

Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome

Emily S. Wan^{1,*}, Weiliang Qiu¹, Andrea Baccarelli², Vincent J. Carey¹, Helene Bacherman¹, Stephen I. Rennard³, Alvar Agusti⁴, Wayne Anderson⁵, David A. Lomas⁶ and Dawn L. DeMeo¹

¹Channing Laboratory and the Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA, USA, ²Exposure Epidemiology and Risk Program, Harvard School of Public Health, Boston, MA, USA, ³Section of Pulmonary and Critical Care, University of Nebraska Medical Center, Omaha, NE, USA, ⁴Thorax Institute, Hospital Clinic, IDIBAPS, Barcelona and CIBERES, Fundacio Caubet-Cimera, Palma de Mallorca, Spain, ⁵GlaxoSmithKline Research and Development, Research Triangle Park, NC, USA and ⁶Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK

Received November 16, 2011; Revised March 22, 2012; Accepted April 2, 2012

The impact of cigarette smoking can persist for extended periods following smoking cessation and may involve epigenetic reprogramming. Changes in DNA methylation associated with smoking may help to identify molecular pathways that contribute to the latency between exposure and disease onset. Cross-sectional cohort data from subjects in the International COPD Genetics Network ($n = 1085$) and the Boston Early-Onset COPD study ($n = 369$) were analyzed as the discovery and replication cohorts, respectively. Genome-wide methylation data on 27 578 CpG sites in 14 475 genes were obtained on DNA from peripheral blood leukocytes using the Illumina HumanMethylation27K Beadchip in both cohorts. We identified 15 sites significantly associated with current smoking, 2 sites associated with cumulative smoke exposure, and, within the subset of former smokers, 3 sites associated with time since quitting cigarettes. Two loci, factor II receptor-like 3 (*F2RL3*) and G-protein-coupled receptor 15 (*GPR15*), were significantly associated in all three analyses and were validated by pyrosequencing. These findings (i) identify a novel locus (*GPR15*) associated with cigarette smoking and (ii) suggest the existence of dynamic, site-specific methylation changes in response to smoking which may contribute to the extended risks associated with cigarette smoking that persist after cessation.

INTRODUCTION

Tobacco smoking is a leading cause of death and disability worldwide. Smoking attributable deaths are projected to increase over the next two decades and will account for 10% of all deaths globally by 2030 (1). The impact of smoking on cardiovascular (2–4), respiratory (5–9) and malignant (10–12) diseases can persist for extended periods after smoking cessation and may involve epigenetic reprogramming. Identifying molecular pathways that contribute to the delay between exposure and disease may offer opportunities for targeted diagnostic and therapeutic interventions.

In mammals, DNA methylation occurs at the 5' carbon of cytosine residues in the context of CpG dinucleotides (13) (e.g. a 'CpG site') and is a highly plastic, tissue-specific phenomenon which can reflect exposure to exogenous stimuli such as smoking (14). Surprisingly, the number of studies examining the effects of smoking on DNA methylation in blood has been limited. Although relative global hypomethylation has been associated with prenatal exposure to tobacco smoke (15–17), the majority of studies have *not* demonstrated a significant impact of adult smoking on global DNA methylation content in blood (18–22) (whether assessed directly or through surrogates such as LINE-1 or Alu measurements).

*To whom correspondence should be addressed at: 181 Longwood Avenue, Boston, MA 02115, USA. Tel: +1 6175252112; Fax: +1 6175250958; Email: reesw@channing.harvard.edu

Limited numbers of studies have investigated differential methylation in response to cigarette smoking in candidate gene promoters in peripheral blood leukocytes; these studies have been largely restricted to malignancy-related genes and have enrolled small numbers of subjects (23–26). In contrast, Breitling *et al.* (14) recently reported the results of a genome-wide analysis in peripheral blood leukocytes using a discovery cohort of 177 subjects followed by replication in a second cohort ($n = 316$); they identified a previously unsuspected CpG site located within the coagulation factor II receptor-like 3 (*F2RL3*) locus as being strongly associated with current smoking.

We hypothesized that a number of smoking metrics will influence DNA methylation patterns. To test this hypothesis, we sought to identify CpG sites that demonstrate differential methylation with respect to the current smoking status, smoking exposure (pack-years) and time since quitting in a large, well-characterized cohort of former and current smoking subjects from the International COPD Genetics Network (ICGN) (27,28) ($n = 1085$) followed by replication in the Boston Severe Early-Onset COPD study (EOCOPD) ($n = 369$) (29). In addition to independently validating the previously reported association of the *F2RL3* locus with current smoking, we identify a novel association at the G-protein-coupled receptor 15 (*GPR15*) locus. We extend our analyses to investigate associations with cumulative smoking exposure (pack-years) and present evidence that suggests the existence of dynamic DNA methylation changes in response to smoking.

RESULTS

The characteristics of the subjects analyzed in the discovery (ICGN) and replication (EOCOPD) cohorts are summarized in Table 1. By design, COPD subjects in the EOCOPD cohort had a lower mean age and FEV₁% predicted than COPD subjects in the ICGN cohort. The proportion of male subjects was greater in the ICGN cohort.

Fifteen CpG sites were significantly associated with current smoking in our cohorts (Table 2). The difference in the mean methylation between current smokers and former and never smokers is illustrated for each cohort. The median and interquartile ranges for each of the 15 sites are listed in Supplementary Material, Table S1. Because our discovery cohort compared current versus former smokers, we repeated our analysis after excluding 68 never smokers from our replication cohort (EOCOPD). Nine sites were significantly associated in this subgroup analysis, eight of which overlap with the primary analysis (Supplementary Material, Table S2). For the two most highly associated sites in both analyses, *F2RL3* and *GPR15*, relative hypomethylation was observed with current smoking. Notably, these two CpG sites were also significantly associated with cumulative exposure to cigarette smoke (Table 3). Increasing cumulative exposure to cigarette smoke is associated with relative hypomethylation at both significant loci (Fig. 1).

Although data in our cohort are cross-sectional, we analyzed the time since quitting cigarettes in former smokers as a surrogate to evaluate the *longitudinal* effects of smoking

Table 1. Cohort characteristics

	ICGN		EOCOPD	
<i>n</i>	1085		369	
Age	57.3 (8.1)		47.5 (7.1)	
Gender				
Female	495 (45.6%)		237 (64.2%)	
Male	590 (54.4%)		132 (35.8%)	
Pack-years	41.7 (26.2)		28.7 (23.6)	
Age started smoking ^a	16.3 (4.1)		16 (3.1)	
Never smokers	0		68 (18.4%)	
Current smokers	396 (36.5%)		103 (27.9%)	
Former smokers	689 (63.5%)		198 (53.7%)	
Time since quitting (years) ^a	12.3 (10.4)		7.7 (7.5)	
Obstructive lung disease ^b	Yes = 657	No = 428	Yes = 233	No = 136
FEV ₁ % predicted	45.4 (17.7)	104.6 (15.8)	38.0 (27.6)	89.7 (16.6)

ICGN, International COPD Genetics Network; EOCOPD, Early-Onset COPD; FEV₁% predicted, forced expiratory volume in the first second expressed as % predicted of normal based on population specific prediction equations.

Data are expressed as mean (SD) or *n* (%).

^aVariable in current or former smokers only (In ICGN, $n = 685$, four subjects missing data).

^bDefined as FEV₁/FVC < 0.7.

on DNA methylation. The significant differentially methylated sites are listed in Table 4 along with the difference in the mean methylation between the highest and lowest quartiles of time since quitting. The three significant loci in the time since quitting analysis had been previously identified in the current smoking analysis; notably, for each site, current smoking was associated with *decreased* methylation, whereas time since smoking cessation was associated with *increased* methylation (Fig. 2). Among former smokers in the ICGN cohort, exploratory analyses revealed a significant interaction between cumulative smoke exposure and the time since quitting under a multiplicative model for *F2RL3* and *GPR15* (Table 5). This is illustrated qualitatively for the *F2RL3* locus in Figure 3; subjects with the highest cumulative smoke exposure and shortest duration of smoking cessation had the lowest mean methylation, whereas subjects with the lowest cumulative smoke exposure and longest duration of cessation had the highest mean methylation.

