



Epigenome-wide association study of chronic obstructive pulmonary disease and lung function in Koreans

Aim: To identify differentially methylated probes (DMPs) and regions (DMRs) in relation to chronic obstructive pulmonary disease (COPD) and lung function traits. **Methods:** We performed an epigenome-wide association study of COPD and spirometric parameters, including forced expiratory volume in 1 s (FEV1), forced vital capacity (FVC) and FEV1/FVC, in blood DNA using the Infinium HumanMethylation450 (n = 100, a Korean COPD cohort). **Results:** We found one significant DMP (cg03559389, *DIP2C*) and 104 significant DMRs after multiple-testing correction. Of these, 34 DMRs mapped to genes differentially expressed with respect to the same trait. Five of the genes were associated with more than two traits: *CTU2*, *USP36*, *ZNF516*, *KLK10* and *CPT1B*. **Conclusion:** We identified novel differential methylation loci related to COPD and lung function in blood DNA in Koreans and confirmed previous findings in non-Asians. Epigenetic modification could contribute to the etiology of these phenotypes.

First draft submitted: 3 January 2017; Accepted for publication: 3 May 2017; Published online: 16 June 2017

Keywords: chronic obstructive pulmonary disease • DNA methylation • epigenome-wide association study • gene expression • lung function

Reduced lung function is not only an important health indicator in respiratory illnesses, but also predicts mortality in the population as a whole [1]. Lung function traits, including forced expiratory volume in 1 s (FEV1), forced vital capacity (FVC) and FEV1/FVC, are useful in the diagnosis or monitoring of lung diseases such as chronic obstructive pulmonary disease (COPD). COPD is one of the leading causes of mortality and morbidity worldwide [2], is a risk factor for lung cancer [3] and impairs health-related quality of life [4].

Numerous genetic loci have been associated with COPD and/or lung function traits; however, a limited proportion of variance of the phenotypes can be explained by these loci [5,6]. This opens the possibility that epigenetic modifications may play a role [7]. DNA methylation is one of the epigenetic mechanisms that can regulate gene expression.

Five epigenome-wide association studies (EWASs) of COPD and/or lung function traits have been published [8–12]. Of these, two [9,12] measured methylation in blood DNA and only one [11] used the Illumina HumanMethylation450 (450k) array. There are few data on COPD or lung function reporting differentially methylated regions (DMRs), involving Asian populations, or connecting methylation alterations in blood to transcriptome data in lung tissue.

The earlier EWASs [8–12] have focused on identification of differentially methylated probes (DMPs) in relation to COPD or lung function metrics. In other words, these studies evaluated associations with methylation levels at individual CpG sites (CpGs). Analysis of DMRs can provide stronger evidence of causality than analysis of individual DMPs [13] and greater power to detect associations with phenotypic traits [14].

Mi Kyeong Lee^{1,2}, Yoonki Hong², Sun-Young Kim³, Woo Jin Kim^{*,2} & Stephanie J London^{*,†,1}

¹Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709, USA

²Department of Internal Medicine & Environmental Health Center, Kangwon National University Hospital, School of Medicine, Kangwon National University, Chuncheon-si, Gangwon-do 19300, South Korea

³Institute of Health & Environment, Seoul National University, Seoul 08826, South Korea

*Author for correspondence: london2@niehs.nih.gov

†Authors contributed equally

We conducted an EWAS of COPD and pulmonary traits in 100 individuals from a Korean COPD cohort using the 450k. Both individual CpGs and regions of differential methylation were examined. For genes to which differential methylation signals are annotated, we checked whether the related trait was associated with gene expression in lung tissue from a separate cohort ($n = 188$, Asan Biobank). In addition, previously identified probes in non-Asians were checked for differential methylation and gene expression in relation to COPD and lung function.

Materials & methods

Study participants & lung function: the Korean COPD cohort

For methylation profiling, study participants ($n = 100$) were sampled from a Korean COPD cohort [15]. Participants in the cohort were recruited from a rural area in Korea. Among 190 participants (135 COPD cases and 55 noncases) enrolled in 2012, 100 (60 COPD cases and 40 noncases) were chosen based on availability of clinical information, computed tomography (CT) data, blood/urine samples and survey questionnaire. We additionally applied approximate frequency matching on age in 10 years categories and smoking status. Survey questionnaires, spirometry measurements and blood/urine samples obtained from each participant were used in further analyses. One trained nurse measured height (cm) and weight (kg) twice for each participant with the body composition analyzer IOI 353 (Aarna Systems., Udaipur, India), and the average of two measurements was used in analyses. We calculated BMI by dividing the weight (kg) by the square of the height (m^2). Smoking status (current, former and never smoking) was self-reported in the questionnaire. Urine cotinine levels (nmol/l) measured by immunoassay (Immulite 2000 Xpi; Siemens Healthcare Diagnostics, NY, USA) were used to confirm the current nonsmoking status. For current and former smokers, pack-years of smoking were calculated by multiplying the number of years smoked by the number of cigarette packs smoked per day.

In this study, we analyzed COPD status and spirometric parameters: FEV1, FVC and FEV1/FVC. The FEV1 and FVC were measured in liters by using the EasyOne spirometer (ndd Medical Technologies, Inc., MA, USA). To assess post-bronchodilator FEV1, spirometry is performed 15 min after inhalation of 400 mcg of salbutamol, through a metered-dose inhaler (MDI) with a spacer. We defined COPD cases based on post-bronchodilator FEV1/FVC < 0.7 [16]. We used questionnaires for COPD assessment test scores and the Modified Medical Research Council dyspnea grade. Emphysema index and wall area (%) were generated based on CT data [17].

DNA methylation profiling

Blood DNA samples from participants' baseline visits were used for methylation profiling. Bisulfite conversion was done by using the EZ DNA methylation kit (Zymo Research, CA, USA). Genome-wide methylation profiles were obtained by using the Infinium HumanMethylation450 BeadChip (Illumina, Inc., CA, USA). The signal extraction and Beta Mixture Quantile dilation [18] normalization to correct for probe design bias was done and followed by ComBat [19] for batch effect corrections. We only included CpGs in autosomal chromosomes for our association analyses. The 450k array provides single nucleotide resolutions of methylation status on 473,864 probes across 22 autosomes. Of these, as a quality control (QC) procedure, we excluded probes that were of low quality (having a detection $p < 0.01$ in any sample or having a bead-count < 3 in 5% or more of samples), non-CpG, non-specific [20], or potentially influenced by nearby genetic variations such as single nucleotide polymorphisms (SNPs) [20]. From the signal extraction to QC procedures, we used the complete pipeline in ChAMP R package [21]. Additionally, we removed 31,831 non-specific probes [22] not included in prior filtering steps. The remaining 402,508 CpGs were included in association analyses. Probe filtering steps are summarized in **Supplementary Table 1**. To reduce potential influences of extreme values on association results, we trimmed methylation values at the outer fences [23], meaning outside of three-times the interquartile range from the first and third quartiles of each probe. This procedure removed 75,549 (0.19%) β values across all participants. The methylation value (β), the proportion of methylation at a given CpG site, was used for statistical analyses. The value ranges from 0 (unmethylated) to 1 (methylated). Cell-type proportions were estimated by applying Houseman algorithm [24] using the Reinius reference panel [25].

