Research paper

Identification of novel differentially methylated sites with potential as clinical predictors of impaired respiratory function and COPD

Mairead L. Bermingham [a,*], Rosie M. Walker [a], Riccardo E. Marioni [a,b], Stewart W. Morris [a], Konrad Rawlik [c], Yanni Zeng [d], Archie Campbell [a,e], Paul Redmond [f], Heather C. Whalley [f], Mark J. Adams [f], Caroline Hayward [d], Ian J. Deary [b,f], David J. Porteous [a,b], Andrew M. McIntosh [a,b,f], Kathryn L. Evans [a,b]

[a] Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
[b] Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK
[c] Division of Genetics and Genomics, The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Roslin, UK
[d] Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
[e] Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK
[f] Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK

ARTICLE INFO

ABSTRACT

Background: The causes of poor respiratory function and COPD are incompletely understood, but it is clear that genes and the environment play a role. As DNA methylation is under both genetic and environmental control, we hypothesised that investigation of differential methylation associated with these phenotypes would permit mechanistic insights, and improve prediction of COPD. We investigated genome-wide differential DNA methylation patterns using the recently released 850 K Illumina EPIC array. This is the largest single population, whole-genome epigenetic study to date.

Methods: Epigenome-wide association studies (EWASs) of respiratory function and COPD were performed in peripheral blood samples from the Generation Scotland: Scottish Family Health Study (GS:SFHS) cohort ($n = 3781$; 274 COPD cases and 2919 controls). In independent COPD incidence data ($n = 149$), significantly differentially methylated sites (DMSs; $p < 3.6 \times 10^{-8}$) were evaluated for their added predictive power when added to a model including clinical variables, age, sex, height and smoking history using receiver operating characteristic analysis. The Lothian Birth Cohort 1936 (LBC1936) was used to replicate association ($n = 895$) and prediction ($n = 178$) results.

Findings: We identified 28 respiratory function and/or COPD associated DMSs, which mapped to genes involved in alternative splicing, JAK-STAT signalling, and axon guidance. In prediction analyses, we observed significant improvement in discrimination between COPD cases and controls ($p < .05$) in independent GS:SFHS ($p = .016$) and LBC1936 ($p = .010$) datasets by adding DMSs to a clinical model.

Interpretation: Identification of novel DMSs has provided insight into the molecular mechanisms regulating respiratory function and aided prediction of COPD risk. Further studies are needed to assess the causality and clinical utility of identified associations.

Fund: Wellcome Trust Strategic Award 10436/Z/14/Z.

© 2019 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Respiratory function is influenced by both environmental factors and genetic factors, with heritability estimates ranging from 39 to 66% [1,2]. Epigenetic modifications are at the interface of genetics and the environment. DNA methylation, the covalent binding of a methyl group to the 5′ carbon of cytosine-phosphate-guanine (CpG) dinucleotide sequences in the genome, is an epigenetic modification of DNA that is associated with gene expression. Epigenome-wide association studies (EWASs) have the potential to provide mechanistic insights into impaired respiratory function and COPD pathogenesis. Previous EWASs of spirometric measures of respiratory function and respiratory disease have however produced inconsistent results, with some identifying significant associations [3–6], and others not [7–9]. Moreover, there has been little consistency between the positive findings reported [9,10].

Research in context

Evidence before this study

We searched for articles in PubMed published in English up to July 25, 2018, with the search terms "DNA methylation" and "respiratory function", or "COPD". We found some evidence for association between differential DNA methylation and both respiratory function and COPD. Of the twelve previous studies identified, eight used peripheral blood samples (sample size [N] range = 100–1,085) and four used lung tissue samples (N range = 24–160). The number of CpG loci analysed range from 27,578 to 485,512. These studies have not identified consistent changes in methylation, most likely due to a combination of factors including small sample sizes, technical issues, phenotypic definitions, and study design. In addition, no previous study has: analysed a sample from a large single cohort; used the recently released Illumina EPIC array (which assesses ~850,000 CpG loci); adjusted both methylation data and phenotype for smoking history; or used both prevalent and incident COPD electronic health record data.

Added value of this study

To our knowledge, this is the largest single cohort epigenome-wide association study (EWAS) of respiratory function and COPD to date ($n = 3,781$). After applying stringent genome-wide significance criteria ($p < 3.6 \times 10^{-8}$), we found that DNA methylation levels at 28 CpG sites in peripheral blood were associated with respiratory function or COPD. Of these 28, seven were testable in an independent population sample: all seven showed consistent direction of effect between the two samples and three showed replication ($p < .007$ [0.05/7 CpG sites tested]). Our results suggest that adjustment of both the phenotypic and the DNA methylation probe data for smoking history, which has not been carried out in previous studies, reduces the confounding effects of smoking, identifies larger numbers of associations, and reduces the heterogeneity of effects across smoking strata. We used gene set enrichment and pathway analyses, together with an approach that combines DNA methylation results with gene expression data to provide evidence for enrichment of differentially methylated sites in genes linked to alternative splicing, and JAK-STAT signalling and axon guidance. Finally, we demonstrated that the inclusion of DNA methylation data improves COPD risk prediction over established clinical variables alone in two independent datasets.

Implications of all the available evidence

There is now accumulating evidence that DNA methylation in peripheral blood is associated with respiratory function and COPD. Our study has shown that DNA methylation levels at 28 CpG sites are robustly associated with respiratory function and COPD, provide mechanistic insights, and can improve prediction of COPD risk. Further studies are warranted to improve understanding of the aetiology of COPD, explore causality and to assess the utility of DNA methylation profiling in the clinical management of this condition.

Studies of lung tissue [5,8] have been constrained by sample availability, with the largest study to date comprising 160 subjects [5]. Inconsistency amongst the results of the peripheral blood-based studies [3,6,7,9,11]

is likely to be due to a number of factors, including small sample size (e.g., two studies had <200 samples) [6,11] and/or investigation of a relatively small number (~27,000) of CpG loci [3,11]. The study with the largest number of samples ($n = 1085$) analysed only 27,000 CpG loci, while the largest study using the 450 K array (the predecessor to the array used here) analysed 920 samples [12]. Differences in spirometric measures, definitions of COPD, study population characteristics and study design, in particular in the method used to adjust for smoking history, are also likely to be important sources of variation [9,10]. Smoking is established as a major risk factor for COPD [13], and previous genome-wide DNA methylation have focused on DNA methylation associated with smoking and COPD [5,14,15]. However, not all smokers develop COPD and >25% of COPD cases occur in never smokers [16]. Results from a growing number of studies suggested that impaired respiratory function and COPD are strongly associated with risk factors other than smoking [17–19], and have a strong genetic component [20–22] that generally acts independently of smoking [23]. To understand the pathological mechanisms of impaired respiratory function and COPD other than smoking we sought to identify robust associations by assessing methylation in a large single cohort sample, applying a more rigorous correction for smoking history and by performing sensitivity analyses. In contrast to prior studies, we used the recently released Illumina EPIC array, which interrogates over 850,000 methylation sites. All 3781 individuals in our sample were from a single cohort with extensive and consistent phenotyping comprising clinical investigation, questionnaire, and linkage to routine medical health records. The cross-sectional design of prior studies has limited their capacity to distinguish cause and effect [10]. To identity predictive biomarkers of COPD, and to provide insights into the causal nature of our findings we tested our findings for their predictive power. We used an independent subpopulation of 150 participants with incident COPD who were disease free at the time of blood sampling. Finally, where data were available, we attempted to replicate our EWAS and prediction findings in an independent cohort, LBC1936, drawn from the same population.

2. Material and methods

A flow chart showing the overall study design is outlined in Fig. 1, and full description of the methods is provided in the appendix.

2.1. Epigenome-wide association study

2.1.1. Cohort information

The Generation Scotland Scottish Family Health Study (GS:SFHS; ≥18 years of age at recruitment) [24] and Lothian Birth Cohort of 1936 (LBC1936; ~70 years of age at recruitment) [25] have extensive clinical, lifestyle, health and genetic data. Medical Research Ethics was obtained for all components of GS:SFHS and LBC1936. Written informed consent was obtained from all participants.

2.1.2. Genome-wide methylation profiling

In the GS:SFHS cohort, DNA methylation data was obtained from 5190 participants using peripheral blood collected at baseline [26]. DNA methylation was assessed using the Infinium MethylationEPIC BeadChip. Quality control procedures were implemented to identify and remove unreliable probes and samples, and probes on the X and Y chromosomes were excluded leaving data for 735,418 methylation loci in 5190 individuals. In the LBC1936, DNA methylation was assessed in whole blood samples from 1004 participants using the Illumina HumanMethylation450 BeadChip. Low-quality probes and samples, and probes on the X and Y chromosomes were removed, leaving 450,726 probes and 920 samples for inclusion in the analysis. M-values were calculated for both datasets. The M-values were then pre-corrected for relatedness (in GS:SFHS), array processing batch and estimated cell counts.