Because our cohorts were ascertained on the presence or absence of COPD, we repeated our analyses with the addition of a covariate to adjust for obstructive lung disease. Using the same criteria for significance employed in our original analysis, 13 of the 15 sites remained significantly associated with current smoking. Both *F2RL3* and *GPR15* remained significantly associated with cumulative smoke exposure as measured in pack-years. *F2RL3* remained significant in the time since quitting analysis after adjustment for obstructive lung disease.

Pyrosequencing was successfully completed in the ICGN cohort for the *F2RL3* and *GPR15* loci; completion rates were 99.6 and 96.9%, respectively. The distributions of the pyrosequencing values with respect to the current smoking status are illustrated in Supplementary Material, Figure S1. Student's *t*-test comparing DNA methylation by the current smoking status remained highly statistically significant for

Table 2. Significant^a differentially methylated CpG sites by current smoking

CpG	Gene	ICGN Change in mean methylation ^b	FDR, <i>P</i> -value	EOCOPD Change in mean methylation ^b	<i>P</i> -value ^c
cg03636183	<i>F2RL3</i>	-0.08	2.36×10^{-46}	-0.06	6.87×10^{-11}
cg19859270	<i>GPR15</i>	-0.02	1.92×10^{-22}	-0.02	3.18×10^{-7}
cg09837977	<i>LRRN3</i>	-0.02	3.88×10^{-11}	-0.02	5.45×10^{-5}
cg01500140	<i>LIM2</i>	0.01	1.88×10^{-10}	0.01	3.83×10^{-5}
cg13247990	<i>MYLK</i>	0.01	3.24×10^{-7}	0.01	2.42×10^{-4}
cg01988129	<i>ADHFE1</i>	0.01	4.16×10^{-6}	0.01	4.12×10^{-5}
cg16254309	<i>CNTNAP2</i>	-3.36×10^{-3}	1.74×10^{-5}	-0.01	3.48×10^{-4}
cg18881723	<i>SLAMF1</i>	-0.01	1.69×10^{-4}	-0.01	3.56×10^{-4}
cg12044210	<i>APBA2</i>	1.74×10^{-3}	3.47×10^{-4}	0.01	5.51×10^{-4}
cg21917349	<i>APBA2</i>	1.66×10^{-3}	5.49×10^{-4}	0.01	5.73×10^{-5}
cg15691199	<i>CEBPE</i>	-3.87×10^{-3}	2.89×10^{-3}	-3.22×10^{-3}	2.66×10^{-4}
cg24262469	<i>TIPARP</i>	-1.96×10^{-3d}	0.01	0.01	7.27×10^{-5}
cg05445326	<i>TMASF19</i>	0.01	0.01	0.01	7.00×10^{-4}
cg15258980	<i>ARHGAP25</i>	-3.11×10^{-3}	0.02	-4.49×10^{-3}	4.14×10^{-4}
cg10161121	<i>FASLG</i>	-0.01	0.02	-2.81×10^{-3}	7.93×10^{-4}

ICGN, International COPD Genetics Network; EOCOPD, Early-Onset COPD; FDR, false-discovery rate.

^aDefined as sites with an FDR adjusted *P*-value <0.05 in the discovery cohort (ICGN) and a one-sided *P*-value <10⁻³ in the replication cohort (EOCOPD).

^bChange in mean methylation = (mean methylation in current smokers) - (mean methylation in former and never smokers). Methylation values may range from 0 (no methylation) to 1 (fully methylated).

^cOne-sided *P*-value.

^dIn model adjusted for covariates direction of effect is positive.

Table 3. Significant^a differentially methylated CpG sites associated with cumulative smoking exposure (pack-years)

CpG	Gene	ICGN			FDR, <i>P</i> -value	EOCOPD			<i>P</i> -value ^b
		Mean methylation in highest pack-years quartile	Mean methylation in lowest pack-years quartile	Difference		Mean methylation in highest pack-years quartile	Mean methylation in lowest pack-years quartile	Difference	
cg03636183	<i>F2RL3</i>	0.76	0.84	-0.07	1.91×10^{-8}	0.79	0.85	-0.07	2.15×10^{-8}
cg19859270	<i>GPR15</i>	0.89	0.90	-0.01	6.35×10^{-6}	0.85	0.89	-0.04	2.09×10^{-8}

ICGN, International COPD Genetics Network; EOCOPD, Early-Onset COPD; FDR, false-discovery rate.

Methylation values may range from 0 (no methylation) to 1 (fully methylated).

Mean value of pack-years by quartile in ICGN: Quartile 1 = 14.4, Quartile 2 = 30.3, Quartile 3 = 44.4, Quartile 4 = 77.9.

Mean value of pack-years by quartile in EOCOPD: Quartile 1 = 1.3, Quartile 2 = 20.0, Quartile 3 = 34.4, Quartile 4 = 60.2.

^aDefined as sites with an FDR adjusted *P*-value <0.05 in the discovery cohort (ICGN) and a one-sided *P*-value <10⁻³ in the replication cohort (EOCOPD).

^bOne-sided *P*-value.

both *F2RL3* (*P*-value 2.2×10^{-16}) and *GPR15* (*P*-value 1.5×10^{-12}) and again demonstrated relative hypomethylation associated with current cigarette smoking for each locus. Consistent with the array-based data, a significant inverse correlation between DNA methylation and pack-years of smoking ($P_{F2RL3} = 8.54 \times 10^{-9}$ and $P_{GPR15} = 7.68 \times 10^{-3}$) and a significant positive correlation between time since quitting ($P_{F2RL3} = 1.35 \times 10^{-11}$ and $P_{GPR15} = 7.58 \times 10^{-5}$) and DNA methylation were also confirmed.

DISCUSSION

Complex chronic exposures such as cigarette smoking can pose challenges in the analysis and interpretation of DNA methylation data. However, the careful analysis of large data sets with detailed phenotype information, along with rigorous

statistical thresholds for significance, can reveal novel loci and pathways which may be involved in mediating the effects of smoking in the development of disease. Significantly associated sites from our analyses have been annotated to genes involved in a wide range of functions including cell signaling/apoptosis (*FASLG*), neurodevelopment (*LRRN3*, *CNTNAP2*) and alcohol metabolism (*ADHFE1*). Myosin light chain kinase (*MYLK*) has been ascribed putative roles in endothelial (30) and epithelial (31–34) barrier functions; variants in this gene have been associated with acute lung injury (35,36), asthma (37–39) and familial aortic dissections (40). Interestingly, cg09837977 (*LRRN3*), cg12044210 and cg21917349 (*APBA2*), and cg10161121 (*FASLG*) map to loci that have previously been included in a smoking ‘quit-success’ genotype score (41).

