

Variable DNA Methylation Is Associated with Chronic Obstructive Pulmonary Disease and Lung Function

Weiliang Qiu¹, Andrea Baccarelli², Vincent J. Carey¹, Nadia Boutaoui³, Helene Bacherman¹, Barbara Klanderman¹, Stephen Rennard⁴, Alvar Agusti⁵, Wayne Anderson⁶, David A. Lomas⁷, and Dawn L. DeMeo^{1,8}

¹Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School and ²Exposure Epidemiology and Risk Program, Harvard School of Public Health, Boston, Massachusetts; ³Division of Pulmonary Medicine, Allergy and Immunology, Department of Pediatrics, Children's Hospital of Pittsburgh and University of Pittsburgh Medical School, Pittsburgh, Pennsylvania; ⁴Section of Pulmonary and Critical Care, University of Nebraska Medical Center, Omaha, Nebraska; ⁵Thorax Institute, Hospital Clinic, Institut d'Investigacions Biomèdiques August Pi i Sunyer, University of Barcelona and Centro de Investigación Biomédica en Red Enfermedades Respiratorias, Fundació d'Investigació Sanitària de les Illes Balears, Palma de Mallorca, Spain; ⁶GlaxoSmithKline Research and Development, Research Triangle Park, Durham, North Carolina; ⁷Cambridge Institute for Medical Research, University of Cambridge, Cambridge, United Kingdom; and ⁸Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts

Rationale: Chronic obstructive pulmonary disease (COPD) is associated with local (lung) and systemic (blood) inflammation and manifestations. DNA methylation is an important regulator of gene transcription, and global and specific gene methylation marks may vary with cigarette smoke exposure.

Objectives: To perform a comprehensive assessment of methylation marks in DNA from subjects well phenotyped for nonneoplastic lung disease.

Methods: We conducted array-based methylation screens, using a test-replication approach, in two family-based cohorts (n = 1,085 and 369 subjects).

Measurements and Main Results: We observed 349 CpG sites significantly associated with the presence and severity of COPD in both cohorts. Seventy percent of the associated CpG sites were outside of CpG islands, with the majority of CpG sites relatively hypomethylated. Gene ontology analysis based on these 349 CpGs (330 genes) suggested the involvement of a number of genes responsible for immune and inflammatory system pathways, responses to stress and external stimuli, as well as wound healing and coagulation cascades. Interestingly, our observations include significant, replicable associations between *SERPINA1* hypomethylation and COPD and lower average lung function phenotypes (combined P-values: COPD, 1.5×10^{-23} ; FEV₁/FVC, 1.5×10^{-35} ; FEV₁, 2.2×10^{-40}).

Conclusions: Genetic and epigenetic pathways may both contribute to COPD. Many of the top associations between COPD and DNA methylation occur in biologically plausible pathways. This large-scale analysis suggests that DNA methylation may be a biomarker of COPD and may highlight new pathways of COPD pathogenesis.

Keywords: chronic obstructive pulmonary disease; epigenetics; DNA methylation; smoking

AT A GLANCE COMMENTARY

Scientific Knowledge on the Subject

Epigenetic variation may explain features of chronic obstructive pulmonary disease (COPD) not explained fully by DNA sequence, such as the variable susceptibility to develop lung disease in smokers, as well as the continued elevated risk for COPD after smoking cessation.

What This Study Adds to the Field

This study represents a large-scale gene-specific investigation of DNA methylation marks that associate with COPD and lower lung function, identifying genes both studied in COPD in the past (such as *SERPINA1*) as well as new candidate genes (such as *FUT7*). Future directions from large-scale epigenetic assessments like this will focus on the genes and pathways to interrogate for functional impact on COPD susceptibility and severity.

(Received in original form August 2, 2011; accepted in final form December 2, 2011)

Supported by U.S. National Institutes of Health grant R01 HL089438; D.L.D. is supported by a Clinician Scientist Development Award from the Doris Duke Foundation and by P01 HL105339. A.B. receives support from the ES00002 New Investigator Funding. The ICGN cohort was funded by GlaxoSmithKline.

Author Contributions: D.L.D., B.J.K., and N.B. conceived and designed experiments; D.L.D. and V.J.C. supervised research; D.L.D., B.J.K., N.B., and H.B. performed the experiments including quality control; D.L.D., W.Q., and V.J.C. performed statistical analysis/analyzed data; and S.R., A.A., W.A., and D.A.L. contributed to cohort enrollment/data collection. All authors contributed to the writing of this manuscript.

Correspondence and requests for reprints should be addressed to Dawn L. DeMeo, M.D., M.P.H., Channing Laboratory, 181 Longwood Avenue, Boston, MA 02115. E-mail: redld@channing.harvard.edu

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

Am J Respir Crit Care Med Vol 185, Iss. 4, pp 373–381, Feb 15, 2012

Copyright © 2012 by the American Thoracic Society

Originally Published in Press as DOI: 10.1164/rccm.201108-1382OC on December 8, 2011

Internet address: www.atsjournals.org

Chronic obstructive pulmonary disease (COPD) is a multifactorial complex human disease of the lungs with world-wide impact (1). Genome-wide genetic association studies have identified variants associated with COPD (2, 3) and lung function (4–6), yet each identified locus explains only a small amount of the risk for COPD. Variation in the α_1 -antitrypsin (AAT) gene (*SERPINA1*) is a monogenic cause of COPD, but even in the presence of AAT deficiency the development of COPD is unpredictable. COPD fits well in the common disease genetic and epigenetic hypothesis, which promulgates that epigenetic variation may be an important mediator between genetic variation and environmental exposures (7). Specifically, DNA methylation may provide further explanation for features of COPD that are not explained fully by DNA sequence variation, such as the variable susceptibility to develop lung disease in smokers, as well as the continued elevated risk for lung function decline years after smoking cessation (8). Epigenetic studies might shed new insights into the pathogenesis of COPD susceptibility and severity.

Cigarette smoking is the major environmental risk factor for lung function decline and COPD. Cigarette smoking has been demonstrated to impact global hypomethylation of repetitive genomic elements (9–11); previous studies have also indicated that cigarette smoking impacts the methylation patterns of individual genes (12–17), and that some of these genes may be relevant

to nicotine addiction (18). Importantly, COPD is considered to be a systemic disease, and identifying molecular signatures is timely; characterizing methylation signatures in DNA from peripheral blood may have important implications as a biomarker for early diagnosis and prognosis, as well as an assay for the systemic impact of COPD. Arrays have transitioned the assessment of DNA methylation from low to high throughput (19), and the availability of genome-wide arrays has allowed epigenetic investigation in epidemiology studies. To date, no study has performed a comprehensive assessment of DNA methylation marks in the peripheral blood of smokers well phenotyped for nonneoplastic lung disease.

