ORIGINAL RESEARCH

Integrative Genomics of Emphysema-Associated Genes Reveals Potential Disease Biomarkers

Ma'en Obeidat[1], Yunlong Nie[1], Nick Fishbane[1], Xuan Li[1], Yohan Bossé[2,3], Philippe Joubert[3,4], David C. Nickle[5], Ke Hao[6], Dirkje S. Postma[7], Wim Timens[8], Marc A. Sze[1], Casey P. Shannon[9], Zsuzsanna Hollander[9], Raymond T. Ng[9], Bruce McManus[1,9], Bruce E. Miller[10], Stephen Rennard[11,12], Avrum Spira[13], Tillie-Louise Hackett[1,14], Wan Lam[15], Stephen Lam[15], Rosa Faner[16], Alvar Agusti[16], James C. Hogg[1,17], Don D. Sin[1,18], and Peter D. Paré[1,18]

[1]The University of British Columbia Centre for Heart Lung Innovation, St. Paul's Hospital, Vancouver, British Columbia, Canada; [2]Department of Molecular Medicine, [3]Institut Universitaire de Cardiologie et de Pneumologie de Québec, and [4]Department of Molecular Biology, Medical Biochemistry, and Pathology, Laval University, Québec, Quebec, Canada; [5]Merck Research Laboratories, Genetics and Pharmacogenomics, Boston, Massachusetts; [6]Icahn Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, New York; [7]Department of Pulmonology, and [8]Department of Pathology and Medical Biology, GRIAC Research Institute, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands; [9]Prevention of Organ Failure (PROOF) Centre of Excellence, Vancouver, British Columbia, Canada; [10]GlaxoSmithKline, King of Prussia, Pennsylvania; [11]Division of Pulmonary and Critical Care Medicine, University of Nebraska Medical Center, Omaha, Nebraska; [12]Clinical Discovery Unit, Early Clinical Development, AstraZeneca, Cambridge, United Kingdom; [13]Division of Computational Biomedicine, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts; [14]Department of Anesthesiology, Pharmacology and Therapeutics, University of British Columbia, Vancouver, British Columbia, Canada; [15]Department of Integrative Oncology, British Columbia Cancer Research Centre, Vancouver, British Columbia, Canada; [16]Fundacio Clinic per a la Recerca Biomedica, Barcelona, Spain; [17]Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia, Canada; and [18]Respiratory Division, Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada

ORCID IDs: 0000-0002-5443-2752 (M.O.); 0000-0002-3067-3711 (Y.B.).

Abstract

Chronic obstructive pulmonary disease is the third leading cause of death worldwide. Gene expression profiling across multiple regions of the same lung identified genes significantly related to emphysema. We sought to determine whether the lung and epithelial expression of 127 emphysema-related genes was also related to lung function in independent cohorts, and whether any of these genes could be used as biomarkers in the peripheral blood of patients with chronic obstructive pulmonary disease. To that end, we examined whether the expression levels of these genes were under genetic control in lung tissue ($n = 1{,}111$). We then determined whether the mRNA levels of these genes in lung tissue ($n = 727$), small airway epithelial cells ($n = 238$), and peripheral blood ($n = 620$) were significantly related to lung function measurements. The expression of 63 of the 127 genes (50%) was under genetic control in lung tissue. The lung and epithelial mRNA expression of a subset of the emphysema-associated genes, including *ASRGL1*, *LPHN2*, and *EDNRB*, was strongly associated with lung function. In peripheral blood, the expression of 40 genes was significantly associated with lung function. Twenty-nine of these genes (73%) were also associated with lung function in lung tissue, but with the opposite direction of effect for 24 of the 29 genes, including those involved in hypoxia and B cell–related responses. The integrative genomics approach uncovered a significant overlap of emphysema genes associations with lung function between lung and blood with opposite directions between the two. These results support the use of peripheral blood to detect disease biomarkers.

Keywords: blood; FEV_1; lung; mRNA; biomarker

(Received in original form September 2, 2016; accepted in final form April 27, 2017)

M.O. is a postdoctoral fellow of the Michael Smith Foundation for Health Research and the Canadian Institute for Health Research Integrated and Mentored Pulmonary and Cardiovascular Training Program. He also received a research grant from the British Columbia Lung Association. K.H. is partially supported by National Natural Science Foundation of China grant no. 21477087.

Author Contributions: Conception and design: M.O., D.D.S., P.D.P.; data acquisition: Y.B., P.J., D.C.N., K.H., D.S.P., W.T., M.A.S., R.T.N., B.M., B.E.M., S.R., A.S., T.-L.H., W.L., S.L., R.F., A.A., J.C.H.; analysis and interpretation: M.O., Y.N., N.F., X.L., C.P.S., Z.H., D.D.S., P.D.P.; drafting of the manuscript: M.O., D.D.S., P.D.P.

Correspondence and requests for reprints should be addressed to Ma'en Obeidat, B.S.Pharm., Ph.D., The University of British Columbia Centre for Heart Lung Innovation, Room 166, 1081 Burrard Street, St. Paul's Hospital, Vancouver, BC, V6Z 1Y6 Canada. E-mail: maen.obeidat@hli.ubc.ca

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org

Am J Respir Cell Mol Biol Vol 57, Iss 4, pp 411–418, Oct 2017

Originally Published in Press as DOI: 10.1165/rcmb.2016-0284OC on May 1, 2017
Internet address: www.atsjournals.org

Clinical Relevance

Chronic obstructive pulmonary disease is currently the third leading cause of death, driven in part by a lack of effective biomarkers and disease-modifying therapies. Genes previously associated with emphysematous destruction were tested for association with lung function in lung tissue, airway epithelium, and peripheral blood. We observed large overlap between lung and blood including genes involved in hypoxia and B cell pathways, yet with opposite direction of association with lung function for most genes. Emphysema-related genes represent potential blood biomarkers in chronic obstructive pulmonary disease.

Chronic obstructive pulmonary disease (COPD) is currently the third leading cause of death worldwide (1), driven in part by a lack of effective biomarkers and disease-modifying therapies. The disease is characterized by a chronic irreversible airflow limitation that is often accompanied by systemic inflammation and comorbidities (2, 3). The pathology of COPD involves two main morphologic phenotypes: small airway obstruction and emphysematous enlargement of airspaces, both of which affect lung function (4, 5). Although the molecular mechanisms underlying small airway obstruction and emphysema may differ, COPD is diagnosed and assessed with the use of lung-function parameters, most commonly the forced expiratory volume in 1 second (FEV_1) and its ratio with the forced vital capacity (FEV_1/FVC).

