ORIGINAL ARTICLE

Human Lung DNA Methylation Quantitative Trait Loci Colocalize with Chronic Obstructive Pulmonary Disease Genome-Wide Association Loci

Jarrett D. Morrow[1], Kimberly Glass[1], Michael H. Cho[1,2], Craig P. Hersh[1,2], Victor Pinto-Plata[3], Bartolome Celli[2], Nathaniel Marchetti[4], Gerard Criner[4], Raphael Bueno[5], George Washko[2], Augustine M. K. Choi[6], John Quackenbush[7], Edwin K. Silverman[1,2], and Dawn L. DeMeo[1,2]

[1]Channing Division of Network Medicine, [2]Division of Pulmonary and Critical Care Medicine, and [5]Division of Thoracic Surgery, Brigham and Women's Hospital, Boston, Massachusetts; [3]Baystate Health, Springfield, Massachusetts; [4]Division of Pulmonary and Critical Care Medicine, Temple University, Philadelphia, Pennsylvania; [6]Department of Medicine, New York Presbyterian/Weill Cornell Medical Center, New York, New York; and [7]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, Massachusetts

ORCID IDs: 0000-0002-5670-5865 (J.D.M.); 0000-0001-9653-0636 (D.L.D.).

Abstract

Rationale: As the third leading cause of death in the United States, the impact of chronic obstructive pulmonary disease (COPD) makes identification of its molecular mechanisms of great importance. Genome-wide association studies (GWASs) have identified multiple genomic regions associated with COPD. However, genetic variation only explains a small fraction of the susceptibility to COPD, and sub–genome-wide significant loci may play a role in pathogenesis.

Objectives: Regulatory annotation with epigenetic evidence may give priority for further investigation, particularly for GWAS associations in noncoding regions. We performed integrative genomics analyses using DNA methylation profiling and genome-wide SNP genotyping from lung tissue samples from 90 subjects with COPD and 36 control subjects.

Methods: We performed methylation quantitative trait loci (mQTL) analyses, testing for SNPs associated with percent DNA methylation and assessed the colocalization of these results with previous COPD GWAS findings using Bayesian methods in the R package coloc to highlight potential regulatory features of the loci.

Measurements and Main Results: We identified 942,068 unique SNPs and 33,996 unique CpG sites among the significant (5% false discovery rate) *cis*-mQTL results. The genome-wide significant and subthreshold ($P < 10^{-4}$) GWAS SNPs were enriched in the significant mQTL SNPs (hypergeometric test $P < 0.00001$). We observed enrichment for sites located in CpG shores and shelves, but not CpG islands. Using Bayesian colocalization, we identified loci in regions near *KCNK3*, *EEFSEC*, *PIK3CD*, *DCDC2C*, *TCERG1L*, *FRMD4B*, and *IL27*.

Conclusions: Colocalization of mQTL and GWAS loci provides regulatory characterization of significant and subthreshold GWAS findings, supporting a role for genetic control of methylation in COPD pathogenesis.

Keywords: methylation QTL; mQTL; epigenetics; chronic obstructive pulmonary disease

(Received in original form July 14, 2017; accepted in final form January 3, 2018)

Supported by NIH grants K25 HL136846 (J.D.M.), P01 HL105339 (E.K.S.), R01 HL111759 (J.Q.), P01 HL114501 (A.M.K.C.), P01 HL132825 (D.L.D.), R01 HL089856 (E.K.S.), R01 HL089897, and K25 HL133599 (K.G.).

Author Contributions: J.D.M., M.H.C., C.P.H., and E.K.S.: analysis and interpretation of data, manuscript preparation, and approval of the final version; K.G. and J.Q.: analysis and interpretation of data and approval of final version of the manuscript; V.P.-P., B.C., N.M., G.C., R.B., G.W., and A.M.K.C.: acquisition of data, review of manuscript, and approval of the final version of manuscript; and D.L.D.: acquisition of data, analysis and interpretation of data, manuscript preparation, and approval of the final version.

Correspondence and requests for reprints should be addressed to Jarrett D. Morrow, Ph.D., M.S., Channing Division of Network Medicine, 181 Longwood Avenue, Boston, MA 02115. E-mail: remdj@channing.harvard.edu.