We propose that the investigation of DNA methylation marks in subjects with and without a history of COPD may provide insights into the variable susceptibility to lung function decline/COPD. To pursue this hypothesis we conducted methylation screens, using the HumanMethylation27 array (Illumina, San Diego, CA) for the assay of 27,578 CpG sites in DNA from 1,454 subjects with and without a history of COPD from two distinct family-based cohorts ($n = 1,085$ and 369), to identify CpG sites associated with the presence and severity of COPD. Some of the results have been reported in the form of an abstract (20).

METHODS

Cohorts and Phenotypes

Study subjects were selected from two distinct family-based cohorts: the International COPD Genetics Network (ICGN) and the Boston Early-Onset COPD Study (EOCOPD). The protocols were approved by the local review board at each of the recruitment sites; all subjects provided written informed consent. Lung function phenotypes used in this analysis included the spirometric measures of FEV₁ and the ratio of FEV₁ to the FVC. For our analysis COPD was defined as present if the FEV₁/FVC ratio was less than 0.7 and FEV₁ was less than 70% predicted.

On the basis of enrollment criteria, all subjects in the ICGN had a history of cigarette smoking, whereas EOCOPD subjects included subjects with and without a history of cigarette smoking. Detailed descriptions of the ICGN and EOCOPD cohorts are provided (*see* the online supplement).

DNA Extraction, Bisulfite Treatment, Methylation Array Methods

DNA extraction from white blood cells was performed with a Gentra Puregene blood kit (Qiagen, Inc., Valencia, CA). Bisulfite modification of 1 μ g of each DNA sample was performed with a 96-well EZ DNA Methylation-Gold deep-well kit (Zymo Research, Orange, CA). Less than 2% of DNA samples failed bisulfite conversion.

We used the Infinium HumanMethylation27 BeadChip (Illumina), which allows for the assay of 27,578 CpG sites in 14,475 consensus coding sequences in the NCBI Database (Genome Build 36). The Illumina software provides a designation of CpG island status of each locus, based on the definition of a CpG island of Takai and Jones (21). Per sample, each locus results in an intensity value for methylated (M) and unmethylated (U) alleles. Methylation status was expressed as the β value (β), which represents a ratio of the M-to-U fluorescent signals, such that $\beta = \text{Max}(M,0)/[\text{Max}(M,0) + \text{Max}(U,0) + 100]$; DNA methylation is represented by a variable between 0 (no methylation) and 1 (complete methylation). Further assay details, including quality control, data preprocessing, and assessment for batch effects, are included in the online supplement. For validation of the associated *SERPINA1* CpGs we used DNA from the ICGN cohort for pyrosequencing (*see* the online supplement).

Statistical Methods

We used a test-replication approach to detect CpG sites associated with the presence and severity of COPD, by using the ICGN cohort as our test cohort and the EOCOPD cohort as our replication cohort. For each CpG site in each cohort, we applied generalized linear mixed models (GLMMs) (22) to assess the impact of DNA methylation (predictor) on outcomes (COPD status, FEV₁/FVC, and FEV₁) with the inclusion of a random family effect. As described previously, percent methylation was expressed quantitatively as the β value (β). We examined models without (model I) and with (model II) adjusting for potential confounders: sex, age, current smoking status, cumulative smoking exposure expressed as pack-years of cigarettes, and indicator variables for batch effects. A CpG site was regarded as statistically significant in a test if the corresponding false-discovery rate (FDR)-adjusted P value was less than 0.05. We computed hypergeometric P values for over- or underrepresentation of each gene ontology (GO) category (the biological process [BP] ontology among the GO annotations) for the genes represented by the 349 overlapping CpG sites, using the Bioconductor GOstats package (23).

RESULTS

Characteristics of both cohorts are presented in Table 1. Both cohort samples included only white subjects.

Table E1 (*see* the online supplement) illustrates the detailed preprocessing steps. In the ICGN cohort one sample had detection P values exceeding 1×10^{-5} at more than 25% of CpG loci and was set to missing; no CpG loci had a median detection P value greater than 0.05. In the EOCOPD cohort no sample had detection P values greater than 1×10^{-5} at more than 25% of CpG loci; four CpG loci had a median detection P value greater than 0.05 and were set to missing. Because we were most

TABLE 1. CHARACTERISTICS OF 1,454 WHITE SUBJECTS FROM ICGN AND EOCOPD COHORTS

n	Subjects with COPD	Subjects without COPD*	Age	Pack-years (mean \pm SD)	FEV ₁ /FVC (mean \pm SD)	FEV ₁ (mean \pm SD)	Female (%)	Current Smoker (%)
ICGN								
1,085	620	325	57.3 \pm 8.1	41.7 \pm 26.2	54.4 \pm 19.8	64.2 \pm 32.7	45.6	36.5
n	Subjects with COPD	Subjects without COPD†	Age	Pack-years (mean \pm SD)‡	FEV ₁ /FVC (mean \pm SD)	FEV ₁ (mean \pm SD)	Female (%)	Current Smoker (%)
EOCOPD								
369	181	109	47.5 \pm 7.1	28.7 \pm 23.6	54.3 \pm 21.1	57.9 \pm 35.2	64.0	27.9

Definition of abbreviations: COPD = chronic obstructive pulmonary disease; EOCOPD = Boston Early-Onset COPD Study; ICGN = International COPD Genetics Network.

*One hundred and forty subjects from ICGN cohort with unclassified COPD status.

†Seventy-nine subjects from the EOCOPD cohort with unclassified COPD status.

‡Sixty-eight subjects among the 369 subjects from the EOCOPD cohort having value 0 for pack-years were removed from the calculation of mean pack-years; these subjects represent lifetime never-smokers.