Gene expression profiling is a powerful approach for unraveling the molecular mechanisms of diseases. A number of studies have performed gene expression profiling in lung tissue of patients with COPD (6–12). One potentially useful method for discovering the molecular mechanisms of COPD is to study different regions across the same lung. This is possible because even within the lungs of patients with COPD, there can be tremendous heterogeneity in the severity of emphysema and burden of small airway disease as measured by micro-computed tomography or quantitative histology (13). To date, two studies have investigated gene expression levels across multiple regions of the same lung and related them to measures of emphysema in the tissue samples (11, 12). This approach has the advantage of identifying genes that correlate with emphysematous destruction in one individual (i.e., the same genetic background), which may reduce the noise and enhance the specificity of the signal. Using this approach, Campbell and colleagues identified 127 genes whose expression levels were significantly associated with regional emphysema severity, which they quantified using the mean linear intercept (Lm) as measured on micro-computed tomography scans (12). Genes whose expression was increased in regions of severe emphysema (versus less affected regions) included those that were involved in B cell receptor signaling. In contrast, the genes whose expression decreased in emphysematous regions were enriched for tissue repair processes, including the transforming growth factor β pathway, actin organization, and integrin signaling.

In this study, we investigated the expression of emphysema-related genes discovered by Campbell and colleagues (12) in four other independent cohorts in which gene profiling was done in whole lung tissue, bronchial epithelial tissue, and peripheral blood, and related the expression levels of these genes to lung function measurements, namely, FEV_1 and the FEV_1/FVC ratio. In lung tissue, we additionally determined whether the expression of emphysema-related genes was under genetic control. Finally, we determined which of the emphysema-related genes could serve as possible biomarkers for COPD by determining the overlap of the significant genes for lung function between lung tissue and peripheral blood.

Materials and Methods

Emphysema Gene Signature Study

The details of the study by Campbell and colleagues were previously published (12). Briefly, the authors profiled gene expression in eight lung tissue samples obtained from eight regions within the same lung with variable emphysematous destruction, which they quantified by determining the mean linear intercept (Lm) in smokers with COPD (12).

Cohorts Used to Validate and Replicate the Emphysema Signature

Lung tissue expression quantitative loci study: To test whether the emphysema gene signature is related to lung function and under genetic control, we used data from the lung expression quantitative loci (eQTL) dataset as described previously (14–17). Briefly, nontumor lung tissues were collected from 1,111 patients who had undergone lung resection surgery, assessments for whole-genome lung gene expression, and genome-wide genotyping. For this analysis, we applied association testing for each variant with messenger RNA (mRNA) expression in either *cis* (within 1 Mb of the transcript start site) or *trans* (all other combinations), and a genome-wide 10% false discovery rate (FDR).

Lung tissue transcriptome study: Details of the study by Faner and colleagues have been previously published (18). Briefly, lung tissue samples were obtained from 70 former smokers with COPD who required thoracic surgery because of cancer or lung transplant. RNA samples were loaded into an Affymetrix GeneChip Human Genome U219 Array Plate (Santa Clara, CA). The microarray data have been deposited in the Gene Expression Omnibus (GSE69818) (18).

Bronchial airway epithelial gene expression study: To determine the robustness of the emphysema-related gene signature, we also examined the expression of these genes in bronchial epithelial cells. In a study by Steiling and colleagues (19), gene expression profiling of the bronchial specimens was performed in 238 current and former smokers with (n = 87) and without (n = 151) COPD.

Evaluation of COPD longitudinally to identify predictive surrogate endpoints (ECLIPSE) study: To determine whether the emphysema-related gene signature is enriched in the peripheral blood of COPD patients, we used gene expression data that were generated in a subset of the ECLIPSE cohort (n = 620) (20).

Association Testing with Lung-Function Measures

Linear regression was used to test for the association of mRNA levels with FEV_1 and FEV_1/FVC in the four cohorts. All analyses were adjusted for the age, sex, and smoking status of the participants. For the lung eQTL and Faner and colleagues (18) studies, the mRNA association results were meta-analyzed using inverse variance weighting in a fixed-effect model.

In a sensitivity analysis, additional adjustments for peripheral blood cell counts were made in the ECLIPSE study. At the

genetic variation level, the SNPs that functioned as eQTLs for the emphysema-associated genes were tested for association with FEV_1% predicted and the FEV_1/FVC ratio, adjusting for age, sex, smoking status, and height.

Additional details regarding the methods used are available in the online supplement.

Results

Gene mRNA Levels Related to Lung Function

Of the 127 genes that comprised the emphysema-related gene signature, 126 had probesets that mapped to the lung eQTL gene expression platform. In the ECLIPSE blood expression data, 124 of the 127 emphysema genes mapped to 302 unique probesets. In the airway epithelium, 104 of the 127 emphysema-related genes mapped to 104 unique probesets, and in the lung tissue study of Faner and colleagues (18), 123 genes with 355 probesets were mapped.

The results for the top 10 genes from each analysis of FEV_1 and FEV_1/FVC in the four studies are shown in Table 1. Based on a meta-analysis of the lung eQTL and Faner and colleagues (18) lung tissue studies, the gene with the strongest association for FEV_1 was *ASRGL1*, which demonstrated a negative association ($\beta = -3.59$, FDR = 5.80E-05), whereas *EDNRB* showed the strongest association with FEV_1/FVC ($\beta = 2.212$, FDR = 2.45E-06). In the airway epithelium, *PRKCE* showed the strongest association with both FEV_1 ($\beta = 0.316$, FDR = 8. 38E-07) and FEV_1/FVC ($\beta = 0.43$, FDR = 8.41E-06).

