Molecular mechanisms underlying variations in lung function: a systems genetics analysis

Ma'en Obeidat, Ke Hao, Yohan Bossé, David C Nickle, Yunlong Nie, Dirkje S Postma, Michel Laviolette, Andrew J Sandford, Denise D Daley, James C Hogg, W Mark Elliott, Nick Fishbane, Wim Timens, Pirro G Hysi, Jaakko Kaprio, James F Wilson, Jennie Hui, Rajesh Rawal, Holger Schulz, Beate Stubbe, Caroline Hayward, Ozren Polasek, Marjo-Riitta Järvelin, Jing Hua Zhao, Deborah Jarvis, Mika Kähönen, Nora Franceschini, Kari E North, Daan W Loth, Guy G Brusselle, Albert Vernon Smith, Vilmundur Gudnason, Traci M Bartz, Jemma B Wilk, George T O'Connor, Patricia A Cassano, Wenbo Tang, Louise V Wain, María Soler Artigas, Sina A Gharib, David P Strachan, Don D Sin, Martin D Tobin, Stephanie J London, Ian P Hall, Peter D Paré

Lancet Respir Med 2015; 3: 782–95

Published Online September 22, 2015 http://dx.doi.org/10.1016/S2213-2600(15)00380-X

See Comment page 739

University of British Columbia Center for Heart Lung Innovation, St Paul's Hospital, Vancouver, BC, Canada (M Obeidat PhD, Y Nie MSc, Prof A J Sandford PhD, D D Daley PhD, Prof J C Hogg MD, W M Elliott PhD, N Fishbane MSc, Prof D D Sin MD, Prof P D Paré MD); **Icahn Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY, USA** (K Hao ScD); **Department of Molecular Medicine** (Y Bossé PhD), **and Institut Universitaire de Cardiologie et de Pneumologie de Québec** (Y Bossé, Prof M Laviolette MD), **Laval University, Québec, QC, Canada; Merck Research Laboratories, Genetics and Pharmacogenomics, Boston, MA, USA** (D C Nickle PhD); **University of Groningen, University Medical Center Groningen, Department of Pulmonology** (Prof D S Postma MD), **and Department of Pathology and Medical Biology** (Prof W Timens MD), **GRIAC Research Institute, University of Groningen, Groningen, Netherlands; Respiratory Division, Department of Medicine** (Prof A J Sandford, D D Daley, Prof D D Sin, Prof P D Paré), **and Department of Pathology and Laboratory Medicine** (Prof J C Hogg, W M Elliott), **University of British Columbia, Vancouver, BC, Canada; Department of Twin Research and Genetic Epidemiology, King's College, London, UK** (P G Hysi PhD);

Summary

Background Lung function measures reflect the physiological state of the lung, and are essential to the diagnosis of chronic obstructive pulmonary disease (COPD). The SpiroMeta-CHARGE consortium undertook the largest genome-wide association study (GWAS) so far (n=48 201) for forced expiratory volume in 1 s (FEV_1) and the ratio of FEV_1 to forced vital capacity (FEV_1/FVC) in the general population. The lung expression quantitative trait loci (eQTLs) study mapped the genetic architecture of gene expression in lung tissue from 1111 individuals. We used a systems genetics approach to identify single nucleotide polymorphisms (SNPs) associated with lung function that act as eQTLs and change the level of expression of their target genes in lung tissue; termed eSNPs.

Methods The SpiroMeta-CHARGE GWAS results were integrated with lung eQTLs to map eSNPs and the genes and pathways underlying the associations in lung tissue. For comparison, a similar analysis was done in peripheral blood. The lung mRNA expression levels of the eSNP-regulated genes were tested for associations with lung function measures in 727 individuals. Additional analyses identified the pleiotropic effects of eSNPs from the published GWAS catalogue, and mapped enrichment in regulatory regions from the ENCODE project. Finally, the Connectivity Map database was used to identify potential therapeutics in silico that could reverse the COPD lung tissue gene signature.

Findings SNPs associated with lung function measures were more likely to be eQTLs and vice versa. The integration mapped the specific genes underlying the GWAS signals in lung tissue. The eSNP-regulated genes were enriched for developmental and inflammatory pathways; by comparison, SNPs associated with lung function that were eQTLs in blood, but not in lung, were only involved in inflammatory pathways. Lung function eSNPs were enriched for regulatory elements and were over-represented among genes showing differential expression during fetal lung development. An mRNA gene expression signature for COPD was identified in lung tissue and compared with the Connectivity Map. This in-silico drug repurposing approach suggested several compounds that reverse the COPD gene expression signature, including a nicotine receptor antagonist. These findings represent novel therapeutic pathways for COPD.

Interpretation The system genetics approach identified lung tissue genes driving the variation in lung function and susceptibility to COPD. The identification of these genes and the pathways in which they are enriched is essential to understand the pathophysiology of airway obstruction and to identify novel therapeutic targets and biomarkers for COPD, including drugs that reverse the COPD gene signature in silico.

Funding The research reported in this article was not specifically funded by any agency. See Acknowledgments for a full list of funders of the lung eQTL study and the Spiro-Meta CHARGE GWAS.

Introduction

Pulmonary function measures reflect the normal and pathological state of the lungs. The most commonly used measures are the forced expiratory volume in 1 s (FEV_1) and the ratio of FEV_1 to forced vital capacity (FEV_1/FVC). These measurements are integral to the diagnosis of chronic obstructive pulmonary disease (COPD), and are also important long-term predictors of population morbidity and all-cause mortality.[1] Pulmonary function is determined by both environmental and genetic factors. Tobacco smoking is the major environmental risk factor for reduced pulmonary function. The genetic contribution to pulmonary function is well established, with estimates of heritability for FEV_1 as high as 50%.[2,3]

The SpiroMeta consortium and the Consortium of Heart and Aging Research in Genomic Epidemiology (CHARGE) have published several genome-wide association studies (GWAS) that identified 26 loci associated with FEV_1 and FEV_1/FVC in the general population.[4–6] Although these findings provide new insights into the genetic architecture of lung function, the exact genes and biological mechanisms underlying these

Research in context

Evidence before this study

We searched PubMed for reports published in English before Nov 1, 2014. We used the search terms "genome-wide association" and "eQTL" and "lung function" or "FEV_1" or "FEV_1/FVC" or "COPD". We additionally searched the published genome-wide association study (GWAS) catalogue. At the time of our search, no large-scale integration reports of GWAS and lung expression quantitative trait loci (eQTLs) existed for lung function measures or chronic obstructive pulmonary disease (COPD). Evidence from GWAS in other diseases suggests that integration of GWAS results with eQTLs in relevant tissue can identify the genes most likely to be responsible for the associations and unravel the molecular mechanisms underlying variation in lung function.

Added value of this study

We show to our knowledge for the first time that single nucleotide polymorphisms (SNPs) associated with lung function are enriched for lung eQTLs. This study identified a large number of SNPs/genes that determine the variation in lung function measures; these were enriched in developmental and inflammatory pathways. Lung function genes are supported by evidence from GWAS, eQTL, and mRNA associations with lung function and as such represent potential therapeutic and biomarker targets. Furthermore, we used the resulting lung gene expression signature to identify potential novel COPD drugs with an in-silico drug repositioning approach.

Implications of all the available evidence

Our study has identified genes underlying the variation in lung function in lung tissue, representing testable hypotheses for future in-vitro and in-vivo studies. Our results serve as a unique resource to the larger scientific community interested in the pathogenesis and genetics of obstructive lung diseases.

associations remain largely unclear. Single nucleotide polymorphisms (SNPs) can determine phenotypic traits by altering the quantity or function of the mRNA or protein for which the gene codes, or both.[7] Recent studies suggest that for complex traits, SNPs in regulatory regions that control the level of gene expression are over-represented in GWAS findings compared with coding variants.[8,9] Genetic loci that control gene expression are called expression quantitative trait loci (eQTLs) and can be identified by testing SNPs for association with mRNA or protein expression.[10] Gene regulation is often tissue-specific,[11,12] and hence to make meaningful discoveries of eQTLs for lung function and COPD, it is informative to study the genetic control of lung-specific gene expression. The lung eQTL consortium identified 468 300 *cis*-acting (affecting expression of genes within 1 Mb of the transcript start site) and 16 677 *trans*-acting (further than 1 Mb away or on a different chromosome) eQTLs out of 2 598 263 SNPs at a 10% false discovery rate (FDR) in a large-scale eQTL mapping study of 1111 human lung tissues.[13–17]

Systems genetics enables a global analysis of molecular mechanisms by integrating data for genetic variation with intermediate phenotypes such as gene expression (mRNA, protein or both), epigenetics changes, or metabolite levels and examining how they interact and converge to alter complex traits and diseases.[18–22] We hypothesised that a subset of SNPs identified in the SpiroMeta-CHARGE GWAS meta-analysis of about 48 000 individuals[5] affect variation in lung function by acting as eQTLs to change the level of expression of their target genes in the lung. In this study, a systems genetics approach was used to overlap results from the SpiroMeta-CHARGE GWAS meta-analysis with the lung eQTLs identified by the lung eQTL consortium and to leverage the SNP-mRNA-lung function correlations to unravel genes and molecular mechanisms underlying lung function variation.

Methods

Study design

The overall study design is shown in figure 1. The first step was the identification of lung function eSNPs: SNPs with an association p value (p_{GWAS}) <0·001 that act as *cis* or *trans* lung eQTLs at an FDR of less than 10%. This step formed the basis for all downstream analyses. At the SNP level, the enrichment of SpiroMeta-CHARGE GWAS hits for lung eQTLs was calculated, and the expected direction of association of mRNA with lung function was inferred given the SNP associations with lung function and mRNA. eSNPs were additionally assessed for functional annotations in the Encyclopedia of DNA Elements (ENCODE) dataset, and for evidence of pleiotropy in the National Human Genome Research Institute (NHGRI) human GWAS Catalog. At the gene level, lung function eSNP-regulated genes in lung tissue were tested for enrichment in pathways and gene ontology processes. Then the pathways were compared with lung function-related genes regulated by eSNPs in blood. The potential developmental role of lung function genes was tested by comparison with a study of the transcriptome in human fetal lung. The levels of expression of lung function genes in lung tissue were tested for association with lung function measures and COPD in participants from the eQTL study. Finally, the Connectivity Map (CMap) database was used to identify potential therapeutics in silico for COPD using the disease's gene expression signature in lung tissue.

Study cohorts

The methods and participant characteristics of the SpiroMeta-CHARGE lung function GWAS have been described in detail elsewhere.[5] Briefly, meta-analyses for cross-sectional lung function measures were undertaken for approximately 2·5 million HapMap II imputed SNPs

Department of Public Health, and Institute for Molecular Medicine (FIMM), University of Helsinki, Helsinki, Finland (Prof J Kaprio MD); **National Institute for Health and Welfare, Helsinki, Finland** (Prof J Kaprio); **Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK** (J F Wilson DPhil, O Polasek PhD); **MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, UK** (J F Wilson, C Hayward PhD); **Busselton Population Medical Research Institute, Busselton, WA, Australia** (J Hui PhD); **PathWest Laboratory Medicine of Western Australia, Nedlands, WA, Australia** (J Hui); **School of Population Health and School of Pahology and Laboratory Medicine, University of Western Australia, Nedlands, WA, Australia** (J Hui); **Research Unit of Molecular Epidemiology** (R Rawal PhD), **Institute of Genetic Epidemiology** (R Rawal), **and Institute of Epidemiology I** (Prof H Schulz MD), **Helmholtz-Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; Comprehensive Pneumology Center Munich (CPC-M), German Center for Lung Research, Munich, Germany** (Prof H Schulz); **University Hospital, Department of Internal Medicine B, Greifswald, Germany** (B Stubbe MD); **Faculty of Medicine, University of Split, Croatia** (O Polasek); **Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health** (Prof M-R Järvelin MD, Prof D Jarvis MD), **and Respiratory Epidemiology and Public Health Group, National Heart and Lung Institute** (Prof D Jarvis), **Imperial College, London, UK; Center for Life Course Epidemiology, Faculty of Medicine, Biocenter Oulu, and Unit of Primary Care, Oulu University Hospital, University of Oulu, Oulu, Finland** (Prof M-R Järvelin); **MRC Epidemiology Unit, University of Cambridge School of Clinical**