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

Am J Respir Crit Care Med Vol 197, Iss 10, pp 1275–1284, May 15, 2018

Originally Published in Press as DOI: 10.1164/rccm.201707-1434OC on January 9, 2018
Internet address: www.atsjournals.org

At a Glance Commentary

Scientific Knowledge on the Subject: Genome-wide association studies (GWASs) have identified multiple genomic regions associated with chronic obstructive pulmonary disease (COPD). However, genetic variation only explains a small fraction of the susceptibility to COPD. Given that an environmentally influenced complex disease like COPD is driven by both genetic and epigenetic variation, understanding the genetic control of DNA methylation in lung tissue may highlight regulatory components of COPD pathogenesis. We include subthreshold GWAS loci, as they may play a role in COPD and may be given priority in further investigations based on our methods, absent higher-powered genetic association studies.

What This Study Adds to the Field: Our integrative analyses using DNA methylation profiling and genome-wide genotyping from lung tissue samples from COPD cases and control subjects, and subsequent Bayesian colocalization with previous COPD GWAS findings, highlight potential regulatory features of the loci, supporting a role for genetic control of methylation in COPD pathogenesis. Our top candidate gene *KCNK3* was near subthreshold loci and could rise to significance in larger GWAS of lung disease, as we provided similar molecular justification for *EEFSEC*, a gene that has only recently been validated by virtue of its proximity to a genome-wide significant locus in a large study. Our methods have also revealed regulatory mechanisms near loci previously associated with COPD.

Chronic obstructive pulmonary disease (COPD) is characterized by progressive airflow obstruction and is the third leading cause of death in the United States (1). Genetic association studies have identified multiple genomic regions that explain only a small fraction of the heritability of COPD; epigenetic association studies in blood and lung tissue have likewise identified associations with COPD affection status. Given that an environmentally influenced complex disease like COPD is driven by both genetic and epigenetic variation, understanding the genetic control of DNA methylation in lung tissue may highlight the regulatory contributions relevant to COPD pathogenesis, using results from prior genome-wide association studies (GWASs) as a guide.

Loci associated with complex disease traits are enriched with epigenetic annotations including DNA methylation (2) and have a role in gene regulation, complex traits, and human development (3, 4). DNA methylation profiling may uncover molecular targets for lung disease perturbed by environmental exposures, such as smoking (5–7) or medications (8), that may also be found suggestive in genetic association studies but not within stringent statistical thresholds adjusting for genome-wide multiple testing. Prioritization of previously identified genomic loci will enhance the molecular understanding of complex disease (9, 10). Genetic control of DNA methylation has been observed in diabetes (11), schizophrenia (12), complex traits such as blood pressure and body mass index (13, 14), and various complex diseases (15). In these contexts, genetic–epigenetic interactions may be linked to the perturbation of regulatory elements, such as transcription factor binding (15, 16) and changes in gene expression (11, 14, 15, 17, 18), potentially with allele-specific effects (19, 20).

Recent studies have demonstrated associations between COPD affection status and DNA methylation in lung (21, 22) and small airway (23) tissues. Cigarette smoke exposure has been associated with DNA methylation variability in the lung (21) and small airways (24). Genetic control of DNA methylation in normal lung tissue (25) and the association of DNA methylation with gene expression in the lungs of subjects with COPD (22, 26) and small airways (23) have also been demonstrated. The genetic control of DNA methylation is observed by testing the association between SNP variants and DNA methylation via methylation quantitative trait loci (mQTL) analyses. *cis*-mQTL analysis involves associations in local genomic regions, whereas *trans*-mQTL analysis involves genome-wide associations. Intersecting mQTL results in normal lung tissue with statistically significant GWAS loci have provided regulatory annotation for lung cancer genes (25) and support the need to investigate the regulatory role of genome-wide DNA methylation in the context of nonneoplastic lung diseases such as COPD. Intersecting whole-blood mQTL results with DNA methylation sites associated with smoke exposure (5) highlighted disease susceptibility factors in the context of environmental exposure and suggested that genetic control of methylation, in addition to smoke exposure and other genetic influences, likely plays a role in COPD pathogenesis.

In this study of genetic control of methylation, we performed *cis*-mQTL analysis in lung tissue from severe COPD cases and ex-smoker control subjects and integrated the findings with prior results from a GWAS (27) and an epigenome-wide association study (EWAS) of COPD (22) using the Infinium HumanMethylation450 BeadChip. We used Bayesian colocalization methods (28) previously applied to mQTL for schizophrenia (12) to test the hypothesis that genetic variants associated with DNA methylation influence COPD pathogenesis and susceptibility. Genetic loci that do not meet genome-wide significance (subthreshold results) may play a role in complex diseases (29) such as COPD. We include these subthreshold loci, as they may be given priority in further investigations based on our methods, absent higher-powered genetic association studies with thousands more subjects. This study in lung tissue provides regulatory information for COPD GWAS loci and informs selection of genomic loci for future *in vitro* validation studies. Some of the results of these studies have been previously reported in the form of an abstract (30).