2.1.3. Trait data

Respiratory function was assessed at the time of blood sampling in 4193 GS:SFHS and 895 LBC1936 participants with methylation data. Forced expiratory volume in 1 s (FEV_1) and forced vital capacity (FVC) were measured in litres, using spirometry. Spirometry was performed three times, and the maximum values of FVC and FEV_1 were used in the analyses. Only pre-bronchodilator spirometry measures were available. Quality control of the phenotype data was undertaken to exclude participants with inaccurate spirometry or covariate data; 4171 and 895 individuals were retained in the GS: SFHS and LBC1936 samples respectively.

Following the Global Initiative for Obstructive Lung Disease (GOLD) criteria. Post-bronchodilator spirometry is used for GOLD COPD diagnosis. Pre-bronchodilator spirometry has been used in other studies [22,27], and has been shown to lead to minimal misclassification of moderate (GOLD 2) to severe (GOLD 4) COPD [27]. Therefore, Individuals with airflow limitation consistent with GOLD 2 or worse (FEV_1/FVC ≤0.7 and percent predicted FEV_1 ≤ 80%) were classified as cases in this study [28]. Individuals with FEV_1 >80% predicted and FEV_1/FVC >0.7 were classified as control subjects. Individuals meeting the criteria for GOLD stage 1 (FEV_1/FVC < 0.7, FEV_1 ≥ 80% predicted) were excluded from the comparison of COPD
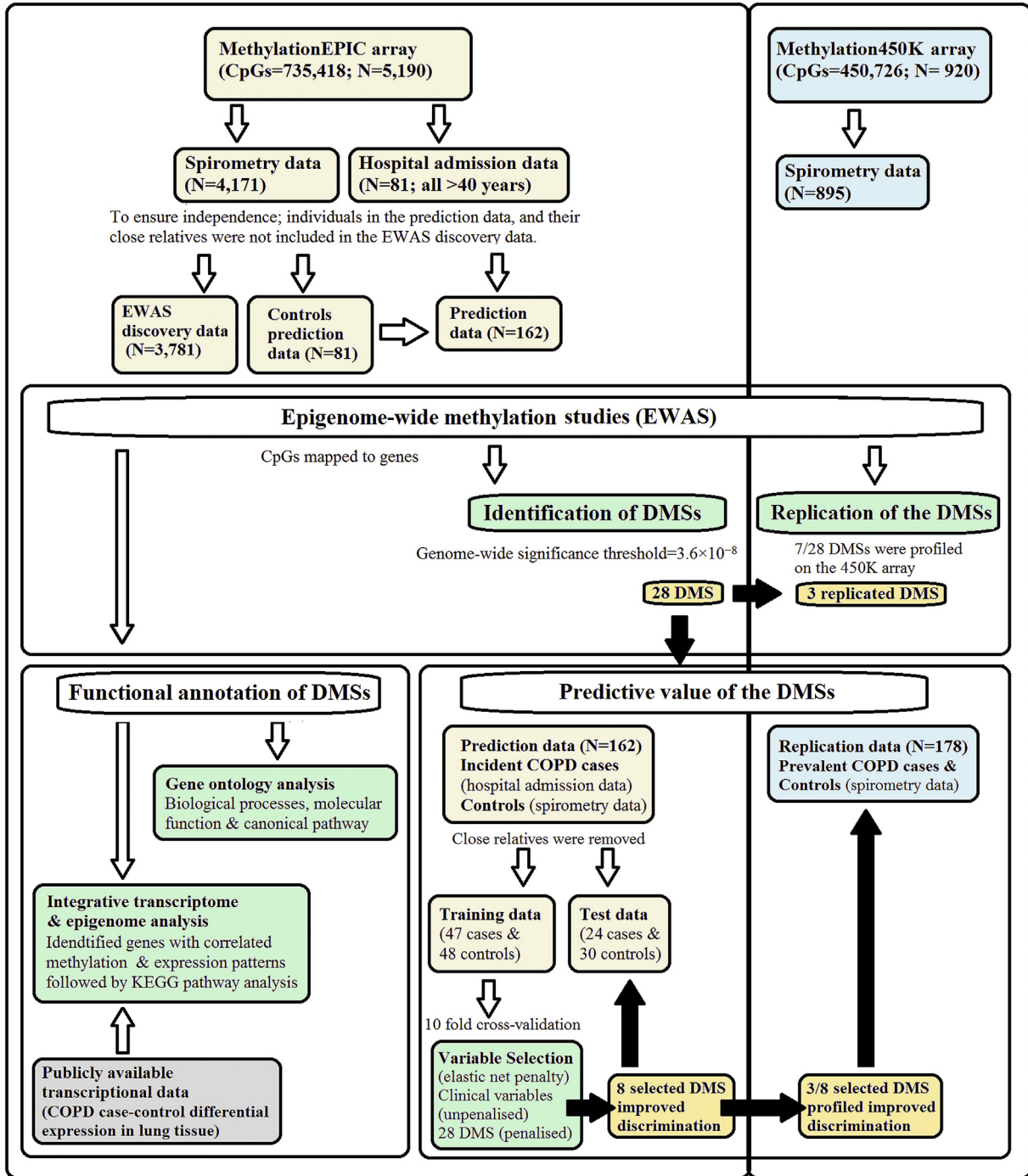

Fig. 1. Flow-chart showing the analysis pipeline. Direction of the arrows represents the workflow of the study design with performed analysis indicated. Lemon and blue boxes represent the in the discovery Generation Scotland: Scottish Family health study cohort and the replication Lothian Birth Cohort of 1936 (LBC1936) data sets respectively. The grey box indicates input data of COPD case-control differential expression in lung tissue. The green boxes indicate the analyses undertaken. The black arrows and gold boxes indicate output of significant results.

cases and controls to minimise potential misclassification of case and control subjects.

The GS:SFHS dataset ($n = 4171$) was then divided into an incident COPD dataset (i.e., where COPD developed after recruitment to the cohort) for prediction analysis and a discovery EWAS dataset. The prediction dataset (described below) comprised of incident COPD cases and matched controls. To ensure independence, individuals in the prediction data and their close relatives (identity by state [IBS] > 0.05) were not included in the discovery dataset; leaving 3781 for inclusion in the EWAS analysis.

2.1.4. Identification of differentially methylated sites (DMSs)

We first corrected the FEV_1, FVC, FEV_1/FVC, and COPD trait data for age, age2, sex, height, height [2], smoking status (current smoker, former smoker [quit <12 months], former smoker [quit ≥12 months], and never smoked), and pack-years using the R *stats* package. FVC data was additionally pre-corrected for weight. Linear regression models were then run in the *limma* package in R, fitting each CpG site (corrected M-values) as the dependent variable, and pre-corrected respiratory function traits or COPD, age, sex, smoking status, pack-years and the

Table 1
Characteristics of Generation Scotland: Scottish Family Health Survey (GS:SFHS) participants ($n = 3781$) in the epigenome-wide association study discovery population.

	Spirometry data			
	COPD cases ($n = 274$)	Controls ($n = 2919$)	GOLD stage 1 ($n = 588$)	Missing ($n = 905$)
Characteristics				
Age, years	53.96 ± 13.59	46.98 ± 13.37	52.15 ± 12.75	49.58 ± 15.56
Sex				
– Male	93 (33.9)	1162 (39.8)	222 (37.8)	335 (37.0))
– Female	181 (66.1)	1757 (60.2)	366 (62.2)	570 (63.0)
Height, cm	166.39 ± 8.76	167.85 ± 9.08	166.98 ± 9.47	166.22 ± 9.75
Weight, kg	73.50 ± 15.63	75.82 ± 16.33	76.60 ± 16.87	75.24 ± 17.45
Smoking status				
– Never	96 (35.0)	1603 (55.0)	249 (42.4)	388 (42.9)
– Former (quit >12 months)	74 (27.0)	715 (24.5)	164 (27.9)	242 (26.8)
– Former (quit <12 months)	3 (1.1)	88 (3.0)	20 (3.4)	31 (3.5)
– Current	89 (32.5)	443 (15.1)	133 (22.6)	206 (22.8)
– Missing records	12 (4.4)	70 (2.4)	22 (3.7)	25 (2.8)
Pack-year				
– Former smokers (quit >12 months)	25.47 ± 31.84	16.67 ± 20.31	22.83 ± 22.85	22.83 ± 22.87
– Former smokers (quit <12 months)	30.00 ± 31.05	15.16 ± 15.76	18.35 ± 20.20	18.35 ± 20.20
– Current smokers	22.97 ± 19.15	15.71 ± 15.95	23.59 ± 16.85	23.59 ± 16.85
Lung function				
– FEV_1, litres/s	2.01 ± 0.60	3.24 ± 0.76	2.58 ± 0.66	–
– FVC, litres/s	3.45 ± 0.93	4.07 ± 0.94	3.72 ± 1.08	–
– FEV_1/FVC	0.59 ± 0.10	0.80 ± 0.06	0.71 ± 0.08	–
– FEV_1 percent predicted	66.96 ± 11.44	99.56 ± 11.17	83.97 ± 12.98	–
– FVC percent predicted	89.90 ± 13.09	99.49 ± 11.90	94.93 ± 20.81	–

Abbreviations: COPD, Chronic obstructive pulmonary disease; FEV_1, Forced expiratory volume in 1 s; FVC, Forced vital capacity. Figures shown are the mean ± standard deviation or n (%).

first 20 principal components from the corrected M-values. The genome-wide significance threshold was set at 3.6×10^{-8} [29].