Epigenome-wide association study

To identify DMPs in relation to COPD, a logistic regression model was used with the response variable of COPD status and the predictor variable of methylation values. To evaluate the association between methylation values and pulmonary traits (FEV1, FVC and FEV1/FVC), we used a robust linear regression model with each trait as the response variable and methylation values as the predictor variable. Covariates in the statistical models for FEV1, FEV1/FVC and COPD were age, sex, height, smoking status, pack-years and the estimated cell-type proportions. Weight was also included in models for FVC. For statistical significance, we set a threshold of $p < 1.2 \times 10^{-7}$ after Bonferroni correction ($0.05/402,508 = 1.2 \times 10^{-7}$) for signifi-

Table 1. Characteristics of the study populations in our epigenome-wide association and transcriptome analysis.				
Characteristics (mean ± standard deviation or n [%])	Epigenome-wide association analysis (n = 100, the Korean COPD [§] cohort)		Transcriptome analysis (n = 188, Asan Biobank)	
	COPD cases (n = 60)	Controls (n = 40)	COPD cases (n = 97)	Controls (n = 91)
Age, years	73.5 ± 5.8	71.8 ± 6.9	67.4 ± 6.4	60.8 ± 9.5
Gender:				
– Male	42 (70.0)	24 (60.0)	97 (100.0)	91 (100.0)
– Female	18 (30.0)	16 (40.0)	0	0
Height, cm	157.1 ± 10.1	158.0 ± 9.0	NA ^{§§}	NA
Weight, kg	55.1 ± 9.2	59.6 ± 8.8	NA	NA
BMI, kg/m ²	22.3 ± 2.8	23.8 ± 2.8	NA	NA
Smoking status:				
– Never	19 (31.7)	20 (50.0)	0	0
– Former	20 (33.3)	10 (25.5)	97 (100.0)	91 (100.0)
– Current	21 (35.0)	10 (25.5)		
Pack-year:				
– Former smokers	24.4 ± 13.7	37.0 ± 26.1	48.3 ± 21.6	35.2 ± 17.2
– Current smokers	37.8 ± 20.8	31.3 ± 15.2		
Lung function				
Pre-bronchodilator:				
– FEV1 [†] , l	1.52 ± 0.47	2.05 ± 0.53	2.16 ± 0.47	2.99 ± 0.59
– FVC [‡] , l	2.51 ± 0.74	2.76 ± 0.66	3.81 ± 0.66	4.00 ± 0.73
– FEV1/FVC	0.61 ± 0.08	0.74 ± 0.05	0.57 ± 0.08	0.75 ± 0.04
– FEV1 percent predicted	0.71 ± 0.19	0.94 ± 0.22	0.72 ± 0.14	0.91 ± 0.12
– FVC percent predicted	0.85 ± 0.18	0.93 ± 0.20	0.90 ± 0.13	0.90 ± 0.13
– FEV1 percent predicted /FVC percent predicted	0.83 ± 0.11	1.01 ± 0.06	0.80 ± 0.12	1.01 ± 0.08
Post-bronchodilator:				
– FEV1 [†] , l	1.61 ± 0.48	2.1 ± 0.54	2.32 ± 0.46	3.10 ± 0.46
– FVC [‡] , l	2.76 ± 0.73	2.8 ± 0.72	3.90 ± 0.60	4.03 ± 0.60
– FEV1/FVC	57.85 ± 6.47	75.26 ± 3.84	0.59 ± 0.07	0.77 ± 0.07
– FEV1 percent predicted	0.75 ± 0.18	0.97 ± 0.22	0.76 ± 0.13	0.94 ± 0.13
– FVC percent predicted	0.94 ± 0.17	0.94 ± 0.20	0.91 ± 0.12	0.91 ± 0.12
– FEV1 percent predicted /FVC percent predicted	0.79 ± 0.09	1.03 ± 0.06	0.84 ± 0.11	1.04 ± 0.11
MMRC scale [¶] (0/1/2/3/4)	13/ 23 / 6 / 12/ 5	NA	NA	NA
CAT [#]	17.95 ± 9.42	NA	NA	NA
CT measurement:				
– Emphysema index ^{††}	9.27 ± 8.02	3.03 ± 3.77	NA	NA
– Wall area (%) ^{†††}	70.45 ± 4.26	69.60 ± 4.88	NA	NA

[†]Forced expiratory volume in one second.
[‡]Forced vital capacity.
[§]Chronic obstructive pulmonary disease.
[¶]Modified Medical Research Council dyspnea scale.
[#]COPD assessment test.
^{††}Emphysema index (lung volume fraction (%) of the lung below -950 HU).
^{†††}Wall area percent (wall area/[wall area + lumen area] × 100).
^{§§}Not available.
CAT: COPD assessment test; COPD: Chronic obstructive pulmonary disease; FEV1: Forced expiratory volume in 1 second; FVC: Forced vital capacity; MMRC: Modified Medical Research Council.