In peripheral blood, *SLC45A3* showed the strongest association with both FEV_1 ($\beta = -10.39$, FDR = 2.23E-02 and FEV_1/FVC ($\beta = -8.59$, FDR = 2.86E-04) after adjustments for cell composition. In the analysis of peripheral blood without adjustment for cell composition, the gene most strongly associated with FEV_1 was *cluster of differentiation 79B* (*CD79B*; $\beta = 0.27$, FDR = 5.23E-04), whereas FEV_1/FVC was most significantly associated with *SLC45A3* ($\beta = -0.09$, FDR = 1.44E-04). Other members of the CD gene family also showed strong positive associations with FEV_1 in peripheral blood, including *CD22* ($\beta = 0.17$, FDR = 9.13E-04) and *CD79A* ($\beta = 0.16$, FDR = 9.13E-04). Interestingly in the lung tissue, *CD22* was negatively associated with FEV_1 ($\beta = -0.06$, FDR = 9.91E-02) and with FEV_1/FVC ($\beta = -1.56$, FDR = 1.38E-02). The lung tissue data are consistent with the direction of association of *CD22* and *CD79A/B* with increased Lm in the emphysema study (i.e., the worse the emphysema, the higher the expression levels of these genes in the lung tissue) (12).

Table 1. The 10 Most Significant Genes for Associations with FEV_1% Predicted and FEV_1/FVC in Lung, Epithelial Cells, and Blood

FEV_1% Predicted					FEV_1/FVC Ratio				
Lung Tissue (Meta-analysis of the eQTL study and Faner and colleagues [20])									
Gene	Estimate	SE	*P* value	FDR	Gene	Estimate	SE	*P* value	FDR
ASRGL1	−3.594	0.713	4.68E-07	5.80E-05	*EDNRB*	2.212	0.394	1.91E-08	2.45E-06
MT1JP	−3.298	0.693	1.98E-06	1.23E-04	*LPHN2*	1.869	0.398	2.66E-06	1.70E-04
EDNRB	2.737	0.713	1.22E-04	3.81E-03	*STXBP6*	1.729	0.394	1.16E-05	4.95E-04
CUGBP2	2.801	0.729	1.23E-04	3.81E-03	*ARHGEF10*	1.705	0.397	1.74E-05	5.57E-04
LPHN2	2.668	0.729	2.51E-04	5.70E-03	*EPAS1*	1.636	0.396	3.65E-05	9.34E-04
FCRLA	−3.210	0.883	2.76E-04	5.70E-03	*DHRS9*	−1.708	0.419	4.48E-05	9.56E-04
QKI	2.572	0.721	3.59E-04	6.35E-03	*CCR7*	−1.667	0.424	8.30E-05	1.52E-03
CXCL13	−2.704	0.840	1.29E-03	1.45E-02	*FCRLA*	−1.885	0.484	9.94E-05	1.54E-03
C1orf55	−2.357	0.735	1.35E-03	1.45E-02	*QKI*	1.540	0.398	1.08E-04	1.54E-03
FOXF1	2.252233	0.706368	1.43E-03	1.45E-02	*CD22*	−1.510	0.422	3.41E-04	4.03E-03
Airway epithelium									
PRKCE	31.985	5.151	3.01E-09	3.22E-07	*PRKCE*	14.347	2.614	1.21E-07	1.29E-05
GPR110	−17.715	3.071	3.01E-08	1.61E-06	*TMEM2*	−11.065	2.120	4.47E-07	2.39E-05
NDEL1	−38.352	7.717	1.44E-06	5.14E-05	*LRRC8A*	−12.346	2.732	1.06E-05	0.000378
LRRC8A	−26.412	5.444	2.46E-06	5.41E-05	*KRT7*	−8.746	2.026	2.49E-05	0.000498
ASRGL1	−12.321	2.543	2.53E-06	5.41E-05	*CTTNBP2NL*	−14.544	3.383	2.68E-05	0.000498
TMEM2	−20.406	4.299	3.93E-06	7.01E-05	*PODXL*	−12.459	2.905	2.79E-05	4.98E-04
HERC3	30.548	6.700	8.93E-06	1.29E-04	*GPR110*	−6.485	1.588	6.41E-05	9.80E-04
CTTNBP2NL	−30.684	6.756	9.62E-06	1.29E-04	*HERC3*	12.95	3.387	1.75E-04	2.15E-03
PNMAL1	17.963	4.210	3.06E-05	3.64E-04	*ASRGL1*	−4.935	1.293	1.81E-04	2.15E-03
DHRS9	−19.416	4.601	3.71E-05	3.97E-04	*NDEL1*	−14.339	3.948	3.58E-04	3.83E-03
Peripheral blood*									
SLC45A3	−10.391	2.756	1.79E-04	2.23E-02	*SLC45A3*	−8.589	1.800	2.29E-06	2.86E-04
PNMAL1	−14.684	4.729	1.99E-03	9.15E-02	*EPAS1*	−8.830	2.047	1.88E-05	1.18E-03
EPAS1	−9.641	3.135	2.20E-03	9.15E-02	*S100A8*	3.872	0.990	1.03E-04	4.28E-03
CD79B	6.003	2.193	6.37E-03	1.52E-01	*NEDD9*	−8.966	2.447	2.71E-04	8.47E-03
TMEM9	−13.952	5.154	6.99E-03	1.52E-01	*PLXNA2*	−9.336	2.716	6.29E-04	1.57E-02
PHLDB1	−12.519	4.650	7.30E-03	1.52E-01	*TMBIM1*	−8.415	3.127	7.32E-03	1.34E-01
WFDC1	−10.283	3.924	9.01E-03	1.58E-01	*TACC1*	−8.490	3.164	7.49E-03	1.34E-01
S1PR1	8.666	3.358	1.01E-02	1.58E-01	*ECHDC3*	−5.154	1.963	8.88E-03	1.39E-01
RALGPS2	4.836	1.978	1.48E-02	2.04E-01	*S1PR1*	5.575	2.209	1.19E-02	1.49E-01
COL4A1	−12.707	5.291	1.66E-02	2.04E-01	*HPCAL1*	−5.797	2.303	1.21E-02	1.49E-01

Definition of abbreviations: eQTL, lung expression quantitative loci; FDR, false discovery rate.
*Results from the cell-count–adjusted analysis.

