genes were found enriched in the genes up-regulated in the bronchial epithelium by disease status. Although not an equivalent comparison, this finding is similar in nature to that of Obeidat et al., [44], where the lung tissue and blood gene expression directionality was opposite across their two tissue cohorts for a majority of the genes of interest.

Within the emphysema results for peripheral blood, we also observed enrichment of the COPD-relevant lung tissue B cell expression module [38] and DNA methylation [57] gene sets. The direction of effect for these enrichments were concordant with respect to disease status. The methylation directionality relationship is more difficult to resolve given the various gene regulation mechanisms [75]. We did not find enrichment for the lung emphysema genes in these phenotype variables. Overall, this suggests a systemic B cell signature observed previously in the lung [38], recapitulated here in peripheral blood. The significant gene at B cell module intersection with the bronchial epithelium results for COPD was *CD28* (*CD28* molecule). This gene may play a role in immunologic senescence [71] and COPD inflammation [76], owing to its role as a co-stimulatory molecule, constitutively expressed by naïve T cells and required for full activation (and survival) of T cells.

Reversal of bronchial epithelium disease signature

Using Connectivity Map, we identified perturbing compounds that produce gene expression signatures in two lung cell lines opposing the disease gene expression signature we observed in the bronchial epithelium. The top chemical pertubagen from the A549 results was lomerizine, a calcium channel blocker, suggesting potential drug repurposing. Others on the list include glucocorticoid receptor agonists, used in the treatment of inflammatory lung diseases [77] through their activation of specific glucocorticoid receptor mechanisms. Among the HCC515 compounds, fluticasone is a current therapeutic for treatment of respiratory disease [78], and as the top result for HCC515, ephedrine is a known bronchodilator.

Some limitations to the current study involve blood and bronchial epithelium cellular heterogeneity. We have partially addressed the heterogeneity in blood using the measured leukocyte counts. However, remaining variation (e.g. lymphocyte composition) may influence the gene expression signatures, as *GPR15* was differentially expressed in our smoking analysis in blood and was found to be expressed in a T cell dependent manner with cigarette smoking [79]. Single cell or single cell type sequencing will better resolve specific gene expression signatures. We have not addressed the polarization of the alveolar macrophages, that increases with COPD severity and cigarette smoke exposure [80]. The study of early and intermediate phenotypes of COPD would help to link the temporal changes in the tissue gene expression overlap with disease progression, as would longitudinally repeated gene expression experiments. Last, despite the use of RNA-seq to improve the resolution of gene expression signatures and use of gene set enrichment to extract signals from all results, our sample size does limit our power to detect these signatures. Given this limited power, our focus was not on the identification of specific biomarkers. Future work will involve larger study cohorts with greater power to also resolve individual biomarkers.

Conclusions

In this integrative genomics study, we have performed RNA-seq profiling of gene expression in three matched COPD-relevant tissues. Using statistical and gene set enrichment methods, we have identified overlapping differentially expressed genes and pathways across the tissues, providing lung disease biomarker insight. We observed no common genes across all three tissues. However, we did observe shared pathways across all three. By integrating the gene expression profiles with previous COPD findings to provide additional disease context, we identified a lung disease signature in our emphysema results in the bronchial epithelium and peripheral blood, while

also suggesting recapitulation of a systemic B cell lung signature in peripheral blood. Together this hints that peripheral blood has the potential to capture relevant lung pathobiology. Connectivity Map provided some translational context, identifying known and putative compounds that elicit a gene expression signature in lung cell lines that opposes the disease signature we observed in the bronchial epithelium.

Additional files

Additional file 1: Supplemental Figures: Additional figures to support the findings of this study. (PDF 780 kb)

Additional file 2: Supplemental Methods. Additional detail regarding the methods in this study. (PDF 56 kb)

Additional file 3: Supplemental Tables: Additional tables to support the findings of this study. **Table S4.** Differential gene expression results in bronchial epithelium. **Table S5.** Differential gene expression results in alveolar macrophages. **Table S6.** Differential gene expression results in peripheral blood. (ZIP 674 kb)