2.1.5. Sensitivity analyses

To assess the impact of pre-correction of the traits for smoking status and pack-years, we undertook sensitivity analyses, in which FEV_1, FVC, FEV_1/FVC, and COPD were not pre-corrected for smoking status and pack-years. To assess the stability of the estimated effects in older adults, according to smoking status (ever smokers and non-smokers), smoking history, and non-restrictive spirometry pattern (individuals with FEV_1/FVC ≥ 0.7 and FVC or FEV_1 < 80% of predicted were excluded). The data was truncated by age (≥ 40 years; an age group with greater risk of COPD), smoking history (≥ 10 pack-years; smokers with a substantial smoking history), non-restrictive spirometry and stratified by smoking status. For each trait-specific genome-wide significant DMS, we conducted separate random-effects meta-analysis to combine regression coefficients and standard errors from analyses across the full dataset and data-subsets using the 'rma' function from the R-package *metafor*. Heterogeneity of effects across analyses was assessed descriptively with the I^2 index. We formally tested heterogeneity of effects via Cochran Q statistic.

2.2. Probe annotation and epigenetic regulation of gene expression

DNA methylation probes were mapped to genes based on the IlluminaHumanMethylationEPICanno.ilm10b2.hg19 library. For each trait, methylation probes were filtered at $p < .001$, and probes that mapped to genes extracted.

For biological processes and molecular function, and canonical pathway enrichment analyses, DMSs were analysed in the Database for Annotation, Visualization and Integrated Discovery (DAVID) database and Ingenuity Pathway Analysis (IPA) software respectively. The Benjamini Hochberg, False Discovery Rate method, was used to correct for multiple-testing with $p < .05$ considered significant. We used the *Significance-based Modules Integrating the Transcriptome and Epigenome* (SMITE) package in R to combine summary statistics from publicly available COPD lung gene expression data [30] with methylation results from this study. Trait-specific gene modules (set of genes with shared regulation; $p < .05$ and 10–500 genes) were then identified and subjected to KEGG pathway enrichment analysis and terms with a $p < .05$ were held as significant.

2.3. Prediction

2.3.1. Case-control data

Incident cases in GS:SFHS were defined as any hospital admission where the primary diagnosis was assigned an ICD-10 J40 to J44 COPD exacerbation code [31]. During follow-up, 81 GS:SFHS participants (all 40 years or older) with DNA methylation data developed COPD. To obtain a balanced dataset for model training an equal number of controls ≥40 years of age were selected at random from those with no-self report, spirometry-defined, or ICD-10 diagnosis of COPD. Participants with missing records and closely related individuals (IBS > 0.05) were excluded; leaving 72 COPD cases and 78 controls. The data was then separated into a training set of 47 COPD cases and 48 controls, and a test set of 25 COPD cases and 30 controls.

As hospital admission data were not available for the LBC1936 cohort, spirometry data were used to define case-control status. In total, 89 participants with DNA methylation data had prevalent COPD [GOLD stage ≥2 cases]. Imbalanced data can negatively impact predictive performance. Controls were therefore selected at random from the participants with DNA methylation data in a 1:1 ratio to case participants. This dataset was used to replicate the prediction findings from GS:SFHS.

2.3.2. Model selection

For the training data, the reduced model, including clinical risk factors, age, age [2], sex, height, height [2], smoking status (current, former and never), and pack-years of smoking [6] was constructed using unpenalized logistic regression. The full model, including DMSs and clinical risk factors, was constructed using penalized logistic regression with an elastic net penalty. Selection of the full model was conducted based on 10-fold cross-validation (appendix p26) using the R package *caret*. The optimal model was selected based on the maximum mean area under the curve (AUC). Final models were constructed using the complete training set and evaluated on the independent test and replication datasets.

2.3.3. Model evaluation

Comparison of the predictive performance of the models was carried out using the AUC in the *pROC* R package. The incremental value of the DMS to predict COPD risk, when added to the model with established clinical predictors was assessed using the integrated discrimination improvement (IDI), and binary net reclassification improvement (NRI) measure. Finally, we performed decision curve analysis to estimate the potential clinical usefulness of the models in the 'rmda' R package.

3. Results

3.1. EWAS sample characteristics

The discovery sample for respiratory function traits comprised 3781 individuals from GS:SFHS. For the COPD analysis, there were 274 cases and 2919 controls (Table 1; Fig. 1).

3.2. Differentially methylated sites

EWASs for the three respiratory function traits (FEV$_1$, FVC and FEV$_1$/FVC) and COPD on the discovery data identified 29 genome-wide significant associations ($p < 3.6 \times 10^{-8}$; Fig. 2; Table 2; appendix p6–7), representing 28 DMSs from 25 annotated genes. Fourteen of the DMS were not associated with smoking status (Table 2). We found only
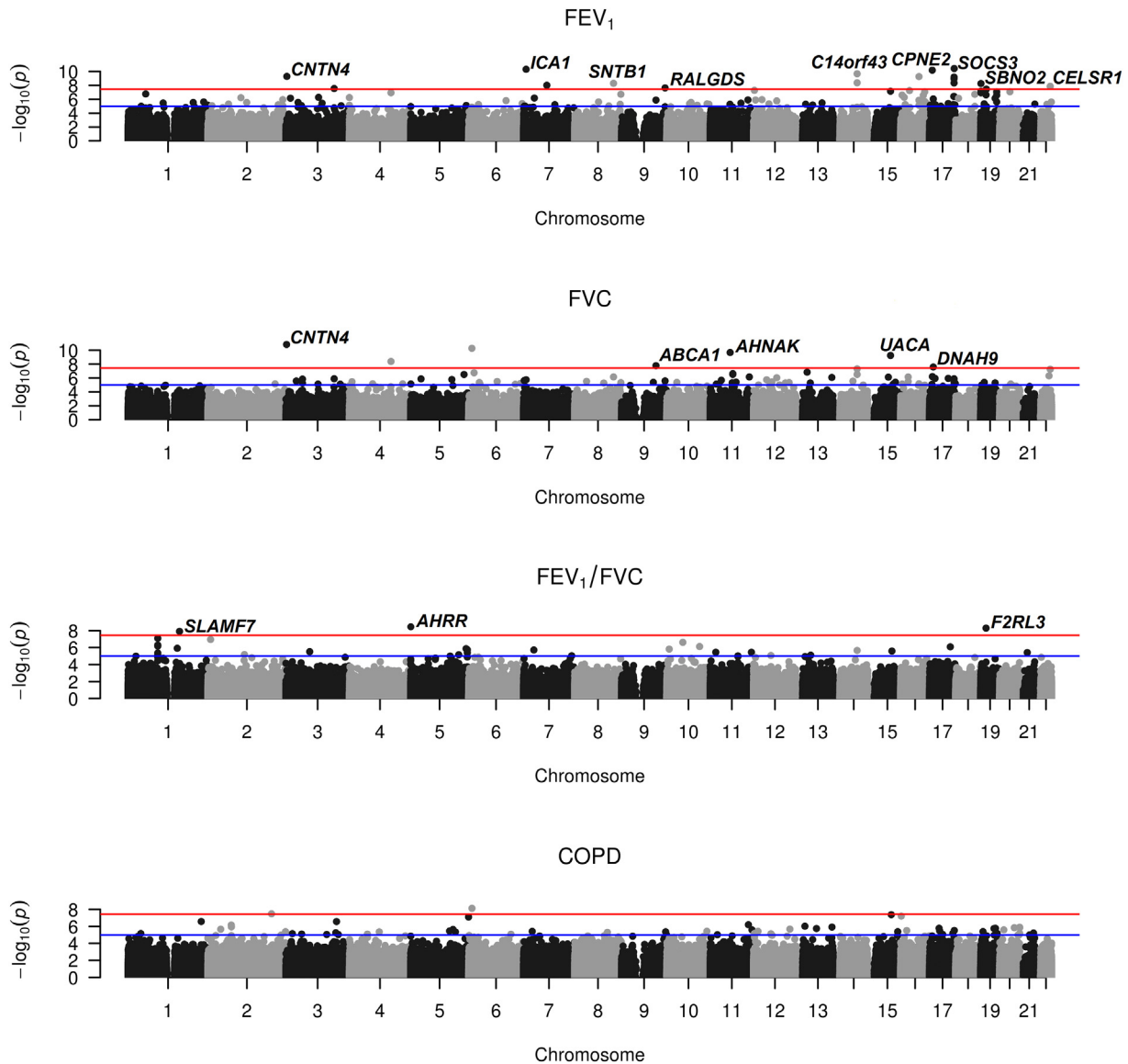