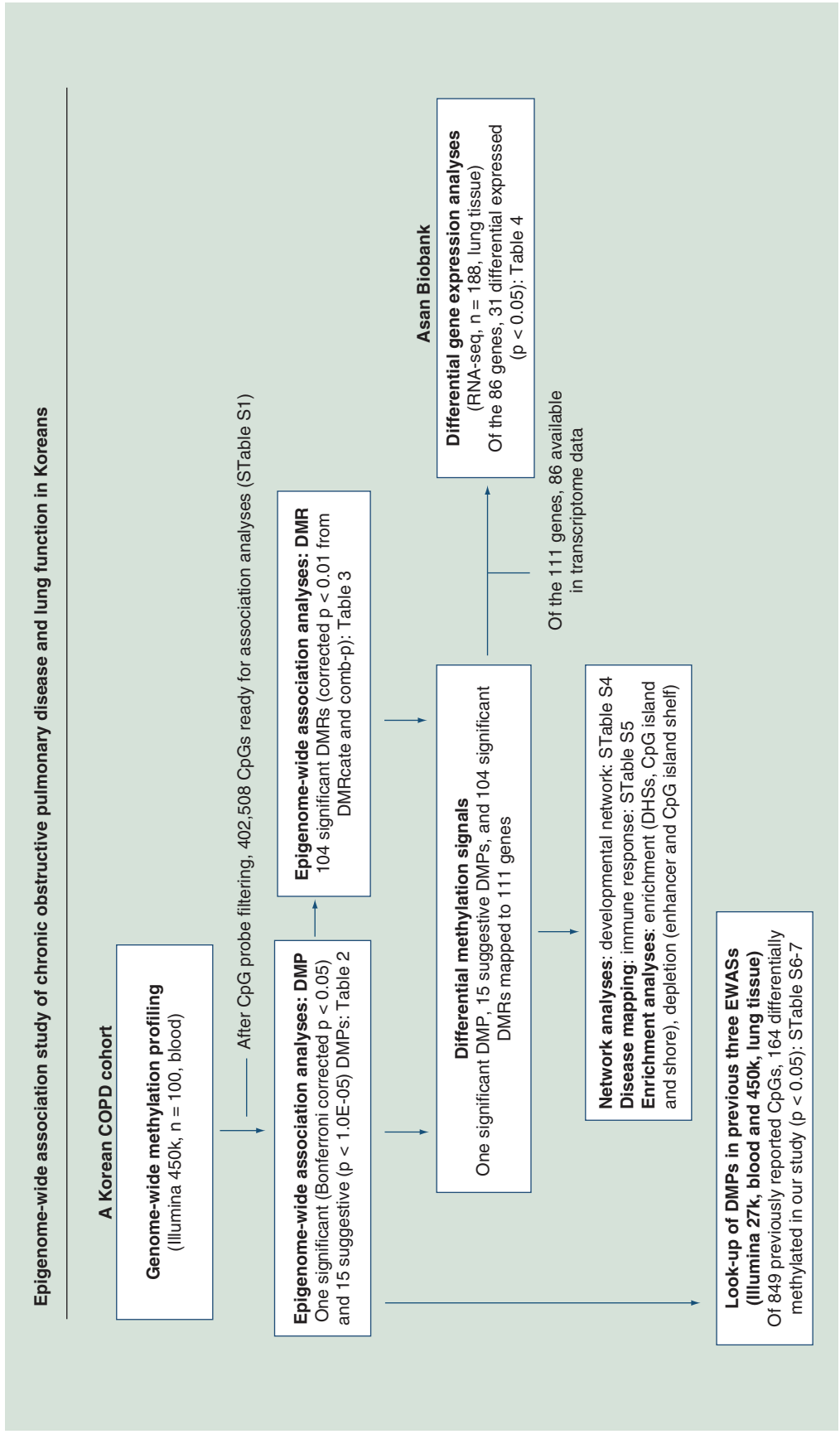


Figure 1. Workflow of our study.

cant DMPs and $p < 1.0 \times 10^{-5}$, an arbitrary threshold, for suggestive DMPs.

To detect regional differential methylation signals, we used two DMR methods: DMRcate [26] and comb-p [27]. These methods identify DMRs using different algorithms. DMRcate identifies DMRs from tunable kernel smoothing process of association signals, whereas comb-p identifies DMRs by regional clustering of low p-values from irregularly spaced p-values. The dmrcate function in the DMRcate R package was used. Input files for both DMR analyses were the epigenome-wide association results: regression coefficients, standard deviations, uncorrected p-values for DMRcate and uncorrected p-values and chromosomal locations for comb-p.

Significant DMRs were defined based on three criteria. First, a DMR should contain more than one probe. Second, regional information can be combined from probes within 1000 basepairs (bp). Third, the region showed multiple-testing corrected $p < 0.01$ in both methods (false-discovery rate [FDR] [28] for DMRcate and Sidak [29] for comb-p). The minimum number of CpGs in a region and the minimum length of a distance were the defaults in DMRcate [30], so the same values were used for comb-p to harmonize results from two approaches. Parameters used for the DMR calling can be found in [Supplementary Table 2](#). A significant DMR can be detected even if there is no genome-wide significant DMP in the region.

Enrichment & functional network analysis & visualization

We performed an enrichment analysis to evaluate over- or under-representation of genomic features in our differential methylation signals compared with all probes in the 450k array. The differential methylation signals contain significant (Bonferroni corrected $p < 0.05$) and suggestive (uncorrected $p < 0.05$) DMPs and nominally significant CpGs (uncorrected $p < 0.05$) within significant DMRs. To assess enrichment or depletion for genomic features, we used the hypergeometric test (two-sided doubling mid-p).

For biological insights of loci from our EWAS, we conducted a functional network analysis and disease mapping with genes to which our significant (corrected $p < 0.05$) and suggestive DMPs ($p < 1.0 \times 10^{-5}$) and significant DMRs (corrected $p < 0.01$ from both DMR methods) were annotated. A list of mapped genes was input to the network analysis and disease mapping in 'core analysis' of Ingenuity Pathway Analysis (IPA; Ingenuity Systems, CA, USA).

For graphical display, we used the University of California, Santa Cruz (UCSC) Genome Browser [31,32] and coMET [33]. To check for presence of regula-

tory elements in loci containing significant DMPs or DMRs, we used the UCSC Genome Browser. The regulatory elements included DNaseI hypersensitivity sites (DHSs), transcription factor binding sites, chromatin state segmentation and histone modification. We added SNPs in publications containing a keyword 'lung' and gene expression in lung tissue. We connected three databases to the Genome Browser [31,32]: 'Roadmap Epigenomics Data Complete Collection at Wash U VizHub (Assembly hg19)' [34], 'ENCODE Analysis Hub (Assembly hg19)' [35] and 'Genotype-Tissue Expression (GTEx) RNA-seq Signal Hub (Assembly hg19)' [36]. For top two DMRs (based on FDR from DMRcate), we used coMET to show statistical significance of association, comethylation patterns, functional annotations and regulation tracks. The functional annotations include genes/transcripts and regulation tracks from Ensembl and CpG islands, Broad Chromatin State Segmentation by hidden Markov model (ChromHMM) domains, Digital DNaseI Hypersensitivity Clusters (DNaseI clusters) and SNPs from the UCSC database.

We used the manufacturer's annotation file [37] for gene annotations in tables and for an enrichment and a functional network analysis.

Replication look-up

We checked differential methylation at DMPs in relation to COPD and lung function in three previous EWASs that used the Illumina 450k in lung tissue from African-Americans [11] or the Illumina 27k in blood from Caucasians [9] or African-Americans [12]. Of 535 CpGs from the 450k [11], 523 were ready for association after probe QC steps in our data. Of 349 CpGs from the 27k [9], 332 probes were available in the 450k, 315 of which were available for association after probe filtering in our data. Of 12 CpGs from the 27k [12], 11 were available for association in our data. For replicated CpGs, we examined whether the annotated genes were differentially expressed in relation to the same trait. The cutoff for statistical significance was set to uncorrected $p < 0.05$ in replication analyses of differential methylation and gene expression analyses.

Transcriptome analysis: Asan Biobank

Transcriptome profiles were from lung tissue of Korean male smokers (97 COPD cases and 91 noncases, Asan Biobank). Details of the RNA-seq (HiSeq 2000 system, Illumina Inc., CA, USA) transcriptome data can be found elsewhere [38]. We included expression data of the genes to which significant (Bonferroni corrected $p < 0.05$) or suggestive (uncorrected $p < 1.0 \times 10^{-5}$) DMPs or significant DMRs (corrected $p < 0.01$ from

Table 2. CpGs with differential methylation ($p < 1.0 \times 10^{-5}$) in relation to lung function in a Korean chronic obstructive pulmonary disease cohort, ordered by chromosomal location.

Chr [†]	Gene (distance to gene [‡])	Trait	Probe	Position [§]	Coef [¶]	SE [#]	p-value ^{††}
1	<i>LDLRAD1</i>	Ratio	cg09175620	54484536	-1.6	0.3	3.4×10^{-6}
	<i>AK5</i>	FVC ^{**}	cg05347985	77751443	-14.0	3.0	8.9×10^{-6}
	<i>TTC24</i>	Ratio	cg03760759	156549533	-1.7	0.4	8.2×10^{-6}
	<i>DUSP5P</i>	FEV1 ^{§§}	cg10107958 ^{¶¶,##}	228783383	-13.3	2.8	8.4×10^{-6}
3	<i>CCDC13</i>	Ratio	cg02828993 ^{¶¶}	42775696	-2.8	0.6	8.2×10^{-6}
	<i>SRPRB</i> (40)	FEV1	cg20276088	133502917	9.7	2.1	8.7×10^{-6}
7	<i>SNX10</i>	Ratio	cg06280210 ^{¶¶,†††}	26331247	-7.1	1.5	7.7×10^{-6}
8	<i>CSMD3</i>	Ratio	cg20564273	114446439	-5.0	1.0	7.3×10^{-6}
9	<i>NXNL2</i> (43179)	Ratio	cg14210862 ^{¶¶,##}	91193195	-3.5	0.7	7.8×10^{-6}
10	<i>DIP2C</i>	Ratio	cg03559389	679322	-6.9	1.2	8.1×10^{-8}
	<i>SMNDC1</i>	Ratio	cg26944434 ^{##}	112064524	-7.3	1.5	5.1×10^{-6}
11	<i>FAR1</i>	Ratio	cg24412501 ^{##}	13690194	-6.1	1.3	9.8×10^{-6}
	<i>PATL1</i> (35101)	Ratio	cg06022671	59439293	-2.6	0.5	3.7×10^{-6}
12	<i>NT5DC3</i>	FEV1	cg15977278 ^{##}	104234975	-31.1	6.6	9.6×10^{-6}
14	<i>SERPINA12</i>	FVC	cg19904425	94984530	-3.7	0.8	7.1×10^{-6}
19	<i>C19orf18</i>	FEV1	cg094440215	58480646	-19.8	4.2	7.7×10^{-6}