Abbreviations

AHRR: Aryl-hydrocarbon receptor repressor; *ALDH1A3*: Aldehyde dehydrogenase 1 family member A3; *APOD*: Apolipoprotein D; BAL: Bronchoalveolar Lavage; CC16: Club Cell Secretory Protein; *CLDN10*: Claudin 10; CMap: Connectivity Map; COPD: Chronic Obstructive Pulmonary Disease; COPDGene: Genetic Epidemiology of COPD; CT: Computed Tomography; *CYP1A1*: Cytochrome P450 family 1 subfamily A member 1; *CYP1B1*: Cytochrome P450 family 1 subfamily A member 1; *CYP2A6*: Cytochrome P450 family 2 subfamily A member 6; DGE: Differential Gene Expression; *FCN1*: Ficolin 1; FDR: False Discovery Rate; GTEx: Genotype-Tissue Expression; GWAS: Genome-wide Association Study; IPF: Idiopathic Pulmonary Fibrosis; RIN: RNA Integrity Number; *SCGB1A1*: Secretoglobulin family 1A member 1; *TMEM2*: *CEMIP2* - cell migration inducing hyaluronidase 2; WBC: White Blood Cell

Acknowledgements

Not applicable

Funding

Funding: NIH grants K25 HL136846, P01 HL105339, P01 HL114501, R01 HL125583, R01 HL130512, P01 HL132825, R01 HL089856, R01 HL089897.

Availability of data and materials

The dataset supporting the conclusions of this article is available in the GEO repository (GSE124180):

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE124180>

Authors' contributions

JM: concept and design, analysis and interpretation of data, manuscript preparation, approval of final manuscript. RC: analysis and interpretation of data, approval of final manuscript. MP: analysis and interpretation of data, approval of final manuscript. KG: analysis and interpretation of data, statistical support, approval of final manuscript. MS: analysis and interpretation of data, statistical support, approval of final manuscript. MD: acquisition of data, approval of final manuscript. CO: acquisition of data, approval of final manuscript. PC: analysis and interpretation of data, approval of final manuscript. DD: analysis and interpretation of data, manuscript preparation, approval of final manuscript. ES: concept and design, acquisition of data, manuscript preparation, approval of final manuscript. CH: concept and design, acquisition of data, analysis and interpretation of data, manuscript preparation, approval of final manuscript.

Ethics approval and consent to participate

Written informed consent was obtained from each subject and the study was approved by the institutional review boards at Partners Healthcare.

Consent for publication

Not applicable

Competing interests

Drs. Morrow, Chase, Parker, Glass, Seo, and Divo declare that they have no competing interests related to this manuscript.

Dr. Owen is currently an employee of Vertex Pharmaceuticals Inc., Boston, MA but has no competing interests related to this manuscript.

Dr. Castaldi has received consulting fees and grant support from GSK.

Dr. DeMeo has received compensation from Novartis.

In the past three years, Edwin K. Silverman received honoraria from Novartis for Continuing Medical Education Seminars and grant and travel support from GlaxoSmithKline.

Dr. Hersh has received consulting fees from AstraZeneca, Concert Pharmaceuticals, Mylan, and 23andMe and grant support from Boehringer Ingelheim and Novartis.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Channing Division of Network Medicine, Brigham and Women's Hospital, 181 Longwood Avenue, Boston, MA 02115, USA. ²Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA.