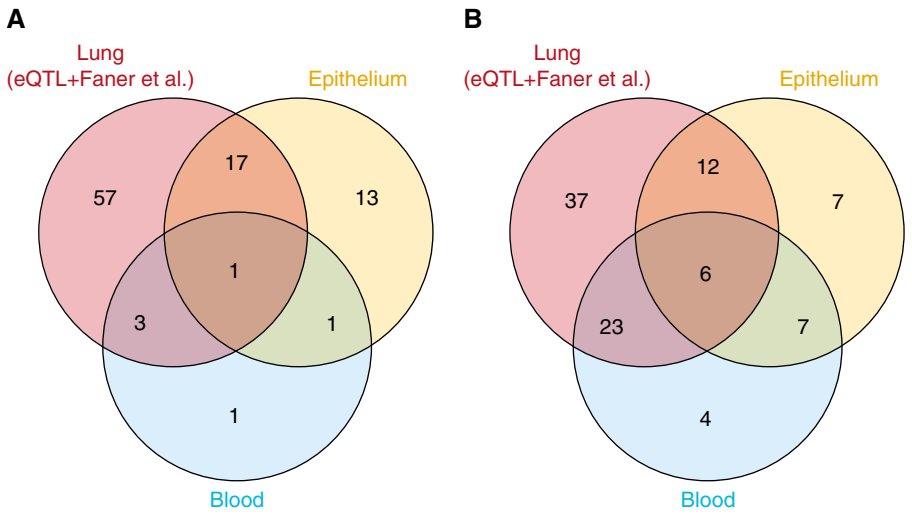

Figure 1. Venn diagram of overlapping genes associated with lung function. The diagram shows the overlap of genes associated with either FEV_1 or FEV_1/FVC at false discovery rate (FDR) < 0.1 in the four studies. (*A*) Blood results with cell-count adjustment. (*B*) Results in peripheral blood without cell-count adjustment. eQTL, lung expression quantitative loci.

Although the adjustment for cell composition generally increased the association FDR values, there was a strong correlation between the cell-adjusted and -unadjusted analyses ($r = 0.80$ for FEV_1 and $r = 0.94$ for FEV_1/FVC, with $P < 2.2 \times 10^{-16}$ for both measures) as shown in Figure 2, and the ranking of the top associated genes remained similar.

The associations of mRNA levels of genes in lung tissue, airway epithelium, and peripheral blood with FEV_1 and FEV_1/FVC are shown in Table E1 in the online supplement.

The overlap of genes associated with either FEV_1 or FEV_1/FVC at 10% FDR in the two lung tissue studies, airway epithelium, and peripheral blood is shown as a Venn diagram in Figure 1. The diagram shows a considerable overlap across the studies: of the 40 lung-function–associated genes in blood, 29 (73%) and 13 (33%) overlapped with genes associated with lung function in lung tissue and airway epithelium, respectively. In the analysis that was not adjusted for cell count, six genes were common to all studies in lung, epithelium, and blood: *FCRLA*, *MYH9*, *RALGPS2*, *S100A8*, *TACC1*, and *TMEM9*. In the analysis that was adjusted for cell count, six genes were associated with lung function in blood, and four (67%) and two (33%) of these overlapped with genes in lung and airway epithelium, respectively. In the cell-adjusted analysis, only one gene, *S100A8* was common to the three sources, and its expression was positively correlated with lung function in lung tissue and airway epithelium, but negatively correlated with lung function in peripheral blood. Interestingly, the directionality in the relationship between gene expression and lung function was discordant between lung tissue and blood for most of the overlapping genes. Of the 29 genes associated with lung function in lung and blood, 25 exhibited an opposite directionality of effect. Similarly, three of the four overlapping genes between lung

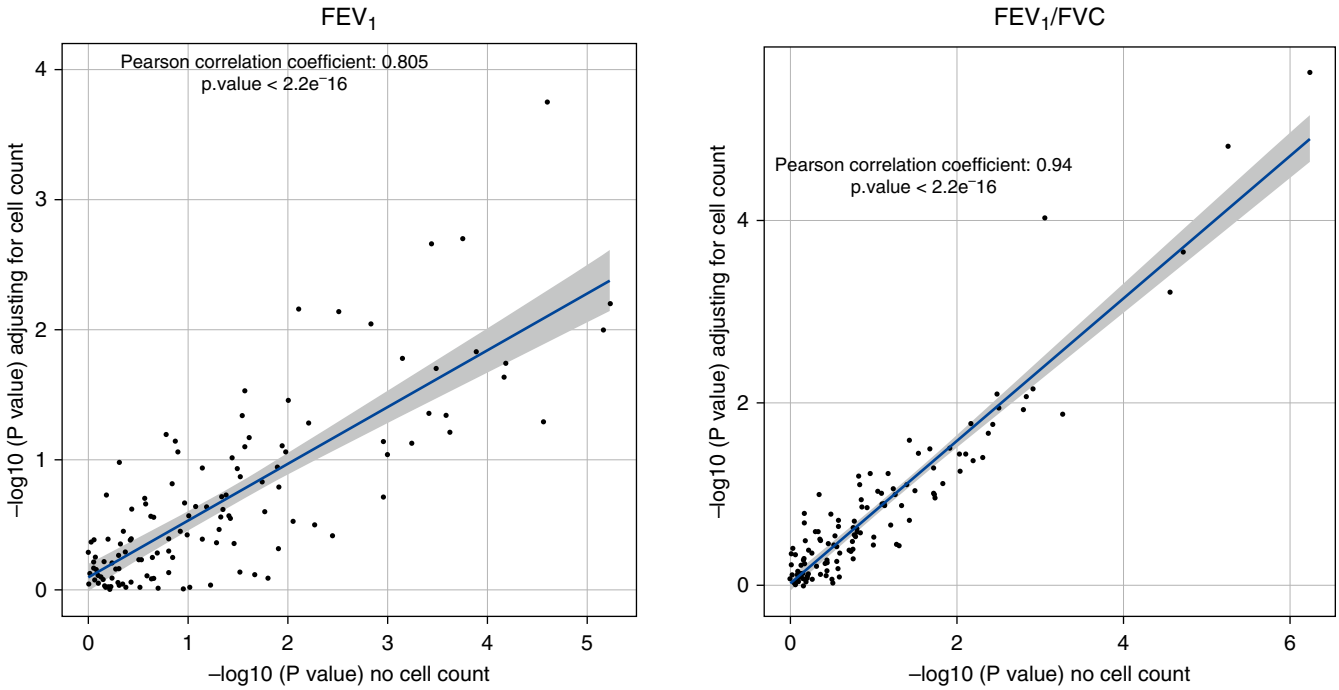

Figure 2. Correlation of association *P* values for emphysema-associated genes with FEV_1 and FEV_1/FVC in peripheral blood with and without adjustment for cell counts. The *x-axis* shows the *P* values for association of emphysema genes in peripheral blood with lung function without adjusting for cell counts. The *y-axis* shows the *P* values after adjustment for cell counts. Pearson's correlation coefficients and *P* values are shown on the graph.