[†]Chromosome.

[‡]Distance to transcription start site of the mapped gene (basepair, bp).

[§]Physical position (bp, National Center for Biotechnology Information human reference genome assembly Build 37.3).

[¶]Regression coefficient from statistical models. Covariates age, sex, height, smoking status, pack-years and estimated cell-type proportions were included in the models for FEV1, FEV1/FVC and COPD. Additionally, weight was included in the model for FVC. FEV1 and FVC were in liters and methylation (β) was ranged between 0 and 1 in the analyses.

[#]Standard error of regression coefficient.

^{††}Uncorrected p-value.

^{**}Forced vital capacity.

^{§§}Forced expiratory volume in 1 s.

^{¶¶}Enhancer.

^{##}Promoter associated.

^{†††}DNaseI hypersensitivity site.

COPD: Chronic obstructive pulmonary disease; FEV1: Forced expiratory volume in 1 second; FVC: Forced vital capacity.

two DMR methods) annotated. Extreme values, outside of three-times the interquartile range away from the first and third quartiles of each gene transcript, were removed to limit their potential impact on association results [23]. This removed 35,607 (1.1%) gene expression values across all participants.

We lacked information on height and weight for the participants, so we analyzed FEV1 percent predicted and FVC percent predicted instead of FEV1 and FVC, respectively. We multiplied the number of cigarette smoked (packs per day) by the number of years smoked for pack-years.

To check differential gene expression in relation to COPD or lung function, a logistic regression model was used for COPD status and robust linear regression models were used for FEV1 percent predicted, FVC percent predicted and FEV1 percent predicted/FVC percent predicted. The gene expression values for each gene transcript were the predictor variables and

COPD or the pulmonary traits were the response variables. Covariates were age and pack-years. The threshold of statistical significance was set to uncorrected $p < 0.05$.

All the methylation data processing, statistical analyses and visualizations were conducted in R (version 3.0.2) [39] except for comb-p, IPA and the UCSC Genome Browser.

Results

Of 100 participants in our EWAS, 60 were COPD cases and 40 were noncases. The average age was 72.8 years. There were 39 never, 30 former and 31 current smokers. Age and height were not significantly different between COPD cases and noncases, whereas weight and BMI were lower in COPD cases than noncases ($p = 0.02$ and 0.008 , respectively). Characteristics of study participants can be found in Table 1. We provide a workflow of our study in Figure 1.

Table 3. Differentially methylated regions in relation to more than one lung function trait or chronic obstructive pulmonary disease (multiple-testing corrected $p < 0.01$ from both DMRcate and comb-p) in a Korean chronic obstructive pulmonary disease cohort, ordered by chromosomal location.

Chr [†]	Gene (distance to gene [‡])	Trait	DMRcate				comb-p			
			Start (bp [§])	End (bp)	# CpGs	FDR [#]	Minimum P ^{††}	Start (bp)	End (bp)	Sidak P ^{‡‡}
3	SRPRB (-337)	FEV1 ^{§§}	133502540	133503437	7 (6)	4.7×10^{-9}	8.7×10^{-6}			3.4×10^{-8}
		FVC ^{¶¶}			7 (6)	1.3×10^{-10}	1.0×10^{-4}			3.8×10^{-8}
5	BHMT	COPD ^{##}	78406928	78407683	9 (6)	0.002	0.004			0.01
		FEV1		78408253	11 (6)	2.0×10^{-4}	3.7×10^{-4}		78408347	0.002
		FVC		78408347	12 (6)	1.0×10^{-6}	1.4×10^{-4}	78407153		1.2×10^{-6}
6	HLA-C (38622)	FEV1	31275148	31276187	13 (8)	6.8×10^{-5}	0.003			2.1×10^{-5}
		COPD		31276667	16 (7)	5.0×10^{-7}	4.7×10^{-4}		31276797	2.7×10^{-5}
8	ZC3H3 (-700)	FEV1	144519125	144519203	3 (3)	3.1×10^{-4}	1.2×10^{-4}			6.5×10^{-4}
		FVC			3 (3)	0.002	3.1×10^{-4}			0.005
10	FAM196A	COPD	128994297	128995192	9 (3)	0.008	0.001	128993810		8.2×10^{-4}
		Ratio	128994702		3 (2)	0.004	0.003			0.009
16	CTU2	COPD	88772396	88772985	6 (4)	0.005	4.2×10^{-4}			0.003
		Ratio			6 (5)	0.002	8.3×10^{-4}			0.009
17	CD300A	COPD	72462164	72463080	8 (6)	1.9×10^{-5}	6.9×10^{-4}			1.2×10^{-5}
		FEV1			8 (7)	2.8×10^{-8}	2.5×10^{-4}			1.2×10^{-8}
		Ratio			8 (5)	3.2×10^{-5}	7.4×10^{-4}			9.9×10^{-6}
	USP36	COPD	76797750	76798776	6 (4)	0.008	0.001			0.003
		Ratio			6 (3)	0.002	0.002			0.002
19	KLK10	FEV1	51520093	51520658	6 (3)	1.5×10^{-4}	9.7×10^{-5}			0.004
		Ratio			6 (4)	3.0×10^{-4}	3.3×10^{-5}			0.008
22	CPT1B	COPD	51016501	51017432	15 (9)	5.3×10^{-5}	0.010			4.7×10^{-4}
		FEV1		51017723	16 (12)	1.0×10^{-14}	6.8×10^{-4}			2.5×10^{-10}
		FVC			16 (13)	5.4×10^{-24}	8.2×10^{-5}			1.5×10^{-13}

[†]Chromosome.

[‡]Minimum distance to transcription start site of the mapped gene (basepair, bp).

[§]Physical position (bp, National Center for Biotechnology Information human reference genome assembly Build 37.3).

^{||}Number of CpGs in a region (number of CpGs with unadjusted $p < 0.05$).

[#]False discovery rate.

^{††}Minimum P in a region.

^{‡‡}P of Sidak multiple-testing correction.

^{§§}Forced expiratory volume in one second.

^{¶¶}Forced vital capacity.

^{##}Chronic obstructive pulmonary disease.

Blank cells in 'Start(bp)' and 'End(bp)' for comb-p represent the same information from DMRcate. Inputs of two packages were results from epigenome-wide association analyses adjusted for age, sex, height, weight (only for FVC), smoking status, pack-years, and estimated cell-type proportions.