Fig. 2. Manhattan plots of epigenome-wide association results for FEV$_1$ (forced expired volume in 1 s), FVC (forced vital capacity; bottom), FEV$_1$/FVC and COPD (chronic obstructive pulmonary disease) from the discovery Generation Scotland: Scottish Family health study cohort data. The red line correspond to the genome-wide ($p = 3.6 \times 10^{-8}$) and suggestive ($p = 1.0 \times 10^{-5}$) significance level. Labels are for the nearest gene to genome-wide significant CpG sites.

Table 2

Genome-wide significant differentially methylated sites (DMSs) associated with the respiratory function traits or chronic obstructive pulmonary disease (COPD) in the Generation Scotland Scottish Family Health Study (GS:SFHS) discovery data. Results are ordered by chromosomal location.

Trait	Chr	Base pair	Gene name/annotation (region)	CpG site (location)	β (*p*-value)	DMSs$_{es}$
FEV$_1$	3	3,010,002	*CNTN4* (Body)	cg13993467 (Open sea)	−0.023 (5.19E-10)	6.31E-01
	3	150,479,084	*SIAH2* (Body)	cg16963852 (North shore)	0.015 (2.88E-08)	1.36E-07
	7	8,201,134	*ICA1* (Body)	cg26804423 (Open sea)	−0.011(4.88E-11)	3.44E-02
	7	72,775,853	[*FKBP6*, 3 kb, 3′ *]	cg26080684 (Open sea)	−0.010 (9.75E-09)	**5.85E-09**
	8	121,597,619	*SNTB1* (Body)	cg01198738 (Open sea)	0.013 (5.12−09)	1.42E-03
	9	136,009,651	*RALGDS* (Body)	cg03770138 (Open sea)	0.010 (2.34E-08)	**1.32E-15**
	14	74,227,431	*ELMSAN1* (TSS1500)	cg18871648 (South shore)	0.014 (4.30E-09)	**1.10E-59**
	14	74,227,441	*C14orf43* (5′UTR)	cg10919522 (South shore)	0.014 (2.14E-10)	**1.14E-47**
	16	57,180,107	*CPNE2* (Body)	cg09018739 (Open sea)	−0.008 (5.49E-10)	**8.97E-14**
	17	8,844,581	*PIK3R5* (5′UTR)	cg07687574 (Open sea)	0.013 (6.68E-11)	**3.15E-27**
	17	76,274,856	*LOC100996291* (TSS1500)	cg19748455 (Open sea)	0.015 (1.11E-9)	5.57E-04
	17	76,354,621	*SOCS3* (Body)	cg18181703 (North shore)	0.011 (4.51E-09)	**7.38E-14**
	17	76,354,934	*SOCS3* (Body)	cg11047325 (Island)	0.019 (3.91E-11)	**7.54E-22**
	17	76,355,061	*SOCS3* (Body)	cg13343932 (Island)	0.014 (6.52E-10)	**8.89E-25**
	19	1,130,866	*SBNO2* (Body)	cg18608055 (Open sea)	0.010 (5.52E-09)	1.15E-04
	19	17,955,786	*JAK3* (5′UTR)	cg02370334 (North Shelf)	0.011 (3.48E-08)	1.76E-01
	22	46,884,476	*CELSR1* (Body)	cg03187361 (Open sea)	0.012 (1.39E-08)	1.49E-05
FVC	3	3,010,002	*CNTN4* (Body)	cg13993467 (Open sea)	−0.025 (1.71E-11)	6.31E-01
	4	129,715,236	[*JADE1*, 16 kb, 5′ *]	cg01620970 (Open sea)	−0.012 (4.11E-09)	**1.71E-08**
	6	10,210,316	[*OFCC1*, Body *]	cg00213822 (Open sea)	−0.020 (5.23E-11)	4.74E-02
	9	107,631,656	*ABCA1* (Body)	cg15659943 (Open sea)	−0.009 (1.68E-08)	**9.96E-09**
	11	62,269,149	*AHNAK* (Body)	cg25465557 (Open sea)	−0.013 (2.11E-10)	2.78E-06
	15	71,041,066	*UACA* (Body)	cg18007249 (Open sea)	−0.013 (5.60E-10)	8.49E-01
	17	11,608,711	*DNAH9* (Body)	cg13108341 (Open sea)	−0.032 (2.57E-08)	7.91E-02
FEV$_1$/FVC	1	160,714,299	*SLAMF7* (5′UTR)	cg00045592 (Open sea)	0.014 (1.22E-08)	**7.91E-255**
	5	373,378	*AHRR* (Body)	cg05575921 (North shore)	0.029 (3.48E-09)	**0.00E+00**
	19	17,000,585	*F2RL3* (Body)	cg03636183 (North shore)	0.016 (5.10E-09)	**0.00E+00**
COPD	2	198,243,567	[*SF3B1*, 13 kb, 3′ *]	cg09455379 (Open sea)	−0.030 (3.28E-08)	4.16E-01
	6	10,466,788	[*GCNT2*, 27 kb, 5′ *]	cg20453862 (Open sea)	−0.039 (7.42E-09)	3.18E-01

Key: Chr: chromosome; Gene name/annotation (region): human genome build 37/Hg19, region relative to the first listed transcript; including the gene body, transcription start site (TSS) 1500 (within 1500 base pairs of a TSS) and 5′-untranslated region (5′-UTR); CpG site location: location relative to CpG island; including shore (±2 kb) and shelf (2 to 4 kb) up (North)- and down (South)-stream from a CpG island, and open sea (>4 kb) from a CpG island; β, regression coefficient; DMSs$_{es}$: associated with having ever smoked in the GS:SFHS cohort, genome-wide significant differentially methylated CpG sites are highlighted in bold.; FEV$_1$, Forced expiratory volume in 1 s; FVC, Forced vital capacity. The CpG sites lacking gene information in the IlluminaHumanMethylationEPICanno.ilm10b2.hg19 library [*] were mapped to the closest gene (in kilobases [kb] from either the 5′ or 3′ end) using the UCSC Genome Browser (http://genome.ucsc.edu). Respiratory function was assessed by spirometry in the GS:SFHS cohort.

marginal evidence of genomic inflation (max = 1.12) across traits (appendix p6 & 27). Ten of the 25 genes that contain DMSs have previously been implicated by genetic, EWAS and functional analysis in respiratory function or disease (excluding cancer; appendix p8–9). Three FEV$_1$ related-DMSs mapped to the *SOCS3* gene (Table 2); DNA methylation levels at these sites are highly correlated (appendix p28).

We next attempted to replicate these findings in 895 individuals from the LBC1936 (Table 3). No other Illumina EPIC dataset was available, but seven of the 28 DMSs identified in the discovery dataset had been profiled using the HumanMethylation450 BeadChip, which had been applied to the LBC1936. Of these seven, two FEV$_1$-associated DMSs (cg18181703 in *SOCS3* and cg18608055 in *SBNO2*) and one FEV$_1$/FVC ratio-associated DMS (cg03636183 in *F2RL3*) replicated in LBC1936 (Bonferroni-corrected $p \leq .00714$). For all seven probes, however, the direction of the effects were the same in the two datasets (Table 4).