Received: 6 January 2019 Accepted: 25 March 2019

Published online: 02 April 2019

References

- Vestbo J, Hurd SS, Agustí AG, Jones PW, Vogelmeier C, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med.* 2013;187(4):347–65.
- Hersh CP, Hokanson JE, Lynch DA, Washko GR, Make BJ, Crapo JD, et al. Family history is a risk factor for COPD. *Chest.* 2011;140(2):343–50.
- McCloskey S, Patel B, Hinchliffe S, Reid E, Wareham N, Lomas D. Siblings of patients with severe chronic obstructive pulmonary disease have a significant risk of airflow obstruction. *Am J Respir Crit Care Med.* 2001;164:1419–24.
- Silverman EK. Genetics of Chronic Obstructive Pulmonary Disease. In: Chadwick D, Goode JA, editors. *Chronic Obstructive Pulmonary Disease: Pathogenesis to Treatment.* New Jersey: Wiley; 2000. p. 45–64.
- Silverman EK, Chapman HA, Drazen JM, Weiss ST, Rosner B, Campbell EJ, et al. Genetic epidemiology of severe, early-onset chronic obstructive pulmonary disease. Risk to relatives for airflow obstruction and chronic bronchitis. *Am J Respir Crit Care Med.* 1998;157(6 Pt 1):1770–8.
- Cho MH, Boutaoui N, Klanderman BJ, Sylvia JS, Ziniti JP, Hersh CP, et al. Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat Genet.* 2010;42(3):200–2.
- Cho MH, Castaldi PJ, Wan ES, Siedlinski M, Hersh CP, Demeo DL, et al. A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. *Hum Mol Genet.* 2012;21(4):947–57.
- Cho MH, McDonald M-LN, Zhou X, Mattheisen M, Castaldi PJ, Hersh CP, et al. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *Lancet Respir Med.* 2014;2(3):214–25.
- Cho MH, Castaldi PJ, Hersh CP, Hobbs BD, Barr RG, Tal-Singer R, et al. A genome-wide association study of emphysema and airway quantitative imaging phenotypes. *Am J Respir Crit Care Med.* 2015;192(5):559–69.
- Hobbs BD, de Jong K, Lamontagne M, Bossé Y, Shrine N, Artigas MS, et al. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet.* 2017;49(3):426–32.
- Sakornsakolpat P, Prokopenko D, Lamontagne M, Reeve NF, Guyatt AL, Jackson VE, et al. Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. *Nat Genet.* 2019;51(3):494–505.
- Wain LV, Shrine N, Artigas MS, Erzurumluoglu AM, Noyvert B, Bossini-Castillo L, et al. Genome-wide association analyses for lung function and chronic