Methods

Lung tissue samples were collected from subjects undergoing thoracic surgery for lung transplantation, lung volume reduction surgery, or lung nodule resection at three medical centers: Brigham and Women's Hospital (Boston, MA), St. Elizabeth's Hospital (Boston, MA), and Temple University Hospital (Philadelphia, PA). Institutional review board approval was obtained at the three centers, and subjects provided written informed consent. All subjects were former smokers by self-report who quit smoking at least 1 month before thoracic surgery (22, 31). Subjects with

severe COPD were defined by Global Initiative for Chronic Obstructive Lung Disease grade 3 to 4 spirometry (FEV_1% predicted $< 50\%$ and $FEV_1/FVC < 0.7$), and control smokers had normal spirometry (FEV_1% predicted $\geq 80\%$ and $FEV_1/FVC \geq 0.7$).

mQTL

Preprocessed methylation data were available for 114 cases and 46 control subjects (22). An overlap of the imputed genotyping and methylation data (*see* online supplement) resulted in a dataset for 126 white subjects (90 cases, 36 control subjects). A *cis*-mQTL analysis was performed using the R/Bioconductor package Matrix eQTL (version 2.1.1) (32). A total window size of 1 million bases was used (testing associations with SNPs 500 kb upstream and downstream from the CpG site). This implementation identifies associations between genotype allele dosage values (coded as 0, 1, and 2; imputed values in the range [0, 2]) and methylation β values (in the range [0, 1]), using the model: methylation = genotype + age + sex + pack-years + population_PCs + batch_number + batch_PCs. Pack-years represents pack-years of smoking, and population_PCs involves two ancestry principal components (PCs) (*see* online supplement). Batch_number is a methylation data batch variable, and batch_PCs represents the methylation principal components described below. We performed 10 analyses after randomly permuting the subject labels on the methylation data and covariates each time. Using the *P* values from these 10 permutations and our mQTL results, we determined that a *P* value of 3.2×10^{-6} corresponds to a false discovery rate (FDR) of 5% (15, 33). A significant association with *P* value below this threshold is presented as an mSite (CpG site) and mQTL (SNP) pair. Before these analyses, an iterative method was used to determine the number of principal components for the matrix of β values to add as covariates to mitigate batch effects. This procedure involves finding the number of principal component covariates that produces a maximum number of significant (FDR < 0.05) *cis*-mQTL (window of 1 Mb). We used linkage disequilibrium–pruned genotyping data (r^2 threshold of 0.3) from chromosomes 21 and 22 (31). The maximum was achieved with 10 PCs, and these were included as additional covariates in the mQTL analyses.

GWAS and mQTL Colocalization

For integration with GWAS, results were extracted from a published COPD case–control GWAS (27) at *P* value thresholds of 0.05, 10^{-4}, and 5×10^{-8}. The total number of SNPs meeting these thresholds was 456,214, 2,252, and 298, respectively. Annotations on the basis of gene proximity for these SNPs were obtained from the Ensembl database (GRCh37) using the R package biomaRt (34). The colocalization of mQTL and GWAS results in local *cis* regions was assessed using the R package coloc, which highlights several posterior probabilities (PP) for consideration (28). This Bayesian framework is used to determine if an overlap of mQTL and GWAS associations is consistent with a shared causal variant. For each CpG mSite, all imputed SNPs in the lung tissue data that overlap with the GWAS and that also have nominal association ($P < 0.05$) with the mSite are included in the colocalization assessment. The effective window size for these SNPs is 1 million bases. A larger value for the PP4 indicates a higher probability of a shared causal variant (fourth hypothesis of the coloc method), implying colocalization in the region for the two association