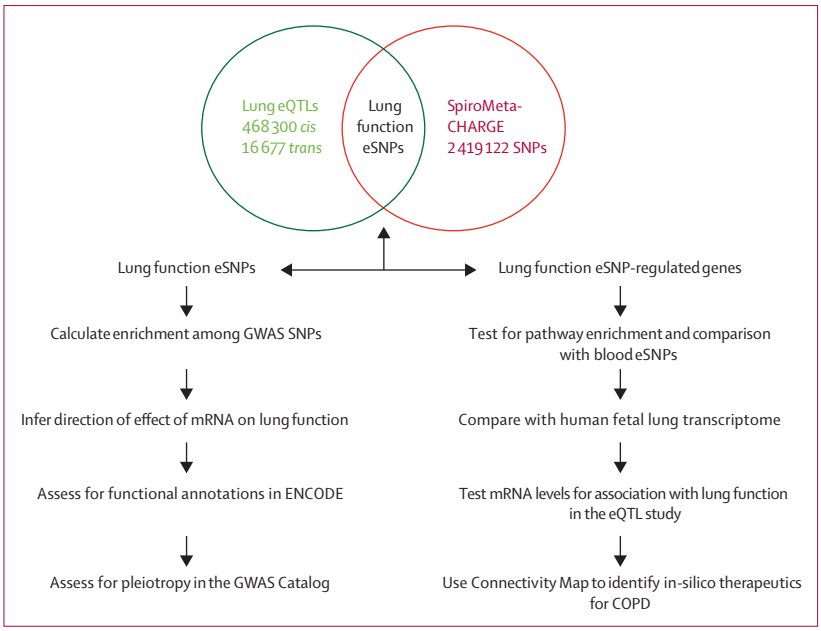

Figure 1: **Study design**
Lung function eSNPs were defined as GWAS SNPs with p_{GWAS}<0·001 that act as *cis* or *trans* lung eQTLs passing the 10% FDR. The enrichment of SpiroMeta-CHARGE GWAS for eSNPs was calculated. The expected direction of effect of mRNA on lung function was inferred. Lung function eSNP-regulated genes in lung tissue were tested for pathway and gene ontology processes enrichment, and the pathways were compared with blood lung function eSNP-regulated genes. The level of expression of lung function eSNP-regulated genes was tested for association with lung function measures in participants from the eQTL study. The lung function eSNPs and their target genes were tested for ENCODE functional enrichment and for transcription factor enrichment in the promoters of lung function eSNP-regulated genes. Furthermore, lung function eSNPs were integrated with the National Human Genome Research Institute human GWAS Catalog, to identify pleiotropic associations. The potential developmental role of lung function eSNP-regulated genes was tested by comparison with a study of the transcriptome of human fetal lung. Finally, the expression pattern of eSNP-regulated genes that were associated with COPD was interrogated using the Connectivity Map database to identify potential therapeutics in silico. eQTLs=expression quantitative trait loci. FDR=false discovery rate. GWAS=genome-wide association studies. SNP=single nucleotide polymorphisms.

Medicine, Institute of Metabolic Science, Cambridge UK (J H Zhao PhD)**; Department of Clinical Physiology, University of Tampere and Tampere University Hospital, Tampere, Finland** (Prof M Kähönen MD)**; Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA** (N Franceschini MD, Prof K E North PhD)**; University of North Carolina Center for Genome Sciences, Chapel Hill, NC, USA** (Prof K E North)**; Departments of Epidemiology and Respiratory Medicine, Erasmus MC, Rotterdam, Netherlands** (D W Loth MD, Prof G G Brusselle MD)**; Department of Respiratory Medicine, Ghent University Hospital, Ghent, Belgium** (Prof G G Brusselle)**; Icelandic Heart Association, Kopavogur, Iceland** (A V Smith PhD,

across 23 individual GWAS with a combined sample size of 48 201 adult individuals of European ancestry. The analyses of FEV_1 and FEV_1/FVC were adjusted for age, age^2, sex, and height, and where appropriate study centre and ancestry principal components as covariates.

The methods and participant characteristics of the lung tissue eQTL study have been described in detail elsewhere.[13] Briefly, lung eQTLs were derived from a meta-analysis of non-tumour lung tissue eQTLs from 1111 patients who underwent lung resection surgery at three participating sites: the University of British Columbia Centre for Heart and Lung Innovation (Vancouver, Canada; n=339), Laval University (Quebec City, Canada; n=409) and the University of Groningen (Groningen, Netherlands; n=363). The expression data are available at NCBI Gene Expression Omnibus repository (accession number GSE23546). Genotyping was performed on DNA extracted from blood or lung tissue with the Illumina Human1M-Duo BeadChip array, and imputation was performed using the HapMap II reference panel providing eQTL data for 2 598 263 million SNPs. With a Benjamini-Hochberg 10% FDR threshold, the study identified 468 300 *cis* (18% of all SNPs tested) and 16 677 *trans* eQTLs representing 0·64% of all SNPs tested.

Appropriate ethics approval for the lung function GWAS and the lung eQTL studies was received from all participating institutions.

Analyses and statistical analyses

To integrate the SNPs that were associated with FEV_1 or FEV_1/FVC, or both, with gene expression in the lung, we first merged the 468 300 *cis*-eQTLs and the 16 677 *trans*-eQTLs with SNPs in the SpiroMeta and CHARGE GWAS (2 419 122 SNPs at N effective >50%). Merged SNPs were then filtered to select those with a p_{GWAS}<0·001 for association with FEV_1 or FEV_1/FVC. Throughout this report we refer to SNPs associated with FEV_1 or FEV_1/FVC that act as eQTLs as lung function eSNPs, and the genes regulated by these SNPs as lung function eSNP-regulated genes. Fold enrichments were calculated from the merged results and significance of the enrichment of eSNPs among SpiroMeta-CHARGE GWAS results was caculated by Fisher's exact test.

The lung function eSNP-regulated genes were tested for enrichment in gene ontology biological processes and pathways with the Web-based Gene Set Analysis Toolkit (WebGestalt).[23]

For lung function eSNP-regulated genes, the relation between mRNA expression levels and lung function measures was examined in 727 of the 1111 participants in the lung eQTL study. The subset of participants who were selected had the appropriate measures of lung function and did not have a diagnosis (other than COPD or lung cancer) likely to affect lung function. Table 1 shows the demographics of this subgroup of participants. A linear regression analysis of the level of expression of each probe set on FEV_1 and FEV_1/FVC, adjusted for age, sex, height, and smoking status was done for each of three cohorts separately, followed by meta-analysis.

The allelic effect of the eSNPs on FEV_1 and FEV_1/FVC from the GWAS, and on mRNA from the lung eQTL study was used to infer the predicted direction of effect between the mRNA and lung function measures; we refer to this as the expected direction. Independently, we regressed the mRNA levels of lung function eSNP-regulated genes on FEV_1 and FEV_1/FVC in the subset of 727 participants of the eQTL study, and we refer to the direction of effect for this association as the observed direction. For lung function eSNP-regulated probe sets that show significant mRNA association with lung function measures, we investigated whether the observed direction of effect was either concordant or discordant with the expected direction.

To identify potential compounds that could reverse (or induce) COPD-associated genes, the Connectivity Map online tool from the Broad Institute was used.[24] The Connectivity Map hosts a publically available database of transcriptional profiles produced by existing drugs.

The database contains more than 7000 genome-wide transcriptomes from cultured human cells treated with 1309 bioactive compounds. Since the Connectivity Map expects a gene signature of upregulated and down-regulated genes as input, we tested the expression levels of lung function eSNP-regulated genes for differential expression between COPD cases (n=428) and controls (n=330) in individuals from the eQTL study. 51 lung function eSNP-regulated genes associated with COPD at nominal p values of <0·05 were mapped into Affymetrix platform HG-U133A probe set IDs (33 genes were remapped) to be used as input for the Connectivity Map to query compounds that could reverse or augment the airway obstruction gene signature.

Lung function eSNPs were tested for pleiotropy (a SNP that influences multiple diseases/traits) by integrating them with the NHGRI GWAS Catalog.[25] The ENCODE ChIP-Seq Significance Tool[26] was used to gain insights into what transcription factors were enriched in the promoters of lung function eSNP-regulated genes. At the SNP level, lung function eSNPs were tested for enrichment in ENCODE functional annotations with HaploReg v2 publically available software.[27]

We assessed whether SpiroMeta-CHARGE SNPs with p_{GWAS}<0·001 act as *cis*-eQTLs meeting 10% FDR in non-transformed peripheral blood mononuclear cells (PBMCs) obtained from 5311 European participants.[28] An eQTL was defined as *cis*-acting if the SNP position is less than 250 Kb away from the midpoint of the probe.

All statistical analyses were done with R version 3.0.1. Additional and detailed methods are available in the appendix (pp 3–6).

Role of the funding source

The funding sources had no roles in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

Of the 2419122 SNPs investigated in the SpiroMeta-CHARGE GWAS, 440665 were *cis*-eQTLs, and 15135 were *trans*-eQTLs that passed the 10% FDR in the lung eQTL study. Figure 2 shows a quantile-quantile (Q-Q) plot of the SpiroMeta-CHARGE associations with FEV_1 and FEV_1/FVC for the *cis*-eQTLs and *trans*-eQTLs. The y axis represents the quantiles of the eQTLs' associations with FEV_1 or FEV_1/FVC in the SpiroMeta-CHARGE GWAS, plotted against an expected distribution derived from the GWAS p values for all 2419122 SNPs. The plots show a large systematic deviation of p values for eQTLs compared with the GWAS association p values for all SNPs. Table 2 shows that of the 440665 *cis*-eQTLs, 3413 were associated with FEV_1 and 2205 were associated with FEV_1/FVC in the SpiroMeta-CHARGE GWAS study at p_{GWAS}<0·001. Of the 15135 *trans*-eQTLs, 1568 were associated with FEV_1 (38-fold enrichment) and 442 with FEV_1/FVC (12·6-fold enrichment) at p_{GWAS}<0·001. We refer to these eQTLs as lung function eSNPs. The enrichment is stronger for *trans*-eQTLs, possibly because of the more stringent statistical threshold needed to define a *trans*-eQTL. Enrichments at different p_{GWAS} cutoffs are shown in the appendix (pp 25–26). The deviation in the Q-Q plots and the enrichment in table 2 suggest that lung eQTLs are enriched for associations with lung function in the SpiroMeta-CHARGE GWAS, and vice versa. That a SNP is associated with both a lung function phenotype and with lung gene expression increases the likelihood that it is involved in a causal pathway,[10] especially since most SNPs identified in GWAS do not affect protein coding.

To validate these findings, lung eQTLs were overlapped with results from a GWAS meta-analysis for type 2 diabetes from Morris and colleagues.[29] The meta-analysis consisted of 12171 cases of type 2 diabetes and 56862 controls across 12 GWAS from European descent populations. Similar analyses to the lung function GWAS overlap were done to obtain Q-Q plots for blood and lung type 2 diabetes eSNPs (appendix p 7). The Q-Q plot shows that the enrichment for lung function SNPs in lung tissue eQTLs is systematically different compared with that for type 2 diabetes.

Since baseline enrichment could be expected for GWAS SNPs proximal to genes irrespective of being eQTLs in the relevant tissue, a comparison was made of SpiroMeta-CHARGE SNPs within 1 Mb of known

	UBC (n=251)	Laval (n=387)	Groningen (n=89)
Age (years)	63·53 (10·22)	63·48 (9·74)	61·54 (9·74)
Sex (male)	119 (47%)	171 (44%)	33 (37%)
Body-mass index (kg/m²)	25·76 (5·4)	26·56 (5·26)	25·17 (4·09)
FEV_1 % predicted	79·42 (23·06)	80·54 (19·04)	72·71 (24·77)
FEV_1/FVC	67·76 (12·46)	67·42 (9·73)	63·07 (16·98)
COPD*	107 (43%)	204 (53%)	53 (60%)
Stage			
1 (mild)	41 (38%)	80 (39%)	15 (28%)
2 (moderate)	57 (53%)	112 (55%)	28 (53%)
3 (severe)	2 (2%)	11 (5%)	2 (4%)
4 (very severe)	7 (7%)	1 (<1%)	8 (15%)
Non-COPD	123 (49%)	159 (41%)	32 (36%)
Smoking			
Smoker	86 (34%)	87 (22%)	31 (35%)
Ex-smoker	150 (60%)	270 (70%)	53 (60%)
Non-smoker	15 (6%)	30 (8%)	5 (6%)
Pack-years smoked	43·02 (30·13)	44·54 (29·48)	36·18 (19·05)

Data are n (%) or mean (SD). UBC=University of British Columbia, Vancouver, Canada. Laval=Laval University, Quebec City, Canada. Groningen=Groningen University, Groningen, Netherlands. COPD=chronic obstructive pulmonary disease. FEV_1=forced expiratory volume in 1 s. FEV_1/FVC=ratio of FEV_1 to forced vital capacity.*Numbers of patients with COPD and without COPD do not add up to the total sample size in each cohort; numbers of patients with unknown COPD status were 21 (8%) in UBC, 24 (6%) in Laval, and 4 (4%) in Groningen.