3.3. Sensitivity analyses

As age and smoking affect both DNA methylation and lung function [13,32],and we do not know if pre-bronchodilator restrictive spirometry patterns were reversible in this study. We therefore undertook sensitivity analyses for each significant DMS for each trait. The associations between each significant DMS and its associated phenotype were there assessed in older adults (>40), smokers with a substantial smoking history and individuals with non-restrictive spirometry, and across smoking strata (ever smokers and non-smokers). Meta-analysis was used to compare regression coefficients from the discovery dataset and the sensitivity analyses. All but four of the associations identified in the discovery dataset were robust to differences in age, spirometry patterns, and smoking status and history ($p < .05/30$ DMSs; appendix p10–11). The associations between FVC and cg00213822, in *OFCC1*,

and cg13108341, in *DNAH9* were primarily driven by non-smokers. Whereas, the associations with FEV$_1$/FVC and the established smoking-associated DMSs, cg05575921, in *AHHR*, and cg03636183, in *F2RL3*, were primarily driven by smokers (Fig. 3).

We carried out sensitivity analyses in which the trait data were not pre-corrected for smoking history (appendix p12–15). Two of the identified associations were affected by pre-correction of the traits (appendix p15 & 29). In addition, an FVC-related DMS was identified at cg10919522 only when the trait data was not pre-corrected for smoking history. Pre-correction for smoking history reduced the heterogeneity of the effect size estimates of association across the age and smoking strata.

3.4. Gene ontology analysis

For each trait, to explore whether genes with DMSs share functional features, we filtered methylation probes at $p < .01$ and performed biological processes (appendix p30), molecular function (appendix p31) and canonical pathway enrichment analyses (appendix p16–23). In the study of molecular functions, we found that each of the four traits were significantly enriched for genes linked to the alternative splicing and phosphoprotein categories (appendix p31). Many of the canonical pathways identified were related to signalling, including apoptosis (appendix p16), cardiovascular signalling (appendix p17) and neurotransmission (appendix p22).

3.5. Integrative analysis of methylation and expression

To investigate the functional relevance of the methylation changes, we integrated transcriptional (COPD case-control differential expression in lung tissue) [30] and epigenetic (from this study) datasets to

Table 3
The characteristics of Lothian Birth Cohort of 1936 (LBC1936) participants ($n = 895$) in the epigenome-wide association study replication population.

	COPD cases ($n = 89$)	Controls ($n = 586$)	GOLD stage 1 ($n = 220$)
Characteristics			
Age, years	69.60 ± 0.79	69.54 ± 0.91	69.53 ± 0.88
Sex			
– Male	44 (46.4)	305 (52.0)	102 (46.4)
– Female	45 (53.6)	281 (48.0)	118 (53.6)
Height, cm	165.75 ± 9.70	166.59 ± 8.84	166.15 ± 8.98
Weight, kg	73.54 ± 14.78	77.29 ± 13.77	77.56 ± 15.44
Smoking status			
– Never	20 (22.5)	324 (55.3)	78 (35.5)
– Former (≥ 12 months)	36 (40.4)	223 (38.0)	109 (49.5)
– Former (< 12 months)	–	1 (0.2)	2 (0.9)
– Current	33 (37.1)	38 (6.5)	31 (14.1)
Pack-year			
– Former smokers (≥ 12 months)	45.91 ± 43.25	22.17 ± 23.73	32.37 ± 30.29
– Former smokers (< 12 months)	–	17.85 ± 0.00	26.63 ± 0.53
– Current smokers	45.64 ± 20.55	42.59 ± 18.94	47.15 ± 22.75
Lung function			
– FEV₁, litres/s	1.52 ± 0.51	2.62 ± 0.57	1.99 ± 0.59
– FVC, litres/s	2.59 ± 0.83	3.21 ± 0.77	2.72 ± 1.01
– FEV1/FVC	0.59 ± 0.09	0.82 ± 0.06	0.75 ± 0.09
– FEV1 percent predicted	58.34 ± 13.92	100.01 ± 11.83	76.58 ± 15.26
– FVC percent predicted	74.00 ± 15.19	90.90 ± 11.91	77.92 ± 23.10

Abbreviations: COPD, Chronic obstructive pulmonary disease; FEV₁, Forced expiratory volume in 1 s; FVC, Forced vital capacity. Figures shown are the mean ± standard deviation or n (%).

identify functional gene modules for the traits under study. This analysis identified two significant modules ($p < .05$) containing 27 and 35 genes with correlated differential methylation associated with FEV₁ and expression in COPD, respectively (appendix p24, p32–33). DMSs mapped to *SIAH2* in module 1 (appendix p32) and *SOCS3* in module 2 (appendix p33). Many of the genes in module 1 also had correlated differential methylation associated with FVC, FEV₁/FVC and COPD and expression in COPD (appendix p32). Gene enrichment analysis revealed that the top pathway for module 1 was axon guidance, while the top pathways for module 2 were cytokine-cytokine receptor interaction and JAK-STAT signalling ($p < .05$; appendix p24).

3.6. Predictive value of the DMS

To determine the predictive value of DMSs in the prognosis (forecasting future risk) of COPD, we used an independent training and test set design to predict COPD risk in GS:SFHS and LBC1936. We calculated the improvement in prediction quality of a model where genome-wide significant DMSs with all traits were added to the reduced model, which included the clinical variables: age, age [2], sex, height, height [2], smoking status and pack-years of smoking [6]. For descriptive statistics of the prediction datasets see appendix p25 & 34. Discrimination of the full model in the GS:SFHS test data was good (AUC = 0.856 [95% CI: 0.757–0.956]; appendix p35) and calibration was fair (appendix p36). Addition of DMSs to the reduced model led to a significant improvement in accuracy (ΔAUC: 0.039 [95%CI: 0.025–0.055; $p = .025$]), discrimination (IDI: 0.048[95%CI: 0.018–0.079; $p = .016$]; appendix p37) and reclassification (NRI: 0.182 [95%CI: 0.030–0.334, $p = .019$]; appendix p38). There was no improvement in prediction accuracy observed when never smokers were removed from the prediction data (ΔAUC: −0.013 [95%CI: 0.007–0.019; $p = .226$]).

We examined glmnet's variable importance measures to determine which DMSs contributed most to the increased discriminatory power. Eight DMSs: cg03770138 (*RALGDS*), cg18181703 (*SOCS3*), cg26804423 (*ICA1*), cg18871648 (*ELMSAN1*), cg11047325 (*SOCS3*), cg01620970 (*JADE1*). cg15659943 (*ABCA1*) and cg18608055 (*SBNO2*); were retained for prediction (appendix p39).

We next assessed the full model in the LBC1936 replication sample, which comprised 89 cases and 89 controls. Due to differences in array coverage, only three of the DMSs retained in the full model built in the GS:SFHS training data could be tested in LBC1936 (cg18181703, cg26804423 and cg18608055). Addition of the three sites to the reduced model led to a significant improvement in accuracy (ΔAUC: 0.029(95%CI: 0.024–0.032; $p = .006$; appendix p35) and discriminative power (IDI: 0.019[95%CI: 0.005–0.033; $p = .010$]), but not reclassification (NRI: 0.045[−0.023–0.113, $p = .196$]).

Decision curve analysis showed that the model incorporating the DMSs had good clinical applicability and was superior to the reduced model over a wide range of threshold probabilities in the discovery and replication data (appendix p 40 & 41).

4. Discussion

We performed EWASs of three respiratory function traits and COPD in DNA extracted from peripheral blood using the high-density Illumina EPIC array, in 3781 individuals from a single cohort. These analyses identified 28 DMSs (27 novel, of which 14 were not associated with smoking status), of which 26 are associated with respiratory function and three with COPD in the discovery GS:SFHS data. Data were available to test seven of the 28 DMSs for replication in an independent dataset; three associations replicated. Incorporation of a subset of the identified DMSs into a model composed of established clinical variables improved discrimination of individuals at-risk of COPD in two independent samples. Finally, functional annotation provided insights into the biology of these phenotypes.