F2RL3 and *GPR15*, the two loci with the strongest associations with current smoking, have also been annotated

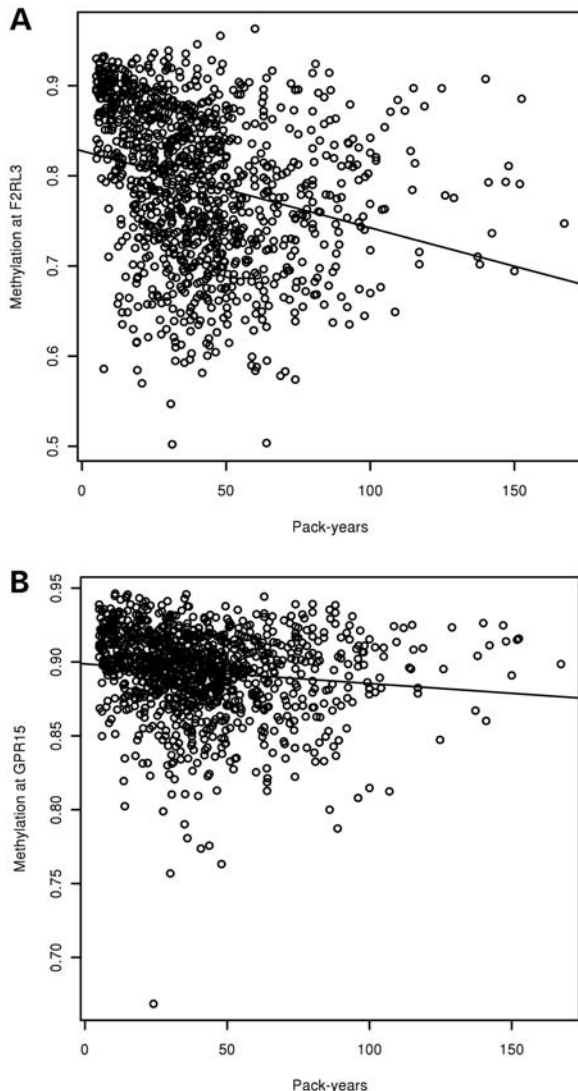


Figure 1. Scatter plots of methylation by pack-years of smoking in the ICGN cohort. X-axis represents cumulative smoke exposure (pack-years), y-axis is the methylation at (A) cg03636183 (*F2RL3*) and (B) cg19859270 (*GPR15*). Methylation values may range from 0 (no methylation) to 1 (complete methylation).

to seemingly disparate functions. *F2RL3*, also known as protease-activated receptor 4, has established roles in platelet activation (42,43) and cell signaling (44); differential methylation at this site with respect to current smoking was first reported by Breitling *et al.* (14). *GPR15* is a protein which shares sequence homology with the angiotensin II AT1 and AT2 receptors (45) and has been established as a co-receptor for the human immunodeficiency virus (46–48). Although there have been no previous statistically significant associations of this locus with smoking, cg19859270 was notably ranked second in the analysis conducted by Breitling *et al.* (14) and demonstrated a similar magnitude and direction of the effect.

We acknowledge that, despite robust statistical significance, the effect estimates of many of the loci reported in this manuscript are modest. Even the effect estimates for *F2RL3* and

GPR15 do not meet the 15% change often cited as a criterion for significance. However, considerable evidence supports the accuracy of our estimates. The differences in the mean methylation between current smokers and non-smokers for *F2RL3* and *GPR15* do not appear to be artifacts of the array-based platform as pyrosequencing produced similar differences in the mean methylation for each site. In addition, the differences in the mean methylation are comparable with those reported in the independent populations studied by Breitling *et al.* (14). We contend, therefore, that large-scale epigenetic studies are a nascent field and that the threshold of change in DNA methylation which correlates with functional significance *on the population level* has not been firmly established.

To examine the possible functional impact of differential methylation at loci reported in this manuscript, we conducted exploratory analyses using publicly available data sets to assess for changes in gene expression by the current smoking status in blood (49–53), histologically normal lung tissue (54) and bronchial epithelium (55) (Supplementary Material, Table S3). Using Student's *t*-test to compare the processed, log-transformed expression values in current smokers versus former/never smokers, several loci demonstrated evidence of nominal association (unadjusted *P*-value < 0.05). Increased gene expression in current smokers was noted for *GPR15* in one of two peripheral blood data sets with transcripts available, whereas *FASLG*, the sole overlapping locus where differential gene expression with respect to the smoking status has previously been reported in blood (50), was found to have decreased expression in smokers in three data sets. Interestingly, increased gene expression in current smokers at *LRRN3*, a locus which has not been previously reported in either methylation or expression studies, was noted in four of five peripheral blood data sets (with the fifth data set potentially limited by small sample size). *FASLG* and *LRRN3* were not significantly associated in either lung or bronchial epithelial tissue while gene expression at *MYLK*, which was reported to demonstrate rapidly reversible changes in bronchial epithelium (55), did not vary by the current smoking status in blood. These differences may be attributable to the tissue specificity of gene expression and DNA methylation patterns.

Although our data are cross-sectional in nature, our analyses suggest that dynamic changes in DNA methylation in response to smoking may exist for *F2RL3* and *GPR15*. The inverse correlation between increasing pack-years and methylation at *F2RL3* and *GPR15* is consistent with progressive demethylation in the setting of higher cumulative smoke exposure. Likewise, the positive correlation between time since quitting and methylation at these loci may be due to remethylation following smoking cessation.

The time course over which DNA methylation changes occur is not known and may vary by CpG site. The concept of differential responses to and recovery from exposure to cigarette smoke was well articulated in a gene expression study conducted in primary bronchial epithelial cells (55); while the majority of gene expression changes due to smoking were found to be rapidly reversible, a subset of loci demonstrated slow or irreversible changes. Studies performed in respiratory epithelial cell cultures exposed to cigarette smoke extract support the concept that distinct loci may be involved in short-term versus chronic exposure to cigarette smoke. In

Table 4. Significant^a differentially methylated sites by time since quitting

CpG	Gene	ICGN Mean methylation in highest time quit quartile	Mean methylation in lowest time quit quartile	Difference	FDR, <i>P</i> -value	EOCOPD Mean methylation in highest time quit quartile	Mean methylation in lowest time quit quartile	Difference	<i>P</i> -value ^b
cg03636183	<i>F2RL3</i>	0.87	0.78	0.09	7.56×10^{-36}	0.85	0.76	0.09	1.60×10^{-6}
cg19859270	<i>GPR15</i>	0.91	0.89	0.02	6.84×10^{-16}	0.89	0.87	0.03	4.24×10^{-4}
cg09837977	<i>LRRN3</i>	0.84	0.82	0.01	3.36×10^{-5}	0.82	0.82	5×10^{-3}	6.69×10^{-4}

ICGN, International COPD Genetics Network; EOCOPD, Early-Onset COPD; FDR, false-discovery rate

Methylation values may range from 0 (no methylation) to 1 (fully methylated).

Mean value of time since quit (in years) by quartiles in ICGN: Quartile 1 = 1.5, Quartile 2 = 6.3, Quartile 3 = 14.2, Quartile 4 = 27.4.

Mean value of time since quit (in years) by quartiles in EOCOPD: Quartile 1 = 1.3, Quartile 2 = 3.5, Quartile 3 = 7.4, Quartile 4 = 18.7.

^aDefined as sites with an FDR adjusted *P*-value <0.05 in the discovery cohort (ICGN) and a one-sided *P*-value <10⁻³ in the replication cohort (EOCOPD).

^bOne-sided *P*-value.

this study, changes in DNA methylation were observable after only 9 months of cigarette smoke exposure (56); none of the differentially methylated sites overlapped with our loci.

To explore whether differential methylation at the loci reported in our study might be involved in rapidly reversible versus slowly reversible-type changes, we conducted additional subgroup analyses in a subset of the public gene expression data sets used in Supplementary Material, Table S3 which had data on former smokers. Although data on time since quitting was not available for former smokers in these cohorts, we hypothesized that genes which were differentially expressed between current and former smokers might represent relatively 'rapidly reversible' loci (Supplementary Material, Table S4). Conversely, we speculated that differential expression between former and never smokers might represent 'slowly reversible' or irreversible changes (Supplementary Material, Table S5). *MYLK*, a locus reported to have 'rapidly reversible' changes in gene expression in bronchial epithelial cells (55), conforms to our hypotheses well; differential expression is seen between current and former smokers, but not between former and never smokers (in bronchial epithelial cells). *LRRN3*, relatively hypomethylated in current smokers (Table 2) and relatively hypermethylated following smoking cessation (Table 4), appears to fit this speculative model of a 'rapidly reversible' locus with differential (increased) expression evident between current and former smokers in peripheral blood, but no differences between former and never smokers. In contrast, differential expression in peripheral blood at *GPR15* and *FASLG* was observed in both the current/former and former/never smoker analyses and suggests that these loci may demonstrate 'slowly reversible' changes. Although intriguing, we acknowledge that only very limited conclusions can be drawn from these unadjusted exploratory analyses conducted in non-contemporaneous data sets and that additional studies are warranted.