Table 1: **Demographics of 727 individuals in whom lung mRNA levels were tested for association with lung function**

Prof V Gudnason MD); **Faculty of Medicine, University of Iceland, Reykjavik, Iceland** (A V Smith, Prof V Gudnason); **Cardiovascular Health Research Unit, Departments of Medicine and Biostatistics** (T M Bartz MSc), **Computational Medicine Core, Center for Lung Biology** (S A Gharib MD), **Division of Pulmonary and Critical Care Medicine, Department of Medicine** (S A Gharib), **University of Washington, Seattle, WA, USA; Human Genetics & Computational Biomedicine, Pfizer Worldwide Research and Development, Cambridge, MA, USA** (J B Wilk PhD); **Pulmonary Center, Boston University School of Medicine, Boston, MA, USA** (Prof G T O'Connor MD); **NHLBI Framingham Heart Study, Framingham, MA, USA** (Prof G T O'Connor); **Division of Nutritional Sciences, Cornell University, Ithaca, NY, USA** (Prof P A Cassano PhD, W Tang PhD); **Division of Biostatistics and Epidemiology, Department of Healthcare Policy and Research, Weill Cornell Medical College, NY, USA** (Prof P A Cassano);

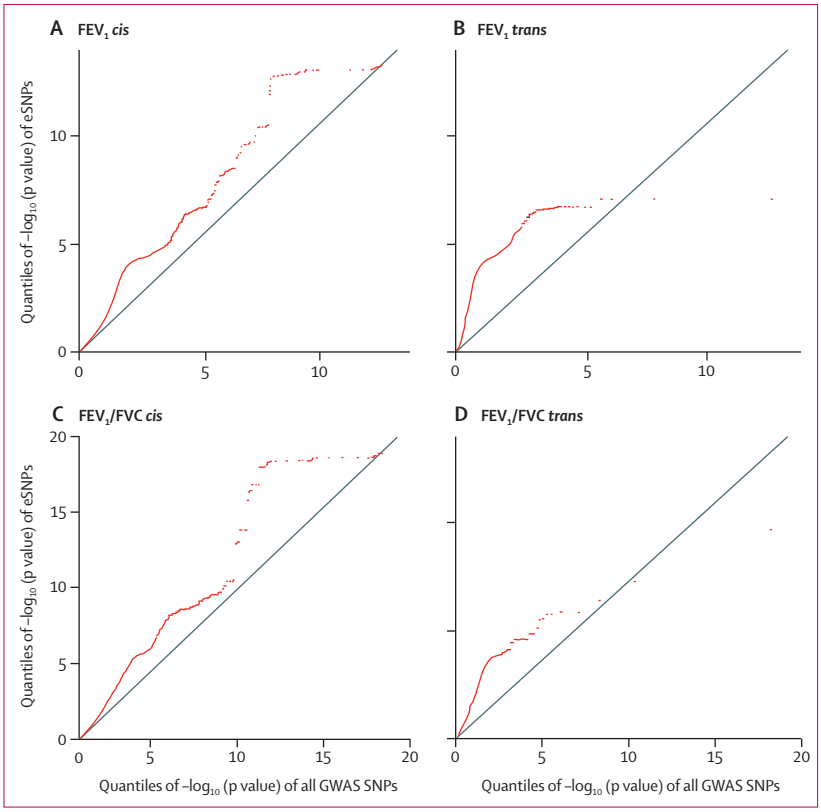

Figure 2: **Quantile-quantile (Q-Q) plot of lung function p_{GWAS} distribution for eSNPs versus lung function p_{GWAS} distribution for all SNPs**

On the y axis are quantiles of the distribution of lung function GWAS p values (p_{GWAS}) for eSNPs that pass 10% FDR. On the x axis are the quantiles of the distribution of lung function GWAS p values for all 2 419 122 SNPs. The systematic deviation of the line from the expected distribution shows lung function associated SNPs are enriched for lung eQTLs. (A) FEV_1 *cis*-eSNPs, (B) FEV_1 *trans*-eSNPs, (C) FEV_1/FVC *cis*-eSNPs, and (D) FEV_1/FVC *trans*-eSNPs. eQTLs=expression quantitative trait loci. FDR=false discovery rate. FEV_1=forced expiratory volume in 1 s. FEV_1/FVC=ratio of FEV_1 to forced vital capacity. GWAS=genome-wide association studies. SNP=single nucleotide polymorphisms.

	FEV_1 *cis*	FEV_1 *trans*	FEV_1/FVC *cis*	FEV_1/FVC *trans*
Number of eSNPs, n/N (%)	3413/6615 (52%)	1568/6615 (23%)	2205/5239 (42%)	442/5239 (8%)
Fold enrichment and p value	2·7*	37·9*	2·2*	12·6*
Number of eSNP-regulated probe sets	496	54	483	38
Number of eSNP-regulated genes	267	29	265	21

eSNPs refer to the SpiroMeta-CHARGE GWAS SNPs with $p_{GWAS}<0{\cdot}001$ acting as *cis*-eQTLs or *trans*-eQTLs at the 10% FDR. The number of genes reflects the conversion of probe sets to genes; on average there were two probe sets per gene. The table also shows the number of probe sets and genes regulated by eSNPs. The decrease in numbers between total eSNPs and eSNP-regulated probe sets reflects the fact that multiple eSNPs can be associated with the same probe set. The decrease in numbers from probe sets to genes reflects that the Affymetrix platform used tested on average two probe sets per gene. eQTLs=expression quantitative trait loci. FDR=false discovery rate. FEV_1=forced expiratory volume in 1 s. FEV_1/FVC=ratio of FEV_1 to forced vital capacity. GWAS=genome-wide association studies. SNP=single nucleotide polymorphisms. *Denotes the enrichment is significant with Fisher's exact test $p<2{\cdot}2\times10^{-16}$.

Table 2: **Number of *cis*-acting and *trans*-acting eSNPs and their relation to genes**

Boehringer Ingelheim Pharmaceuticals, Ridgefield,CT, USA (W Tang)**; University of Leicester, Genetic Epidemiology Group, Department of Health Sciences,**

genes to all SNPs. The Q-Q plots (appendix p 8) show a modest deviation for proximal SNPs but only for SNPs with low p_{GWAS}. However, the deviation is less than what is seen in the Q-Q plots of lung eSNPs (figure 2), suggesting that genomic distance alone does not explain the deviation noted for lung eSNPs and that this large deviation is driven by the tissue type used for eQTL discovery and its relevance to the phenotype under investigation.

The full list of lung function *cis* and *trans* eSNP-regulated genes is provided in the appendix (pp 27–68). A Venn diagram illustrating the extent of overlap for *cis* and *trans* FEV_1 and FEV_1/FVC eSNP-regulated genes is shown in the appendix (p 9). For example, there are 63 lung function *cis* eSNP-regulated genes associated with both FEV_1 and FEV_1/FVC.

Table 3 shows lung eQTL integration results for 50 SNPs within the 26 loci reported in three of the consortia's published meta-analyses that met the genome-wide significance threshold.[4-6] Of the 50 lung function-associated SNPs reported in the three meta-analyses, 25 (50%) act as lung eSNPs at 10% FDR. Table 3 shows the GWAS p values and the genes suggested to underlie the associations in the published reports (usually the closest gene), as well as the lung eQTL-regulated genes and the eQTL p values for the SNPs. For ten of the 25 eSNPs, the GWAS-suggested gene (based on SNP position) and the eSNP-regulated gene were the same. Another ten of the 25 eSNP-regulated genes were different from the ones suggested in the GWAS. For four of the 25, the eSNP regulated the expression of more than one gene, including the gene suggested by the GWAS. In these instances, the level of eQTL statistical significance with the alternative genes was usually stronger. Finally, one eSNP (rs2857595) regulated the expression of 11 genes, which did not include the GWAS suggested gene—*NCR3* in this case.[5] Some of the reported SNPs reside at the same locus and are in linkage disequilibrium.

The pathway analyses of lung function *cis* eSNP-regulated genes identified several enriched gene ontology processes that were common for FEV_1 and FEV_1/FVC (figure 3). These processes were related to development and inflammatory or immune responses such as natural killer T-cell differentiation, immune system processes, and positive regulation of the immune system. No gene ontology processes were significantly enriched (at any FDR value) among the 29 FEV_1 and 21 FEV_1/FVC *trans*-regulated genes. *Cis* and *trans* lung function eSNP-regulated genes were enriched for several Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (appendix p 10 and p 23).

To test whether the genes and pathways identified by integration of the GWAS and lung eQTL data were unique to lung tissue, we compared the results with those derived from integration of the SpiroMeta-CHARGE GWAS SNPs that had $p_{GWAS}<0{\cdot}001$ with *cis*-eQTL data derived from non-transformed PBMCs reported in a study of 5311 participants.[28] 3002 (45%) of the 6615 FEV_1-associated SNPs act as blood *cis*-eQTLs and 1958 (37%) of the 5239 FEV_1/FVC-associated SNPs act as blood *cis*-eQTLs (appendix p 69). These SNPs regulate the expression of 306 and 276 blood