COPD: Chronic obstructive pulmonary disease; FDR: False discovery rate; FEV: Forced expiratory volume in 1 second; FVC: Forced vital capacity.

From our DMP analyses, we found a significant association between cg03559389 (*DIP2C*) and FEV1/FVC after Bonferroni correction ($p = 8.1 \times 10^{-8}$) (Table 2). There was no significant DMP for FEV1, FVC nor COPD after multiple-testing correction. We found suggestive associations ($p < 1. \times 10^{-5}$) at 15 CpGs: four

CpGs for FEV1, two for FVC and nine for FEV1/FVC. The top CpG for FEV1 was cg09440215 (*C19orf18*) and the top for FVC was cg19904425 (*SERPINA12*). There was no systematic inflation in our association results based on genomic inflation factor (λ) values: 1.00 for FEV1, 0.93 for FVC, 1.16 for FEV1/FVC

Table 4. Associations with gene expression (unadjusted $p < 0.05$) in the Asan Biobank for genes to which differentially methylated probes or genes annotated, ordered by chromosomal location.

Differentially methylated probes in relation to lung function (the Korean COPD cohort)		Differentially methylated probe		Differentially expressed genes in relation to lung function in lung tissue (Asan Biobank)					
Chr [†]	Trait	Probe or region [‡]	Coef	p-value [§]	Gene	Transcript	Coef	p-value	
10	Ratio	cg03559389	-6.9	8.1×10^{-8}	Body	<i>DIP2C</i>	NM_014974	-0.029	6.5×10^{-5}
11		cg06022671	-2.6	3.7×10^{-6}		<i>PATL1</i>	NM_152716	-0.015	1.3×10^{-6}
Differentially methylated region									
		Start–end (bp)	#CpGs	FDR [¶]	Genomic feature				
1	Ratio	5353149-55353706	5 (4)	0.002	TSS1500; promoter	<i>DHCR24</i>	NM_014762	-3.9×10^{-4}	0.005
	COPD#	110752952-110754257	10 (6)	1.9×10^{-4}	TSS1500; TSS200; 1stExon; promoter	<i>KCNK4</i>	NM_001039574	-0.424	0.036
2	COPD	74347215-74347776	4 (3)	0.010		<i>BOLA3</i> (-15313)	NM_001035505	0.110	5.3×10^{-5}
	Ratio	102313644-102314757	8 (2)	0.004	TSS1500; TSS200; 1stExon; body; promoter (cell type specific)	<i>MAP4K4</i>	NM_145686	-0.006	7.5×10^{-5}
	Ratio	113992762-113993313	8 (8)	1.8×10^{-5}	Body; promoter (cell type specific); DHS	<i>PAX8</i>	NM_013952	0.048	1.7×10^{-4}
	COPD	220083306-220083951	10 (5)	0.006	1stExon; 5'UTR; TSS200; TSS1500; promoter	<i>ABCB6</i>	NM_005689	0.578	8.9×10^{-4}
3	COPD	167452484-167453518	13 (3)	5.4×10^{-5}	5'UTR; TSS200; TSS1500; promoter; DHS	<i>PDCD10</i>	NM_007217	-0.097	0.002
4	COPD	20253130-20254519	8 (7)	1.9×10^{-4}	TSS1500	<i>SLIT2</i>	NM_004787	-0.087	5.0×10^{-5}
5	COPD	140800398-140800929	8 (4)	8.6×10^{-4}	Body; DHS	<i>PCDHGA4</i>	NM_032053	-2.093	1.6×10^{-4}
7	COPD	27182637-27185512	48 (26)	3.5×10^{-14}	Body; 1stExon; 5'UTR; TSS200; TSS1500; 3'UTR; body; DHS	<i>HOXA5</i>	NM_019102	0.049	0.021
8	Ratio	98610507-98610888	3 (3)	0.007		<i>MTDH</i> (-45900)	NM_178812	-0.006	0.002
	Ratio	99960843-99961768	6 (4)	0.005	5'UTR; body; enhancer	<i>OSR2</i>	NM_053001	-0.012	0.036
	COPD	144358043-144359316	5 (5)	3.1×10^{-4}	Body; 3'UTR; promoter	<i>GLI4</i> (9709)	NM_138465	0.761	1.2×10^{-5}
	COPD	144659627-144661051	11 (8)	0.003	Body; TSS200; TSS1500; promoter	<i>NAPRT1</i>	NM_145201	0.135	0.005
10	COPD	105212236-105213309	11 (5)	1.5×10^{-4}	TSS200; TSS1500; 3'UTR; enhancer; Promoter	<i>CALHM2</i>	NM_015916	0.173	0.003
11	Ratio	94277826-94279068	11 (5)	0.002	1stExon; enhancer; Promoter (cell type specific)	<i>FUT4</i>	NM_002033	-0.043	0.004
	COPD	108408907-108409825	7 (5)	5.8×10^{-4}	Body; TSS1500	<i>EXPH5</i>	NM_015065	-0.385	0.002

Table 4. Associations with gene expression (unadjusted $p < 0.05$) in the Asan Biobank for genes to which differentially methylated probes or genes annotated, ordered by chromosomal location (cont.).

Differentially methylated probes in relation to lung function (the Korean COPD cohort)		Differentially methylated probe		Differentially expressed genes in relation to lung function in lung tissue (Asan Biobank)					
Chr [†]	Trait	Probe or region [‡]	Coef	p-value [§]	Genomic feature	Gene	Transcript	Coef	p-value
12	COPD	2943902-2944493	6 (6)	2.9×10^{-4}	1stExon; 5'UTR; TSS200; TSS1500; Enhancer; DHS	<i>NR1P2</i>	NM_031474	-0.840	0.004
16	COPD	67232921-67233983	6 (4)	2.2×10^{-4}	TSS200; 1stExon; body; enhancer; DHS	<i>ELMO3</i>	NM_024712	0.535	2.1×10^{-6}
	COPD	88772396-88772985	6 (4)	0.005	TSS1500; 1stExon; promoter	<i>CTU2</i>	NM_001012759	0.877	1.8×10^{-5}
	Ratio		6 (5)	0.002				0.026	5.2×10^{-5}
17	COPD	33759484-33760527	12 (10)	3.1×10^{-4}	1stExon; TSS200; TSS1500; DHS	<i>SLFN12</i>	NM_018042	-0.539	1.6×10^{-4}
	COPD	73824354-73825227	6 (4)	0.003	Body; enhancer; promoter	<i>UNC13D</i>	NM_199242	0.072	0.010
	COPD	76797750-76798776	6 (4)	0.008	Body	<i>USP36</i>	NM_025090	-0.262	3.5×10^{-5}
	Ratio		6 (3)	0.002				-0.006	8.7×10^{-4}
18	Ratio	74114570-74114728	3 (3)	4.8×10^{-4}	Body	<i>ZNF516</i>	NM_014643	-0.042	2.7×10^{-7}
	COPD		3 (3)	0.003				-1.276	1.1×10^{-6}
19	COPD	11517079-11517436	4 (4)	0.003	Body	<i>RGL3</i>	NM_001161616	0.084	0.003
	Ratio	37958346-37959104	8 (4)	7.2×10^{-4}	TSS200; TSS1500; promoter	<i>ZNF570</i>	NM_144694	-0.032	0.008
	FEV1 ^{††}	51520093-51520658	6 (3)	1.5×10^{-4}	Body; DHS	<i>KLK10</i>	NM_002776	1.381	0.010
	Ratio		6 (4)	3.0×10^{-4}				0.008	0.015
21	COPD	38362230-38363160	8 (6)	0.003	5'UTR; TSS200; TSS1500; promoter	<i>HLC5</i>	NM_001242785	-1.268	6.7×10^{-7}
22	COPD	51016501-51017432	15 (9)	5.3×10^{-5}	5'UTR; TSS200; 1stExon; Enhancer; DHS	<i>CPT1B</i>	NM_001145135	0.197	0.001
	FEV1	51016501-51017723	16 (12)	1.0×10^{-14}				0.820	0.027

[†]Chromosome.