We recognize the limitation of association studies as providing only correlative data; we cannot establish whether the changes in methylation reported herein are primary (i.e. established directly by smoking exposure) or secondary to more generalized processes (i.e. established as a consequence of perturbations in other cellular processes). Gene expression (55) and DNA methylation (56) changes may both persist in non-malignant tissues for extended periods following smoking cessation; the mechanism by which these changes

are established and maintained is a topic of active research (57). Changes in methylation marks at specific loci may arise from a change in DNA methyltransferase expression, abundance or activity. Indeed, changes in DNA methyltransferase expression and levels have been reported in neurons (58) and human bronchial epithelial cell cultures (56) exposed to nicotine or cigarette smoke extract. However, we found no compelling evidence for differential expression in the DNA methyltransferases by current smoking status in blood (49–53,55), histologically normal lung tissue (54) or bronchial epithelium (55) based on our exploration of public data sets (Supplementary Material, Table S6). Alternative mechanisms, beyond changes in methyltransferase gene expression, may contribute to the establishment of locus-specific differential methylation observed in our study; additional *in vitro* and *in vivo* studies are needed to explore this question.

Potential limitations of this study include the inability to quantify the degree to which cell type heterogeneity in blood contributes to our results and the reliance upon self-reported smoking metrics. Epidemiological studies have found self-reported smoking to be reliable except in cases where non-smoking is perceived as socially desirable (59–61). Because misclassification due to under-reporting of current smoking would bias our results towards the null, we believe our results are robust even in the absence of biochemical confirmation of current smoking status (i.e. serum cotinine measurements).

Lastly, the frequency and impact of genetic and structural variants on the array-based platforms used to assess DNA methylation in our study is unknown and is an area of active research (62). We performed sequence alignment of the probes of significantly associated sites against a consensus human genome (NCBI37/hg19) and noted a number of single nucleotide polymorphisms, copy number variants and repetitive elements overlapping with our probes (Supplementary Material, Table S7). To assess the impact of these variants on our methylation data, we plotted the distribution of the raw methylation values of current and non-smokers (Supplementary Material, Fig. S2) and the distribution of the raw methylation values against the continuous variable of pack-years of smoking (Supplementary Material, Fig. S3). We saw no evidence of distinct subpopulations indicative of genetic or structural variants among significantly associated sites.

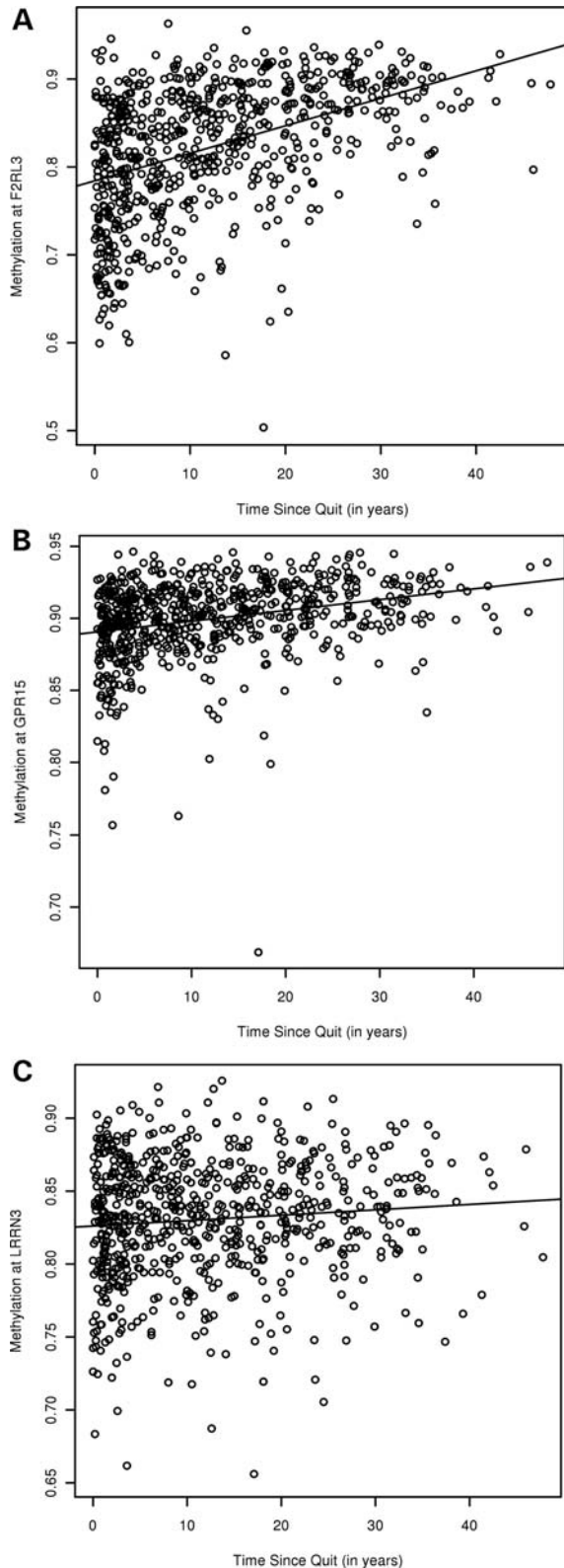


Figure 2. Scatter plots of methylation by time since quitting in the ICGN cohort. X-axis represents time since smoking cessation (in years), y-axis is the methylation at (A) cg03636183 (*F2RL3*), (B) cg19859270 (*GPR15*) and (C) cg09837977 (*LRRN3*). Methylation values may range from 0 (no methylation) to 1 (complete methylation).

Table 5. *P*-values for multiplicative interaction term between pack-years and time since quitting in ICGN

CpG	Gene <i>P</i> -value (unadjusted model)	Interaction term	Interaction term <i>P</i> -value (adjusted model) ^a
cg03636183	<i>F2RL3</i>	7.17×10^{-3}	4.56×10^{-5}
cg19859270	<i>GPR15</i>	0.03	0.01

^aModel adjusted for age, sex and surrogate variables.

Additionally, despite the report of structural variants involving the *GPR15* locus, the association of this locus remained robust when assessed through a second technology (pyrosequencing).

Given the profound physiological impact of smoking as well as considerable epidemiological evidence linking smoking to a number of systemic diseases beyond directly exposed organs, the effects of smoking on peripheral blood DNA methylation appear to impact a limited number of loci in our cohorts. In addition, since gene expression changes were *not* noted in our exploratory analyses at the majority of loci reported in our study, the impact of differential methylation at these loci remains to be established. Thus, the integration of epigenetic data with genetic and genomic lines of investigation will likely yield additional biological insights into clinical challenges such as smoking cessation and smoking-related complex diseases.