	SNP	Locus	Lung function trait	Lung function GWAS p value*	GWAS-proposed gene	eQTL-regulated gene	eQTL p value	Reference
1p36.13	rs2284746	1p36.13	FEV_1/FVC	$7·5×10^{-16}$	*MFAP2*	*MFAP2*	$1·83×10^{-7}$	5
1p36.13	rs2284746	1p36.13	FEV_1/FVC	$7·5×10^{-16}$	*MFAP2*	*PADI2*	$2·31×10^{-10}$	5
1q41	rs993925	1q41	FEV_1/FVC	$1·16×10^{-8}$	*TGFB2*	*TGFB2*	$1·38×10^{-5}$	5
2q35	rs2571445	2q35	FEV_1	$1·11×10^{-12}$	*TNS1*	*TNS1*	$1·95×10^{-5}$	4
2q36.3	rs1435867	2q36.3	FEV_1/FVC	$1·53×10^{-5}$†	*PID1*	..	..	6
2q36.3	rs10498230	2q36.3	FEV_1/FVC	$1·46×10^{-5}$†	*PID1*	..	..	6
2q37.3	rs12477314	2q37.3	FEV_1/FVC	$1·68×10^{-12}$	*HDAC4*	..	..	5
3p24.2	rs1529672	3p24.2	FEV_1/FVC	$3·97×10^{-14}$	*RARB*	..	..	5
3q26.2	rs1344555	3q26.2	FEV_1	$2·65×10^{-5}$	*MECOM*	..	..	5
4q22.1	rs2869967	4q22.1	FEV_1/FVC	$1·91×10^{-7}$†	*FAM13A*	*FAM13A*	$1·74×10^{-6}$	6
4q22.1	rs6830970	4q22.1	FEV_1/FVC	$6·63×10^{-7}$†	*FAM13A*	..	..	6
4q24	rs10516526	4q24	FEV_1	$2·18×10^{-23}$	*GSTCD*	*NPNT*	$3·89×10^{-5}$	4
4q24	rs17331332	4q24	FEV_1	$5·69×10^{-15}$	*NPNT*	*NPNT*	$6·73×10^{-6}$	6
4q24	rs17036341	4q24	FEV_1	$2·18×10^{-15}$	*NPNT*	*NPNT*	$1·67×10^{-5}$	6
4q24	rs11727189	4q24	FEV_1	$4·66×10^{-17}$	*INTS12*	*NPNT*	$4·70×10^{-5}$	6
4q24	rs17036090	4q24	FEV_1	$5·61×10^{-15}$	*INTS12*	..	..	6
4q24	rs17036052	4q24	FEV_1	$1·83×10^{-15}$	*ARHGEF38 (FLJ20184)*	..	..	6
4q24	rs17035960	4q24	FEV_1	$9·42×10^{-14}$	*ARHGEF38 (FLJ20184)*	..	..	6
4q24	rs11097901	4q24	FEV_1	$3·26×10^{-16}$	*GSTCD*	*NPNT*	$3·85×10^{-5}$	6
4q24	rs11728716	4q24	FEV_1	$7·20×10^{-16}$	*GSTCD*	*NPNT*	$2·11×10^{-5}$	6
4q31.21	rs12504628	4q31.21	FEV_1/FVC	$6·48×10^{-13}$	*HHIP*	*HHIP*	$2·00×10^{-6}$	4
4q31.21	rs1980057	4q31.21	FEV_1/FVC	$3·21×10^{-20}$	*HHIP*	*HHIP*	$4·65×10^{-7}$	6
4q31.22	rs1032295	4q31.22	FEV_1/FVC	$4·37×10^{-15}$	*HHIP*	..	..	6
5q15	rs153916	5q15	FEV_1/FVC	$2·12×10^{-8}$	*SPATA9*	*RHOBTB3*	$5·77×10^{-16}$	5
5q32	rs3995090	5q32	FEV_1	$4·29×10^{-9}$	*HTR4*	..	..	4
5q33.1	rs6889822	5q33.1	FEV_1	$8·17×10^{-9}$	*HTR4*	*FBXO38*	$4·33×10^{-5}$	4
5q33.1	rs11168048	5q33.1	FEV_1/FVC	$1·08×10^{-11}$	*HTR4*	..	..	6
5q33.1	rs7735184	5q33.1	FEV_1/FVC	$6·23×10^{-11}$	*HTR4*	..	..	6
5q33.3	rs2277027	5q33.3	FEV_1/FVC	$9·93×10^{-11}$	*ADAM19*	*ADAM19*	$4·41×10^{-6}$	6
5q33.3	rs1422795	5q33.3	FEV_1/FVC	$2·62×10^{-10}$	*ADAM19*	..	..	6
6q21	rs2798641	6q21	FEV_1/FVC	$8·35×10^{-9}$	*ARMC2*	..	..	5
6p22.1	rs6903823	6p22.1	FEV_1	$2·18×10^{-10}$	*ZKSCAN3 (ZNF323)*	*BTN3A2*	$2·26×10^{-19}$	5
6p22.1	rs6903823	6p22.1	FEV_1	$2·18×10^{-10}$	*ZKSCAN3 (ZNF323)*	*HCG4P6*	$4·80×10^{-18}$	5
6p22.1	rs6903823	6p22.1	FEV_1	$2·18×10^{-10}$	*ZKSCAN3 (ZNF323)*	*HLA-A*	$3·46×10^{-28}$	5
6p22.1	rs6903823	6p22.1	FEV_1	$2·18×10^{-10}$	*ZKSCAN3 (ZNF323)*	*PGBD1*	$2·48×10^{-8}$	5
6p22.1	rs6903823	6p22.1	FEV_1	$2·18×10^{-10}$	*ZKSCAN3 (ZNF323)*	*ZFP57*	$1·39×10^{-25}$	5
6p22.1	rs6903823	6p22.1	FEV_1	$2·18×10^{-10}$	*ZKSCAN3 (ZNF323)*	*ZSCAN26 (ZNF187)*	$3·16×10^{-5}$	5
6p22.1	rs6903823	6p22.1	FEV_1	$2·18×10^{-10}$	*ZKSCAN3 (ZNF323)*	*ZSCAN31 (ZNF323)*	$3·17×10^{-20}$	5
6p21.2	rs2395730	6p21.2	FEV_1/FVC	$7·98×10^{-8}$	*DAAM2*	..	..	4
6p21.32	rs2070600‡	6p21.32	FEV_1/FVC	$3·07×10^{-15}$	*AGER*	*AGER*	$1·30×10^{-5}$	4
6p21.32	rs2070600‡	6p21.32	FEV_1/FVC	$3·07×10^{-15}$	*AGER*	*HLA-DRB6*	$2·62×10^{-14}$	4
6p21.32	rs2070600‡	6p21.32	FEV_1/FVC	$3·07×10^{-15}$	*AGER*	*HLA-DRB1*	$2·21×10^{-6}$	4
6p21.32	rs2070600‡	6p21.32	FEV_1/FVC	$3·07×10^{-15}$	*AGER*	*AGER*	$4·09×10^{-6}$	4
6p21.32	rs2070600‡	6p21.32	FEV_1/FVC	$3·07×10^{-15}$	*AGER*	*HLA-DQA2*	$7·18×10^{-12}$	4
6p21.32	rs2070600‡	6p21.32	FEV_1/FVC	$3·07×10^{-15}$	*AGER*	*HLA-DRB4*	$9·24×10^{-22}$	4
6p21.32	rs2070600‡	6p21.32	FEV_1/FVC	$3·07×10^{-15}$	*AGER*	*AGER*	$5·86×10^{-7}$	4
6p21.32	rs10947233	6p21.32	FEV_1/FVC	$6·66×10^{-12}$	*PPT2*	*HLA-DRB6*	$1·06×10^{-11}$	6
6p21.32	rs10947233	6p21.32	FEV_1/FVC	$6·66×10^{-12}$	*PPT2*	*HLA-DQA2*	$2·10×10^{-9}$	6
6p21.32	rs10947233	6p21.32	FEV_1/FVC	$6·66×10^{-12}$	*PPT2*	*AGER*	$1·92×10^{-5}$	6
6p21.32	rs10947233	6p21.32	FEV_1/FVC	$6·66×10^{-12}$	*PPT2*	*HLA-DRB1*	$3·43×10^{-5}$	6
6p21.33	rs2857595	6p21.33	FEV_1/FVC	$2·28×10^{-10}$	*NCR3*	*APOM*	$7·79×10^{-6}$	5

(Table 3 continues on next page)

Leicester, UK (L V Wain PhD, M Soler Artigas PhD, Prof M D Tobin PhD)**; National Institute for Health Research (NIHR) Leicester Respiratory Biomedical Research Unit, Glenfield Hospital, Leicester, UK** (L V Wain, M Soler Artigas, Prof M D Tobin)**; Population Health Research Institute, St George's, University of London, London, UK** (Prof D P Strachan MD)**; Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, NC, USA** (Prof S J London MD)**; and University of Nottingham Division of Respiratory Medicine, University Hospital of Nottingham, Nottingham, UK** (Prof I P Hall MD)

Correspondence to:
Prof Peter D Paré, The University of British Columbia Center for Heart Lung Innovation, St Paul's Hospital, Vancouver, BC, V6Z 1Y6, Canada
peter.pare@hli.ubc.ca

For **NCBI Gene Expression Omnibus** see http://www.ncbi.nlm.nih.gov/geo

For the **Connectivity Map** see http://www.broadinstitute.org/cmap/

See **Online** for appendix

	SNP	Locus	Lung function trait	Lung function GWAS p value*	GWAS-proposed gene	eQTL-regulated gene	eQTL p value	Reference
(Continued from previous page)								
6p21.33	rs2857595	6p21.33	FEV_1/FVC	$2{\cdot}28\times10^{-10}$	*NCR3*	*ATP6V1G2*	$1{\cdot}32\times10^{-6}$	5
6p21.33	rs2857595	6p21.33	FEV_1/FVC	$2{\cdot}28\times10^{-10}$	*NCR3*	*BTN3A2*	$2{\cdot}36\times10^{-11}$	5
6p21.33	rs2857595	6p21.33	FEV_1/FVC	$2{\cdot}28\times10^{-10}$	*NCR3*	*CDSN*	$1{\cdot}14\times10^{-6}$	5
6p21.33	rs2857595	6p21.33	FEV_1/FVC	$2{\cdot}28\times10^{-10}$	*NCR3*	*HLA-C*	$6{\cdot}60\times10^{-20}$	5
6p21.33	rs2857595	6p21.33	FEV_1/FVC	$2{\cdot}28\times10^{-10}$	*NCR3*	*HLA-DRB1*	$2{\cdot}72\times10^{-6}$	5
6p21.33	rs2857595	6p21.33	FEV_1/FVC	$2{\cdot}28\times10^{-10}$	*NCR3*	*HLA-DRB3*	$4{\cdot}47\times10^{-44}$	5
6p21.33	rs2857595	6p21.33	FEV_1/FVC	$2{\cdot}28\times10^{-10}$	*NCR3*	*HLA-DRB6*	$1{\cdot}36\times10^{-5}$	5
6p21.33	rs2857595	6p21.33	FEV_1/FVC	$2{\cdot}28\times10^{-10}$	*NCR3*	*PSORS1C1*	$3{\cdot}27\times10^{-14}$	5
6p21.33	rs2857595	6p21.33	FEV_1/FVC	$2{\cdot}28\times10^{-10}$	*NCR3*	*TUBB*	$2{\cdot}94\times10^{-8}$	5
6p21.33	rs2857595	6p21.33	FEV_1/FVC	$2{\cdot}28\times10^{-10}$	*NCR3*	*ZFP57*	$8{\cdot}56\times10^{-15}$	5
6q24.1	rs3817928	6q24.1	FEV_1/FVC	$1{\cdot}17\times10^{-9}$	*ADGRG6 (GPR126)*	..	..	6
6q24.1	rs7776375	6q24.1	FEV_1/FVC	$6{\cdot}71\times10^{-9}$	*ADGRG6 (GPR126)*	..	..	6
6q24.1	rs6937121	6q24.1	FEV_1/FVC	$1{\cdot}25\times10^{-8}$	*ADGRG6 (GPR126)*	..	..	6
6q24.1	rs11155242	6q24.1	FEV_1/FVC	$1{\cdot}45\times10^{-7}$†	*ADGRG6 (GPR126)*	..	..	6
9q22.32	rs16909898	9q22.32	FEV_1/FVC	$5{\cdot}34\times10^{-7}$†	*PTCH1*	..	..	6
9q22.32	rs10512249	9q22.32	FEV_1/FVC	$5{\cdot}75\times10^{-7}$†	*PTCH1*	..	..	6
10p13	rs7068966	10p13	FEV_1/FVC and FEV_1	$6{\cdot}13\times10^{-13}$	*CDC123*	*CAMK1D*	$9{\cdot}65\times10^{-18}$	5
10p13	rs7068966	10p13	FEV_1/FVC and FEV_1	$6{\cdot}13\times10^{-13}$	*CDC123*	*CDC123*	$9{\cdot}78\times10^{-7}$	5
10q22.3	rs11001819	10q22.3	FEV_1	$2{\cdot}98\times10^{-12}$	*C10orf11*	*C10orf11*	$1{\cdot}05\times10^{-8}$	5
12q13.3	rs11172113	12q13.3	FEV_1/FVC	$1{\cdot}24\times10^{-8}$	*LRP1*	*STAT6*	$3{\cdot}31\times10^{-11}$	5
12q23.1	rs1036429	12q23.1	FEV_1/FVC	$2{\cdot}3\times10^{-11}$	*CCDC38*	*SNRPF*	$4{\cdot}70\times10^{-16}$	5
15q23	rs12899618	15q23	FEV_1/FVC	$7{\cdot}24\times10^{-15}$	*THSD4*	..	..	4
16q21	rs12447804	16q21	FEV_1/FVC	$3{\cdot}59\times10^{-8}$	*MMP15*	..	..	5
16q23.1	rs2865531	16q23.1	FEV_1/FVC	$1{\cdot}77\times10^{-11}$	*CFDP1*	..	..	5
21q22.11	rs9978142	21q22.11	FEV_1/FVC	$2{\cdot}65\times10^{-8}$	*KCNE2*	..	..	5

The table reports lung eQTL results for genome-wide associated SNPs (listed by locus) in the three published FEV_1 and FEV_1/FVC consortia manuscripts.[4-6] Gene names shown in parentheses are previous gene symbols or synonyms. FEV_1=forced expiratory volume in 1 s. FEV_1/FVC=ratio of FEV_1 to forced vital capacity. GWAS=genome-wide association studies. SNP=single nucleotide polymorphisms. ..=SNP is not a lung eQTL at the 10% FDR. *Denotes the GWAS p value from the meta-analysis of discovery and replication stages. †Denotes that the SNP achieved genome-wide significance in the discovery cohort only in the CHARGE study. ‡SNP rs2070600 was also reported in the CHARGE study[6] ($p=3{\cdot}15\times10^{-14}$ for association with FEV_1/FVC). eQTLs=expression quantitative trait loci.

Table 3: **Results of integrating lung eQTLs with FEV_1 and FEV_1/FVC GWAS reported associations in the SpiroMeta and CHARGE consortia reports[4-6]**

eSNP-regulated genes for FEV_1 and FEV_1/FVC, respectively. A list of the blood eSNPs and the genes under their genetic control for both FEV_1 and FEV_1/FVC is provided in the appendix (pp 70–83). The Q-Q plots for blood *cis*-eSNPs are shown in the appendix (p 11), and for comparison we also show the Q-Q plot for lung eSNPs on the same graph. The plots show a deviation for FEV_1 blood *cis*-eSNPs from the distribution of p values for the whole GWAS that is very similar to the lung FEV_1 *cis*-eSNPs. Some FEV_1/FVC blood *cis*-eSNPs have higher p values compared with the GWAS result and a very different distribution from FEV_1/FVC lung eSNPs. Although many lung function eSNP-regulated genes overlapped between the two tissues, some were unique to either lung or blood (appendix p 12). For example, 88 FEV_1 *cis*-regulated genes were common to both lung and blood, whereas 127 and 154 FEV_1 *cis* genes were only under genetic control in lung and blood, respectively.

Similar to lung function eSNP-regulated genes derived from the lung tissue analysis, lung function eSNP-regulated genes derived from blood were enriched in processes related to development, maturation, and inflammatory processes (appendix p 13). To extend the analysis, we performed gene ontology process enrichment analyses on genes from three groups: lung-only regulated genes, blood-only regulated genes, and shared lung and blood regulated genes. The results (appendix p 84) showed that lung function eSNP-regulated genes restricted to blood were enriched solely for inflammatory processes, whereas eSNP-regulated genes in lung tissue and those shared between lung and blood were enriched for both developmental and inflammatory processes. In concordance with these results, lung function eSNP-regulated genes were over-represented among genes differentially expressed during human fetal lung development (appendix pp 86–116).