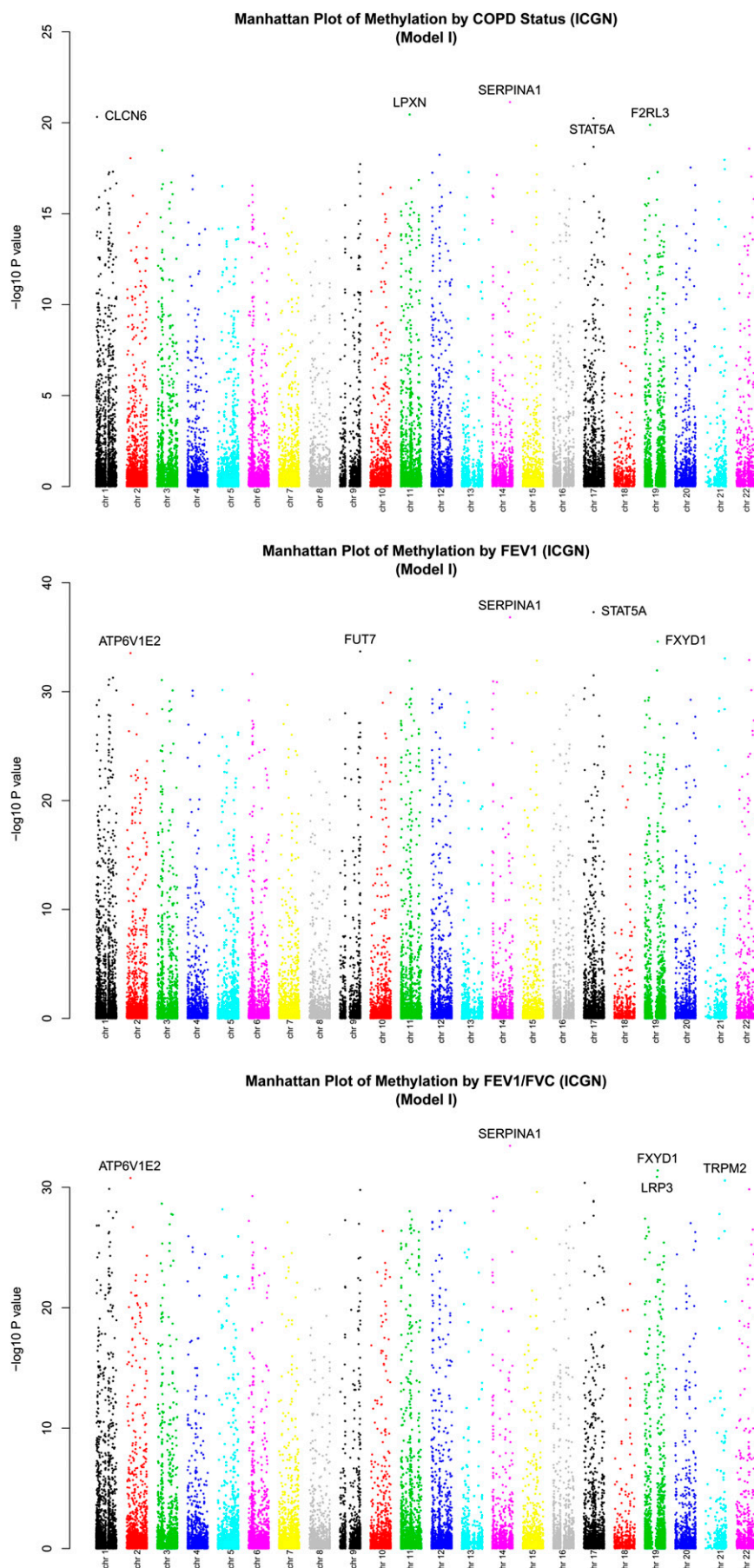


Figure 1. (A–C) Manhattan plots for generalized linear mixed model (GLMM) model I with outcome variables chronic obstructive pulmonary disease (COPD) status, FEV₁, and FEV₁/FVC ratio, respectively, based on International COPD Genetics Network (ICGN) data. The genes represented by the five CpG sites with the smallest *P* values are labeled in the plots. cg02181506 in the *SERPINA1* gene ranked first for COPD and FEV₁/FVC ($P = 7.3 \times 10^{-22}$ and 3.5×10^{-34} , respectively) and second for FEV₁ (1.5×10^{-37}).

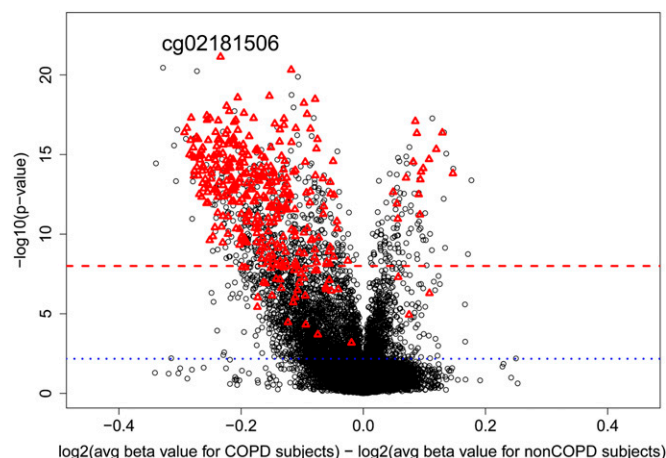


Figure 2. Volcano plot based on model I for International COPD Genetics Network (ICGN) data. The 349 overlapping CpG sites are indicated by red triangles. (Model I is the generalized linear mixed model [GLMM] with outcome variable chronic obstructive pulmonary disease [COPD] [or FEV₁/FVC, or FEV₁], predictor CpG methylation level, and a random family effect.) Red dashed line, $P = 1 \times 10^{-8}$; blue dotted line, false discovery rate-adjusted $P = 0.049939$.

interested in methylation status at autosomal loci, we excluded CpG loci on the sex chromosomes. A total of 26,486 CpG markers passed quality control criteria in the ICGN data set of 1,085 subjects and 26,485 CpG markers passed in the EOCOPD data set of 369 subjects, and these clean data formed the core of our analysis; 26,485 CpGs overlapped between the ICGN and EOCOPD data sets. Each batch was designated by a variable to evaluate for any batch-related technical effects (see METHYLATION ARRAY METHODS in the online supplement).

COPD Status

In the analysis of the test (ICGN) cohort, multiple sites were statistically significant (FDR-adjusted P value < 0.05) based on model I. The results of model I, using the test (ICGN) cohort, are summarized (Manhattan plots; Figure 1). For the association with COPD status in the ICGN cohort, 3,565 CpG sites were statistically significant (FDR-adjusted P value < 0.05) across the genome; the top associations for COPD are presented (Figure 1). A CpG site (cg02181506) in the *SERPINA1* gene on chromosome 14 was top-ranked ($P = 7.3 \times 10^{-22}$) with a plausible fold change (Figure 2) and second-ranked by P value but with a robust fold change in model II ($P = 2.6 \times 10^{-14}$) (Figure E1). (This CpG site was also the top-ranked site associated with the FEV₁/FVC ratio, using model I, and was the second highest ranked association with FEV₁ in both models I and II.) The direction of the effect suggested relatively lower methylation of this CpG promoter site in the presence of COPD. This CpG site was significant at $P < 0.001$ with the same direction of effect for all quantitative models in our EOCOPD replication cohort (Table E2) and for model II for the presence of COPD. The Liptak combined P value for association with the three phenotypes of interest (presence of COPD, FEV₁/FVC, and FEV₁) across both cohorts remained significant (Table E2). Methylation percent for another CpG site (cg24621042) in the *SERPINA1* gene was also statistically significantly associated with all three phenotypes. The combined results for cg02181506 and cg24621042 are summarized (Table E2). In addition to the association in *SERPINA1*, the top associations with COPD in the ICGN test cohort included

associations with CpGs in *L PXN*, *STAT5A*, *CLCN6*, and *F2RL3* (Figure 1). The P value for testing the association of CpG cg02181506 (gene *SERPINA1*) with COPD disease status, based on GLMM for the ICGN cohort, was 7.33×10^{-22} , which is much smaller than the Bonferroni correction cutoff $0.05/26486 = 1.89 \times 10^{-6}$ for genome-wide significance. To verify the magnitude ($\sim 1.0 \times 10^{-22}$) of the observed significance, we applied the extended Wilcoxon rank-sum test (24) that incorporates correlations among subjects in the same family. The Z-score of the extended Wilcoxon rank-sum test was -9.95 with a P value $= 1.28 \times 10^{-23}$, which is close to the GLMM-based P value 7.33×10^{-22} . We also applied the generalized estimating equation with independent and exchangeable correlation structure, respectively. The Z-scores (P values) were -10.89 (1.29×10^{-27}) and -10.09 (6.12×10^{-24}), respectively.