MATERIALS AND METHODS

Subjects/cohorts

The discovery cohort consisted of subjects participating in the ICGN (27,28), a family-based study which recruited probands in the age range of 45–65, with ≥ 5 pack-years of cigarette smoking and airflow limitation (defined as an $FEV_1 < 60\%$ predicted and an $FEV_1/VC < 90\%$ predicted). Subjects with alpha-1 antitrypsin genotypes PiZZ, Z-Null, Null-Null or SZ were excluded from the study. In addition, all probands were required to have at least one eligible sibling with ≥ 5 pack-years of cigarette smoking. All subjects completed spirometry, questionnaire data and had whole blood drawn for DNA analysis. Subjects included in the current analysis were of self-reported Caucasian ancestry. The protocol was approved by the local Internal Review Board at each of the recruitment sites and all subjects provided written informed consent. Details regarding the number of samples excluded at each step during data cleaning have been previously described (63) and are summarized in Supplementary Material, Methods; the number of subjects analyzed from this cohort was 1085.

The replication cohort consisted of subjects recruited from the Boston Severe EOCOPD (29), a family-based study which recruited probands who had an $FEV_1 < 40\%$ predicted, by age < 53 in the absence of severe alpha-1 antitrypsin deficiency. Although extended pedigrees were recruited, DNA methylation data were obtained only on probands and their siblings. Subjects completed spirometry, questionnaire data and provided blood for DNA analysis. The protocol was

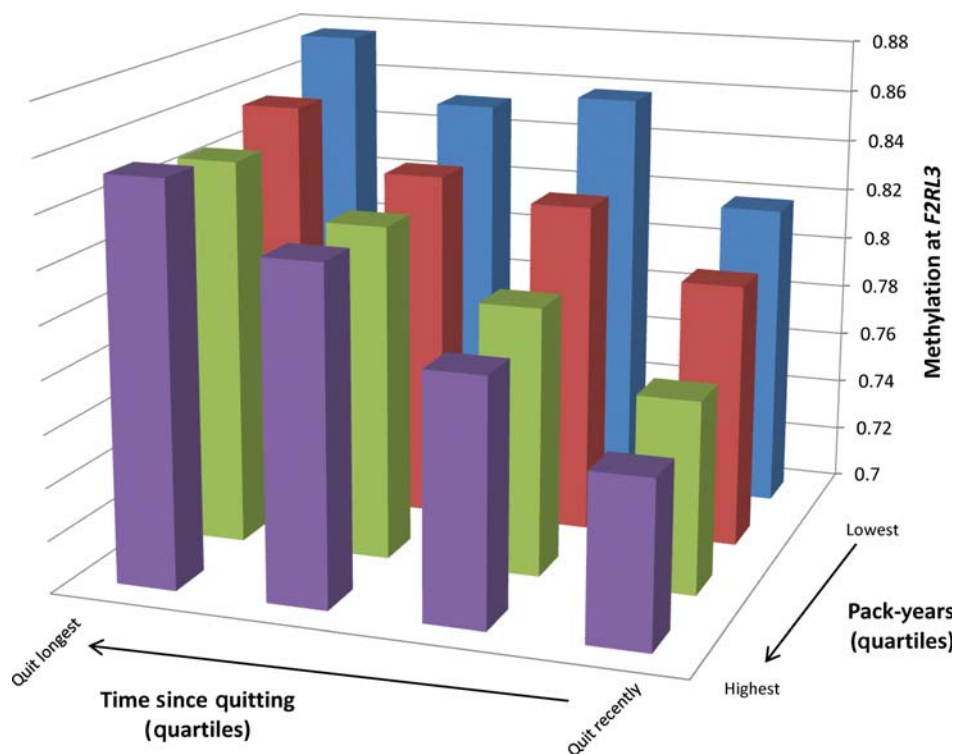


Figure 3. Qualitative illustration of interaction between cumulative smoke exposure and time since smoking cessation at the *F2RL3* locus in the ICGN cohort. Quartiles of cumulative smoke exposure (pack-years) are plotted on the x-axis, quartiles of time since smoking cessation (in years) are plotted on the z-axis, while the mean methylation of each quartile is plotted on the y-axis. Methylation values may range from 0 (no methylation) to 1 (complete methylation). The mean methylation for each cell is as follows (from left to right): Blue columns: 0.88 (lowest pack-years, longest duration since cessation), 0.85, 0.86, 0.82. Red columns: 0.86, 0.83, 0.83, 0.80. Green columns: 0.85, 0.83, 0.80, 0.77. Purple columns: 0.85, 0.83, 0.79, 0.76 (highest pack-years, shortest duration since cessation).

approved by the Brigham and Women's Hospital Human Research Committee; all subjects provided written informed consent. Sample processing and quality control measures were identical to those employed for the ICGN cohort as outlined above; the number of subjects from the EOCOPD cohort included in our analysis was 369.

Variable definitions

Current smoking was defined as an affirmative response to the question: 'Do you currently smoke cigarettes (as of 1 month ago)?' The average number of cigarettes smoked per day was the self-reported number of cigarettes smoked per day during the period when the subject smoked. Pack-years was calculated according to the number of years smoked multiplied by the average number of cigarettes smoked per day divided by 20. In former smokers, the time since quitting was calculated as the subject's age minus the age when they stopped smoking (in years). Obstructive lung disease was defined as an FEV₁/FVC ratio <0.7.

DNA extraction and bisulfite treatment

For both cohorts, DNA extraction from white blood cells was performed using the Gentra Puregene blood kit (Qiagen, Inc., Valencia, CA, USA). Bisulfite modification of 1 mg of DNA was performed using the 96-well EZ DNA Methylation Gold

Deep-well kit (Zymo Research, Orange, CA, USA). The protocol is outlined in Supplementary Material, Methods. Less than 2% of DNA samples failed bisulfite conversion.

DNA methylation array

Both cohorts were assessed using the Illumina (San Diego, CA, USA) Infinium HumanMethylation27K BeadChip, which assays 27 578 CpG sites in 14 475 consensus coding sequences in the NCBI Database (Genome Build 36). Bisulfite converted DNA was amplified, fragmented and hybridized to locus-specific 50-mer probes on the arrays, followed by single base extension using biotin (C and G) or dinitrophenyl (A and T) labeled nucleotides, per routine. All arrays were imaged using an Illumina BeadArray Reader. Beadstudio software (methylation module version 3.2) was used to assemble the workspace. Each locus results in an intensity value for methylation (*M*) and unmethylated (*U*) alleles. The methylation status was expressed as a beta value (β), which represents a ratio of the *M* to *U* fluorescent signals such that $\beta = \text{Max}(M,0)/[\text{Max}(M,0) + \text{Max}(U,0) + 100]$. Using this metric, DNA methylation is represented by a variable between 0 (no methylation) to 1 (complete methylation). Raw average beta values were exported and analyzed without normalization. CpG sites will be discussed using the industry assigned identifier ('cg') as outlined in the Illumina HumanMethylation27K

manifest. Quality control measures for array data are described in Supplementary Material, Methods.

Pyrosequencing

We performed pyrosequencing using the Pyromark96MD (Qiagen, Inc.) for technical validation of cg03636183 (*F2RL3*) and cg19859270 (*GPR15*) in the ICGN cohort. Details are provided in Supplementary Material. CpG methylation percentages were calculated using the height of the T and C peaks at the methylation site and applying the formula $C/(C + T) \times 100$ as implemented in the Pyromark CpG software (version 1.0.11).