There were 868 unique lung function eSNP-regulated probe sets when overlap between FEV_1 and FEV_1/FVC,

including both *cis* and *trans*, was accounted for. Of these, 193 (22%), which mapped to 134 genes, showed significant associations of lung mRNA levels with either FEV_1 or FEV_1/FVC (at nominal p values of $p<0{\cdot}05$) in 727 participants from the lung eQTL study (table 4 and appendix pp 116–145). The expression of genes such as *SLC35A1, ARIH2, ZNRD1, PADI2, PABPC4, TRIM38, LINC00310, SPINK6, PTCH1*, and *TGFB2* showed the strongest associations with lung function. These genes are regulated by eSNPs that fell below genome-wide significance in the GWAS (the p_{GWAS} values for these eSNPs ranged from $9{\cdot}78\times10^{-04}$ to $1{\cdot}52\times10^{-06}$, and as such they did not meet the GWAS cutoff for significance at $p<5\times10^{-8}$). Furthermore, the Q-Q plots (appendix p 14) show a systematic deviation of lung function eSNP-regulated probe set associations with FEV_1 and FEV_1/FVC compared with the expected distribution under the null hypothesis of no association. This finding is expected in view of how these genes were identified; by GWAS and lung eQTL analyses, and the deviation supports the integrative approach for finding genes that underlie the phenotype. For 56% of the lung function eSNP-regulated probe sets, the direction of effect in the eQTL participants was concordant with the expected finding. The concordance at different p_{GWAS} cutoffs remained modest (appendix p 146).

The identification of eSNPs helps unravel interesting biological relations, such as the significant interaction detected on lung function between eSNPs for the receptor encoding gene *PTCH1* and its ligand competitor encoding *HHIP* (appendix p 24). Additionally, eSNPs in the *HHIP* locus were in high linkage disequilibrium with two functional SNPs identified in vitro by Zhou and colleagues[30] to regulate the expression of *HHIP* (appendix pp 147–148). Furthermore, lung eSNPs in the 4q31 (*HHIP*) and 15q25 (*CHRNA3/5*) loci were also recently identified by Castaldi and colleagues as COPD eSNPs in blood and sputum tissues (appendix p 149).[31]

The appendix (p 15) shows a Circos plot[32] to summarise results from the multiple associations for lung function eSNPs: the GWAS p values, lung eQTL p values, and the p values for the lung mRNA associations with FEV_1 and FEV_1/FVC in the lung eQTL study.

Lung function eSNPs were associated at GWAS Catalog p values ranging from 9×10^{-6} to 4×10^{-186} with several diseases and traits (appendix pp 150–204) that can be broadly classified into four categories: (1) inflammatory (including asthma, ulcerative colitis, type 1 diabetes, leprosy, C reactive protein levels); (2) developmental (including age at menopause, bone mineral density, age at onset of menarche, height, and primary tooth development); (3) neuropsychiatric (including schizophrenia, autism, Parkinson's disease, bipolar disorder, nicotine dependence, smoking behaviour); and (4) cardiovascular and obesity-related traits.

By use of ChIP-Seq data from the ENCODE project, lung function eSNP-regulated genes were seen to be enriched (at FDR $<0{\cdot}05$) in binding sites for several transcription factors. The top ranked transcription factors are c-Myc, NF-κb, P300, Pol2, Ebf1, and Sin3a. At the SNP level, lung function eSNPs were associated with several ENCODE functional annotations consistent with their role as regulatory SNPs. The enhancer and DNA hypersensitivity site enrichment for FEV_1 and FEV_1/FVC *cis*-eSNPs are shown in the appendix (pp 16–20). Additional results and discussion of integration of lung eQTL results with ENCODE data are provided in the appendix (p 23 and pp 205–209).

To gain insights and generate hypotheses about potential therapeutic agents for COPD, the mRNA levels of lung function eSNP-regulated genes were tested for association with COPD in individuals from the eQTL study (appendix pp 209–233). The resulting COPD gene

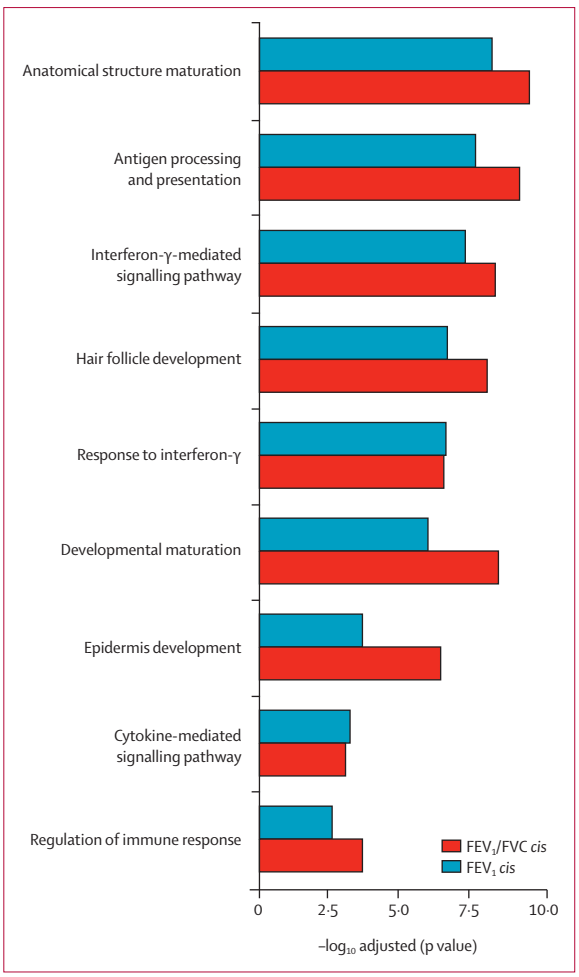

***Figure 3*: Gene ontology enrichment analyses of lung tissue eSNP-regulated genes**
The graph shows the enriched gene ontology processes on the y axis and their corresponding FDR-adjusted p values for enrichment on the x axis. There were no significant gene ontology processes enriched among FEV_1 and FEV_1/FVC *trans*-regulated genes. FDR=false discovery rate. FEV_1=forced expiratory volume in 1 s. FEV_1/FVC=ratio of FEV_1 to forced vital capacity. SNP=single nucleotide polymorphisms.