Pyrosequencing Validation of *SERPINA1* Findings

Among the top associations, the CpG in *SERPINA1* (cg02181506) had a robust fold change for COPD (Figure 2 and Figure E1) in the setting of high overall ranking for all three phenotypes (Table E2) and biological plausibility; thus, we focused on CpGs in *SERPINA1* for technical validation. We performed technical validation for the two CpG sites (cg02181506 and cg24621042) in the *SERPINA1* gene by pyrosequencing DNA from the ICGN cohort. A total of 943 DNA samples passed quality control for pyrosequencing. Although the mean and median values for percent methylation from pyrosequencing varied from the Illumina percent methylation, the trend was the same for lower methylation in DNA from subjects with COPD ($P = 4.3 \times 10^{-17}$ for cg02181506 and $P = 9.5 \times 10^{-8}$ for cg24621042) (Figure 3).

Lung Function

Using the same approach as that used for the analysis of COPD status, we also analyzed the association of DNA methylation with the continuous phenotypes for lung function. A total of 4,798 CpG sites for the FEV₁/FVC ratio and 4,899 CpG sites for FEV₁ were statistically significantly associated (FDR-adjusted P value < 0.05), based on model I. These results for model I, using the test (ICGN) cohort, are summarized (Manhattan plots in Figure 1), identifying top associations for CpG sites in *SERPINA1*, *ATP6V1E2*, *FXYD1*, *TRPM2*, and *LRP3* for the FEV₁/FVC ratio and for CpG sites in *SERPINA1*, *ATP6V1E2*, *FXYD1*, *FUT7*, and *STAT5A* for FEV₁.

Overlap of Test Replication Data Sets

Because COPD and low lung function are correlated, we evaluated the intersection of all CpG sites that met an FDR-adjusted P value less than 0.05 for all three phenotypes of interests (presence of COPD, FEV₁/FVC, and FEV₁), based on model I in the ICGN test cohort and a P value threshold of at least 10^{-3} in our replication cohort. This resulted in a subset of 349 CpG sites representing 330 genes that met the stringent P value threshold in both cohorts for all three phenotypes of interest in both cohorts (Figure E2). The direction of association revealed that 95% of CpG sites (330 of 349) were relatively hypomethylated in the presence of lower lung function and COPD. The direction of the associations in the test and replication cohorts was the same for each of the 349 CpG sites. We observed that approximately 70% of these 349 CpG sites are outside of CpG islands. Data for the top five associated sites ranked by combined Liptak P values are presented (Table 2); the P values for all 349 CpG sites are listed (Figure E2 and Tables E3 and E4). In

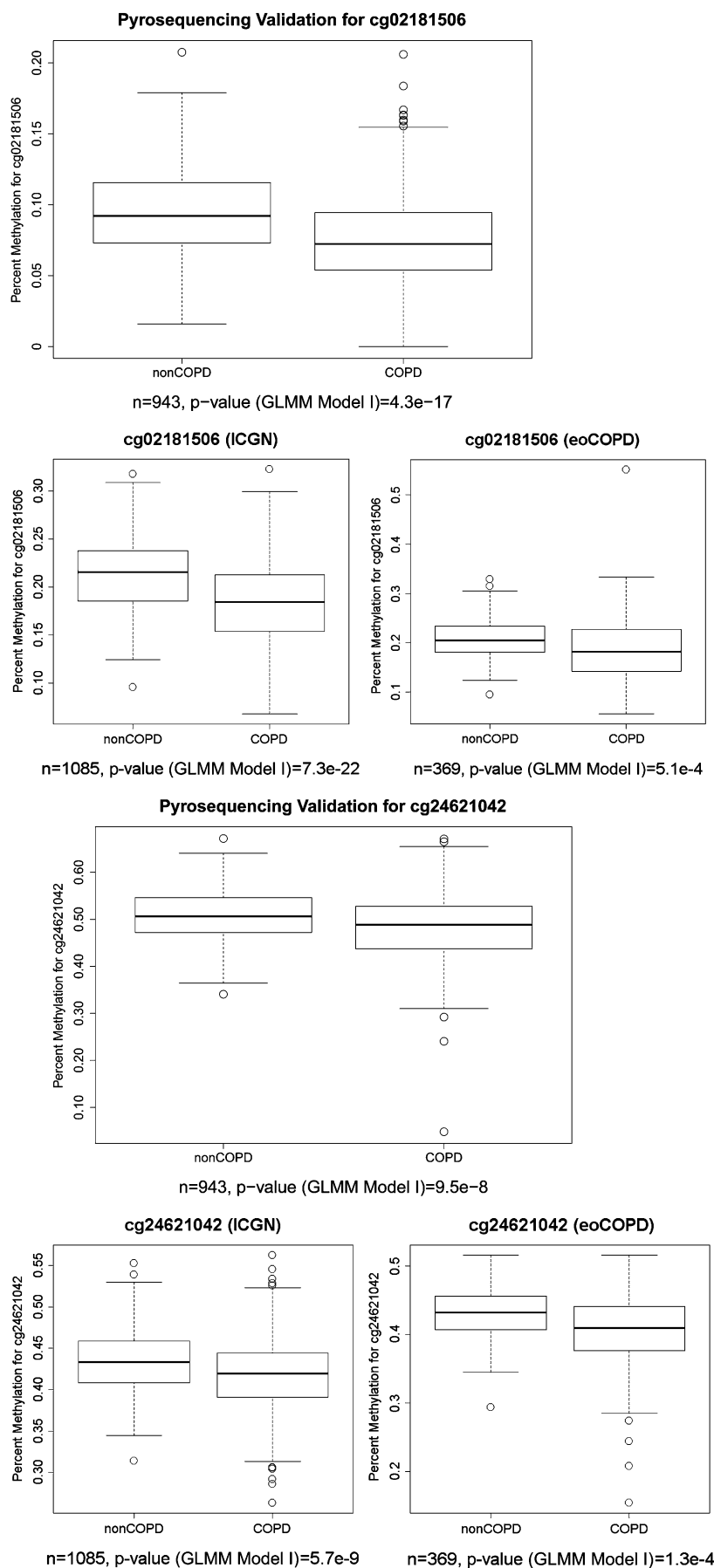


Figure 3. (A–F) Parallel box plots of percent DNA methylation obtained from pyrosequencing followed by Illumina array percent methylation for chronic obstructive pulmonary disease (COPD) status for *SERPINA1* cg02181506 and cg24621042. GLMM = generalized linear mixed model.

a univariate model of smoking intensity as the predictor of methylation percent in our data set, only 40 of the 349 top associated CpGs demonstrated a statistically significant association in both the test and replication cohorts (Table E5).