Smoking is a major risk for impaired respiratory function and COPD [28], and has been shown to impact DNA methylation [3,8]. We for the first time, adjusted both the phenotypic and DNA methylation data for
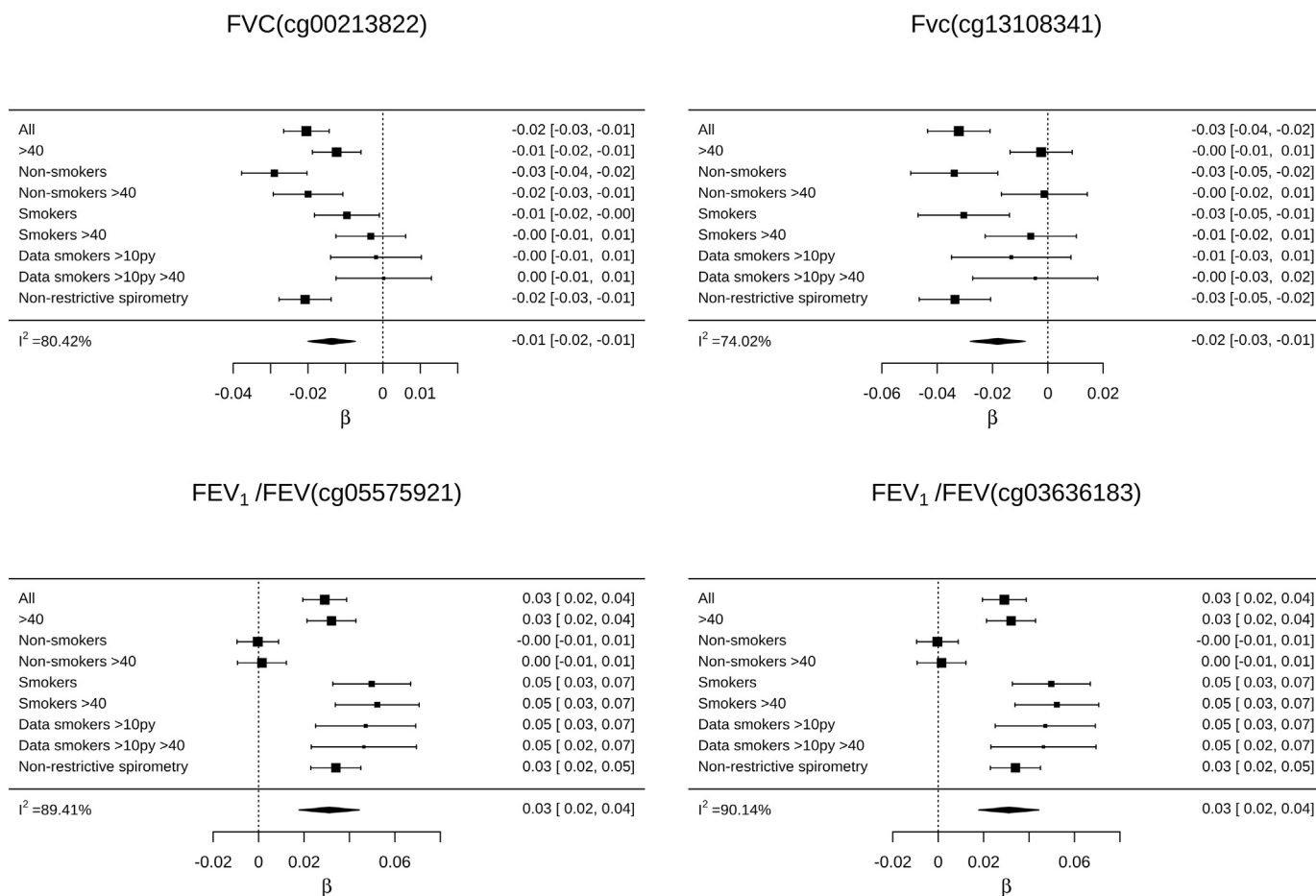
Table 4
Replication of the genome-wide significant differentially methylated sites (DMSs) associated with FEV₁ and FEV₁/FVC from the discovery Generation Scotland: Scottish Family Health Survey (GS:SFHS) cohort in the Lothian Birth Cohort of 1936 (LBC1936). Results are ordered by trait and chromosomal location.

Trait	Chr	Base pair	Gene name/annotation (region)	CpG site (location)	β (p-value)
FEV₁	7	8,201,134	*ICA1* (Body)	cg26804423 (Open sea)	−0.017 (6.32E-02)
	14	74,227,441	*C14orf43* (5'UTR)	cg10919522 (South shore)	0.011 (2.95E-01)
	16	57,180,107	*CPNE2* (Body)	cg09018739 (Open sea)	−0.015 (3.04E-02)
	17	76,354,621	*SOCS3* (Body)	cg18181703 (North shore)	0.034 (**1.04E-04**)
	19	1,130,866	*SBNO2* (Body)	cg18608055 (Open sea)	0.018 (**6.54E-03**)
FEV₁/FVC	5	373,378	*AHRR* (Body)	cg05575921 (North shore)	0.041 (4.83E-02)
	19	17,000,585	*F2RL3* (Body)	cg03636183 (North shore)	0.028 (**8.75E-03**)

Key: Chr: chromosome; Gene name/annotation (region): human genome build 37/Hg19, region relative to the first listed transcript; including the gene body, transcription start site (TSS) 1500 (within 1500 base pairs of a TSS) and 5′-untranslated region (5′-UTR); CpG site location: location relative to CpG island; including shore (±2 kb) and shelf (2 to 4 kb) up (North)- and down (South)-stream from a CpG island, and open sea (>4kb) from a CpG island; β, regression coefficient; FEV₁, Forced expiratory volume in 1 s; FVC, Forced vital capacity.

Fig. 3. Forest plot and meta-analysis across the discovery, older adult (> 40 years), smokers with a substantial smoking history, individuals with non-restrictive spirometry pattern, and stratified smoking status datasets from the Generation Scotland: Scottish Family Health Survey (GS:SFHS) cohort for differentially methylated sites associated with FVC and FEV_1/FVC that showed high heterogeneity in older adults and across the smoking strata. The sizes of the forest plot squares are proportional to the amount of information each dataset contains. Key: All, discovery data; >40, data from participants aged 40 or greater; >10py, data from participants with a smoking history of 10 pack years or greater.

smoking history. This approach appeared to reduce the confounding effects of smoking, identify more associations, and reduce the heterogeneity of effect estimates across smoking strata. It identified 14 DMS that did not associate with smoking status.

Ten of the 25 genes, harbouring the novel DMSs, have been previously linked to respiratory function or disease (appendix p8–9). In four cases, these links come from studies in lung tissue: DNA methylation changes in *ABCA1* in lung tissue has been reported to be associated with pulmonary arterial hypertension; differential expression of *ABCA1* and *DNAH9* has been reported in lung tissue of patients with COPD and primary ciliary dyskinesia respectively and pathological changes in lung tissue have been reported following knockdown and knockout of *SLAMF7*, *ABCA1*, and *SOCS3*.

Three DMSs showed replication. The first, cg18181703, is one of three FEV_1-associated DMS in *SOCS3*, which has been associated with infection and autoimmunity [33], modulates the lung inflammatory response [34], and JAK-STAT signal transduction [35]. Transcriptional down-regulation of *SOCS3* has been observed in COPD [36] and asthma [37] patients. Differentially methylated sites in *SOCS3* within a FEV_1-related gene module in this study were correlated with differential gene expression in lung tissue of COPD patients [30]. DMSs in *SOCS3* were the second and fifth most predictive DMSs in the prognostic model for COPD in the GS:SFHS cohort. Inclusion of this DMS in the prediction model also improved the prediction of prevalent COPD risk in the LBC1936. However, this was one of the three DMS that could be tested in the LBC1936.

The second replicated DMS, cg18608055, is a FEV_1-associated DMS in *SBNO2*, which encodes a transcriptional co-regulator of the pro-

inflammatory cascade [38]. Inclusion of this DMS in the prediction model improved the prediction of COPD risk in the GS:SFHS cohort and LBC1936. The third, cg03636183, is a FEV_1/FVC ratio and smoking-associated DMS in *F2RL3* [39,40], which has been previously reported to be associated with respiratory function [3].

As discussed above, *SOCS3* and *SBNO2* were found to provide discriminatory power in the prediction analysis. The inclusion of three other DMSs in *RALGDS, ICA1, ELMSAN1, JADE1* and *ABCA1* improved the prediction of incident COPD risk in the GS:SFHS. *JADE1* is a negative regulator of Wnt signalling, which has been linked to the pathogenesis and progression of COPD [41]. *RALGDS* is a Ras effector and regulates cellular processes such as vesicular trafficking, endocytosis, and migration [42].

Functional annotation of differently methylated genes identified enrichment of the molecular function alternative splicing (Fig. 4). This finding is consistent with earlier reports that genes associated with COPD (unlike those associated with other complex traits examined) have greater transcriptional complexity due to a disproportionately high level of alternative splicing [43]. In addition, many such genes are spliced differently in COPD patients and controls [43].