[‡]Start-end position of each region (bp).

[§]Statistical significance from statistical model.

[¶]False discovery rate from DMRcate.

^{‡‡}Chronic obstructive pulmonary disease.

^{†††}Forced expiratory volume in one second.

Genomic features were based on Illumina's Annotation file and those for DMRs were based on CpGs at start and end position of each region. Categories for the features includes (1) body, gene body; (2) 5'UTR, 5 prime untranslated region; (3) 3'UTR, 3 prime untranslated region; (4) TSS200, 200 basepair within transcription start site; (5) TSS1500, 1500 basepair within transcription start site; and (6) DHS, DNase I hypersensitivity site.

Differentially methylated probes or regions in this Table were located inside each mapped gene.

COPD: Chronic obstructive pulmonary disease; DHS: DNaseI hypersensitivity site; FEV1: Forced expiratory volume in 1 second; FVC: Forced vital capacity.

and 1.37 for COPD. We provide Manhattan plots (Supplementary Figure 1) and quantile–quantile (qq) plots (Supplementary Figure 2).

From our DMR analyses, we identified 104 regions of significant differential methylation (corrected $p < 0.01$) in relation to either the COPD or spirometric measures (Supplementary Table 3). Of the 104, ten (*SRPRB*, *ZC3H3*, *HLA-C*, *KLK10*, *FAM196A*, *CTU2*, *USP36*, *BHMT*, *CPT1B* and *CD300A*) were associated with more than one trait (Table 3).

For the enrichment analysis, we included one significant and 15 suggestive DMPs and among the 845 CpGs within 104 DMRs we examined 505 nominally significant ones. After taking two overlapping CpGs into account we analyzed 519 CpGs. We found enrichment for CpG islands (50 vs 31% overall from the array; $p = 7.0 \times 10^{-21}$), CpG island shores (29 vs 23% overall; $p = 0.0004$) and DHSs (25 vs 12% overall; $p = 1.2 \times 10^{-14}$) in Supplementary Table 8.

Our functional network analysis highlighted embryonic/organ development, cell death/survival and gene expression (Supplementary Table 4). Inflammatory response was highlighted from the disease mapping (Supplementary Table 5).

We visualized regulatory elements at 11 loci containing the genome-wide significant DMP (Supplementary Figure 3) and the ten significant DMRs associated with more than one trait (Supplementary Figure 4). Regulatory elements appear to be well-represented in the 11 loci (Supplementary Table 10). For example, a locus ± 10 kb around a DMR ‘chr22:51016501–51017432’ (Supplementary Figure 4J) contains potential functional elements including DHSs, open chromatin, active promoter, active transcription start sites, enhancers and histone modification. Of the ten DMRs, we visualized top two loci (*CPT1B* and *SRPRB*) based on the FDR values from DMRcate to show regional association results and functional annotation tracks (Supplementary Figure 5).

For 111 genes to which the 16 DMPs or 104 DMRs mapped, we had expression data at 86 genes from the Asan Biobank. Of these 86, we identified differential gene expression in relation to the same trait (uncorrected $p < 0.05$) at 31 genes (Table 4). Of the 31 genes, five (*CTU2*, *USP36*, *ZNF516*, *KLK10* and *CPT1B*) were associated with more than one trait. Of the 31 genes, 21 showed differential gene expression in relation to COPD. Of these 21, four (*CTU2*, *USP36*, *ZNF516* and *CPT1B*) were associated with FEV1 or FEV1/FVC, as well as COPD.

From a look-up of differential methylation at 315 probes from an EWAS in blood DNA [9] and 523 probes from an EWAS in lung tissue [11], 77 and 87 showed differential methylation in relation to COPD

or lung function (uncorrected $p < 0.05$) in our data, respectively. Of the 77 replicated probes from the EWAS [9] in blood, 72 probes mapped to 71 genes for which we had gene expression values in our data. Of the 71 genes, differential gene expression in relation to the same trait of methylation differences was observed at 29 genes (Supplementary Table 6). Of the 87 replicated probes from the EWAS [11] in lung tissue, 65 probes mapped to 62 genes with gene expression in our data. Of the 62 genes, 31 genes were differentially expressed (Supplementary Table 7). No CpG of the 11 from an EWAS in blood DNA [12] was replicated in our study.

Discussion

Our study identified novel differential methylation signals associated with COPD and lung function traits in blood DNA and linked the methylation alterations to differential gene expression in lung tissue. Previous studies of COPD or lung function used DNA methylation in blood from smokers using the Infinium HumanMethylation27 BeadChip (27k) [9,12], in small airways epithelia from former smokers using the 27k [10], in lung tissue using the Nimblegen array [8], or in lung tissue from former smokers using the 450k [11]. This is the first study to use the Illumina 450k in blood DNA.

We found one genome-wide significant CpG and 15 suggestive CpGs in relation to lung function traits. These DMPs are novel, meaning never reported in the previous EWASs of COPD or lung function. Cg03559389 (*DIP2C*) showed a significant association with FEV1/FVC after Bonferroni correction. *DIP2C* has a DMAP1 binding domain. Mutations in *DIP2C* have been identified in lung cancer samples [40]. Among the 15 suggestive DMPs, cg19904425 (*SERPINA12*) is 127kb away from *SERPINA1*, an important gene for COPD [41] and lung function [42]. A recent exome-wide association analysis [43] reported a rare variant (rs140198372) in *SERPINA12* 30kb away from cg19904425 was associated with airflow limitation.

Several loci from our DMR study overlap loci associated with COPD or pulmonary function in previous genome-wide association studies (GWASs), candidate gene studies or gene expression analyses. Four loci have been reported in GWASs: *NCR3* [44], *AGER-PPT2* [45], *CCL18* [46] and *UPK3A* [47]. Three overlap loci from candidate gene studies: *ILIRN* [48], *F2R* [49] and *MMP9* [50]. Two were reported in gene expression studies: *PSORS1C1* [51] and *USP36* [52]. Two DMRs in our study were near *HLA-DQB1/HLA-DQA2*, a locus from a SNP-by-smoking interaction study of lung function [53]. This is notable because we

excluded 28,652 CpGs at which methylation can be influenced by neighboring SNPs [20]. We found differential methylation in a locus previously associated with a putative biomarker of COPD, surfactant protein D. A GWAS [54] of surfactant protein D reported five genome-wide significant SNPs in/near *PSORS1C1*. Our DMR resides between the SNPs. This may suggest an epigenetic contribution modulating the association between the genetic variants and pulmonary traits or COPD.

Of interest, the highlighted fundamental pathways in our study support the emerging hypothesis of the importance of fundamental developmental processes in COPD pathogenesis [55,56]. This study supports a possible role for epigenetic modifications in this process.