	GWAS p value	eQTL p value	Probe set ID	Gene Symbol	Expected direction*	Expected *vs* observed direction†	FEV_1 beta‡	FEV_1 SE§	FEV_1 p value	FEV_1/FVC beta¶	FEV_1/FVC SE‖	FEV_1/FVC p value
FEV_1 *cis*												
rs4489748	$3{\cdot}43\times10^{-4}$	$1{\cdot}79\times10^{-23}$	100122809_TGI_at	*ROM1*	Negative	Concordant	−1·54	0·42	$2{\cdot}21\times10^{-04}$	−15·55	7·81	$4{\cdot}64\times10^{-2}$
rs17339831	$9{\cdot}45\times10^{-4}$	$3{\cdot}22\times10^{-27}$	100122984_TGI_at	*NR3C1*	Negative	Discordant	0·13	0·035	$1{\cdot}92\times10^{-04}$	1·69	0·63	$7{\cdot}56\times10^{-3}$
rs17339831	$9{\cdot}45\times10^{-4}$	$5{\cdot}24\times10^{-13}$	100129487_TGI_at	*NR3C1*	Negative	Discordant	0·15	0·0406	$1{\cdot}42\times10^{-04}$	1·88	0·74	$1{\cdot}12\times10^{-2}$
rs3824658	$6{\cdot}74\times10^{-4}$	$1{\cdot}90\times10^{-15}$	100132638_TGI_at	*ITIH5*	Negative	Discordant	0·29	0·084	$5{\cdot}56\times10^{-04}$	5·929	1·56	$1{\cdot}49\times10^{-4}$
rs17638781	$6{\cdot}21\times10^{-4}$	$8{\cdot}23\times10^{-8}$	100137671_TGI_at	*SPINK6*	Negative	Concordant	−6·84	1·90	$3{\cdot}15\times10^{-04}$	−117·50	34·24	$6{\cdot}01\times10^{-4}$
rs1264709	$5{\cdot}18\times10^{-4}$	$1{\cdot}38\times10^{-8}$	100158228_TGI_at	*ZNRD1*	Positive	Discordant	−0·69	0·18	$1{\cdot}18\times10^{-04}$	−12·97	3·26	$7{\cdot}08\times10^{-5}$
rs2293476	$9{\cdot}58\times10^{-4}$	$2{\cdot}01\times10^{-8}$	100302787_TGI_at	*PABPC4*	Positive	Discordant	−0·12	0·03	$3{\cdot}59\times10^{-04}$	−2·42	0·61	$7{\cdot}70\times10^{-5}$
rs3932521	$6{\cdot}81\times10^{-4}$	$1{\cdot}08\times10^{-5}$	100306590_TGI_at	*SLC35A1*	Positive	Concordant	0·22	0·055	$4{\cdot}46\times10^{-05}$	3·06	1·00	$2{\cdot}17\times10^{-3}$
rs2804529	$8{\cdot}18\times10^{-4}$	$6{\cdot}24\times10^{-6}$	100309207_TGI_at	*VDAC2*	Negative	Concordant	−0·17	0·05	$4{\cdot}97\times10^{-04}$	−1·77	0·92	$5{\cdot}51\times10^{-2}$
rs8897	$9{\cdot}61\times10^{-4}$	$1{\cdot}72\times10^{-13}$	100314021_TGI_at	*ARIH2*	Positive	Discordant	−0·43	0·121	$1{\cdot}67\times10^{-04}$	−3·57	2·11	$9{\cdot}05\times10^{-2}$
FEV_1 *trans*												
rs1794282	$9{\cdot}70\times10^{-4}$	$1{\cdot}22\times10^{-10}$	100126143_TGI_at	*APOM*	Positive	Concordant	0·67	0·40	$9{\cdot}46\times10^{-2}$	−5·90	7·22	$4{\cdot}14\times10^{-1}$
rs3093975	$2{\cdot}54\times10^{-4}$	$9{\cdot}54\times10^{-11}$	100139688_TGI_at	*HCG18*	Positive	Concordant	0·16	0·09	$8{\cdot}33\times10^{-2}$	1·86	1·69	$2{\cdot}72\times10^{-1}$
rs9257744	$7{\cdot}61\times10^{-4}$	$4{\cdot}82\times10^{-16}$	100143968_TGI_at	*ZNF323*	Negative	Discordant	0·13	0·06	$4{\cdot}65\times10^{-2}$	1·43	1·16	$2{\cdot}18\times10^{-1}$
rs199439	$4{\cdot}72\times10^{-4}$	$5{\cdot}86\times10^{-12}$	100300682_TGI_at	*BRWD1*	Negative	Discordant	0·59	0·38	$1{\cdot}16\times10^{-1}$	11·09	6·81	$1{\cdot}03\times10^{-1}$
rs9257744	$7{\cdot}61\times10^{-4}$	$9{\cdot}21\times10^{-16}$	100301615_TGI_at	*ZSCAN31* (*ZNF323*)	Negative	Discordant	0·13	0·057	$2{\cdot}45\times10^{-2}$	1·39	1·03	$1{\cdot}78\times10^{-1}$
rs13193532	$1{\cdot}16\times10^{-5}$	$1{\cdot}30\times10^{-56}$	100303994_TGI_at	*ZFP57*	Negative	Concordant	−0·57	0·27	$3{\cdot}37\times10^{-2}$	−8·12	4·95	$1{\cdot}01\times10^{-1}$
rs3130825	$7{\cdot}60\times10^{-4}$	$7{\cdot}39\times10^{-12}$	100308127_TGI_at	*HCG18*	Positive	Concordant	0·34	0·16	$3{\cdot}97\times10^{-2}$	3·87	3·004	$1{\cdot}97\times10^{-1}$
rs538628	$4{\cdot}90\times10^{-4}$	$2{\cdot}55\times10^{-284}$	100308628_TGI_at	*LRRC37A4*	Positive	Concordant	0·10	0·06	$6{\cdot}30\times10^{-2}$	1·78	1·00	$7{\cdot}40\times10^{-2}$
rs9468413	$7{\cdot}06\times10^{-6}$	$2{\cdot}20\times10^{-10}$	100311443_TGI_at	*LRRC16A*	Positive	Concordant	0·60	0·38	$1{\cdot}22\times10^{-1}$	1·90	7·00	$7{\cdot}86\times10^{-1}$
rs3130453	$8{\cdot}58\times10^{-4}$	$3{\cdot}52\times10^{-11}$	100311825_TGI_at	*HLA-F-AS1*	Negative	Concordant	−0·7	0·31	$2{\cdot}46\times10^{-2}$	−15·69	5·61	$5{\cdot}17\times10^{-3}$
FEV_1/FVC *cis*												
rs9467772	$8{\cdot}22\times10^{-4}$	$1{\cdot}55\times10^{-5}$	100127556_TGI_at	*TRIM38*	Positive	Discordant	−0·16	0·06	$4{\cdot}87\times10^{-3}$	−3·85	1·03	$1{\cdot}96\times10^{-4}$
rs2834440	$1{\cdot}55\times10^{-4}$	$4{\cdot}19\times10^{-5}$	100128165_TGI_at	*LINC00310* (*C21orf82*)	Negative	Concordant	−3·038	0·86	$4{\cdot}31\times10^{-4}$	−51·77	15·71	$9{\cdot}83\times10^{-4}$
rs17638781	$9{\cdot}78\times10^{-4}$	$8{\cdot}23\times10^{-8}$	100137671_TGI_at	*SPINK6*	Negative	Concordant	−6·848	1·90	$3{\cdot}15\times10^{-4}$	−117·50	34·25	$6{\cdot}01\times10^{-4}$
rs10512248	$9{\cdot}27\times10^{-5}$	$1{\cdot}11\times10^{-5}$	100141529_TGI_at	*PTCH1*	Positive	Concordant	0·888	0·31	$4{\cdot}43\times10^{-3}$	19·17	5·59	$6{\cdot}07\times10^{-4}$
rs2235910	$1{\cdot}33\times10^{-4}$	$2{\cdot}28\times10^{-43}$	100143443_TGI_at	*PADI2*	Negative	Concordant	−0·068	0·02	$5{\cdot}68\times10^{-4}$	−1·24	0·31	$7{\cdot}56\times10^{-5}$
rs925284	$9{\cdot}79\times10^{-4}$	$8{\cdot}83\times10^{-6}$	100154282_TGI_at	*HOXB7*	Negative	Concordant	−0·56	0·20	$4{\cdot}18\times10^{-3}$	−10·74	3·52	$2{\cdot}27\times10^{-3}$
rs3094622	$4{\cdot}47\times10^{-4}$	$1{\cdot}31\times10^{-7}$	100158228_TGI_at	*ZNRD1*	Positive	Discordant	−0·69	0·18	$1{\cdot}18\times10^{-4}$	−12·97	3·26	$7{\cdot}08\times10^{-5}$
rs2834463	$6{\cdot}64\times10^{-4}$	$9{\cdot}42\times10^{-10}$	100301107_TGI_at	*LINC00310* (*C21orf82*)	Negative	Concordant	−1·16	0·52	$2{\cdot}42\times10^{-2}$	−32·48	9·30	$4{\cdot}79\times10^{-4}$
rs12037222	$1{\cdot}54\times10^{-4}$	$2{\cdot}67\times10^{-9}$	100302787_TGI_at	*PABPC4*	Positive	Discordant	−0·12	0·03	$3{\cdot}59\times10^{-4}$	−2·43	0·61	$7{\cdot}70\times10^{-5}$
rs1481345	$1{\cdot}52\times10^{-6}$	$2{\cdot}79\times10^{-6}$	100303767_TGI_at	*TGFB2*	Negative	Discordant	0·27	0·10	$8{\cdot}93\times10^{-3}$	5·81	1·83	$1{\cdot}48\times10^{-3}$
FEV_1/FVC *trans*												
rs13193532	$7{\cdot}71\times10^{-4}$	$2{\cdot}73\times10^{-18}$	100123483_TGI_at	*HCG4P6*	Negative	Concordant	−1·29	1·4	$3{\cdot}70\times10^{-1}$	−51·55	25·90	$4{\cdot}65\times10^{-2}$
rs1264376	$5{\cdot}33\times10^{-4}$	$4{\cdot}91\times10^{-10}$	100141210_TGI_at	..	Positive	Concordant	0·47	0·30	$1{\cdot}13\times10^{-1}$	11·88	5·40	$2{\cdot}78\times10^{-2}$
rs4713279	$7{\cdot}06\times10^{-4}$	$4{\cdot}88\times10^{-10}$	100155331_TGI_at	..	Positive	Concordant	0·32	0·15	$3{\cdot}52\times10^{-2}$	5·49	2·80	$5{\cdot}14\times10^{-2}$
rs1491106	$6{\cdot}02\times10^{-5}$	$4{\cdot}00\times10^{-20}$	100161662_TGI_at	*RBP2*	Positive	Concordant	0·06	0·08	$4{\cdot}55\times10^{-1}$	2·93	1·53	$5{\cdot}53\times10^{-2}$
rs3132450	$6{\cdot}75\times10^{-5}$	$7{\cdot}20\times10^{-16}$	100300398_TGI_at	*HLA-DQB1*	Positive	Discordant	−0·0006	0·006	$9{\cdot}25\times10^{-1}$	−0·18	0·11	$9{\cdot}92\times10^{-2}$
rs13193532	$7{\cdot}71\times10^{-4}$	$1{\cdot}30\times10^{-56}$	100303994_TGI_at	*ZFP57*	Negative	Concordant	−0·57	0·27	$3{\cdot}37\times10^{-2}$	−8·12	4·95	$1{\cdot}01\times10^{-1}$
rs3132450	$6{\cdot}75\times10^{-5}$	$4{\cdot}78\times10^{-16}$	100304009_TGI_at	*HLA-DQB1*	Positive	Discordant	−0·0005	0·006	$9{\cdot}29\times10^{-1}$	−0·17	0·11	$1{\cdot}02\times10^{-1}$
rs1264376	$5{\cdot}33\times10^{-4}$	$3{\cdot}07\times10^{-11}$	100304197_TGI_at		Positive	Concordant	0·29	0·20	$1{\cdot}55\times10^{-1}$	7·84	3·71	$3{\cdot}48\times10^{-2}$
rs1491106	$6{\cdot}02\times10^{-5}$	$9{\cdot}92\times10^{-19}$	100309242_TGI_at	*RBP2*	Positive	Concordant	0·09	0·09	$3{\cdot}03\times10^{-1}$	3·62	1·62	$2{\cdot}54\times10^{-2}$
rs3131784	$8{\cdot}63\times10^{-4}$	$1{\cdot}93\times10^{-17}$	100311825_TGI_at	*HLA-F-AS1*	Negative	Concordant	−0·70	0·31	$2{\cdot}46\times10^{-2}$	−15·69	5·61	$5{\cdot}17\times10^{-3}$

Gene names shown in parentheses are previous gene symbols or synonyms. *The direction of effect inferred from the SNP's association with lung function and mRNA expression. †Indicates whether the observed direction of effect of mRNA versus lung function in the 727 participants from the eQTL study was concordant or discordant with the expected direction. ‡Effect estimate of probe set's regression on FEV_1. §Standard error of probe set›s regression on FEV_1. ¶Effect estimates of probe set's regression on FEV_1/FVC. ‖Standard error of probe set's regression on FEV_1/FVC. eQTLs=expression quantitative trait loci. FEV_1=forced expiratory volume in 1 s. FEV_1/FVC=ratio of FEV_1 to forced vital capacity. GWAS=genome-wide association studies. ID=identifier. SNP=single nucleotide polymorphisms.

Table 4: **Results of lung function eSNP-regulated genes' mRNA regression on FEV_1 and FEV_1/FVC in 727 participants from the lung eQTL study**

signature was then used to query the Connectivity Map database of drug gene expression profiles to identify potential COPD therapeutics. Several agents had a negative enrichment score suggesting that they are predicted to reverse the COPD gene signature. These agents included the local anaesthetic and non-selective nicotinic receptor blocker adiphenine; disulfiram, which is used as a treatment for alcohol dependence and as a possible treatment for cocaine dependence; perphenazine, a dopamine D1 and D2 receptor antagonist; GABAA receptor antagonists (SR-95531; Gabazine), and anti-inflammatory and antioxidant agents (hecogenin and withaferin A). Among the agents that had a positive enrichment score (ie, predicted to induce COPD) were vorinostat and trichostatin A, both of which act as histone deacetylase inhibitors; alsterpaullone, a cyclin-dependent kinase inhibitor; and the chemotherapeutic agent doxorubicin (appendix pp 234–235).

Discussion

Studies of genetic associations and examination of gene expression in relevant tissues are important steps in unravelling the molecular mechanisms underlying common diseases. In this study, a systems genetics approach was used to integrate the largest published GWAS on the two major lung function parameters used clinically—FEV_1 and FEV_1/FVC—with a powerful lung tissue eQTL resource. The main findings are that lung function associated SNPs are enriched for lung eQTLs and vice versa; integration identifies the specific genes that are more likely to be responsible for the GWAS signal; lung function eSNP-regulated genes in lung tissue are involved in developmental and inflammatory pathways whereas lung function eSNP-regulated genes in PBMCs are associated only with inflammatory pathways; and the importance of developmental pathways is emphasised by the strong overrepresentation of lung function eSNP-regulated genes among genes that are differentially expressed during fetal lung development. Additionally, we noted that the associations of mRNA levels of lung function eSNP-regulated genes with lung function showed stronger associations to what is expected by chance in individuals from the lung eQTL study, and in-silico analysis showed that several compounds are predicted to reverse (nicotine receptor antagonists) or to induce (histone deacetylase inhibitors) the COPD gene signature and could guide discovery of new therapeutics.

Restricting susceptibility loci to those that achieve genome-wide significance in GWAS is recognised as overly conservative; however, including SNPs with higher p values will identify false-positive associations unless there is additional evidence to implicate these variants.[33] We chose to interrogate all SNPs with $p_{GWAS}<0{\cdot}001$, because this threshold might uncover biologically relevant yet statistically modest associations. The usefulness of this approach is suggested by the fact that 47% of the SNPs associated with FEV_1 and 42% of the SNPs associated with FEV_1/FVC were *cis*-acting or *trans*-acting eSNPs. This finding represents substantial enrichment since only 18% and 0·6% of all SNPs acted as *cis* and *trans* lung eQTLs, respectively.

The discovery of eSNPs could help to identify genes in causal pathways, especially in regions containing multiple genes. For example, intronic SNP rs10516526 in *GSTCD* on chromosome 4q24, which showed one of the strongest associations with FEV_1 ($p=4{\cdot}75\times10^{-14}$) maps to a linkage disequilibrium rich region containing the *GSTCD*, *INTS12*, and *NPNT* genes. In lung tissue, rs10516526 is an eQTL for *NPNT* but not for *GSTCD* or *INTS12*. However, in PBMCs[28] and lymphoblastoid cell lines,[15] the same SNP is an eQTL for *INTS12*, reflecting the tissue specificity of eQTLs. These results suggest that one mechanism through which this susceptibility locus could have its effect is through modulation of the lung tissue levels of *NPNT*, a conclusion strengthened by the association between FEV_1/FVC and mRNA levels for *NPNT* in the lung eQTL participants ($p=0{\cdot}01$) and by the strong NPNT staining of pulmonary endothelial cells and alveolar epithelial cells (appendix p 21).

The identification of *trans*-eSNPs in particular reveals novel genes that would not be implicated because of their genomic position. An example is the retinol binding protein (*RBP2*; appendix p 23). Alternatively, integration of GWAS and eQTL results provides additional support for commonly associated genes. For example, the intergenic SNP rs13141641 on 4q31, which showed one of the strongest GWAS associations with FEV_1/FVC ($p=8{\cdot}451\times10^{-18}$) is a lung (but not blood) eQTL for *HHIP*. *HHIP* eSNPs identified in this study were in strong linkage disequilibrium with functional SNPs reported by Zhou and colleagues[30] to affect the expression of *HHIP* in vitro (appendix pp 147–148).