Gene Ontology

We computed hypergeometric *P* values for over- or underrepresentation of each gene ontology (GO) category in the biological process (BP) ontology among the GO annotations for the genes represented by the 349 overlapping CpG sites, using the Bioconductor GStats package. The computations were done conditionally on the basis of the structure of the GO graph. There were 30 overrepresented GO BP categories having a *P* value less than 0.001, which included immune and inflammatory system pathways, responses to stress and external stimuli, as well as wound-healing and coagulation cascades (Figure 4).

DISCUSSION

This study represents the largest study of methylation profiles from peripheral blood DNA in family-based cohorts of subjects with a history of smoking and well phenotyped for noncancerous lung disease. We have shown that DNA methylation status at distinct CpG loci is associated with both the presence and severity of COPD as measured by spirometry in DNA from 1,454 subjects, using a test-replication approach. Studies of DNA methylation in samples from family cohorts allow some control for shared genetics while exploring other potential biological explanations. None of the subjects in our study were homozygous recessive for AAT deficiency, a mendelian recessive form of COPD. It was unexpected that one of the highest ranked methylation marks associated with the presence of COPD was for a CpG site in the *SERPINA1* gene. Although at present we do not know whether changes at this particular site (cg02181506) in the gene result in functional variation, AAT is an acute-phase reactant and hypomethylation of the AAT gene has been associated with increased gene expression in rat models (25). The identification of this as a top-ranked association, using an unbiased approach, is significant, as this gene encodes the most potent circulating antielastase active in the lung; whether relative hypomethylation in the human AAT gene similarly results in increased gene expression requires further investigation.

COPD is considered by many to be a disease with systemic impact and gene ontology assessment identified a cadre of pathways that are biologically plausible. Identification of associated methylation variations in DNA from white blood cells has been successfully performed for other nonneoplastic processes (reviewed by Terry and colleagues [26]) and supports the importance of assaying blood as an additional biomarker of COPD. The subset of 349 methylation marks provides a host of new genes to investigate for both COPD susceptibility and severity. Further evidence that these marks may predict the development of COPD and/or regulate its severity requires an assessment of longitudinal changes in lung function in association with methylation marks and smoking, together with functional assessment of the impact of DNA methylation on gene expression. We observed that approximately 70% of the overlapping 349 CpG sites are outside of CpG islands, which is a pattern that has also been identified in other complex diseases (27, 28). In addition to the *SERPINA1* association, one of the top five associated CpG sites for all three phenotypes in both cohorts was with a CpG in the fucosyltransferase-7 (*FUT7*) gene (OMIM 602030), which was relatively hypomethylated in the presence of COPD and lower lung function. *FUT7* is an interesting candidate for lung disease as it encodes sialyl Lewis X, a receptor component that promotes leukocyte migration to inflamed tissue (29). In the case of COPD, neutrophilic inflammation is a prominent feature, and neutrophil adhesion molecule expression has been observed to be altered in subjects with COPD (30, 31). Sialyl Lewis X does serve as a ligand for E-selectin, and E-selectin has been observed to have increased expression in subjects with chronic bronchitis and airflow obstruction (32). Further work is needed to understand whether relative hypomethylation of *FUT7* alters peripheral expression of sialyl Lewis X, thus assisting in neutrophil migration to the lung—if so, this would be another provocative and biologically plausible pathway uncovered.

Methylation profiling in white blood cell source DNA may provide important insights into the systemic impact of COPD and smoking, and may provide a relevant role for the identification of smokers at high risk to develop nonneoplastic lung disease. Smoking has been noted to impact the methylation marks in the blood of specific genes in past and current smokers. To date, studies have examined one to a few genes at a time in

TABLE 2. COMBINED RESULTS FOR ICGN AND EOCOPD COHORTS FOR TOP FIVE ASSOCIATED CpG SITES FOR CHRONIC OBSTRUCTIVE PULMONARY DISEASE, FEV₁/FVC, AND FEV₁ FOR MODEL 1*

Phenotype	Gene	CpG	ICGN		EOCOPD		Liptak <i>P</i> Value
			Statistic	<i>P</i> Value	Statistic	<i>P</i> Value	
COPD	<i>C3orf18</i>	cg23320649	−9.336748558	3.31×10^{-19}	−5.030909969	1.36×10^{-06}	5.00×10^{-24}
	<i>SERPINA1</i>	cg02181506	−10.07843765	7.33×10^{-22}	−3.551771539	0.000509	1.47×10^{-23}
	<i>CLCN6</i>	cg05228408	−9.85450366	4.80×10^{-21}	−3.605002098	0.000422	6.33×10^{-23}
	<i>CBFA2T3</i>	cg13745346	−9.085306768	2.47×10^{-18}	−4.650616791	7.11×10^{-06}	1.94×10^{-22}
	<i>FUT7</i>	cg02679745	−9.1177428	1.90×10^{-18}	−4.278923879	3.30×10^{-05}	7.82×10^{-22}
FEV ₁	<i>FXD1</i>	cg27461196	13.20562673	2.49×10^{-35}	5.908397863	1.23×10^{-08}	7.43×10^{-42}
	<i>FUT7</i>	cg02679745	13.00255827	2.01×10^{-34}	5.843217447	1.73×10^{-08}	8.11×10^{-41}
	<i>SERPINA1</i>	cg02181506	13.69019842	1.51×10^{-37}	4.696318512	4.55×10^{-06}	1.07×10^{-40}
	<i>TRPM2</i>	cg06812844	12.85620513	9.07×10^{-34}	5.575724122	6.86×10^{-08}	1.67×10^{-39}
	<i>FERMT3</i>	cg14654385	12.00868971	4.59×10^{-30}	6.689597668	1.67×10^{-10}	1.08×10^{-38}
FEV ₁ /FVC ratio	<i>SERPINA1</i>	cg02181506	12.94839154	3.52×10^{-34}	4.026095358	7.69×10^{-05}	7.21×10^{-36}
	<i>FXD1</i>	cg27461196	12.48502033	3.99×10^{-32}	4.623399599	6.28×10^{-06}	1.56×10^{-35}
	<i>FUT7</i>	cg02679745	12.11231648	1.63×10^{-30}	5.085411115	7.58×10^{-07}	3.48×10^{-35}
	<i>TRPM2</i>	cg06812844	12.29556593	2.64×10^{-31}	4.731029483	3.89×10^{-06}	4.82×10^{-35}
	<i>LRP3</i>	cg08700306	12.3633689	1.34×10^{-31}	4.171614996	4.28×10^{-05}	6.53×10^{-34}

Definition of abbreviations: COPD = chronic obstructive pulmonary disease; EOCOPD = Boston Early-Onset COPD Study; ICGN = International COPD Genetics Network.