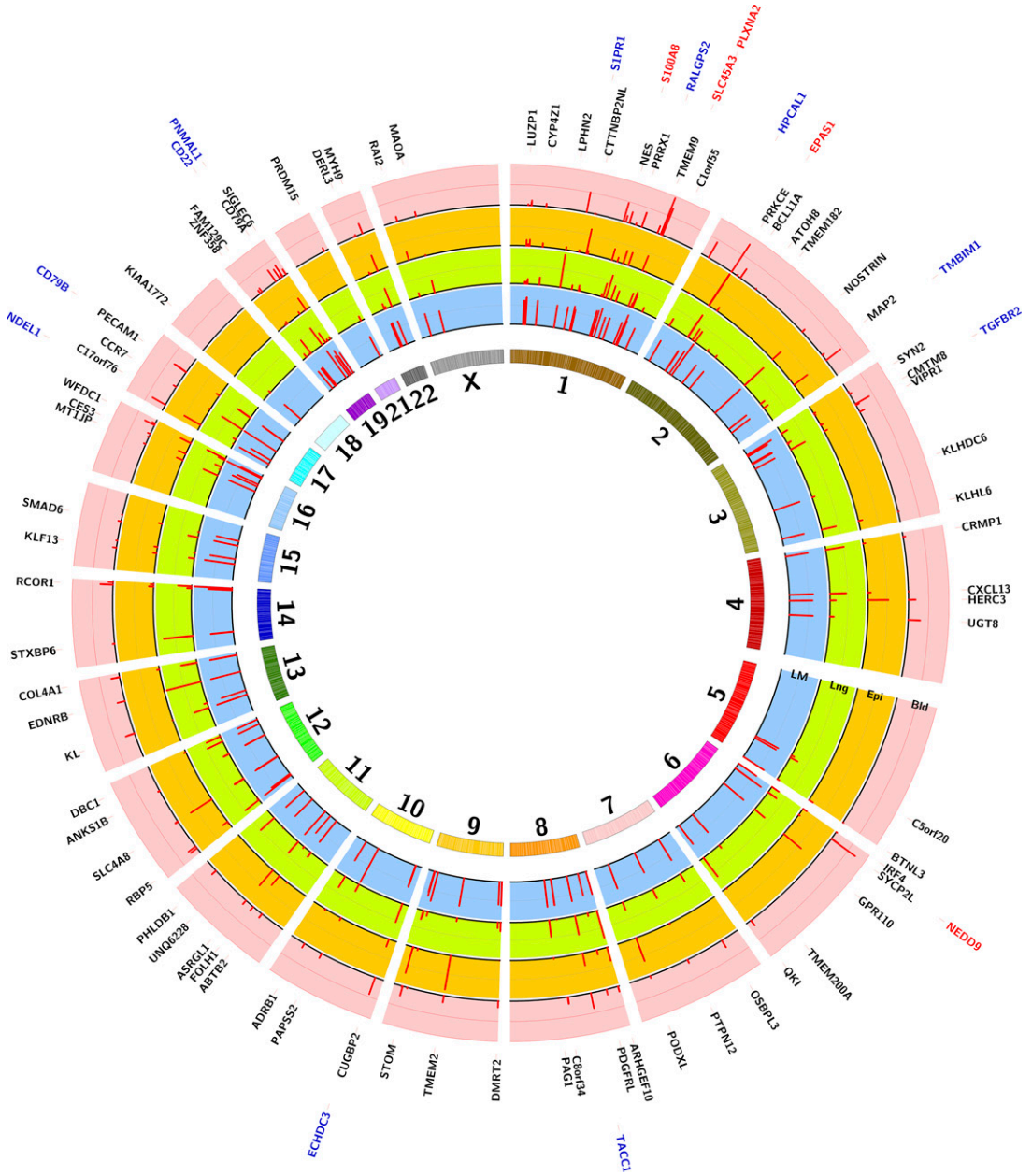

Figure 3. Circos plots depicting the mRNA association *P* values with FEV_1/FVC for 127 genes across lung, epithelium, and blood. The chromosomal positions are shown inside the circles. The first circle (*light blue*) represents results from the mean linear intercept associations. The second circle (*green*) represents association results from the lung tissue meta-analysis (lung eQTL and Faner and colleagues [20]). The third circle (*yellow*) represents results from the airway epithelium study. The outermost circle (*pink*) represents the association results in peripheral blood. Gene symbols are shown outside the circles. Genes highlighted in blue are those that showed significant associations (FDR < 10%) in peripheral blood in the model without cell-count adjustment, and genes in red are those that passed the 10% FDR threshold in the cell-count–adjusted analysis.

and blood in the cell-count–adjusted analysis showed an opposite direction of effect.

A Circos plot is shown in Figure 3 to visualize the relationship of the expression of the 127 genes with Lm and with FEV_1/FVC in lung tissue and epithelial and blood cells. The 127 genes are organized by their chromosomal positions, with the significance of the association in each study represented by the length of the red lines inside the four circles.

Integrative Genomics in Lung Tissue. Because the lung eQTL study measured genome-wide gene variants and gene expression in the same 1,111 individuals (15), we were able to conduct an integrative genomics analysis to determine the associations of SNPs with mRNA levels in lung tissues and with lung function measurements in the same subjects.

In this analysis, we found that 63 (50%) of the 127 genes whose expression was related to emphysema were under genetic control in lung tissue (i.e., had one or more SNP that was related to the expression level at a 10% FDR). Among the 63 genes, 58 and

Table 2. The Five Most Significant eQTL Associations with Lung Function Measures in Individuals from the Lung eQTL Study

eQTL SNP ID	chr	*cis/trans*	eQTL estimate	eQTL *P* value	Gene	Lung function estimate	Lung function *P* value
FEV_1% Predicted							
rs11404483	8	*cis*	−0.2717	1.84E-08	*PAG1*	2.030	1.90E-03
rs10411704	19	*cis*	−0.6428	1.81E-40	*CD22*	3.519	5.03E-03
rs1106729	1	*cis*	−0.5928	4.03E-50	*TMEM9*	1.634	2.80E-02
rs17265731	2	*cis*	0.6281	2.83E-22	*NOSTRIN*	−3.577	3.01E-02
rs4275	22	*cis*	0.5212	6.87E-44	*TPST2*	2.032	3.53E-02
FEV_1/FVC							
rs13230759*	7	*cis*	0.697	2.39E-13	*OSBPL3*	−3.112	3.33E-06
rs9356119*	6	*cis*	−0.997	7.23E-176	*QKI*	1.206	1.13E-04
rs10411704*	19	*cis*	−0.643	1.81E-40	*CD22*	2.460	4.66E-04
rs6580837	12	*cis*	0.291	4.53E-12	*SLC4A8*	1.361	1.75E-02
rs4700772	5	*cis*	−0.663	4.25E-64	*BTNL3*	0.910	3.12E-02

Definition of abbreviations: chr, chromosome; *cis/trans*, whether the eQTL is acting in *cis* (< 1 Mb away from the genes) or *trans*; eQTL, lung expression quantitative loci; SNP, single-nucleotide polymorphism.
*Indicates that this SNP met the significance threshold after adjusting for multiple testing.