- obstructive pulmonary disease identify new loci and potential druggable targets. *Nat Genet.* 2017;49(3):416–25.
13. Castaldi PJ, Dy J, Ross J, Chang Y, Washko GR, Curran-Everett D, et al. Cluster analysis in the COPDGen study identifies subtypes of smokers with distinct patterns of airway disease and emphysema. *Thorax.* 2014;69(5):416–23.
 14. McDonough JE, Yuan R, Suzuki M, Seyednejad N, Elliott WM, Sanchez PG, et al. Small-airway obstruction and emphysema in chronic obstructive pulmonary disease. *N Engl J Med.* 2011;365(17):1567–75.
 15. Hogg JC, Chu F, Utokaparch S, Woods R, Elliott WM, Buzatu L, et al. The nature of small-airway obstruction in chronic obstructive pulmonary disease. *N Engl J Med.* 2004;350(26):2645–53.
 16. Suzuki M, Sze MA, Campbell JD, Brothers JF, Lenburg ME, McDonough JE, et al. The cellular and molecular determinants of emphysematous destruction in COPD. *Sci Rep.* 2017;7(1):9562.
 17. Chang Y, Glass K, Liu Y-Y, Silverman EK, Crapo JD, Tal-Singer R, et al. COPD subtypes identified by network-based clustering of blood gene expression. *Genomics.* 2016;107(2–3):51–8.
 18. Bhattacharya S, Tyagi S, Srisuma S, Demeo DL, Shapiro SD, Bueno R, et al. Peripheral blood gene expression profiles in COPD subjects. *J Clin Bioinforma.* 2011;1(1):12.
 19. Morrow JD, Qiu W, Chhabra D, Rennard SI, Belloni P, Belousov A, et al. Identifying a gene expression signature of frequent COPD exacerbations in peripheral blood using network methods. *BMC Med Genet.* 2015;8(1):1.
 20. Oudijk E-JD. Systemic inflammation in COPD visualised by gene profiling in peripheral blood neutrophils. *Thorax.* 2005;60(7):538–44.
 21. Hobbs BD, Hersh CP. Integrative genomics of chronic obstructive pulmonary disease. *Biochem Biophys Res Commun.* 2014;452(2):276–86.
 22. Spira A, Beane J, Shah V, Liu G, Schembri F, Yang X, et al. Effects of cigarette smoke on the human airway epithelial cell transcriptome. *Proc Natl Acad Sci.* 2004;101(27):10143–8.
 23. Tilley AE, O'Connor TP, Hackett NR, Strulovici-Barel Y, Salit J, Amoroso N, et al. Biologic Phenotyping of the Human Small Airway Epithelial Response to Cigarette Smoking. *Königshoff M, editor. PLoS ONE.* 2011;6(7):e22798.
 24. Steiling K, Lenburg ME, Spira A. Airway gene expression in chronic obstructive pulmonary disease. *Proc Am Thorac Soc.* 2009;6(8):697–700.
 25. Steiling K, van den Berge M, Hijazi K, Florido R, Campbell J, Liu G, et al. A dynamic bronchial airway gene expression signature of chronic obstructive pulmonary disease and lung function impairment. *Am J Respir Crit Care Med.* 2013;187(9):933–42.
 26. Yeo J, Morales DA, Chen T, Crawford EL, Zhang X, Blomquist TM, et al. RNAseq analysis of bronchial epithelial cells to identify COPD-associated genes and SNPs. *BMC Pulm Med.* 2018;18:42.
 27. Obeidat M, Nie Y, Chen V, Shannon CP, Andiappan AK, Lee B, et al. Network-based analysis reveals novel gene signatures in peripheral blood of patients with chronic obstructive pulmonary disease. *Respir Res.* 2017;18(1):72.
 28. Reinhold D, Morrow JD, Jacobson S, Hu J, Ringel B, Seibold MA, et al. Meta-analysis of peripheral blood gene expression modules for COPD phenotypes. *Chotimall SH, editor. PLoS ONE.* 2017;12(10):e0185682.
 29. Beineke P, Fitch K, Tao H, Elashoff MR, Rosenberg S, Kraus WE, et al. A whole blood gene expression-based signature for smoking status. *BMC Med Genet.* 2012;5:58.
 30. Huan T, Joehanes R, Schurmann C, Schramm K, Pilling LC, Peters MJ, et al. A whole-blood transcriptome meta-analysis identifies gene expression signatures of cigarette smoking. *Hum Mol Genet.* 2016;25(21):4611–23.
 31. Parker MM, Chase RP, Lamb A, Reyes A, Saferali A, Yun JH, et al. RNA sequencing identifies novel non-coding RNA and exon-specific effects associated with cigarette smoking. *BMC Med Genet.* 2017;10:58.
 32. Hiemstra PS, McCray PB, Bals R. The innate immune function of airway epithelial cells in inflammatory lung disease. *Eur Respir J.* 2015;45(4):1150–62.
 33. Xue J, Schmidt SV, Sander J, Draffehn A, Krebs W, Quester I, et al. Transcriptome-based network analysis reveals a Spectrum model of human macrophage activation. *Immunity.* 2014;40(2):274–88.
 34. Heguy A, O'Connor TP, Luettich K, Worgall S, Ciecuch A, Harvey B-G, et al. Gene expression profiling of human alveolar macrophages of phenotypically normal smokers and nonsmokers reveals a previously unrecognized subset of genes modulated by cigarette smoking. *J Mol Med.* 2006;84(4):318–28.