*Generalized linear mixed model (GLMM) with outcome variable COPD (or FEV₁/FVC, or FEV₁), predictor CpG methylation level, and random family effect.

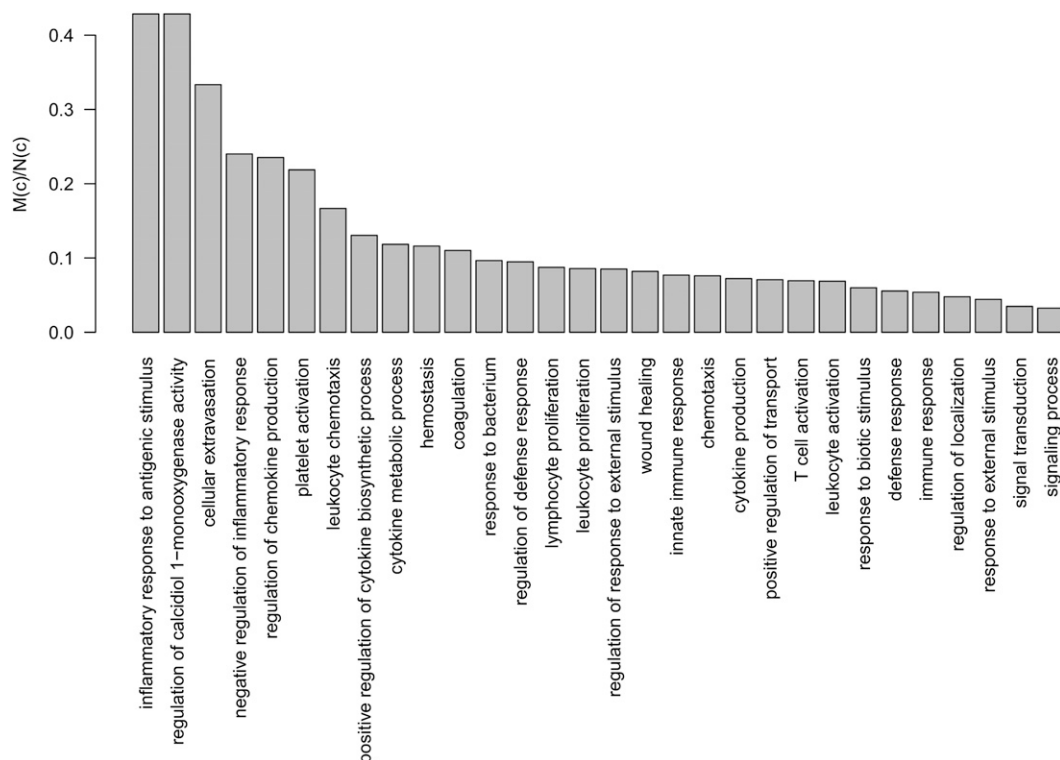


Figure 4. Gene ontology analysis identified 30 overrepresented categories. Plot of M_c/N_c versus category c of the top 30 overrepresented GO BP categories, where N_c is the number of genes annotated to category c and M_c is the number of genes among the 323 genes represented by the 349 overlapping CpG sites, which were annotated to category c . BP = biological process; GO = gene ontology.

small cohorts of subjects with a history of cigarette smoking and without cancer. For example, methylation of p16 has been observed in sputum, bronchial epithelium, and bronchial alveolar lavage specimens from cigarette smokers from multiple studies (13–17). Interestingly, in a univariate model of smoking intensity as the predictor of methylation percent in our data set, only 40 of the 349 top associated CpGs demonstrated a statistically significant association in both cohorts, potentially suggesting that an effect of DNA methylation in the majority of associated sites is not due solely to an effect of cigarette smoking. Using the same platform to investigate DNA methylation across the genome, a CpG site in *F2RL3* was implicated in association with smoking status (12); although this CpG site is one of our top associated sites with COPD and lung function in the ICGN cohort, it did not replicate in the EOCOPD cohort and thus is not included in our subset of 349 top associated methylation marks.

COPD is considered to be a disease of aging, and potentially a disease of accelerated aging—both systemically and of the lung specifically. Low global DNA methylation was measured in aged mammalian tissues in a study published decades ago (33). More recently, aging has been associated with hypomethylation of specific methylation marks in CpG sites in DNA from blood (34, 35). The associations we observed with CpG hypomethylation may be due in part to an overall accelerated impact of aging-related processes. In addition, DNA methylation profiling in blood may provide important insights into the systemic inflammatory impact of COPD and smoking (36), and may reveal a methyl biomarker for the identification of smokers at high risk to develop lung disease and/or as a biomarker of disease activity—both timely endeavors in understanding pathogenesis of this devastating lung disease (37). COPD has been associated with elevation in markers of systemic inflammation (38). Variable global DNA methylation has been linked with systemic inflammatory markers assayed in the peripheral blood, such as that of long interspersed nucleotide element (LINE)-1 hypomethylation associated with higher vascular cell adhesion

molecule-1. Cigarette smoking (10) and aging (39) have been associated with global DNA hypomethylation. Although one cannot assume that the average single-site methylation across the genome recapitulates what is occurring in Alu and LINE elements, it is relevant to note that the trend in the most highly associated CpG sites is for relative hypomethylation in the setting of COPD and lower lung function; the same trend we have reported for Alu and LINE elements in a longitudinal study of lung function (40). One explanation for this trend could be due to the down-regulation of DNA methyltransferase 1 by nicotine over time (41). COPD, a nonneoplastic disease, has been observed to co-occur with lung cancer (42, 43). One explanation for this could be that COPD and lung cancer represent an epigenetic continuum, characterized by an accumulation of methylation marks over time. It is interesting to note that some of the 349 CpG sites that we identified as most highly associated with COPD in our two cohorts have been implicated in some capacity as either oncogenes or tumor suppressor genes. Our findings suggest that large-scale methylation studies may facilitate the uncovering of details regarding an epigenetic continuum from COPD to lung cancer, an insight of high public health importance.