The enrichment for CpG shores (regions within 2kb from a CpG island) in differentially methylation signals related to lung function is consistent with the findings from a previous EWAS of COPD [11]. To assess enrichment for gene ontology terms in the 111 genes to which significant and suggestive DMPs and significant DMRs mapped, we used WEB-based Gene Set Analysis Toolkit (WebGestalt) [57] with databases for biological process, cellular process and molecular function. No gene ontology terms were significantly enriched.

Among epigenetic modifications, we examined methylation alterations in relation to COPD and lung function in this study. The well-represented regulatory elements including DNase Hypersensitivity sites/clusters, histone modification and chromatin state segmentation in our significant loci suggest that these loci may play functional roles in control of COPD or lung function.

From our replication look-up, we confirmed differential methylation in relation to COPD and lung function previously identified in non-Asian populations (Caucasians [9,11] and African Americans [11]). Of interest, differential methylation from an EWAS [11] in lung tissue was also replicated in our study in blood, supporting the possibility of blood DNA methylation biomarkers for lung function.

In our data, there were 19 mild and 41 moderate to severe COPD cases [16]. To check whether there were CpGs associated with moderate to severe COPD, we performed an EWAS of COPD using 41 moderate to severe cases and 40 noncases. There were no significant DMPs in relation to moderate to severe COPD after multiple-testing correction. In our study, there was no genome-wide significant DMP in relation to COPD. However, all 14 DMPs related to FEV1 or FEV1/FVC (unadjusted $p < 1 \times 10^{-5}$) were also related to COPD ($p < 0.05$) as shown in **Supplementary Table 9**.

Our study has limitations. First, the data were cross-sectional for this analysis that hampers the causal inference between methylation alterations and COPD status or lung function levels. We cannot ascertain if these might have predated the COPD or are the result of the COPD. Second, we did not have an available population for replication to decrease the chance of reporting false positives at the lower significance threshold used for the DMP analysis. As strengths, we explored regional differential methylation alterations by using two different methods (DMRcate and comb-p) in addition to examining individual probes. Further, the gene expression profiles in lung tissue strengthen the biologic evidence that the DNA methylation alterations observed in blood are related to lung function and thus provide a type of functional replication. Finally, our study was conducted in Asian populations; there are few methylation and transcriptome data on these populations.

Although GWASs of COPD and lung function have identified numerous loci, these findings only explain the small proportion of variance in these traits [5,6]. Epigenetics may explain some of the remaining variance [58]. Several EWASs reported candidate loci in relation to these phenotypes [8–12]. Studies in blood or lung tissue involving Asians, Caucasians or African Americans could help understand underlying mechanisms of COPD and lung function, and provide candidate target genes for clinical purposes.

Conclusion

We report novel differential methylation alterations associated with COPD or pulmonary traits in blood DNA, and link methylation alterations to differential gene expression related to lung function in lung tissue. Our finding of differential methylation near reported loci from genetic studies of COPD and pulmonary traits supports a possible epigenetic role in control of lung function.

Future perspective

The methylation alterations in blood related to COPD and lung function traits could play a role in development of biomarkers of COPD and lung function. Future studies would be necessary to confirm these findings and understand their mechanistic basis.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.futuremedicine.com/doi/full/10.2217/epi-2017-0002

Acknowledgements

The authors appreciate all of the study participants for their contribution to this research. The authors thank F Day, T Wang

and S Reese of NIEHS and J Jin of Westat, Inc. for expert computational assistance.

No writing assistance was utilized in the production of this manuscript.

Financial & competing interest disclosure

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (2013R1A1A1057961, 2017R1A2B4003790), the Ministry of Education (2013R1A6A3A04059017) and grants from the Environmental Health Center funded by the Ministry of Environment, Republic of Korea. Supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (ZO1 ES043012). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Ethical conduct of research

The Institute Review Board of the Kangwon National University Hospital approved analyses of the data (Institutional Review Board of Kangwon National University Hospital 2012-06-007-001 and KNUH-2016-05-003-001). Informed written consent was obtained from all participants. The study adhered to the tenets of the Helsinki Declaration of 1975, as revised in 2008.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Summary points

- This is the first epigenome-wide association study of chronic obstructive pulmonary disease and lung function traits in an Asian population.
- Differentially methylated probes from earlier epigenome-wide association studies in non-Asian populations were differentially methylated in our study.
- Our study revealed differentially methylated probes and differentially methylated regions in relation to chronic obstructive pulmonary disease and lung function traits in blood DNA. Candidate loci includes *DIP2C*, *SRPRB*, *ZC3H3*, *HLA-C*, *KLK10*, *FAM196A*, *CTU2*, *USP36*, *BHMT*, *CPT1B* and *CD300A*.
- Genes including *CTU2*, *USP36*, *ZNF516*, *KLK10* and *CPT1B* to which the differential methylation alterations mapped were also differentially expressed in lung tissue with respect to the same traits.

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

- 1 Schunemann HJ, Dorn J, Grant BJ, Winkelstein W, Jr., Trevisan M. Pulmonary function is a long-term predictor of mortality in the general population: 29-year follow-up of the Buffalo Health Study. *Chest* 118(3), 656–664 (2000).
- 2 Mannino DM, Buist AS. Global burden of COPD: risk factors, prevalence, and future trends. *Lancet* 370(9589), 765–773 (2007).
- 3 Koshiol J, Rotunno M, Consonni D *et al.* Chronic obstructive pulmonary disease and altered risk of lung cancer in a population-based case-control study. *PLoS ONE* 4(10), e7380 (2009).
- 4 Carrasco Garrido P, De Miguel Diez J, Rejas Gutierrez J *et al.* Negative impact of chronic obstructive pulmonary disease on the health-related quality of life of patients. Results of the EPIDEPOC study. *Health Qual. Life Outcomes* 4, 31 (2006).
- 5 Loth DW, Soler Artigas M, Gharib SA *et al.* Genome-wide association analysis identifies six new loci associated with forced vital capacity. *Nat. Genet.* 46(7), 669–677 (2014).
- 6 Cho MH, Castaldi PJ, Wan ES *et al.* A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. *Hum. Mol. Genet.* 21(4), 947–957 (2012).
- 7 Boezen HM, Mannino DM. The future of nature versus nurture in understanding chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* 188(8), 891–892 (2013).
- 8 Yoo S, Takikawa S, Geraghty P *et al.* Integrative analysis of DNA methylation and gene expression data identifies EPAS1 as a key regulator of COPD. *PLoS Genet.* 11(1), e1004898 (2015).
- 9 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.* 185(4), 373–381 (2012).
- 10 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.* 50(5), 912–922 (2014).
- 11 Morrow JD, Cho MH, Hersh CP *et al.* DNA methylation profiling in human lung tissue identifies genes associated with COPD. *Epigenetics* doi:10.1080/15592294.2016.1226451 (2016) (Epub ahead of print).
- Identified DNA methylation signals related to chronic obstructive pulmonary disease in lung tissue using the 450K Illumina array.
- 12 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.* 17(1), 143 (2016).