HHIP blocks the hedgehog signalling pathway by binding to PTCH1 ligands, a membrane receptor for the hedgehog (HH) proteins, Sonic (SHH), Desert (DHH), and Indian hedgehog (IHH). Interestingly, a SNP (rs10512248) in *PTCH1* was associated with FEV_1/FVC ($p=9{\cdot}2\times10^{-05}$) and is an eQTL for *PTCH1* ($p=1{\cdot}1\times10^{-05}$). The availability of eQTL data thus allows the suggestion of a biologically plausible causal pathway for the *PTCH1*/*HHIP* combination of susceptibility alleles. The *HHIP* allele that is associated with lower lung function substantially increases the mRNA levels of *HHIP*, suggesting that decreased hedgehog signalling adversely affects lung development. Similarly, the *PTCH1* allele that is associated with lower lung function substantially decreases the expression of the receptor, which would also be expected to decrease hedgehog signalling. There was a strong interaction between the *HHIP* and *PTCH1* eSNPs on FEV_1 and FEV_1/FVC in the participants of the eQTL study (appendix p 22 and p 24). The availability of mRNA data from the lung tissue samples of the participants of the eQTL study provides additional support for this pathway; higher levels of

HHIP (p=0·01 for FEV_1 and p=0·003 for FEV_1/FVC) and lower levels of *PTCH1* (p=0·004 for FEV_1 and p=0·0006 for FEV_1/FVC) were associated with worse lung function in this cohort. Lung function eSNP-regulated genes derived from analysis of lung tissue, and to a lesser extent, lung function eSNP-regulated genes derived from the analysis of PBMCs were enriched for genes involved in development, maturation, and inflammatory processes. Lung function eSNP-regulated genes that were unique only to blood cells were enriched for inflammatory but not developmental processes. A link between lung development and growth in utero or in infancy and impaired lung function and COPD in adults has been previously proposed,[34–39] and indeed genes determining adult lung function have also been associated with reduced airway calibre in early childhood.[40] The molecular mechanisms explaining this link remain poorly understood. Lung function eSNP-regulated genes such as *TGFB2*, *HHIP*, *PTCH1*, *NOTCH4*, and *RBP2* are members of families that are well known for their role in lung development and growth.[34,41–49] Additional support for the role of developmental genes as determinants of lung function is the finding that lung function eSNP-regulated genes were enriched among genes that vary in expression during human fetal lung development.

Inflammation was the other enriched pathway. Inflammation is inextricably linked to tissue remodelling and repair processes, which can affect the lung parenchyma and airways to alter lung function.[50] Although inflammation and lung development could operate independently of each other, it is also feasible that the two processes interact. Inflammation may lead to activation of tissue repair and remodelling processes that reactivate genes involved in lung development and growth. Alternatively, genetically determined variation in lung development and growth could alter lung structure to affect particle deposition and the inflammatory response to toxic inhalants, such as tobacco smoke.

In this study, gene expression could be related to lung function in the same participants whose lung tissue was used to generate the lung eQTL data. The mRNA expression of 193 probe sets (22% of all probe sets tested) showed significant association with either FEV_1 or FEV_1/FVC at a nominal p value. This level of association is more than would be expected by chance as shown in the Q-Q plot. Of the 193 probe sets, 109 (56%) showed an observed direction of effect that was concordant with the expected direction of effect. There are several potential mechanisms that could explain discordant relations for the other 84 probe sets (44%). First, changes in gene expression may be a response to the disease process and the magnitude of this response may overwhelm the effect of the eSNP. Second, although the expected direction is based on the effect of the top eSNP for that gene, the same gene could also be under genetic control of other weaker eSNPs that collectively have an opposite direction of effect. Third, epigenetic modifications or microRNAs could also affect the levels of expression. Finally, the relation could simply be a false-positive finding, in view of our relaxed nominal p values threshold.

The identification of genes that underlie lung function variation in lung tissue could be an essential step for COPD drug and biomarker development. In-silico drug repositioning can identify existing drugs that recapitulate or reverse the gene signature associated with COPD. Such screening has proven to be a valuable method for drug repurposing.[51–53] The Connectivity Map data suggest that vorinostat and trichostatin A, both of which act as histone deacetylase inhibitors, can reproduce the COPD gene signature. Histone deacetylase activators such as theophylline have been suggested as a treatment for COPD, especially in combination with corticosteroids.[54] Trichostatin A has also been shown to cause emphysema in rats, and was associated with decreased vascular endothelial growth factor (VEGF) expression.[55] Another Connectivity Map candidate inducer of COPD, alsterpaullone, downregulates VEGF and fibroblast growth factor.[56] Connectivity Map analysis showed several compounds that reversed the COPD gene signature, including adiphenine, a local anaesthetic that non-selectively inhibits at least four different nicotinic acetylcholine receptor (nAChR) subtypes.[57,58] SNPs at the 15q25 locus containing genes encoding the nAChRs CHRNA3 and CHRNA5 have been associated with lung cancer, COPD, and lung function.[59–62] nAChR in the lung has been reported to have a role in cell proliferation and apoptosis in response to carcinogens,[63,64] as well as in inflammation.[65] Other compounds that reverse the COPD gene signature affect dopamine; these include disulfiram, which is used as a treatment for alcohol dependence and is being explored for use in cocaine dependence.[66] Disulfiram inhibits dopamine β-hydroxylase, which converts dopamine to norepinephrine.[66] Perphenazine is an antagonist of dopamine D1 and D2 receptors,[67] which also binds to the α-adrenergic receptor.[68] One study suggested that dopamine receptor and transporter genes (*DRD2* and *SLC6A3*) might have a role in the progression of COPD.[69]

To our knowledge, this study is the first to report large-scale integration of lung function GWAS and lung eQTLs. The datasets used are the largest so far and the results represent a resource to the scientific community as a whole. The availability of lung tissue mRNA levels, eQTLs, and lung function measures on the same individuals provided a unique complementary resource to lung function GWAS findings, and aids the translation of SNP associations into actionable targets.

Our study has some limitations. First, such large-scale integration of GWAS and eQTLs can lead to coincidental overlap that is not necessarily reflective of disease biology. Recent methods have been proposed to integrate GWAS and eQTLs such as colocalisation,[70] a weighted approach,[71] and signature matching.[72] Additionally, as the number of

respiratory-related tissue eQTLs increases, approaches employing joint analysis of eQTLs in multiple tissues[73] will be needed to increase power. Second, although the eQTL discovery and mRNA associations were adjusted for smoking status, this adjustment could have undermined the power to detect SNPs and mRNAs that show markedly different associations in smokers versus non-smokers (interaction). Additionally, no adjustments for pack-years were made. Projects that investigate these relations represent future avenues to follow up findings from this study. Finally, the Connectivity Map drug gene expression profiles were measured in non-respiratory cancer cell lines, which could be different from lung or airway-related expression profiles. Similar drug repurposing approaches in respiratory-related cells or tissues are warranted.

In summary, the systems genetics approach identified genes and molecular mechanisms that underlie the variation in lung function measures, generating hypotheses for future in-vitro and in-vivo studies. This study emphasises the importance of lung development and inflammatory pathways for lung function variation in adults. The finding that existing drugs can reverse the lung tissue gene signature associated with airflow obstruction suggests attractive candidates for interfering with the pathogenesis of COPD.

Contributors

PDP, IPH, MDT, SJL, SAG, and MO conceived and designed the study. KH, YB, DCN, DSP, ML, AJS, DDD, JCH, WTi, NF, MO, YN, and DDS participated in the lung eQTL data collection and analysis. PGH, JK, JFW, JH, RR, HS, BS, CH, OP, M-RJ, JHZ, DJ, MK, NF, KEN, DWL, GGB, AVS, VG, TMB, JBW, GTOC, PAC, WTa, MO, LVW, MSA, SAG, DPS, MDT, SJL, and IPH participated in GWAS data collection and analysis. MO, YN, NF, KH, and PDP participated in study data analyses. WME and PDP undertook functional analyses. MDT and SJL provided statistical support and advice. IPH, MDT, SJL, DDS, PDP, and MO provided advice on study conduct. MO and PDP wrote the manuscript. All authors discussed the results and implications and commented on the manuscript at all stages.

Declaration of interests

DCN is an employee of Merck and Co. DSP reports grants to the university and consultancy fees to the university from AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, Takeda, and TEVA outside of the submitted work. DDS reports personal fees from Amgen, grants and personal fees from AstraZeneca, personal fees from Boehringer Ingelheim, grants from Novartis, outside of the submitted work. DJ reports grants from the European Union. IPH reports grants from MRC during the conduct of the study. JK reports grants from the Academy of Finland and from the European Union during the conduct of the study, and personal fees from Pfizer, outside of the submitted work. JBW is an employee of Pfizer. PAC reports grants from National Institutes of Health during the conduct of the study. ML has received payments from Boston Scientific, AstraZeneca, and Merck for lectures and from GSK for a consultants' meeting. WTi reports personal fees from Pfizer, GSK, Chiesi, and Roche Diagnostics/Ventana, and grants from Dutch Asthma Fund, outside of the submitted work. WTa reports personal fees from Boehringer Ingelheim Pharmaceuticals, outside of the submitted work. All other authors declare no competing interests.

Acknowledgments

TwinsUK: The study was funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007-2013). The study also receives support from the National Institute for Health Research (NIHR) BioResource Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. SNP Genotyping was performed by The Wellcome Trust Sanger Institute and National Eye Institute via NIH/CIDR. SpiroMeta consortium: The research undertaken by MDT, MSA, and LVW was part-funded by the NIHR. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health. MDT holds a Medical Research Council Senior Clinical Fellowship (G0902313). JH: The Busselton Health Study acknowledges the generous support for the 1994/5 follow-up study from Healthway, Western Australia, and the Busselton community volunteers who assisted with data collection and the study participants from the Shire of Busselton. The Busselton Health Study is supported by The Great Wine Estates of the Margaret River region of Western Australia. MR-J: NFBC1966 received financial support from the Academy of Finland (project grants 104781, 120315, 129269, 1114194, Center of Excellence in Complex Disease Genetics and SALVE), University Hospital Oulu, Biocenter, University of Oulu, Finland (75617), the European Commission (EURO-BLCS, Framework 5 award QLG1-CT-2000-01643), NHLBI grant 5R01HL087679-02 through the STAMPEED programme (1RL1MH083268-01), NIH/NIMH (5R01MH63706:02), ENGAGE project and grant agreement HEALTH-F4-2007-201413, and the Medical Research Council, UK (G0500539, G0600705, PrevMetSyn/SALVE), EU Framework Programme 7 small-scale focused research collaborative project EurHEALTHAgeing 277849. The DNA extractions, sample quality controls, biobank up-keeping, and aliquotting was performed in the National Public Health Institute, Biomedicum Helsinki, Finland, and supported financially by the Academy of Finland and Biocentrum Helsinki. We thank Paula Rantakallio (launch of NFBC1966 and 1986), Outi Tornwall, and Minttu Jussila (DNA biobanking). IPH receives funding from the Medical Research Council (G1000861). EPIC-Norfolk study: We are grateful for the support and participation of participants in the study. We thank many colleagues who have worked on the EPIC-Norfolk study. We thank the staff at the Respiratory Health Network Tissue Bank of the FRQS for their valuable assistance collecting the lung tissue specimens for the eQTL study at Laval University. ARIC: We thank the staff and participants of the ARIC study for their important contributions. Funding: MO is a Postdoctoral Fellow of the Michael Smith Foundation for Health Research (MSFHR) and the Canadian Institute for Health Research (CIHR) Integrated and Mentored Pulmonary and Cardiovascular Training programme (IMPACT). JK: Wellcome Trust Sanger Institute. European Commission: ENGAGE--European Network for Genetic and Genomic Epidemiology. Academy of Finland (grant numbers 213506, 129680, 265240, 263278). JFW: Chief Scientist Office of the Scottish Executive. Royal Society. EU FP6. MDT: MRC, G0902313. JH: HealthWay, Western Australia; The Great Wine Estates of the Margaret River region of Western Australia. CH and OP: Medical Research Council; Ministry of Science, Education, and Sport of the Republic of Croatia; European Union framework program 6 EUROSPAN project. MRJ: Academy of Finland; University Hospital Oulu, Biocenter, University of Oulu, Finland; European Commission, EURO-BLCS Framework 5 award; NHLBI grant through the STAMPEED programme; NIH/NIMH; European Commission, ENGAGE project and grant agreement; Medical Research Council, UK, PrevMetSyn/SALVE; European Commission, EU Framework Programme 7 small-scale focused research collaborative project EurHEALTHAgeing. SJL: Supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences. JHZ is funded by the Medical Research Council, UK. DJ: Coordination of the ECRHS-phenotype acquisition, EU; Gabriel project-genotyping, EU. MK: Medical Research Fund of the Tampere University Hospital. The lung eQTL study at Laval University was supported by the Chaire de pneumologie de la Fondation JD Bégin de l'Université Laval, the Fondation de l'Institut Universitaire de Cardiologie et de Pneumologie de Québec, the Respiratory Health Network of the FRQS, the Canadian Institutes of Health Research (MOP - 123369), and the Cancer Research Society and Read for the Cure. YB is the recipient of a Junior 2 Research Scholar award from the Fonds de recherche Québec – Santé (FRQS). British 1958 Birth Cohort DNA collection: use of phenotype and genotype data, funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02. Genotyping for the B58C-WTCCC subset was funded by the Wellcome Trust grant 076113/B/04/Z. The B58C-T1DGC genotyping used resources provided by the Type 1 Diabetes Genetics Consortium, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney

Diseases, National Institute of Allergy and Infectious Diseases, National Human Genome Research Institute, National Institute of Child Health and Human Development, and Juvenile Diabetes Research Foundation International, and supported by U01 DK062418. B58C-T1DGC GWAS data were deposited by the Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research (CIMR), University of Cambridge, which is funded by Juvenile Diabetes Research Foundation International, the Wellcome Trust, and the National Institute for Health Research Cambridge Biomedical Research Centre; the CIMR is in receipt of a Wellcome Trust Strategic Award (079895). The B58C-GABRIEL genotyping was supported by a contract from the European Commission Framework Programme 6 (018996) and grants from the French Ministry of Research. ARIC: The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367, and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.