There are several novel features of our present analysis. First, this analysis represents the first study of genome-wide methylation marks in the DNA from subjects well phenotyped for COPD. Second, we have used a test-replication approach in two distinct family-based data sets. An important point about our replication cohort is that the probands in this cohort were ascertained on the basis of the most severe disease at a young age—thus replicated CpG associations may be more representative of those genes that impact COPD independent of age-related changes in the epigenome. However, there are several important limitations to address. Any study of complex trait epigenetics must address these challenges, which include but are not limited to (1) inherent phenotypic heterogeneity, (2) need for replication of findings, (3) influence of population substructure, (4) optimal inclusion of covariates and adjustment for

ascertainment strategies, (5) generalizability, (6) target tissue (lung) availability, (7) functional relevance, and (8) sequence context dependence. Our study has included subjects phenotyped for COPD by spirometry—although computed tomography scans are not available to address contributions to COPD from underlying emphysema, the replication of our top associations with careful lung function phenotyping addresses the first two issues. We have focused our initial study in 1,454 white subjects, limiting the impact of population substructure—similar studies are underway in a cohort of African American subjects to assess generalizability. We chose to highlight the unadjusted model to avoid overadjustment for age and sex (included in the FEV₁ percent predicted and thus the definition of COPD). In addition, our large test cohort was ascertained on the basis of smoking for at least 5 pack-years; best adjustments for ascertainment in large-scale complex trait epigenetic studies are in evolution. For comparison, we have included the top 100 associations for the covariate-adjusted models (Tables E6 and E7). Specific to our study, the outcome of interest is COPD, so one argument is that the target organ tissue should include sampling of lung tissue. We fully acknowledge the limitation of linking the peripheral blood methylome as a surrogate of the lung methylome; however, for those of us who are interested in COPD as a systemic disease the white blood cell source DNA may represent a biomarker of the aggregate inflammatory impact of advanced lung disease and smoking. Supporting this is a small study by Russo and colleagues (44), who observed concordance between the promoter methylation status of six genes in matched blood and bronchial epithelial samples and suggested that DNA from peripheral lymphocytes may be a surrogate for bronchial epithelial tissue in smokers. We should note that our data are not adjusted for cell count differential, which may impact the magnitude of the observed effect. Issues of reverse causality are also crucial to consider. In our model, we were interested in DNA methylation as a predictor of lung function and COPD, but it is just as imperative to consider the impact of lung function and COPD on DNA methylation. Longitudinal studies of change in lung function and change in DNA methylation status over time would be informative. Last, the array used in this study has multiple probes that are potentially impacted by single-nucleotide polymorphism and repeat regions. The locations of these variants range from the CpG site itself to anywhere along the 50–base pair probe. Although many analyses exclude these *a priori*, we chose to analyze all content and have provided an annotation of the probes for our top associations (Tables E8 and E9). Importantly, methylation associations should be validated on a second platform; as such, we pyrosequenced the *SERPINA1* association—the single-nucleotide polymorphism 23 base pairs away from the assayed CpG, using the Illumina probe, did not impact our findings.

This study represents the first genome-wide methylation analysis in the DNA from subjects carefully phenotyped for non-neoplastic obstructive lung disease. Furthermore, we have used a test-replication approach in two distinct family-based data sets, followed by validation. The demonstration of differential methylation in the blood by COPD affection status and in association with lower lung function provides a proof of concept that interrogation of high-throughput arrays can result in the identification of methyl biomarkers of complex human disease. The design of the array used in this study is limited, and our findings support the high importance of next-generation sequencing approaches for the most comprehensive interrogation of the human methylome. Although our current study is observational, gene ontology assessment revealed plausible genes and pathways, using this hypothesis free test-replication strategy. These pathways (including immune and inflammatory system pathways, responses to stress and

external stimuli, and wound-healing and coagulation cascades) are all provocative for further study. Whether these findings represent simply statistical associations or true biological effects remains unknown and supports the importance of integrated genomics and functional genetics to further characterize epigenetic biomarkers of COPD.

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

References

1. Miniño AM, Xu J, Kochanek KD. Deaths: preliminary data for 2008. *Natl Vital Stat Rep* 2010;59:17.
2. Cho MH, Boutaoui N, Klanderman BJ, Sylvia JS, Ziniti JP, Hersh CP, DeMeo DL, Hunninghake GM, Litonjua AA, Sparrow D, *et al*. Variants in *FAM13A* are associated with chronic obstructive pulmonary disease. *Nat Genet* 2010;42:200–202.
3. Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, Feng S, Hersh CP, Bakke P, Gulsvik A, *et al*. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet* 2009;5:e1000421.
4. Wilk JB, Chen TH, Gottlieb DJ, Walter RE, Nagle MW, Brandler BJ, Myers RH, Borecki IB, Silverman EK, Weiss ST, *et al*. A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS Genet* 2009;5:e1000429.
5. Hancock DB, Eijgelsheim M, Wilk JB, Gharib SA, Loehr LR, Marcianti KD, Franceschini N, van Durme YM, Chen TH, Barr RG, *et al*. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet* 2010;42:45–52.
6. Repapi E, Sayers I, Wain LV, Burton PR, Johnson T, Obeidat M, Zhao JH, Ramasamy A, Zhai G, Vitart V, *et al*. Genome-wide association study identifies five loci associated with lung function. *Nat Genet* 2010; 42:36–44.
7. Björnsson HT, Fallin MD, Feinberg AP. An integrated epigenetic and genetic approach to common human disease. *Trends Genet* 2004;20: 350–358.
8. Hogg JC. Why does airway inflammation persist after the smoking stops? *Thorax* 2006;61:96–97.
9. Hsiung DT, Marsit CJ, Houseman EA, Eddy K, Furniss CS, McClean MD, Kelsey KT. Global DNA methylation level in whole blood as a biomarker in head and neck squamous cell carcinoma. *Cancer Epidemiol Biomarkers Prev* 2007;16:108–114.
10. Smith IM, Mydlarz WK, Mithani SK, Califano JA. DNA global hypomethylation in squamous cell head and neck cancer associated with smoking, alcohol consumption and stage. *Int J Cancer* 2007;121:1724–1728.
11. Hillemacher T, Frieling H, Moskau S, Muschler MA, Semmler A, Kornhuber J, Klockgether T, Bleich S, Linnebank M. Global DNA methylation is influenced by smoking behaviour. *Eur Neuropsychopharmacol* 2008;18:295–298.
12. Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am J Hum Genet* 2011;88:450–457.
13. Kersting M, Friedl C, Kraus A, Behn M, Pankow W, Schuermann M. Differential frequencies of p16^{INK4a} promoter hypermethylation, p53 mutation, and K-ras mutation in exfoliative material mark the development of lung cancer in symptomatic chronic smokers. *J Clin Oncol* 2000;18:3221–3229.
14. Belinsky SA, Palmisano WA, Gilliland FD, Crooks LA, Divine KK, Winters SA, Grimes MJ, Harms HJ, Tellez CS, Smith TM, *et al*. Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers. *Cancer Res* 2002;62:2370–2377.
15. Soria JC, Rodriguez M, Liu DD, Lee JJ, Hong WK, Mao L. Aberrant promoter methylation of multiple genes in bronchial brush samples from former cigarette smokers. *Cancer Res* 2002;62:351–355.
16. Zochbauer-Muller S, Lam S, Toyooka S, Virmani AK, Toyooka KO, Seidl S, Minna JD, Gazdar AF. Aberrant methylation of multiple genes in the upper aerodigestive tract epithelium of heavy smokers. *Int J Cancer* 2003;107:612–616.
17. Destro A, Bianchi P, Alloisio M, Laghi L, Di Gioia S, Malesci A, Cariboni U, Gribaudo G, Bulfamante G, Marchetti A, *et al*. K-ras and p16^{INK4A}