Statistical analysis

Surrogate variable analysis (64,65), which identifies and estimates the effects of known and unknown sources of variation in high-throughput experiments, was performed to account for batch effects. To account for familial correlations in the data, generalized linear mixed models (66–69) as implemented in R (release 2.12) were used to analyze the methylation data. We used an identity link function in our models and the penalized quasi-likelihood method (66) to estimate the model parameters. The Illumina ‘beta’ value was modeled as the dependent variable, whereas the smoking metric (current smoking, pack-years and time since quitting cigarettes) was included as the independent variable. All models were adjusted for age, gender and significant surrogate variables. CpG sites with a false-discovery rate (FDR) adjusted *P*-value <0.05 (70) in the discovery cohort (ICGN) were examined using the same models in the replication cohort (EOCOPD). CpG sites demonstrating a *P*-value <10^{−3} in the replication cohort with the same direction of effect were considered significant. To assess the impact of COPD on our results, we repeated our analysis with the addition of a covariate to account for obstructive lung disease.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

Conflict of Interest statement. E.S.W., H.B. W.Q., V.J.C., A.B. and D.L.D. have no conflicts of interest to report. D.A.L. has received grant funding, consultancy fees and honoraria from GlaxoSmithKline. W.A. was an employee of GlaxoSmithKline at the time of the analyses. S.R. has consulted or participated in advisory boards for: ABIM, Adelphi Research, Almirall/Forest, APT, AstraZeneca, Boehringer Ingelheim/Pfizer, BoomCom, Capital Research, CommonHealth, Decision Resources, Easton Associates, Equinox, Forest (Cory Paeth), Fulcrum, Gerson Lehman, GSK Eclipse, Guidepoint, Informed, Insyght, MedImmune, Novartis, Oriol, Pearl, Pfizer (Varenicline), PharmaVentures, Prescott, Propagate Pulmonary Reviews, Roche, Schering Plough, Smith Research, Talecris, UBC. S.R. has given lectures for: American College of Osteopathic Physicians, Asan Medical Center, American Thoracic Society, COPD Foundation, Creative Education Concepts, Duke, Information TV, Nycomed, Otsuka,

University of Washington, University of Alabama-Birmingham. S.R. has received industry-sponsored grants from: AstraZeneca, Otsuka, Pfizer, Nabi, Merck. A.A. is a board member of GlaxoSmithKline, Almirall, Boehringer-Ingelheim, Astra-Zeneca, Esteve, Novartis, Nycomed and Roche. A.A. has grants (current or pending) from GlaxoSmithKline, Almirall and Nycomed. A.A. has received payment for lectures and educational presentations from GlaxoSmithKline, Almirall, Boehringer-Ingelheim, Astra-Zeneca, Esteve and Nycomed.

FUNDING

This work was supported by the National Institutes of Health (R01 HL089438, R01 HL075478, T32 HL007427, K12 HL089990, P01 HL105339); the Doris Duke Foundation Clinician Scientist Development Award (to D.L.D.); and the National Institute of Environmental Health Sciences-Harvard School of Public Health New Investigator Fund (ES00002 to A.B.). None of the funding sources played a role in the management, analysis or interpretation of the data, nor did they influence the preparation, review or decision to submit for publication.

REFERENCES

- Mathers, C.D. and Loncar, D. (2006) Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med.*, **3**, e442.
- Conen, D., Everett, B.M., Kurth, T., Creager, M.A., Buring, J.E., Ridker, P.M. and Pradhan, A.D. (2011) Smoking, smoking status, and risk for symptomatic peripheral artery disease in women: a cohort study. *Ann. Intern. Med.*, **154**, 719–726.
- Kawachi, I., Colditz, G.A., Stampfer, M.J., Willett, W.C., Manson, J.E., Rosner, B., Speizer, F.E. and Hennekens, C.H. (1993) Smoking cessation and decreased risk of stroke in women. *JAMA*, **269**, 232–236.
- Rea, T.D., Heckbert, S.R., Kaplan, R.C., Smith, N.L., Lemaitre, R.N. and Psaty, B.M. (2002) Smoking status and risk for recurrent coronary events after myocardial infarction. *Ann. Intern. Med.*, **137**, 494–500.
- Bouloukaki, I., Tsiligianni, I.G., Tsoumakidou, M., Mitrouska, I., Prokopoulos, E.P., Mavroudi, I., Siafakas, N.M. and Tzanakis, N. (2011) Sputum and nasal lavage lung-specific biomarkers before and after smoking cessation. *BMC Pulm. Med.*, **11**, 35.
- Braber, S., Henricks, P.A., Nijkamp, F.P., Kraneveld, A.D. and Folkerts, G. (2010) Inflammatory changes in the airways of mice caused by cigarette smoke exposure are only partially reversed after smoking cessation. *Respir. Res.*, **11**, 99.
- Louhelainen, N., Stark, H., Mazur, W., Ryttila, P., Djukanovic, R. and Kinnula, V.L. (2010) Elevation of sputum matrix metalloproteinase-9 persists up to 6 months after smoking cessation: a research study. *BMC Pulm. Med.*, **10**, 13.
- Vernooy, J.H., Kucukaycan, M., Jacobs, J.A., Chavannes, N.H., Buurman, W.A., Dentener, M.A. and Wouters, E.F. (2002) Local and systemic inflammation in patients with chronic obstructive pulmonary disease: soluble tumor necrosis factor receptors are increased in sputum. *Am. J. Respir. Crit. Care Med.*, **166**, 1218–1224.
- Willemse, B.W., ten Hacken, N.H., Rutgers, B., Lesman-Leege, I.G., Postma, D.S. and Timens, W. (2005) Effect of 1-year smoking cessation on airway inflammation in COPD and asymptomatic smokers. *Eur. Respir. J.*, **26**, 835–845.
- Luo, J., Margolis, K.L., Wactawski-Wende, J., Horn, K., Messina, C., Stefanick, M.L., Tindle, H.A., Tong, E. and Rohan, T.E. (2011) Association of active and passive smoking with risk of breast cancer among postmenopausal women: a prospective cohort study. *BMJ*, **342**, d1016.
- Vineis, P., Alavanja, M., Buffler, P., Fontham, E., Franceschi, S., Gao, Y.T., Gupta, P.C., Hackshaw, A., Matos, E., Samet, J. *et al.* (2004)