5 genes were *cis* and *trans* eQTLs, respectively. The expression of the 63 genes was regulated by a total of 6,756 eQTL SNPs.

After the analysis was restricted to the best eQTL per gene (i.e., eQTL with the lowest *P* value), 63 eQTLs were tested for association with FEV_1 and FEV_1/FVC in individuals from the lung eQTL study. The top five eQTLs associated with FEV_1 and/or with FEV_1/FVC are shown in Table 2, and the complete results are shown in Table E2.

The three most significantly associated eQTLs for FEV_1 were *1*) SNP rs11404483 (P = 1.90E-03), which was an eQTL for *PAG1* (P = 1.84E-08); *2*) SNP rs10411704 (P = 5.03E-03), which was an eQTL for *CD22* (P = 1.81E-40); and *3*) SNP rs113812546 (P = 2.80E-02), which was an eQTL for *TMEM9* (P = 4.03E-50). The SNPs that were most significantly associated with FEV_1/FVC were *1*) rs13230759 (P = 3.33E-06), which was an eQTL for *OSBPL3* (P = 2.39E-13); *2*) SNP rs9356119 (P = 1.13E-04), which was an eQTL for *QKI* (P = 7.23E-176); and *3*) the *CD22* eQTL SNP rs10411704 (P = 4.66E-04). The three SNP associations for FEV_1/FVC were the only ones that met the adjusted significance threshold of $P < 8.1 \times 10^{-04}$. The alleles that reduced *CD22* and *QKI* gene expression were associated with higher FEV_1 and FEV_1/FVC, and the allele that increased *OSBPL3* expression was associated with lower lung function. These data suggest that lower CD22, *QKI*, and *OSBPL3* mRNA expression is associated with higher lung function. For *CD22*, this allelic association was consistent with the mRNA association in lung tissue; *CD22* expression was negatively associated with FEV_1 and FEV_1/FVC.

Discussion

The identification of gene expression changes in tissues relevant to a disease is an important step toward improving our understanding of disease pathogenesis and identifying new therapeutic and biomarker targets. To our knowledge, this is the first study to use a targeted approach and comprehensively evaluate whether lung tissue, airway epithelium, and blood expression of genes that were previously associated with emphysema are related to lung function.

The main findings of this study were that *1*) many of the genes related to emphysema were expressed in whole lung tissue, epithelial cells, and blood, and were associated with FEV_1 and FEV_1/FVC across four different cohorts, providing additional confidence that these genes (and their associated pathways) may play a role in the pathogenesis of COPD; *2*) the expression of half of the emphysema-associated genes (63 out of 127) was under genetic control in lung tissue (i.e., had eQTLs) and a number of these eQTLs, including ones related to the expression of *OSBPL3*, *QKI*, and the B cell marker *CD22*, were also associated with lung function, supporting a potential causal pathway; *3*) in peripheral blood, the expression levels of 40 of the 127 emphysema genes were significantly related to lung function, with *SLC45A3*, the hypoxia-inducible factor *EPAS1*, and the B cell marker CD22 demonstrating the strongest associations with lung function; and *4*) of these 40 genes discovered in peripheral blood, 29 (73%) were also significantly associated with lung function in whole lung tissue or bronchial epithelium. Interestingly, however, only 5 of the 29 exhibited a direction of effect similar to that observed in blood.

In this study, the most strongly associated genes in lung tissue were *EDNRB* and *LPHN2*, which were both positively correlated, and *ASRGL1*, which was negatively correlated. The endothelin receptor type B (EDNRB) encodes a G protein–coupled receptor that activates a phosphatidylinositol/calcium second-messenger system (21) and is linked to the hypoxia-inducible factor pathway and associated with high-altitude adaptation (22). The adhesion G protein–coupled receptor L2 (LPHN2) is a member of the latrophilin family of genes, which were recently proposed as potential bronchodilator targets in asthma (23). Asparaginase-like 1 (ASRGL1) is involved in reducing cell proliferation (24) and inducing apoptosis (25).

In peripheral blood, the strongest association of emphysema-related genes with lung function was observed for *SLC45A3*, *EPAS1*, and *PNMAL1*, which were all negatively correlated with lung function. Endothelial PAS domain protein 1 (EPAS1) encodes the hypoxia-inducible factor 2α (HIF-2α) transcription factor, whose expression is induced in hypoxic conditions (26). In an integrative study of methylation, gene expression, and

disease phenotypes, *EPAS1* was identified as a key regulator of COPD in lung tissue (27). The role of the paraneoplastic Ma antigen family like 1 (*PNMAL1*) gene is not well understood, but a recent study of the lung transcriptome reported upregulation of PNMAL1 in severe emphysema versus bronchiolitis in COPD patients with the same degree of airflow limitation (18).

We observed a considerable overlap of the emphysema-related gene associations with lung function across whole lung tissue, bronchial epithelium, and blood. The greatest overlap was between lung and blood samples: 73% of the genes that were significant in blood, as determined by an analysis not adjusted for cell count, were significantly related to lung function in lung tissue. This number decreased slightly to 66% in the analysis that was adjusted for cell count. There is considerable interest in discovering novel blood biomarkers because blood is a convenient and easily accessible tissue (unlike lungs). However, because blood cells arise from the bone marrow and are processed largely in extrapulmonary organs, there is significant skepticism regarding the concept of "blood as a window to the lung." The findings of the present study suggest that although it is clear that the lung gene expression signature related to lung function is not fully recapitulated in blood cells, there is a sizable overlap with the lung tissue signature in blood. However, for most of these associations, the direction of effect is in the opposite direction. Thus, for selected genes/proteins, blood may be used as a potential source for biomarker discovery. Further studies are needed to fully validate this notion and to advance discoveries of easily accessible biomarkers to track disease severity and progression in COPD, which are critically needed (28).