- 13 Lister R, Pelizzola M, Dowen RH *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462(7271), 315–322 (2009).
- **Reported single-base-resolution DNA methylation maps throughout the human genome to understand genome-wide patterns and composition of DNA methylation.**
- 14 Jaffe AE, Murakami P, Lee H *et al.* Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int. J. Epidemiol.* 41(1), 200–209 (2012).
- 15 Kim WJ, Hong Y. Methodology of an observational cohort study for subjects with chronic obstructive pulmonary disease in dusty areas near cement plants. *J. Pulm. Respir. Med.* 04(01), 169–174 (2014).
- 16 Global Initiative for Chronic Obstructive Lung Disease. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease (2016). <http://goldcopd.org>
- 17 Ji W, Lim MN, Bak SH *et al.* Differences in chronic obstructive pulmonary disease phenotypes between non-smokers and smokers. *Clin. Respir. J.* doi:10.1111/crj.12577 (2016) (Epub ahead of print).
- 18 Teschendorff AE, Marabita F, Lechner M *et al.* A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* 29(2), 189–196 (2013).
- 19 Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8(1), 118–127 (2007).
- 20 Nordlund J, Backlin CL, Wahlberg P *et al.* Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. *Genome Biol.* 14(9), r105 (2013).
- 21 Morris TJ, Butcher LM, Feber A *et al.* ChAMP: 450k chip analysis methylation pipeline. *Bioinformatics* 30(3), 428–430 (2014).
- 22 Price ME, Cotton AM, Lam LL *et al.* Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. *Epigenetics Chromatin* 6(1), 4 (2013).
- 23 What are outliers in the data?
www.itl.nist.gov
- 24 Houseman EA, Accomando WP, Koestler DC *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 13, 86 (2012).
- 25 Reinius LE, Acevedo N, Joerink M *et al.* Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS ONE* 7(7), e41361 (2012).
- 26 Peters TJ, Buckley MJ, Statham AL *et al.* De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin* 8, 6 (2015).
- 27 Pedersen BS, Schwartz DA, Yang IV, Kechris KJ. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated p-values. *Bioinformatics* 28(22), 2986–2988 (2012).
- 28 Benjamini Y, Hochberg Y. Controlling the false discovery rate – a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series. B Stat. Methodol.* 57(1), 289–300 (1995).
- 29 Sidak Z. Rectangular confidence regions for the means of multivariate normal distributions. *J. Am. Stat. Assoc.* 62(318), 8 (1967).
- 30 Peters T. Package ‘DMRcate’ (2016).
- 31 Kent WJ, Sugnet CW, Furey TS *et al.* The human genome browser at UCSC. *Genome Res.* 12(6), 996–1006 (2002).
- 32 UCSC Genome Browser.
<http://genome.ucsc.edu>
- 33 Martin TC, Yet I, Tsai PC, Bell JT. coMET: visualisation of regional epigenome-wide association scan results and DNA co-methylation patterns. *BMC Bioinformatics* 16, 131 (2015).
- 34 Roadmap Epigenomics Visualization Hub (VizHub).
<http://vizhub.wustl.edu/>
- 35 Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414), 57–74 (2012).
- 36 Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* 45(6), 580–585 (2013).
- 37 Illumina I. Illumina Infinium HumanMethylation450 BeadChip Annotation.
<https://support.illumina.com>
- 38 Kim WJ, Lim JH, Lee JS, Lee SD, Kim JH, Oh YM. Comprehensive analysis of transcriptome sequencing data in the lung tissues of COPD subjects. *Int. J. Genomics* 2015, 206937 (2015).
- 39 R Foundation for Statistical Computing. R: A language and environment for statistical computing (2013).
www.gbif.org/resource/81287
- 40 Integrative analysis of complex cancer genomics and clinical profiles.
www.cbioportal.org/beta/index.do
- 41 Kumar M, Bhadoria DP, Dutta K *et al.* Combinatorial effect of TIMP-1 and alpha1AT gene polymorphisms on development of chronic obstructive pulmonary disease. *Clin. Biochem.* 44(13), 1067–1073 (2011).
- 42 Perez-Rubio G, Jimenez-Valverde LO, Ramirez-Venegas A *et al.* Prevalence of alpha-1 antitrypsin high-risk variants in Mexican mestizo population and their association with lung function values. *Arch. Bronconeumol.* 51(2), 80–85 (2015).
- 43 Jackson VE, Ntalla I, Sayers I *et al.* Exome-wide analysis of rare coding variation identifies novel associations with COPD and airflow limitation in MOCS3, IFIT3 and SERPINA12. *Thorax* 71(6), 501–509 (2016).
- 44 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.* 43(11), 1082–1090 (2011).
- 45 Hancock DB, Eijgelsheim M, Wilk JB *et al.* Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat. Genet.* 42(1), 45–52 (2010).

- 46 Yao TC, Du G, Han L *et al.* Genome-wide association study of lung function phenotypes in a founder population. *J. Allergy Clin. Immunol.* 133(1), 248–255; e241–e210 (2014).
- 47 Li X, Hawkins GA, Ampleford EJ *et al.* Genome-wide association study identifies TH1 pathway genes associated with lung function in asthmatic patients. *J. Allergy Clin. Immunol.* 132(2), 313e315–320e315 (2013).
- 48 Hegab AE, Sakamoto T, Saitoh W *et al.* Polymorphisms of TNFalpha, IL1beta, and IL1RN genes in chronic obstructive pulmonary disease. *Biochem. Biophys. Res. Commun.* 329(4), 1246–1252 (2005).
- 49 Plate M, Lawson PJ, Hill MR *et al.* Impact of a functional polymorphism in the PAR-1 gene promoter in COPD and COPD exacerbations. *Am. J. Physiol. Lung Cell Mol. Physiol.* 307(4), L311–L316 (2014).
- 50 Lee SY, Kim MJ, Kang HG *et al.* Polymorphisms in matrix metalloproteinase-1, -9 and -12 genes and the risk of chronic obstructive pulmonary disease in a Korean population. *Respiration* 80(2), 133–138 (2010).
- 51 Qiu W, Cho MH, Riley JH *et al.* Genetics of sputum gene expression in chronic obstructive pulmonary disease. *PLoS ONE* 6(9), e24395 (2011).
- 52 Sun XX, He X, Yin L, Komada M, Sears RC, Dai MS. The nucleolar ubiquitin-specific protease USP36 deubiquitinates and stabilizes c-Myc. *Proc. Natl Acad. Sci. USA* 112(12), 3734–3739 (2015).
- 53 Hancock DB, Soler Artigas M, Gharib SA *et al.* Genome-wide joint meta-analysis of SNP and SNP-by-smoking interaction identifies novel loci for pulmonary function. *PLoS Genet.* 8(12), e1003098 (2012).
- 54 Kim DK, Cho MH, Hersh CP *et al.* Genome-wide association analysis of blood biomarkers in chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* 186(12), 1238–1247 (2012).
- 55 Boucherat O, Morissette MC, Provencher S, Bonnet S, Maltais F. Bridging lung development with chronic obstructive pulmonary disease. Relevance of developmental pathways in chronic obstructive pulmonary disease pathogenesis. *Am. J. Respir. Crit. Care Med.* 193(4), 362–375 (2016).
- 56 Martinez FD. Early-life origins of chronic obstructive pulmonary disease. *N. Engl. J. Med.* 375(9), 871–878 (2016).
- 57 Wang J, Duncan D, Shi Z, Zhang B. WEB-based GENE SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res.* 41, W77–W83 (2013).
- 58 Manolio TA, Collins FS, Cox NJ *et al.* Finding the missing heritability of complex diseases. *Nature* 461(7265), 747–753 (2009).