- Tobacco and cancer: recent epidemiological evidence. *J. Natl Cancer Inst.*, **96**, 99–106.
12. Newcomb, P.A. and Carbone, P.P. (1992) The health consequences of smoking. *Cancer Med. Clin. North Am.*, **76**, 305–331.
 13. Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M. *et al.* (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, **462**, 315–322.
 14. Breitling, L.P., Yang, R., Korn, B., Burwinkel, B. and Brenner, H. (2011) Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am. J. Hum. Genet.*, **88**, 450–457.
 15. Breton, C.V., Byun, H.M., Wenten, M., Pan, F., Yang, A. and Gilliland, F.D. (2009) Prenatal tobacco smoke exposure affects global and gene-specific DNA methylation. *Am. J. Respir. Crit. Care Med.*, **180**, 462–467.
 16. Guerrero-Preston, R., Goldman, L.R., Brebi-Mieville, P., Ili-Gangas, C., Lebron, C., Witter, F.R., Apelberg, B.J., Hernandez-Roystacher, M., Jaffe, A., Halden, R.U. *et al.* (2010) Global DNA hypomethylation is associated with *in utero* exposure to cotinine and perfluorinated alkyl compounds. *Epigenetics*, **5**, 539–546.
 17. Terry, M.B., Ferris, J.S., Pilsner, R., Flom, J.D., Tehranifar, P., Santella, R.M., Gamble, M.V. and Susser, E. (2008) Genomic DNA methylation among women in a multiethnic New York City birth cohort. *Cancer Epidemiol. Biomarkers Prev.*, **17**, 2306–2310.
 18. Choi, J.Y., James, S.R., Link, P.A., McCann, S.E., Hong, C.C., Davis, W., Nesline, M.K., Ambrosone, C.B. and Karpf, A.R. (2009) Association between global DNA hypomethylation in leukocytes and risk of breast cancer. *Carcinogenesis*, **30**, 1889–1897.
 19. Zhang, F.F., Cardarelli, R., Carroll, J., Fulda, K.G., Kaur, M., Gonzalez, K., Vishwanatha, J.K., Santella, R.M. and Morabia, A. (2011) Significant differences in global genomic DNA methylation by gender and race/ethnicity in peripheral blood. *Epigenetics*, **6**, 623–629.
 20. Hsiung, D.T., Marsit, C.J., Houseman, E.A., Eddy, K., Furniss, C.S., McClean, M.D. and Kelsey, K.T. (2007) Global DNA methylation level in whole blood as a biomarker in head and neck squamous cell carcinoma. *Cancer Epidemiol. Biomarkers Prev.*, **16**, 108–114.
 21. Moore, L.E., Pfeiffer, R.M., Poscablo, C., Real, F.X., Kogevinas, M., Silverman, D., Garcia-Closas, R., Chanock, S., Tardon, A., Serra, C. *et al.* (2008) Genomic DNA hypomethylation as a biomarker for bladder cancer susceptibility in the Spanish Bladder Cancer Study: a case-control study. *Lancet Oncol.*, **9**, 359–366.
 22. Zhu, Z.Z., Hou, L., Bollati, V., Tarantini, L., Marinelli, B., Cantone, L., Yang, A.S., Vokonas, P., Lissowska, J., Fustinoni, S. *et al.* (2012) Predictors of global methylation levels in blood DNA of healthy subjects: a combined analysis. *Int. J. Epidemiol.*, **41**, 126–139.
 23. Ally, M.S., Al-Ghannem, R. and Pufulete, M. (2009) The relationship between gene-specific DNA methylation in leukocytes and normal colorectal mucosa in subjects with and without colorectal tumors. *Cancer Epidemiol. Biomarkers Prev.*, **18**, 922–928.
 24. Belinsky, S.A., Klinge, D.M., Dekker, J.D., Smith, M.W., Bocklage, T.J., Gilliland, F.D., Crowell, R.E., Karp, D.D., Stidley, C.A. and Picchi, M.A. (2005) Gene promoter methylation in plasma and sputum increases with lung cancer risk. *Clin. Cancer Res.*, **11**, 6505–6511.
 25. Brait, M., Ford, J.G., Papaiahgari, S., Garza, M.A., Lee, J.I., Loyo, M., Maldonado, L., Begum, S., McCaffrey, L., Howerton, M. *et al.* (2009) Association between lifestyle factors and CpG island methylation in a cancer-free population. *Cancer Epidemiol. Biomarkers Prev.*, **18**, 2984–2991.
 26. Umemura, S., Fujimoto, N., Hiraki, A., Gemba, K., Takigawa, N., Fujiwara, K., Fujii, M., Umemura, H., Satoh, M., Tabata, M. *et al.* (2008) Aberrant promoter hypermethylation in serum DNA from patients with silicosis. *Carcinogenesis*, **29**, 1845–1849.
 27. Patel, B.D., Coxson, H.O., Pillai, S.G., Agusti, A.G., Calverley, P.M., Donner, C.F., Make, B.J., Muller, N.L., Rennard, S.L., Vestbo, J. *et al.* (2008) Airway wall thickening and emphysema show independent familial aggregation in chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.*, **178**, 500–505.
 28. Zhu, G., Warren, L., Aponte, J., Gulsvik, A., Bakke, P., Anderson, W.H., Lomas, D.A., Silverman, E.K. and Pillai, S.G. (2007) The SERPINE2 gene is associated with chronic obstructive pulmonary disease in two large populations. *Am. J. Respir. Crit. Care Med.*, **176**, 167–173.
 29. Silverman, E.K., Chapman, H.A., Drazen, J.M., Weiss, S.T., Rosner, B., Campbell, E.J., O'Donnell, W.J., Reilly, J.J., Ginns, L., Mentzer, S. *et al.* (1998) Genetic epidemiology of severe, early-onset chronic obstructive pulmonary disease. Risk to relatives for airflow obstruction and chronic bronchitis. *Am. J. Respir. Crit. Care Med.*, **157**, 1770–1778.
 30. Dudek, S.M., Jacobson, J.R., Chiang, E.T., Birukov, K.G., Wang, P., Zhan, X. and Garcia, J.G. (2004) Pulmonary endothelial cell barrier enhancement by sphingosine 1-phosphate: roles for cortactin and myosin light chain kinase. *J. Biol. Chem.*, **279**, 24692–24700.
 31. Clayburgh, D.R., Rosen, S., Witkowski, E.D., Wang, F., Blair, S., Dudek, S., Garcia, J.G., Alverdy, J.C. and Turner, J.R. (2004) A differentiation-dependent splice variant of myosin light chain kinase, MLCK1, regulates epithelial tight junction permeability. *J. Biol. Chem.*, **279**, 55506–55513.
 32. Connell, L.E. and Helfman, D.M. (2006) Myosin light chain kinase plays a role in the regulation of epithelial cell survival. *J. Cell Sci.*, **119**, 2269–2281.
 33. Russo, J.M., Florian, P., Shen, L., Graham, W.V., Tretiakova, M.S., Gitter, A.H., Mrsny, R.J. and Turner, J.R. (2005) Distinct temporal-spatial roles for rho kinase and myosin light chain kinase in epithelial purse-string wound closure. *Gastroenterology*, **128**, 987–1001.
 34. Wang, F., Graham, W.V., Wang, Y., Witkowski, E.D., Schwarz, B.T. and Turner, J.R. (2005) Interferon-gamma and tumor necrosis factor-alpha synergize to induce intestinal epithelial barrier dysfunction by up-regulating myosin light chain kinase expression. *Am. J. Pathol.*, **166**, 409–419.
 35. Gao, L., Grant, A., Halder, I., Brower, R., Sevransky, J., Maloney, J.P., Moss, M., Shanholtz, C., Yates, C.R., Meduri, G.U. *et al.* (2006) Novel polymorphisms in the myosin light chain kinase gene confer risk for acute lung injury. *Am. J. Respir. Cell Mol. Biol.*, **34**, 487–495.
 36. Christie, J.D., Ma, S.F., Aplenc, R., Li, M., Lanke, P.N., Shah, C.V., Fuchs, B., Albelda, S.M., Flores, C. and Garcia, J.G. (2008) Variation in the myosin light chain kinase gene is associated with development of acute lung injury after major trauma. *Crit. Care Med.*, **36**, 2794–2800.
 37. Flores, C., Ma, S.F., Maresso, K., Ober, C. and Garcia, J.G. (2007) A variant of the myosin light chain kinase gene is associated with severe asthma in African Americans. *Genet. Epidemiol.*, **31**, 296–305.
 38. Lee, S.O., Cheong, H.S., Park, B.L., Bae, J.S., Sim, W.C., Chun, J.Y., Isbat, M., Uh, S.T., Kim, Y.H., Jang, A.S. *et al.* (2009) MYLK polymorphism associated with blood eosinophil level among asthmatic patients in a Korean population. *Mol. Cells.*, **27**, 175–181.
 39. Leguillet, R., Laviolette, M., Bergeron, C., Zitouni, N., Kogut, P., Solway, J., Kachmar, L., Hamid, Q. and Lauzon, A.M. (2009) Myosin, transgelin, and myosin light chain kinase: expression and function in asthma. *Am. J. Respir. Crit. Care Med.*, **179**, 194–204.
 40. Wang, L., Guo, D.C., Cao, J., Gong, L., Kamm, K.E., Regalado, E., Li, L., Shete, S., He, W.Q., Zhu, M.S. *et al.* (2010) Mutations in myosin light chain kinase cause familial aortic dissections. *Am. J. Hum. Genet.*, **87**, 701–707.
 41. Rose, J.E., Behm, F.M., Drgon, T., Johnson, C. and Uhl, G.R. (2010) Personalized smoking cessation: interactions between nicotine dose, dependence and quit-success genotype score. *Mol. Med.*, **16**, 247–253.
 42. Kahn, M.L., Nakanishi-Matsui, M., Shapiro, M.J., Ishihara, H. and Coughlin, S.R. (1999) Protease-activated receptors 1 and 4 mediate activation of human platelets by thrombin. *J. Clin. Invest.*, **103**, 879–887.
 43. Xu, W.F., Andersen, H., Whitmore, T.E., Presnell, S.R., Yee, D.P., Ching, A., Gilbert, T., Davie, E.W. and Foster, D.C. (1998) Cloning and characterization of human protease-activated receptor 4. *Proc. Natl Acad. Sci. USA*, **95**, 6642–6646.
 44. Chen, H.T., Tsou, H.K., Tsai, C.H., Kuo, C.C., Chiang, Y.K., Chang, C.H., Fong, Y.C. and Tang, C.H. (2010) Thrombin enhanced migration and MMPs expression of human chondrosarcoma cells involves PAR receptor signaling pathway. *J. Cell Physiol.*, **223**, 737–745.
 45. Heiber, M., Marchese, A., Nguyen, T., Heng, H.H., George, S.R. and O'Dowd, B.F. (1996) A novel human gene encoding a G-protein-coupled receptor (GPR15) is located on chromosome 3. *Genomics*, **32**, 462–465.
 46. Blaak, H., Boers, P.H., Gruters, R.A., Schuitemaker, H., van der Ende, M.E. and Osterhaus, A.D. (2005) CCR5, GPR15, and CXCR6 are major coreceptors of human immunodeficiency virus type 2 variants isolated from individuals with and without plasma viremia. *J. Virol.*, **79**, 1686–1700.
 47. Cilliers, T., Willey, S., Sullivan, W.M., Patience, T., Pugach, P., Coetzer, M., Papanathanopoulos, M., Moore, J.P., Trkola, A., Clapham, P. *et al.* (2005) Use of alternate coreceptors on primary cells by two HIV-1 isolates. *Virology*, **339**, 136–144.