Table 1. Top 20 *cis*–Methylation Quantitative Trait Loci Results

mSite (CpG Site ID)	mQTL (SNP rsID or Chr:coord)	*P* Value	Regression Coefficient	Gene Symbol	Chr	mQTL to mSite Distance (Bases for hg19)
cg08477332*	rs11548102	7.50×10^{-110}	−0.378	*S100A14*	chr1	2,987
cg11268327	rs1063630	4.05×10^{-106}	0.412	*MICA*	chr6	1,840
cg22216157*	rs4716790	2.51×10^{-98}	−0.213	*PTPRN2*	chr7	2,643
cg05084668*	rs4414796	5.92×10^{-96}	0.432	*ALG1L*	chr3	1,4542
cg12454169†	rs1662955	1.02×10^{-94}	−0.385	*LCLAT1*	chr2	1,376
cg12897067	rs9327741	8.37×10^{-92}	−0.498	<NA>	chr5	14,819
cg26893861*	rs15359	2.77×10^{-89}	0.353	*DUSP3*	chr17	12
cg19178509	rs2844664	6.55×10^{-88}	0.306	*DDR1*	chr6	34,636
cg12606694†	rs3903759	1.56×10^{-84}	−0.317	*AKAP7*	chr6	2
cg17099656	rs5769127	8.47×10^{-84}	−0.228	*CERK*	chr22	101
cg26850117	rs7756483	5.57×10^{-83}	0.226	*LINC00574*	chr6	9,046
cg24786174†	rs689902	4.56×10^{-82}	−0.232	*ZNF516*	chr18	1,726
cg05988603	rs17180521	4.22×10^{-81}	0.396	*ZNRD1-AS1*	chr6	53,529
cg23649088*	rs12693903	2.57×10^{-78}	−0.305	*C2orf69*	chr2	31,584
cg20409752*	rs6680259	2.60×10^{-78}	0.443	*CAMTA1*	chr1	418
cg19169023	rs62001448	5.25×10^{-76}	−0.290	*TYRO3*	chr15	12,179
cg05269323	rs3735165	1.27×10^{-75}	0.165	*REPIN1*	chr7	659
cg10639368	rs925738	2.77×10^{-70}	−0.169	*HDAC4*	chr2	37
cg16906346*	rs9467249	7.80×10^{-70}	0.274	*KIAA0319*	chr6	2,379
cg08344634†	rs4410948	3.72×10^{-69}	−0.158	*KBTBD11*	chr8	1,244

Definition of abbreviations: Chr = chromosome; coord = coordinate; mQTL = methylation quantitative trait loci; NA = not available; rsID = reference SNP ID number.

Results for the most significant mQTL/SNP only.

*This mSite and a different significant mQTL were associated in a previous normal lung tissue *cis*-mQTL analysis (false discovery rate < 5%) (25).

†This mSite and mQTL association identified in previous normal lung tissue *cis*-mQTL analysis (false discovery rate < 5%) (25).

profiles. A larger PP3 value indicates a higher probability of separate mQTL and GWAS associations (28). In addition to the value of PP4, the ratio of PP4 to PP3 may also be used as an indicator of colocalization significance, with a ratio greater than five suggesting a strong shared signal over strong individual signals (12, 35).

Results

DNA methylation data were available for lung tissue from 90 patients with severe COPD (mean FEV_1, 26.1% predicted) and 36 control subjects with normal spirometry (*see* Table E1 in the online supplement). There were no significant differences between cases and control subjects by sex or age. The COPD cases had higher cumulative pack-years of cigarettes smoked and quit smoking more recently, on average 8 fewer years in the past ($P = 0.0002$).

mQTL

Using the DNA methylation profiling and imputed genotyping data for lung tissue from the white cases and control subjects, we performed a *cis*-mQTL analysis, testing for SNPs associated with DNA methylation levels. We identified 942,068 unique mQTL and 33,996 unique mSites (5% FDR). The mQTL and mSites are the significant SNPs and CpG sites from the mQTL analysis, respectively. This represents 17% of the SNPs and 10% of the CpG sites tested. We intersected these results with a previous study of mQTL in normal tissue (25). Of the 26,564 statistically significant CpG sites from the Shi and colleagues study in our final DNA methylation data, 15,649 are found among our significant results (59%) (25). Our top results, selecting the most significantly associated mQTL for each CpG mSite, are shown in Table 1 (5% FDR, Manhattan and volcano plots in Figure 1). We found replication for 11 of these 20 mSites in the Shi and colleagues study; four also had the same associated mQTL (25).

We examined the distribution of CpG island location annotations (36) for the 33,996 significant CpG mSites (Figure E1); 11,719 mSites do not have an island-related annotation. Significant enrichment in the shore and shelf regions was observed (hypergeometric test P value < 0.0001). Of the 9,055 genes in the annotation of the 33,996 mSites, 478 have significant (hypergeometric test, FDR q value < 0.05) mSite enrichment (Table E2), when using the number of CpG sites assayed for each gene as the reference. Among these 478 multi-mSite genes, 36 were from the MHC (major histocompatibility complex) region (bases 29,640,000 to 33,120,000 on chromosome 6). Pathway analysis using the KEGG (Kyoto Encyclopedia of Genes and Genomes) database and ConsensusPathDB (37) revealed enrichment for immune system pathways and diseases, including asthma (Table E3). These enrichments for CpG sites with genetic associations suggest that genetic control of methylation may be more pronounced for some genes or genomic regions.