We also observed that there was less overlap in the genes related to lung function between blood and airway epithelium versus blood and whole lung tissue. This may be driven in part by the reduced sample size of the airway epithelial cell study compared with the lung eQTL study. It may also reflect the fact that the emphysema-associated signature in the lung is derived from resident or infiltrating immune cells as well as structural and mesenchymal cells.

The integrative approach uncovered a widespread "flipping" of expression changes for overlapping genes between lung and blood (24 of 29 overlapping genes had an opposite direction of effect), which persisted even when the blood associations were adjusted for cell count (three of four overlapping genes were flipped), providing novel insights into disease pathogenesis. The reasons for this discordance are unknown but may lie in the pathways and processes in which these genes are involved. Alternatively, the discordance may reflect differences between the lung and blood in the number of cell types that express these genes. The 40 genes significantly related to lung function in peripheral blood were enriched in only one Kyoto Encyclopedia of Genes and Genomes pathway: B cell receptor signaling (FDR = 1.32E-06). Enrichment in B cell signaling was similarly reported in studies of lung tissue, such as those conducted by Campbell and colleagues (12) and more recently by Faner and colleagues (18). Little is known about the role of B lymphocytes in the blood of COPD patients. Brandsma and colleagues showed lower levels of memory B cells in peripheral blood from COPD patients, but increased numbers of (class-switched) memory B cells in the peripheral blood of current smokers compared with ex-smokers and never smokers, irrespective of COPD (29). It is possible that as B cells are recruited into lung tissue during inflammation, their numbers in the peripheral circulation decrease. Our finding that B cell–specific genes (*CD79A* and *CD79B*) showed a discordant pattern of mRNA association with lung function in blood and lung tissue lends support to the hypothesis that alterations in the number of B cells may be responsible for these associations. Indeed, *CD79A* and *CD79B* have been suggested as stable "housekeeping" genes related to B cell numbers in peripheral blood (30). However, the genetic association of the *CD22* SNP rs10411704 with lung function and with *CD22* expression supports a causal rather than a reactive role for this gene in COPD. Another possibility, which is not mutually exclusive of the other explanations for the opposite relationships, is chance (false-positive results).

The current study has a number of limitations. First, lung-function measures cannot distinguish between obstruction in the small airways and emphysematous destruction of the lung parenchyma, or provide information about regional differences in disease severity. On the other hand, FEV_1 is easily measured and is highly reproducible, and the 29 genes that overlap between the lung and blood offer a potential for using blood samples as biomarkers of disease severity in the lungs. Second, the exact cellular composition of the lung tissue samples used in the lung tissue studies is unknown; thus, the cellular heterogeneity of the lung tissues may have undermined the discovery of biomarkers that show cell-type–specific expression patterns related to lung function. Third, the reported associations with disease severity may be reactive rather than causal, and changes in gene expression may not necessarily translate into changes into protein levels (31). Fourth, mRNA associations in lung tissue and blood were measured in separate individuals, and therefore the discordant expression changes observed between blood and lung could be due to other factors, such as patient demographics and disease severity, or could be false-positive results. Fifth, peripheral whole blood is a heterogeneous tissue composed of many different immune cell subsets, and the cellular composition varies in response to physiological or pathological processes. Although it is common practice to adjust statistically for the peripheral blood cell composition, as we did in this study by including the white blood cell count and its differential as covariates in the analysis, regression methods may not fully take into account cell-specific gene expression and thus may conceal important cell-specific signatures. However, because changes in cell composition can be causally related to disease, adjustments for cell counts are likely overly conservative and may lead to false negatives. On the other hand, not adjusting for cell count may lead to false-positive findings. Further work using blood-cell–specific gene expression signatures in the context of COPD is warranted to address this issue. Finally, the cross-sectional lung-function measures do not fully assess disease activity, and because not all studies captured postbronchodilator values, we used a mixture of pre- and postbronchodilator lung-function measurements, which likely reduced the power of the present analysis.

In conclusion, using a hypothesis-driven integrative genomics approach, we found that expression in the lung, epithelium, and blood of genes associated with emphysematous destruction in lung tissue was related to lung function in large (and completely independent) cohorts of patients with COPD. In particular, the

greatest overlap of genes was found between lung and blood samples, with most genes showing the opposite direction of association with lung function across these sites. Together, the data suggest that the "emphysema"-related genes harbor molecular drivers of COPD progression and represent potential therapeutic and biomarker targets in COPD that may be detected in peripheral blood. Such targets are urgently needed to combat the growing burden of COPD worldwide. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