Functional analysis also identified two cellular pathways (Fig. 4). JAK-STAT signalling (appendix p42) was highlighted in the EWAS data at both the DMS level and in the gene ontology analysis. A module comprising genes (including SOCS3) that showed both differential methylation with FEV_1 in this study, and COPD-based gene expression in lung tissue [30] was enriched for JAK-STAT signalling genes. Thus our data add further support to previous evidence [44] for the importance of this pathway in respiratory function and COPD. Modules comprising
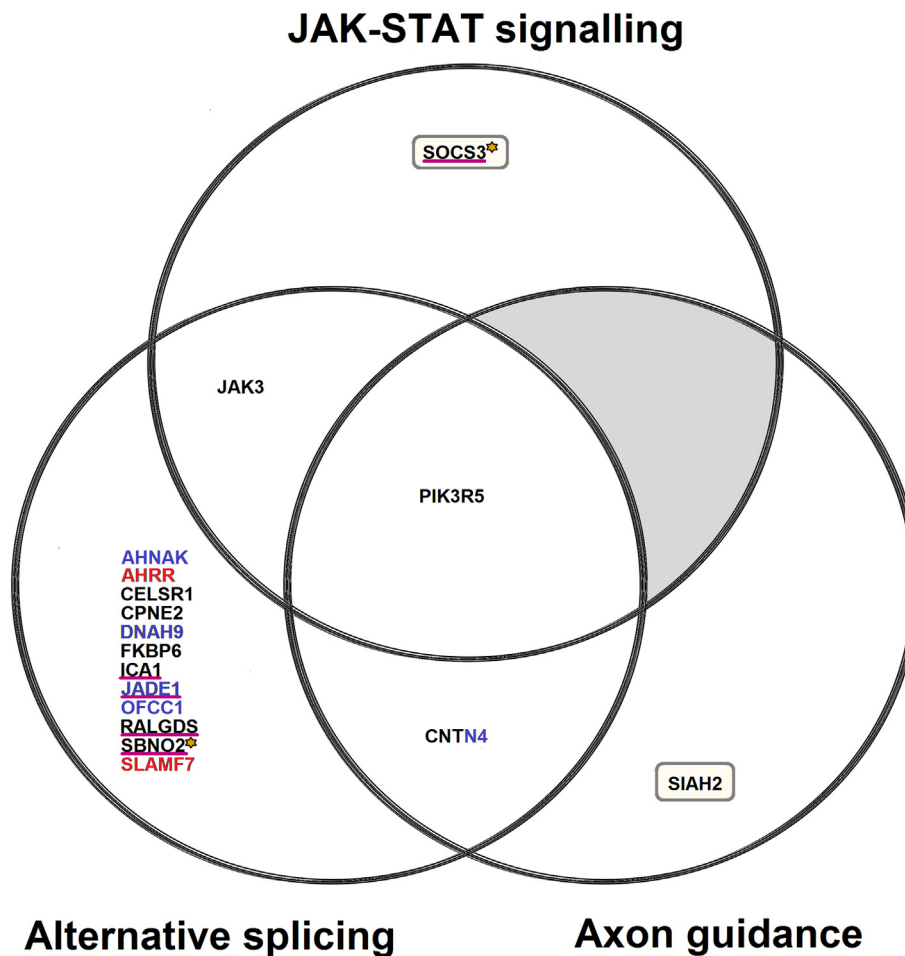
Fig. 4. Venn diagram showing additional sources of evidence for the functional annotation categories showing enrichment in the epigenome-wide association study (EWAS) data: JAK-STAT signalling, axon guidance and alternative splicing. Key: the black, blue and red text represent genes enriched for FEV_1, FVC and FEV_1/FVC ratio-related differentially methylated positions (DMPs) respectively. A single FEV_1 and FVC-related DMP mapped to the CNTN4 gene. The genes with the gold asterisk, represent the FEV_1-related DMPs that replicated in the Lothian Birth Cohort of 1936 (LBC1936). The genes underlined with magenta represent the DMPs that improved the prediction of incident COPD risk in an independent Generation Scotland Scottish Family Health Study (GS:SFHS) dataset when included in the full model. The inclusion of a DMPs in SOCS3 and *SBNO2* in the full model additionally improved the prediction of COPD risk in the LBC1936 replication data. The genes in yellow rectangles represent those with correlated FEV_1-related differential methylation in whole blood with the altered gene expression in COPD in lung tissue. Gene lists were extracted from Database for Annotation, Visualization and Integrated Discovery (DAVID; http://david. abcc.ncifcrf.gov/tools.jsp) Bioinformatics resources database and Ingenuity Pathway Analysis (IPA; http://www.ingenuity.com) functional enrichment and *Significance-based Modules Integrating the Transcriptome and Epigenome (SMITE)* result tables.

genes (including *SIAH2*) that showed both differential methylation with all three respiratory function traits and COPD in this study, and COPD-based gene expression in lung tissue [30] were enriched for axon guidance signalling genes. *SIAH2* upregulation mediates the ubiquitination of *NRF2* which has been previously associated with respiratory function [45] and COPD [46]. Axon guidance signalling (appendix p43) was highlighted by two of our analyses, adding support to the neuropathology hypothesis of COPD [47].

To investigate the potential clinical implications of our findings we assessed the predictive properties of DMSs in the prognosis of COPD. The inclusion of DMSs provided added clinical value to established clinical variables in both the discovery and replication datasets. Clinical studies are needed to provide formal proof that changes in DNA methylation at these sites contribute causally to the pathogenesis, and can impact prognosis of COPD.

There are three main limitations to this study. Firstly, DNA methylation was quantified in peripheral blood. There is ongoing debate about whether DNA methylation from peripheral blood can serve as a surrogate marker for DNA methylation in lung tissue [48]. The overlap between our findings and previous studies performed in lung tissue (appendix p8–9) suggest that, for at least some loci, the study of DNA

methylation in blood may yield mechanistic insights. Moreover, our data demonstrate that DMSs from peripheral blood have both predictive and clinical value. As such, blood may be an appropriate tissue for the development of biomarkers, as it is easily and repeatedly accessible.

The second limitation is that DNA methylation was measured in blood samples collected at the same time that spirometry tests were performed. As such, our reported associations are subject to reverse causality. However, the integrated alterations in DNA methylation in this study and gene expression profiles in COPD [30] and prospective predictive value of the selected DMSs provide indications that the DNA methylation alterations observed in blood may play a causal role in respiratory function. Nevertheless, longitudinal studies, with serial measurements of DNA methylation will be required to address causality formally.

The third limitation is that only seven of the 29 could be tested for replication due the Illumina HumanMethylation450 BeadChip array been used to profile participants in the LBC1936.

Another limitation of this study is that COPD cases in the EWAS were classified based on pre-bronchodilator spirometry data. It is therefore not possible to determine if their airflow limitation was reversible, and so a proportion of these cases may have been suffering from other

respiratory diseases, such as asthma. Nevertheless, sensitivity analysis demonstrated that restrictive spirometry had little impact on our results. Also, GOLD COPD diagnosis is based on post-bronchodilator spirometry. Hence, COPD defined cases in this study might not have met the GOLD stage 2–4 criteria if given bronchodilators. Nonetheless, pre-bronchodilator spirometry classification of COPD has been used previously [22,27], and has been shown to lead to minimal misclassification of moderate to severe (GOLD stage ≥2) COPD [27].

A further limitation is that incident COPD was defined based on ICD-10 COPD exacerbation codes. We were not able to remove never smokers in prediction analyses, as the small sample size leads to overfitting of the training data and no improvement in accuracy. We were therefore not able to rule out confounding by other respiratory conditions in never smokers.

In conclusion, using a large dataset and a robust methodological approach, we have identified DMSs associated with respiratory function and COPD, provided new mechanistic insights and supported previous hypotheses into impaired respiratory function and the pathogenesis of COPD. We also demonstrated that DMSs can be incorporated into existing models for predicting COPD risk, yielding better prediction than established clinical variables alone.

Author contributions

Declarations of interests

Dr. McIntosh reports grants from The Sackler Trust and Dr. Deary reports grants from Wellcome and Medical Research Council outside the submitted work. The remaining authors have nothing to disclose.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ebiom.2019.03.072.

References

[1] Palmer LJ, Knuiman MW, Divitini ML, et al. Familial aggregation and heritability of adult lung function: results from the Busselton Health Study. Eur Respir J 2001;17: 696–702.

[2] Klimentidis YC, Vazquez AI, de Los Campos G, Allison DB, Dransfield MT, Thannickal VJ. Heritability of pulmonary function estimated from pedigree and whole-genome markers. Front Genet 2013;4:174.

[3] Qiu W, Baccarelli A, Carey VJ, et al. Variable DNA methylation is associated with chronic obstructive pulmonary disease and lung function. Am J Respir Crit Care Med 2012;185:373–81.