References

1 Schunemann HJ, Dorn J, Grant BJ, et al. Pulmonary function is a long-term predictor of mortality in the general population: 29-year follow-up of the Buffalo Health Study. *Chest* 2000; **118:** 656–64.

2 Lewiiter FI, Tager IB, McGue M, et al. Genetic and environmental determinants of level of pulmonary function. *Am J Epidemiol* 1984; **120:** 518–30.

3 Wilk JB, Djousse L, Arnett DK, et al. Evidence for major genes influencing pulmonary function in the NHLBI Family Heart Study. *Genet Epidemiol* 2000; **19:** 81–94.

4 Repapi E, Sayers I, Wain LV, et al. Genome-wide association study identifies five loci associated with lung function. *Nat Genet* 2010; **42:** 36–44.

5 Soler Artigas M, Loth DW, Wain LV, et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet* 2011; **43:** 1082–90.

6 Hancock DB, Eijgelsheim M, Wilk JB, et al. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet* 2010; **42:** 45–52.

7 Hindorff LA, Sethupathy P, Junkins HA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* 2009; **106:** 9362–67.

8 Nicolae DL, Gamazon E, Zhang W, et al. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet* 2010; **6:** e1000888.

9 Maurano MT, Humbert R, Rynes E, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science* 2012; **337:** 1190–95.

10 Schadt EE, Molony C, Chudin E, et al. Mapping the genetic architecture of gene expression in human liver. *PLoS Biol* 2008; **6:** e107.

11 Dimas AS, Deutsch S, Stranger BE, et al. Common regulatory variation impacts gene expression in a cell type–dependent manner. *Science* 2009; **325:** 1246–50.

12 Emilsson V, Thorleifsson G, Zhang B, et al. Genetics of gene expression and its effect on disease. *Nature* 2008; **452:** 423–28.

13 Hao K, Bossé Y, Nickle DC, et al. Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet* 2012; **8:** e1003029.

14 Lamontagne M, Couture C, Postma DS, et al. Refining susceptibility loci of chronic obstructive pulmonary disease with lung eQTLs. *PLoS ONE* 2013; **8:** e70220.

15 Obeidat Me, Miller S, Probert K, et al. GSTCD and INTS12 regulation and expression in the human lung. *PLoS ONE* 2013; **8:** e74630.

16 Dijkstra AE, Smolonska J, van den Berge M, et al. Susceptibility to chronic mucus hypersecretion, a genome wide association study. *PLoS ONE* 2014; **9:** e91621.

17 Thun GA, Imboden M, Ferrarotti I, et al. Causal and synthetic associations of variants in the *SERPINA* gene cluster with alpha1-antitrypsin serum levels. *PLoS Genet* 2013; **9:** e1003585.

18 Nadeau JH, Dudley AM. Systems genetics. *Science* 2011; **331:** 1015–16.

19 Civelek M, Lusis AJ. Systems genetics approaches to understand complex traits. *Nat Rev Genet* 2014; **15:** 34–48.

20 Li H. Systems genetics in "-omics" era: current and future development. *Theory Biosci* 2013; **132:** 1–16.

21 Powell JE, Henders AK, McRae AF, et al. The Brisbane Systems Genetics Study: genetical genomics meets complex trait genetics. *PLoS ONE* 2012; **7:** e35430.

22 Taneera J, Lang S, Sharma A, et al. A systems genetics approach identifies genes and pathways for type 2 diabetes in human islets. *Cell Metabolism* 2012; **16:** 122–34.

23 Wang J, Duncan D, Shi Z, et al. WEB-based gene set analysis toolkit (WebGestalt): update 2013. *Nucleic Acids Res* 2013; **41:** W77–W83.

24 Lamb J, Crawford ED, Peck D, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 2006; **313:** 1929–35.

25 Hindorff LA MacArthur J, Morales J, et al. A catalog of published genome-wide association studies. http://www.genome.gov/gwastudies (accessed April 25, 2014).

26 Auerbach RK, Chen B, Butte AJ. Relating genes to function: identifying enriched transcription factors using the ENCODE ChIP-Seq significance tool. *Bioinformatics* 2013; **29:** 1922–24.

27 Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 2012; **40:** D930–D934.

28 Westra H-J, Peters MJ, Esko T, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* 2013; **45:** 1238–43.

29 Morris AP, Voight BF, Teslovich TM, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* 2012; **44:** 981–90.

30 Zhou X, Baron RM, Hardin M, et al. Identification of a chronic obstructive pulmonary disease genetic determinant that regulates HHIP. *Hum Mol Genet* 2012; **21:** 1325–35.

31 Castaldi PJ, Cho MH, Zhou X, et al. Genetic control of gene expression at novel and established chronic obstructive pulmonary disease loci. *Hum Mol Genet* 2015; **24:** 1200–10.

32 Krzywinski MI, Schein JE, Birol I, et al. Circos: an information aesthetic for comparative genomics. *Genome Research* 2009; **19:** 1639–45.

33 Liu C. Brain expression quantitative trait locus mapping informs genetic studies of psychiatric diseases. *Neurosci Bull* 2011; **27:** 123–33.

34 Warburton D, Gauldie J, Bellusci S, et al. Lung development and susceptibility to chronic obstructive pulmonary disease. *Proc Am Thorac Soc* 2006; **3:** 668–72.

35 Checkley W, West KP, Wise RA, et al. Maternal vitamin A supplementation and lung function in offspring. *N Engl J Med* 2010; **362:** 1784–94.

36 Massaro D, Massaro DG. Lung development, lung function, and retinoids. *N Engl J Med* 2010; **362:** 1829–31.

37 Shi W, Warburton D. Is COPD in adulthood really so far removed from early development? *Eur Respir J* 2010; **35:** 12–13.

38 Ranganathan S. Lung development, lung growth and the future of respiratory medicine. *Eur Respir J* 2010; **36:** 716–17.

39 Shi W, Chen F, Cardoso WV. Mechanisms of lung development. *Proc Am Thorac Soc* 2009; **6:** 558–63.

40 Kerkhof M, Boezen HM, Granell R, et al. Transient early wheeze and lung function in early childhood associated with chronic obstructive pulmonary disease genes. *J Allergy Clin Immunol* **133:** 68–76.e4.

41 Chuang P-T, Kawcak TN, McMahon AP. Feedback control of mammalian Hedgehog signaling by the Hedgehog-binding protein, Hip1, modulates Fgf signaling during branching morphogenesis of the lung. *Genes Dev* 2003; **17:** 342–47.

42 Pepicelli CV, Lewis PM, McMahon AP. Sonic hedgehog regulates branching morphogenesis in the mammalian lung. *Curr Biol* 1998; **8:** 1083–86.

43 Alejandre-Alcázar MA, Michiels-Corsten M, Vicencio AG, et al. TGF-β signaling is dynamically regulated during the alveolarization of rodent and human lungs. *Dev Dyn* 2008; **237:** 259–69.

44 Bartram U, Speer CP. The role of transforming growth factor beta in lung development and disease. *Chest* 2004; **125:** 754–65.

45 Morimoto M, Liu Z, Cheng H-T, et al. Canonical Notch signaling in the developing lung is required for determination of arterial smooth muscle cells and selection of Clara versus ciliated cell fate. *J Cell Sci* 2010; **123:** 213–24.

46 Collins BJ, Kleeberger W, Ball DW. Notch in lung development and lung cancer. *Semin Cancer Biol* 2004; **14:** 357–64.

47 Chytil F. Vitamin A and lung development. *Pediatr Pulmonol* 1985; **1** (suppl): S115–17.

48 Biesalski HK, Nohr D. Importance of vitamin-A for lung function and development. *Mol Aspects Med* 2003; **24:** 431–40.

49 Malpel S, Mendelsohn C, Cardoso WV. Regulation of retinoic acid signaling during lung morphogenesis. *Development* 2000; **127:** 3057–67.

50 Hogg JC, Pierce RA. Remodelling of peripheral lung tissue in COPD. *Eur Respir J* 2008; **31:** 913–14.

51 Campbell J, McDonough J, Zeskind J, et al. A gene expression signature of emphysema-related lung destruction and its reversal by the tripeptide GHK. *Genome Med* 2012; **4:** 67.

52 Sirota M, Dudley JT, Kim J, et al. Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Sci Transl Med* 2011; **3:** 96ra77.

53 Jahchan NS, Dudley JT, Mazur PK, et al. A drug repositioning approach identifies tricyclic antidepressants as inhibitors of small cell lung cancer and other neuroendocrine tumors. *Cancer Discov* 2013; **3:** 1364–77.

54 Barnes PJ. Role of HDAC2 in the pathophysiology of COPD. *Annu Rev Physiol* 2009; **71:** 451–64.

55 Mizuno S, Yasuo M, Bogaard HJ, et al. Inhibition of histone deacetylase causes emphysema. *Am J Physiol Lung Cell Mol Physiol* 2011; **300:** L402–13.

56 Krishnapati LS, Ghaskadbi S. Identification and characterization of VEGF and FGF from Hydra. *Int J Dev Biol* 2013; **57:** 897–906.

57 Spitzmaul G, Gumilar F, Dilger JP, et al. The local anaesthetics proadifen and adiphenine inhibit nicotinic receptors by different molecular mechanisms. *Br J Pharmacol* 2009; **157:** 804–17.

58 Gentry CL, Lukas RJ. Local anesthetics noncompetitively inhibit function of four distinct nicotinic acetylcholine receptor subtypes. *J Pharmacol Exp Ther*2001; **299:** 1038–48.

59 Zhou H, Yang J, Li D, et al. Association of IREB2 and CHRNA3/5 polymorphisms with COPD and COPD-related phenotypes in a Chinese Han population. *J Hum Genet* 2012; **57:** 738–46.

60 Hardin M, Zielinski J, Wan ES, et al. CHRNA3/5, IREB2, and ADCY2 are associated with severe chronic obstructive pulmonary disease in Poland. *Am J Respir Cell Mol Biol* 2012; **47:** 203–08.

61 Pillai SG, Ge D, Zhu G, et al. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet* 2009; **5:** e1000421.

62 Cho MH, McDonald M-LN, Zhou X, et al. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *Lancet Respir Med* 2014; **2:** 214–25.

63 Schuller HM. Is cancer triggered by altered signalling of nicotinic acetylcholine receptors? *Nature Rev Cancer* 2009; **9:** 195–205.

64 Paleari L, Negri E, Catassi A, et al. Inhibition of nonneuronal alpha7-nicotinic receptor for lung cancer treatment. *Am J Respir Crit Care Med* 2009; **179:** 1141–50.

65 Gwilt CR, Donnelly LE, Rogers DF. The non-neuronal cholinergic system in the airways: an unappreciated regulatory role in pulmonary inflammation? *Pharmacol Ther* 2007; **115:** 208–22.

66 Kosten TR, Wu G, Huang W, et al. Pharmacogenetic randomized trial for cocaine abuse: disulfiram and dopamine beta-hydroxylase. *Biol Psychiatry* 2013; **73:** 219–24.

67 Ozdemir E, Bagcivan I, Gursoy S. Role of D(1)/D(2) dopamin receptors antagonist perphenazine in morphine analgesia and tolerance in rats. *Bosn J Basic Med Sci* 2013; **13:** 119–25.

68 Hals PA, Hall H, Dahl SG. Phenothiazine drug metabolites: dopamine D2 receptor, alpha 1- and alpha 2-adrenoceptor binding. *Eur J Pharmacol* 1986; **125:** 373–81.

69 Tashkin DP, Rabinoff M, Noble EP, et al. Association of dopamine-related gene alleles, smoking behavior and decline in FEV1 in subjects with COPD: findings from the Lung Health Study. *COPD* 2012; **9:** 620–28.

70 Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* 2014; **10:** e1004383.

71 Li L, Kabesch M, Bouzigon E, et al. Using eQTL weights to improve power for genome-wide association studies: a genetic study of childhood asthma. *Front Genet* 2013; **4:** 103.

72 He X, Fuller CK, Song Y, et al. Sherlock: detecting gene-disease associations by matching patterns of expression QTL and GWAS. *Am J Hum Genet* 2013; **92:** 667–80.

73 Flutre T, Wen X, Pritchard J, et al. A statistical framework for joint eQTL analysis in multiple tissues. *PLoS Genet* 2013; **9:** e1003486.