References

1. World Health Organization. The top 10 causes of death. Geneva, Switzerland: World Health Organization; 2014 [accessed 2016 Jun 1]. Available from: http://www.who.int/mediacentre/factsheets/fs310/en/index4.html
2. The Global Initiative for Chronic Obstructive Lung Disease. Global initiative for chronic obstructive pulmonary disease and comorbidities. GOLD; 2011 [accessed 2016 May 5]. Available from: http://goldcopd.org/
3. Decramer M, Janssens W. Chronic obstructive pulmonary disease and comorbidities. *Lancet Respir Med* 2013;1:73–83.
4. Barnes PJ. New concepts in chronic obstructive pulmonary disease. *Annu Rev Med* 2003;54:113–129.
5. Hogg JC, Timens W. The pathology of chronic obstructive pulmonary disease. *Annu Rev Pathol* 2009;4:435–459.
6. Golpon HA, Coldren CD, Zamora MR, Cosgrove GP, Moore MD, Tuder RM, Geraci MW, Voelkel NF. Emphysema lung tissue gene expression profiling. *Am J Respir Cell Mol Biol* 2004;31:595–600.
7. Bhattacharya S, Srisuma S, Demeo DL, Shapiro SD, Bueno R, Silverman EK, Reilly JJ, Mariani TJ. Molecular biomarkers for quantitative and discrete COPD phenotypes. *Am J Respir Cell Mol Biol* 2009;40:359–367.
8. Ning W, Li C-J, Kaminski N, Feghali-Bostwick CA, Alber SM, Di YP, Otterbein SL, Song R, Hayashi S, Zhou Z, *et al*. Comprehensive gene expression profiles reveal pathways related to the pathogenesis of chronic obstructive pulmonary disease. *Proc Natl Acad Sci USA* 2004;101:14895–14900.
9. Spira A, Beane J, Pinto-Plata V, Kadar A, Liu G, Shah V, Celli B, Brody JS. Gene expression profiling of human lung tissue from smokers with severe emphysema. *Am J Respir Cell Mol Biol* 2004;31:601–610.
10. Wang IM, Stepaniants S, Boie Y, Mortimer JR, Kennedy B, Elliott M, Hayashi S, Loy L, Coulter S, Cervino S, *et al*. Gene expression profiling in patients with chronic obstructive pulmonary disease and lung cancer. *Am J Respir Crit Care Med* 2008;177:402–411.
11. Stepaniants S, Wang IM, Boie Y, Mortimer J, Kennedy B, Elliott M, Hayashi S, Luo H, Wong J, Loy L, *et al*. Genes related to emphysema are enriched for ubiquitination pathways. *BMC Pulm Med* 2014;14:187.
12. Campbell JD, McDonough JE, Zeskind JE, Hackett TL, Pechkovsky DV, Brandsma C-A, Suzuki M, Gosselink JV, Liu G, Alekseyev YO, *et al*. A gene expression signature of emphysema-related lung destruction and its reversal by the tripeptide GHK. *Genome Med* 2012;4:67.
13. McDonough JE, Yuan R, Suzuki M, Seyednejad N, Elliott WM, Sanchez PG, Wright AC, Gefter WB, Litzky L, Coxson HO, *et al*. Small-airway obstruction and emphysema in chronic obstructive pulmonary disease. *N Engl J Med* 2011;365:1567–1575.
14. Lamontagne M, Couture C, Postma DS, Timens W, Sin DD, Paré PD, Hogg JC, Nickle D, Laviolette M, Bossé Y. Refining susceptibility loci of chronic obstructive pulmonary disease with lung eqtls. *PLoS One* 2013;8:e70220.
15. Hao K, Bossé Y, Nickle DC, Paré PD, Postma DS, Laviolette M, Sandford A, Hackett TL, Daley D, Hogg JC, *et al*. Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet* 2012;8:e1003029.
16. Obeidat M, Miller S, Probert K, Billington CK, Henry AP, Hodge E, Nelson CP, Stewart CE, Swan C, Wain LV, *et al*. GSTCD and INTS12 regulation and expression in the human lung. *PLoS One* 2013;8:e74630.
17. Obeidat Me, Hao K, Bossé Y, Nickle DC, Nie Y, Postma DS, Laviolette M, Sandford AJ, Daley DD, Hogg JC, *et al*. Molecular mechanisms underlying variations in lung function: a systems genetics analysis. *Lancet Respir Med* 2015;3:782–795.
18. Faner R, Cruz T, Casserras T, López-Giraldo A, Noell G, Coca I, Tal-Singer R, Miller B, Rodriguez-Roisin R, Spira A, *et al*. Network analysis of lung transcriptomics reveals a distinct b-cell signature in emphysema. *Am J Respir Crit Care Med* 2016;193:1242–1253.
19. Steiling K, van den Berge M, Hijazi K, Florido R, Campbell J, Liu G, Xiao J, Zhang X, Duclos G, Drizik E, *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;187:933–942.
20. Vestbo J, Anderson W, Coxson HO, Crim C, Dawber F, Edwards L, Hagan G, Knobil K, Lomas DA, MacNee W, *et al*.; ECLIPSE investigators. Evaluation of copd longitudinally to identify predictive surrogate end-points (eclipse). *Eur Respir J* 2008;31:869–873.
21. Tykocki NR, Watts SW. The interdependence of endothelin-1 and calcium: a review. *Clin Sci (Lond)* 2010;119:361–372.
22. Stobdan T, Zhou D, Ao-Ieong E, Ortiz D, Ronen R, Hartley I, Gan Z, McCulloch AD, Bafna V, Cabrales P, *et al*. Endothelin receptor B, a candidate gene from human studies at high altitude, improves cardiac tolerance to hypoxia in genetically engineered heterozygote mice. *Proc Natl Acad Sci USA* 2015;112:10425–10430.
23. Faiz A, Donovan C, Nieuwenhuis MA, van den Berge M, Postma DS, Yao S, Park CY, Hirsch R, Fredberg JJ, Tjin G, *et al*. Latrophilin receptors: novel bronchodilator targets in asthma. *Thorax* 2017;72:74–82.
24. Chuang JC, Yu CL, Wang SR. Modulation of lymphocyte proliferation by enzymes that degrade amino acids. *Clin Exp Immunol* 1990;82:469–472.
25. Bussolati O, Belletti S, Uggeri J, Gatti R, Orlandini G, Dall'Asta V, Gazzola GC. Characterization of apoptotic phenomena induced by treatment with L-asparaginase in NIH3T3 cells. *Exp Cell Res* 1995;220:283–291.
26. Tian H, McKnight SL, Russell DW. Endothelial PAS domain protein 1 (EPAS1), a transcription factor selectively expressed in endothelial cells. *Genes Dev* 1997;11:72–82.
27. Yoo S, Takikawa S, Geraghty P, Argmann C, Campbell J, Lin L, Huang T, Tu Z, Foronjy RF, Spira A, *et al*. Integrative analysis of DNA methylation and gene expression data identifies EPAS1 as a key regulator of COPD. *PLoS Genet* 2015;11:e1004898.
28. Vestbo J, Rennard S. Chronic obstructive pulmonary disease biomarker(s) for disease activity needed–urgently. *Am J Respir Crit Care Med* 2010;182:863–864.
29. Brandsma CA, Hylkema MN, Geerlings M, van Geffen WH, Postma DS, Timens W, Kerstjens HA. Increased levels of (class switched) memory B cells in peripheral blood of current smokers. *Respir Res* 2009;10:108.
30. Singh A, Yamamoto M, Ruan J, Choi JY, Gauvreau GM, Olek S, Hoffmueller U, Carlsten C, FitzGerald JM, Boulet L-P, *et al*. Th17/Treg ratio derived using DNA methylation analysis is associated with the late phase asthmatic response. *Allergy Asthma Clin Immunol* 2014;10:32–32.
31. Tebaldi T, Re A, Viero G, Pegoretti I, Passerini A, Blanzieri E, Quattrone A. Widespread uncoupling between transcriptome and translatome variations after a stimulus in mammalian cells. *BMC Genomics* 2012;13:220.