[4] Vucic EA, Chari R, Thu KL, et al. DNA methylation is globally disrupted and associated with expression changes in chronic obstructive pulmonary disease small airways. Am J Respir Cell Mol Biol 2014;50:912–22.

[5] Morrow JD, Cho MH, Hersh CP, et al. DNA methylation profiling in human lung tissue identifies genes associated with COPD. Epigenetics 2016;11:1–10.

[6] Lee MK, Hong Y, Kim S-Y, Kim WJ, London SJ. Epigenome-wide association study of chronic obstructive pulmonary disease and lung function in Koreans. Epigenomics 2017;9:971–84.

[7] Marioni RE, Shah S, McRae AF, et al. The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. Int J Epidemiol 2015;44: 1388–96.

[8] Sundar IK, Yin Q, Baier BS, et al. DNA methylation profiling in peripheral lung tissues of smokers and patients with COPD. Clin Epigenetics 2017;9:38.

[9] de Vries M, van der Plaat DA, Vonk JM, Boezen HM. No association between DNA methylation and COPD in never and current smokers. BMJ Open Respir Res 2018; 5:e000282.

[10] Machin M, Amaral AFS, Wielscher M, et al. Systematic review of lung function and COPD with peripheral blood DNA methylation in population based studies. BMC Pulm Med 2017;17:54.

[11] Bell JT, Tsai P-C, Yang T-P, et al. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. PLoS Genet 2012;8:e1002629.

[12] Marioni RE, Shah S, McRae AF, et al. DNA methylation age of blood predicts all-cause mortality in later life. Genome Biol 2015;16:25.

[13] Allinson JP, Hardy R, Donaldson GC, Shaheen SO, Kuh D, Wedzicha JA. Combined impact of smoking and early-life exposures on adult lung function trajectories. Am J Respir Crit Care Med 2017;196:1021–30.

[14] Busch R, Qiu W, Lasky-Su J, Morrow J, Criner G, DeMeo D. Differential DNA methylation marks and gene comethylation of COPD in African-Americans with COPD exacerbations. Respir Res 2016;17:143.

[15] Clifford RL, Fishbane N, Patel J, et al. Altered DNA methylation is associated with aberrant gene expression in parenchymal but not airway fibroblasts isolated from individuals with COPD. Clin Epigenetics 2018;10:32.

[16] Salvi SS, Barnes PJ. Chronic obstructive pulmonary disease in non-smokers. Lancet 2009;374:733–43.

[17] Gan WQ, FitzGerald JM, Carlsten C, Sadatsafavi M, Brauer M. Associations of ambient air pollution with chronic obstructive pulmonary disease hospitalization and mortality. Am J Respir Crit Care Med 2013;187:721–7.

[18] Gershon A, Campitelli MA, Hwee J, et al. Socioeconomic status, sex, age and access to medications for COPD in Ontario, Canada. COPD 2015;12:668–79.

[19] Hayden LP, Cho MH, Raby BA, et al. Childhood asthma is associated with COPD and known asthma variants in COPDGene: a genome-wide association study. Respir Res 2018;19:209.

[20] Shrine N, Guyatt AL, Erzurumluoglu AM, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. Nat Genet 2019;51:481–93.

[21] Wyss AB, Sofer T, Lee MK, et al. Multiethnic meta-analysis identifies ancestry-specific and cross-ancestry loci for pulmonary function. Nat Commun 2018;9: 2976.

[22] Wain LV, Shrine N, Artigas MS, et al. Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. Nat Genet 2017;49:416–25.

[23] Wain LV, Shrine N, Miller S, et al. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. Lancet Respir Med 2015; 3:769–81.

[24] Smith BH, Campbell A, Linksted P, et al. Cohort profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. Int J Epidemiol 2013;42: 689–700.

[25] Deary IJ, Gow AJ, Pattie A, Starr JM. Cohort profile: the Lothian Birth Cohorts of 1921 and 1936. Int J Epidemiol 2012;41:1576–84.

[26] McCartney DL, Hillary RF, Stevenson AJ, et al. Epigenetic prediction of complex traits and death. Genome Biol 2018;19:136.

[27] Soler Artigas M, Wain LV, Repapi E, et al. Effect of five genetic variants associated with lung function on the risk of chronic obstructive lung disease, and their joint effects on lung function. Am J Respir Crit Care Med 2011;184:786–95.

[28] GOLD. Pocket Guide to COPD Diagnosis, Management, and Prevention – 2017. goldcopd.org; 2017.

[29] Saffari A, Silver MJ, Zavattari P, et al. Estimation of a significance threshold for epigenome-wide association studies. Genet Epidemiol 2018;42:20–33.

[30] Brandsma C-A, van den Berge M, Postma DS, et al. A large lung gene expression study identifying fibulin-5 as a novel player in tissue repair in COPD. Thorax 2015; 70:21–32.

[31] Poder TG, Carrier N, Bélanger M, et al. Eosinophil counts in first COPD hospitaliza-
tions: a 1-year cost analysis in Quebec, Canada. Int J Chron Obstruct Pulmon Dis
2018;13:3065–76.

[32] Guida F, Sandanger TM, Castagné R, et al. Dynamics of smoking-induced genome-
wide methylation changes with time since smoking cessation. Hum Mol Genet
2015;24:2349–59.

[33] Yin Y, Liu W, Dai Y. SOCS3 and its role in associated diseases. Hum Immunol 2015;
76:775–80.

[34] Ding F-M, Liao R-M, Chen Y-Q, et al. Upregulation of SOCS3 in lung CD4+ T cells in a
mouse model of chronic PA lung infection and suppression of Th17-mediated neu-
trophil recruitment in exogenous SOCS3 transfer in vitro. Mol Med Rep 2017;16:
778–86.

[35] Murray PJ. The JAK-STAT signaling pathway: input and output integration. J
Immunol 2007;178:2623–9.

[36] Dong R, Xie L, Zhao K, Zhang Q, Zhou M, He P. Cigarette smoke-induced lung inflam-
mation in COPD mediated via LTB4/BLT1/SOCS1 pathway. Int J Chron Obstruct
Pulmon Dis 2016;11:31–41.

[37] Sun W, Xiao B, Jia A, et al. MBD2-mediated Th17 differentiation in severe asthma is
associated with impaired SOCS3 expression. Exp Cell Res 2018;371:196–204.

[38] El Kasmi KC, Smith AM, Williams L, et al. Cutting edge: a transcriptional repressor
and corepressor induced by the STAT3-regulated anti-inflammatory signaling path-
way. J Immunol 2007;179:7215–9.

[39] Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related dif-
ferential DNA methylation: 27K discovery and replication. Am J Hum Genet 2011;
88:450–7.

[40] Zhang Y, Yang R, Burwinkel B, Breitling LP, Brenner H. F2RL3 methylation as a bio-
marker of current and lifetime smoking exposures. Environ Health Perspect 2014;
122:131–7.

[41] Baarsma HA, Königshoff M. WNT-er is coming ': WNT signalling in chronic lung dis-
eases. Thorax 2017;72:746–59.

[42] Ferro E, Trabalzini L. RalGDS family members couple Ras to Ral signalling and that's
not all. Cell Signal 2010;22:1804–10.

[43] Lackey L, McArthur E, Laederach A. Increased transcript complexity in genes associ-
ated with chronic obstructive pulmonary disease. PLoS One 2015;10:e0140885.

[44] Huan W, Tianzhu Z, Yu L, Shumin W. Effects of ergosterol on COPD in mice via JAK3/
STAT3/NF-κB pathway. Inflammation 2017;40:884–93.

[45] Ban WH, Kang HH, Kim IK, et al. Clinical significance of nuclear factor erythroid 2-
related factor 2 in patients with chronic obstructive pulmonary disease. Korean J In-
tern Med 2018;33:745–52.

[46] Cui W, Zhang Z, Zhang P, et al. Nrf2 attenuates inflammatory response in COPD/em-
physema: crosstalk with Wnt3a/β-catenin and AMPK pathways. J Cell Mol Med
2018;22:3514–25.

[47] Oancea C, Tudorache E, Tudorache V. Neurocognitive impairment as systemic effects
of COPD. In: McCarthy C, editor. COPD - an update in pathogenesis and clinical man-
agement. InTech; 2018. https://doi.org/10.5772/intechopen.70580.

[48] Stueve TR, Li W-Q, Shi J, et al. Epigenome-wide analysis of DNA methylation in lung
tissue shows concordance with blood studies and identifies tobacco smoke-
inducible enhancers. Hum Mol Genet 2017;26:3014–27.