 35. Bhattacharya S, Srisuma S, DeMeo DL, Shapiro SD, Bueno R, Silverman EK, et al. Molecular biomarkers for quantitative and discrete COPD phenotypes. *Am J Respir Cell Mol Biol.* 2009;40(3):359–67.
 36. Campbell JD, McDonough JE, Zeskind JE, Hackett TL, Pechkovsky DV, Brandsma C-A, et al. A gene expression signature of emphysema-related lung destruction and its reversal by the tripeptide GHK. *Genome Med.* 2012;4(8):67.
 37. Faner R, Cruz T, Casserras T, López-Giraldo A, Noell G, Coca I, et al. Network analysis of lung transcriptomics reveals a distinct B cell signature in emphysema. *Am J Respir Crit Care Med.* 2016;193(11):1242–53.
 38. Morrow JD, Zhou X, Lao T, Jiang Z, DeMeo DL, Cho MH, et al. Functional interactors of three genome-wide association study genes are differentially expressed in severe chronic obstructive pulmonary disease lung tissue. *Sci Rep.* 2017;7:44232.
 39. Spira A, Beane J, Pinto-Plata V, Kadar A, Liu G, Shah V, et al. Gene expression profiling of human lung tissue from smokers with severe emphysema. *Am J Respir Cell Mol Biol.* 2004;31(6):601–10.
 40. Kim WJ, Lim JH, Lee JS, Lee S-D, Kim JH, Oh Y-M. Comprehensive analysis of transcriptome sequencing data in the lung tissues of COPD subjects. *Int J Genomics.* 2015;2015:1–9.
 41. Jeong I, Lim J-H, Oh DK, Kim WJ, Oh Y-M. Gene expression profile of human lung in a relatively early stage of COPD with emphysema. *Int J Chron Obstruct Pulmon Dis.* 2018;13:2643–55.
 42. Imkamp K, Berg M, Vermeulen CJ, Heijink IH, Guryev V, Kerstjens HAM, et al. Nasal epithelium as a proxy for bronchial epithelium for smoking-induced gene expression and expression quantitative trait loci. *J Allergy Clin Immunol.* 2018;142(1):314–7.
 43. Boudewijn IM, Faiz A, Steiling K, van der Wiel E, Telenga ED, Hoonhorst SJM, et al. Nasal gene expression differentiates COPD from controls and overlaps bronchial gene expression. *Respir Res.* 2017;18(1):213.
 44. Obeidat M, Nie Y, Fishbane N, Li X, Bossé Y, Joubert P, et al. Integrative genomics of emphysema-associated genes reveals potential disease biomarkers. *Am J Respir Cell Mol Biol.* 2017;57(4):411–8.
 45. Poliska S, Csanky E, Szanto A, Szatmari I, Mesko B, Szeles L, et al. Chronic obstructive pulmonary disease-specific gene expression signatures of alveolar macrophages as well as peripheral blood monocytes overlap and correlate with lung function. *Respiration.* 2011;81(6):499–510.
 46. Mele M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, et al. The human transcriptome across tissues and individuals. *Science.* 2015;348(6235):660–5.
 47. Pierson E, the GTEx Consortium, Koller D, Battle A, Mostafavi S. Sharing and Specificity of Co-expression Networks across 35 Human Tissues. *Rigoutsos I, editor. PLOS Comput Biol.* 2015;11(5):e1004220.
 48. Saha A, Kim Y, Gewirtz ADH, Jo B, Gao C, McDowell IC, et al. Co-expression networks reveal the tissue-specific regulation of transcription and splicing. *Genome Res.* 2017;27(11):1843–58.
 49. Sonawane AR, Platig J, Fagny M, Chen C-Y, Paulson JN, Lopes-Ramos CM, et al. Understanding tissue-specific gene regulation. *Cell Rep.* 2017;21(4):1077–88.
 50. Halloran JW, Zhu D, Qian DC, Byun J, Gorlova OY, Amos CI, et al. Prediction of the gene expression in normal lung tissue by the gene expression in blood. *BMC Med Genet.* 2015;8:77.
 51. Morrow JD, Cho MH, Platig J, Zhou X, DeMeo DL, Qiu W, et al. Ensemble genomic analysis in human lung tissue identifies novel genes for chronic obstructive pulmonary disease. *Hum Genomics.* 2018;12:1.
 52. Civelek M, Lusis AJ. Systems genetics approaches to understand complex traits. *Nat Rev Genet.* 2013;15(1):34–48.
 53. Regan EA, Hokanson JE, Murphy JR, Make B, Lynch DA, Beaty TH, et al. Genetic epidemiology of COPD (COPDGen) study design. *COPD J Chronic Obstr Pulm Dis.* 2011;7(1):32–43.
 54. Diaz AA, Valim C, Yamashiro T, Estépar RSJ, Ross JC, Matsuoka S, et al. Airway count and emphysema assessed by chest CT imaging predicts clinical outcome in smokers. *Chest.* 2010;138(4):880–7.
 55. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
 56. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* 2007;3(9):e161.
 57. Morrow JD, Cho MH, Hersh CP, Pinto-Plata V, Celli B, Marchetti N, et al. DNA methylation profiling in human lung tissue identifies genes associated with COPD. *Epigenetics.* 2016;11(10):730–9.
 58. Buro-Auriemma LJ, Salit J, Hackett NR, Walters MS, Strulovici-Barel Y, Staudt MR, et al. Cigarette smoking induces small airway epithelial epigenetic changes with corresponding modulation of gene expression. *Hum Mol Genet.* 2013;22(23):4726–38.