48. Farzan, M., Choe, H., Martin, K., Marcon, L., Hofmann, W., Karlsson, G., Sun, Y., Barrett, P., Marchand, N., Sullivan, N. *et al.* (1997) Two orphan seven-transmembrane segment receptors which are expressed in CD4-positive cells support simian immunodeficiency virus infection. *J. Exp. Med.*, **186**, 405–411.
49. Buttner, P., Mosig, S. and Funke, H. (2007) Gene expression profiles of T lymphocytes are sensitive to the influence of heavy smoking: A pilot study. *Immunogenetics*, **59**, 37–43.
50. Charlesworth, J.C., Curran, J.E., Johnson, M.P., Goring, H.H.H., Dyer, T.D., Diego, V.P., Kent, J.W., Mahaney, M.C., Almasy, L., MacClue, J.W. *et al.* (2010) Transcriptomic epidemiology of smoking: the effect of smoking on gene expression in lymphocytes. *BMC Med. Genomics*, **3**, 29.
51. Dumeaux, V., Olsen, K.S., Nuel, G., Paulssen, R.H., Borresen-Dale, A.L. and Lund, E. (2010) Deciphering normal blood gene expression variation—the NOWAC postgenome study. *PLoS Genet.*, **6**, e1000873.
52. Rotunno, M., Hu, N., Su, H., Wang, C., Goldstein, A.M., Bergen, A.W., Consonni, D., Pesatori, A.C., Bertazzi, P.A., Wacholder, S. *et al.* (2011) A gene expression signature from peripheral whole blood for stage I lung adenocarcinoma. *Cancer Prev. Res. (Phila)*, **4**, 1599–1608.
53. Showe, M.K., Vachani, A., Kossenkov, A.V., Yousef, M., Nichols, C., Nikonova, E.V., Chang, C., Kucharczuk, J., Tran, B., Wakeam, E. *et al.* (2009) Gene expression profiles in peripheral blood mononuclear cells can distinguish patients with non-small cell lung cancer from patients with nonmalignant lung disease. *Cancer Res.*, **69**, 9202–9210.
54. Landi, M.T., Dracheva, T., Rotunno, M., Figueroa, J.D., Liu, H., Dasgupta, A., Mann, F.E., Fukuoka, J., Hames, M., Bergen, A.W. *et al.* (2008) Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. *PLoS ONE*, **3**, e1651.
55. Beane, J., Sebastiani, P., Liu, G., Brody, J.S., Lenburg, M.E. and Spira, A. (2007) Reversible and permanent effects of tobacco smoke exposure on airway epithelial gene expression. *Genome Biol.*, **8**, R201.
56. Liu, F., Killian, J.K., Yang, M., Walker, R.L., Hong, J.A., Zhang, M., Davis, S., Zhang, Y., Hussain, M., Xi, S. *et al.* (2010) Epigenomic alterations and gene expression profiles in respiratory epithelia exposed to cigarette smoke condensate. *Oncogene*, **29**, 3650–3664.
57. Schembri, F., Sridhar, S., Perdomo, C., Gustafson, A.M., Zhang, X., Ergun, A., Lu, J., Liu, G., Bowers, J., Vaziri, C. *et al.* (2009) MicroRNAs as modulators of smoking-induced gene expression changes in human airway epithelium. *Proc. Natl Acad. Sci. USA*, **106**, 2319–2324.
58. Satta, R., Maloku, E., Zhubi, A., Pibiri, F., Hajos, M., Costa, E. and Guidotti, A. (2008) Nicotine decreases DNA methyltransferase 1 expression and glutamic acid decarboxylase 67 promoter methylation in GABAergic interneurons. *Proc. Natl Acad. Sci. USA*, **105**, 16356–16361.
59. Murray, R.P., Connett, J.E., Lauger, G.G. and Voelker, H.T. (1993) Error in smoking measures: effects of intervention on relations of cotinine and carbon monoxide to self-reported smoking. The Lung Health Study Research Group. *Am. J. Public Health*, **83**, 1251–1257.
60. Patrick, D.L., Cheadle, A., Thompson, D.C., Diehr, P., Koepsell, T. and Kinne, S. (1994) The validity of self-reported smoking: a review and meta-analysis. *Am. J. Public Health*, **84**, 1086–1093.
61. Wagenknecht, L.E., Burke, G.L., Perkins, L.L., Haley, N.J. and Friedman, G.D. (1992) Misclassification of smoking status in the CARDIA study: a comparison of self-report with serum cotinine levels. *Am. J. Public Health*, **82**, 33–36.
62. Bock, C., Tomazou, E.M., Brinkman, A.B., Muller, F., Simmer, F., Gu, H., Jager, N., Gnirke, A., Stunnenberg, H.G. and Meissner, A. (2010) Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nat. Biotechnol.*, **28**, 1106–1114.
63. Qiu, W., Baccarelli, A., Carey, V.J., Boutaoui, N., Bacherman, H., Klanderman, B., Rennard, S., Agusti, A., Anderson, W., Lomas, D.A. *et al.* (2012) Variable DNA methylation is associated with COPD and lung function. *Am J Respir Crit Care Med*, **185**, 373–381.
64. Leek, J.T. and Storey, J.D. (2007) Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.*, **3**, 1724–1735.
65. Leek, J.T. and Storey, J.D. (2008) A general framework for multiple testing dependence. *Proc. Natl Acad. Sci. USA*, **105**, 18718–18723.
66. Breslow, N. and Clayton, D. (1993) Approximate inference in generalized linear mixed models. *J. Am. Stat. Assoc.*, **88**, 9–25.
67. Scall, R. (1991) Estimation in generalized linear models with random effects. *Biometrika*, **78**, 719–727.
68. Wolfinger, R. and O’Connell, M. (1993) Generalized linear mixed models: a pseudo-likelihood approach. *J. Stat. Comp. Simulat.*, **48**, 233–243.
69. Venables, W.N. and Ripley, B.D. (2002) *Modern Applied Statistics with S-PLUS*, 4th edn. Springer, New York, NY.
70. Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Royal Stat. Soc.*, **57**, 289–300.