- alterations in sputum of NSCLC patients and in heavy asymptomatic chronic smokers. *Lung Cancer* 2004;44:23–32.
18. Launay JM, Del Pino M, Chironi G, Callebort J, Peoc'h K, Megnier JL, Mallet J, Simon A, Rendu F. Smoking induces long-lasting effects through a monoamine-oxidase epigenetic regulation. *PLoS ONE* 2009;4:e7959.
 19. Laird PW. Principles and challenges of genome-wide DNA methylation analysis. *Nat Rev Genet* 2010;11:191–203.
 20. DeMeo DL, Boutaoui N, Klanderman BJ, Baccarelli A, Silverman EK. Variable DNA methylation patterns may identify smokers susceptible to develop COPD. *Am J Respir Crit Care Med* 2009;179:A3981.
 21. Takai D, Jones PA. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci USA* 2002;99:3740–3745.
 22. Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. *JASA* 1993;88:9–25.
 23. Falcon S, Gentleman R. Using GOSTats to test gene lists for GO term association. *Bioinformatics* 2007;23:257–258.
 24. Rosner B, Glynn RJ, Lee ML. Extension of the rank sum test for clustered data: two-group comparisons with group membership defined at the subunit level. *Biometrics* 2006;62:1251–1259.
 25. Barton DE, Francke U. Activation of human α_1 -antitrypsin genes in rat hepatoma \times human fibroblast hybrid cell lines is correlated with demethylation. *Somat Cell Mol Genet* 1987;13:635–644.
 26. Terry MB, Delgado-Cruzata L, Vin-Raviv N, Wu HC, Santella RM. DNA methylation in white blood cells: association with risk factors in epidemiologic studies. *Epigenetics* 2011;6:828–837.
 27. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M, *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* 2009;41:178–186.
 28. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Gayther SA, Apostolidou S, Jones A, Lechner M, Beck S, Jacobs IJ, *et al.* An epigenetic signature in peripheral blood predicts active ovarian cancer. *PLoS ONE* 2009;4:e8274.
 29. Miyashiro M, Furuya S, Fujishige K, Sugita T. Highly sensitive cell-based assay system to monitor the sialyl Lewis X biosynthesis mediated by α_{1-3} fucosyltransferase-VII. *Biochem Biophys Res Commun* 2004;324:98–107.
 30. Noguera A, Busquets X, Sauleda J, Villaverde JM, MacNee W, Agusti AG. Expression of adhesion molecules and G proteins in circulating neutrophils in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 1998;158:1664–1668.
 31. Noguera A, Batle S, Miralles C, Iglesias J, Busquets X, MacNee W, Agusti AG. Enhanced neutrophil response in chronic obstructive pulmonary disease. *Thorax* 2001;56:432–437.
 32. Di Stefano A, Maestrelli P, Roggeri A, Turato G, Calabro S, Potena A, Mapp CE, Ciaccia A, Covacev L, Fabbri LM, *et al.* Upregulation of adhesion molecules in the bronchial mucosa of subjects with chronic obstructive bronchitis. *Am J Respir Crit Care Med* 1994;149:803–810.
 33. Wilson VL, Jones PA. DNA methylation decreases in aging but not in immortal cells. *Science* 1983;220:1055–1057.
 34. Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, Nelson HH, Karagas MR, Padbury JF, Bueno R, *et al.* Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet* 2009;5:e1000602.
 35. Bjornsson HT, Sigurdsson MI, Fallin MD, Irizarry RA, Aspelund T, Cui H, Yu W, Rongione MA, Ekstrom TJ, Harris TB, *et al.* Intra-individual change over time in DNA methylation with familial clustering. *JAMA* 2008;299:2877–2883.
 36. Agusti A. Systemic effects of chronic obstructive pulmonary disease: what we know and what we don't know (but should). *Proc Am Thorac Soc* 2007;4:522–525.
 37. Vestbo J, Rennard S. Chronic obstructive pulmonary disease biomarker (s) for disease activity needed—urgently. *Am J Respir Crit Care Med* 2010;182:863–864.
 38. Lomas DA, Silverman EK, Edwards LD, Miller BE, Coxson HO, Tal-Singer R. Evaluation of serum CC-16 as a biomarker for COPD in the eclipse cohort. *Thorax* 2008;63:1058–1063.
 39. Bollati V, Schwartz J, Wright R, Litonjua A, Tarantini L, Suh H, Sparrow D, Vokonas P, Baccarelli A. Decline in genomic DNA methylation through aging in a cohort of elderly subjects. *Mech Ageing Dev* 2009;130:234–239.
 40. Lange N, Sordillo JE, Tarantini L, Bollati V, Sparrow D, Vokonas P, Zanolletti A, Schwartz J, Baccarelli A, Litonjua A, *et al.* Global DNA methylation and lung function in the Normative Aging Study [abstract]. *Am J Respir Crit Care Med* 2011;183:A5694.
 41. Satta R, Maloku E, Zhubi A, Pibiri F, Hajos M, Costa E, Guidotti A. Nicotine decreases DNA methyltransferase 1 expression and glutamic acid decarboxylase 67 promoter methylation in GABAergic interneurons. *Proc Natl Acad Sci USA* 2008;105:16356–16361.
 42. Mannino DM, Aguayo SM, Petty TL, Redd SC. Low lung function and incident lung cancer in the United States: data from the first National Health and Nutrition Examination Survey follow-up. *Arch Intern Med* 2003;163:1475–1480.
 43. Tockman MS, Anthonisen NR, Wright EC, Donithan MG. Airways obstruction and the risk for lung cancer. *Ann Intern Med* 1987;106:512–518.
 44. Russo AL, Thiagalingam A, Pan H, Califano J, Cheng KH, Ponte JF, Chinnappan D, Nemani P, Sidransky D, Thiagalingam S. Differential DNA hypermethylation of critical genes mediates the stage-specific tobacco smoke-induced neoplastic progression of lung cancer. *Clin Cancer Res* 2005;11:2466–2470.