59. Boyle JO, Gumus ZH, Kacker A, Choksi VL, Bocker JM, Zhou XK, et al. Effects of cigarette smoke on the human Oral mucosal transcriptome. *Cancer Prev Res (Phila Pa)*. 2010;3(3):266–78.
60. Bosse Y, Postma DS, Sin DD, Lamontagne M, Couture C, Gaudreault N, et al. Molecular signature of smoking in human lung tissues. *Cancer Res*. 2012; 72(15):3753–63.
61. Fukumoto J, Soundararajan R, Leung J, Cox R, Mahendrasah S, Muthavarapu N, et al. The role of club cell phenoconversion and migration in idiopathic pulmonary fibrosis. *Aging*. 2016;8(11):3091–109.
62. de Jong K, Vonk JM, Imboden M, Lahousse L, Hofman A, Brusselle GG, et al. Genes and pathways underlying susceptibility to impaired lung function in the context of environmental tobacco smoke exposure. *Respir Res*. 2017; 18(1):142.
63. Thorgeirsson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, Geller F, et al. Sequence variants at CHRN3–CHRNA6 and CYP2A6 affect smoking behavior. *Nat Genet*. 2010;42(5):448–53.
64. Bu X-L, Cao G-Q, Shen L-L, Xiang Y, Jiao S-S, Liu Y-H, et al. Serum amyloid-Beta levels are increased in patients with chronic obstructive pulmonary disease. *Neurotox Res*. 2015;28(4):346–51.
65. Jin M, Lee EC, Ra SW, Fishbane N, Tam S, Criner GJ, et al. Relationship of absolute telomere length with quality of life, exacerbations, and mortality in COPD. *Chest*. 2018;154(2):266–73.
66. Meiners S, Eickelberg O, Königshoff M. Hallmarks of the ageing lung. *Eur Respir J*. 2015;45(3):807–27.
67. Thannickal VJ, Murthy M, Balch WE, Chandel NS, Meiners S, Eickelberg O, et al. Blue journal conference. Aging and susceptibility to lung disease. *Am J Respir Crit Care Med*. 2015;191(3):261–9.
68. Tsuji T, Aoshiba K, Nagai A. Alveolar cell senescence in patients with pulmonary emphysema. *Am J Respir Crit Care Med*. 2006;174(8):886–93.
69. Lomas DA, Silverman EK, Edwards LD, Miller BE, Coxson HO, Tal-Singer R, et al. Evaluation of serum CC-16 as a biomarker for COPD in the ECLIPSE cohort. *Thorax*. 2008;63(12):1058–63.
70. Zemans RL, Jacobson S, Keene J, Kechris K, Miller BE, Tal-Singer R, et al. Multiple biomarkers predict disease severity, progression and mortality in COPD. *Respir Res*. 2017;18:117.
71. Barnes PJ. Inflammatory mechanisms in patients with chronic obstructive pulmonary disease. *J Allergy Clin Immunol*. 2016;138(1):16–27.
72. Yang IV, Luna LG, Cotter J, Talbert J, Leach SM, Kidd R, et al. The peripheral blood transcriptome identifies the presence and extent of disease in idiopathic pulmonary fibrosis. *PLoS One*. 2012;7(6):e37708.
73. Munthe-Fog L, Hummelshoj T, Honoré C, Moller ME, Skjoedt MO, Palsgaard I, et al. Variation in FCN1 affects biosynthesis of ficolin-1 and is associated with outcome of systemic inflammation. *Genes Immun*. 2012;13(7):515–22.
74. Bahr TM, Hughes GJ, Armstrong M, Reisdorph R, Coldren CD, Edwards MG, et al. Peripheral blood mononuclear cell gene expression in chronic obstructive pulmonary disease. *Am J Respir Cell Mol Biol*. 2013;49(2):316–23.
75. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet*. 2012;13(7):484–92.
76. Sharma G, Hanania NA, Shim YM. The aging immune system and its relationship to the development of chronic obstructive pulmonary disease. *Proc Am Thorac Soc*. 2009;6(7):573–80.
77. Sundahl N, Bridelance J, Libert C, De Bosscher K, Beck IM. Selective glucocorticoid receptor modulation: new directions with non-steroidal scaffolds. *Pharmacol Ther*. 2015;152:28–41.
78. Calverley PMA, Jones PW. Salmeterol and fluticasone propionate and survival in chronic obstructive pulmonary disease. *N Engl J Med*. 2007; 356(8):775–89.
79. Bauer M, Fink B, Seyfarth H-J, Wirtz H, Frille A. Tobacco-smoking induced GPR15-expressing T cells in blood do not indicate pulmonary damage. *BMC Pulm Med*. 2017;17:159.
80. Bazzan E, Turato G, Tinè M, Radu CM, Balestro E, Rigobello C, et al. Dual polarization of human alveolar macrophages progressively increases with smoking and COPD severity. *Respir Res*. 2017;18:40.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