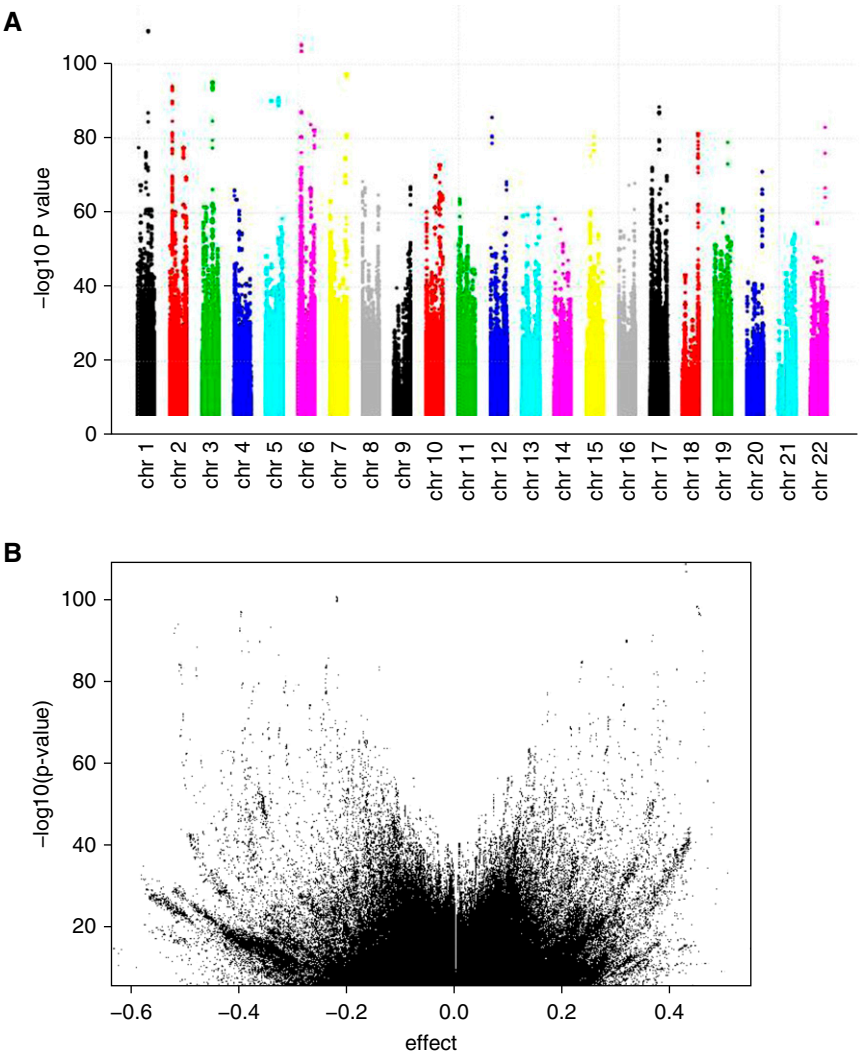

Figure 1. Summary plots for the *cis*–methylation quantitative trait loci (mQTL) results. (*A*) Manhattan plot and (*B*) volcano plot. The negative $\log_{10}$-transformed *P* values for the significant (false discovery rate < 5%) *cis*-mQTL results are plotted across all chromosomes in a Manhattan plot (*A*), and against the mQTL regression β values (effect) (*B*). chr = chromosome.

Disease Relevance of mQTL

We intersected all 811,739,658 *cis*-mQTL results with the previously published COPD case–control GWAS (27) results ($P < 0.05$; *see* METHODS) by the rsIDs of the SNPs/mQTL, and the overlapping SNP count was 277,643. A scatter plot of the *cis*-mQTL *P* values (Figure 2) with an overlay of COPD EWAS significance (22) graphically highlights the CpG mSites and mQTL with prior COPD associations. We observed that mQTL with COPD GWAS associations generally are not associated with CpG sites identified in the COPD EWAS, as regions of at least subthreshold GWAS associations ($P < 10^{-4}$) and significant mQTL *P* values are effectively

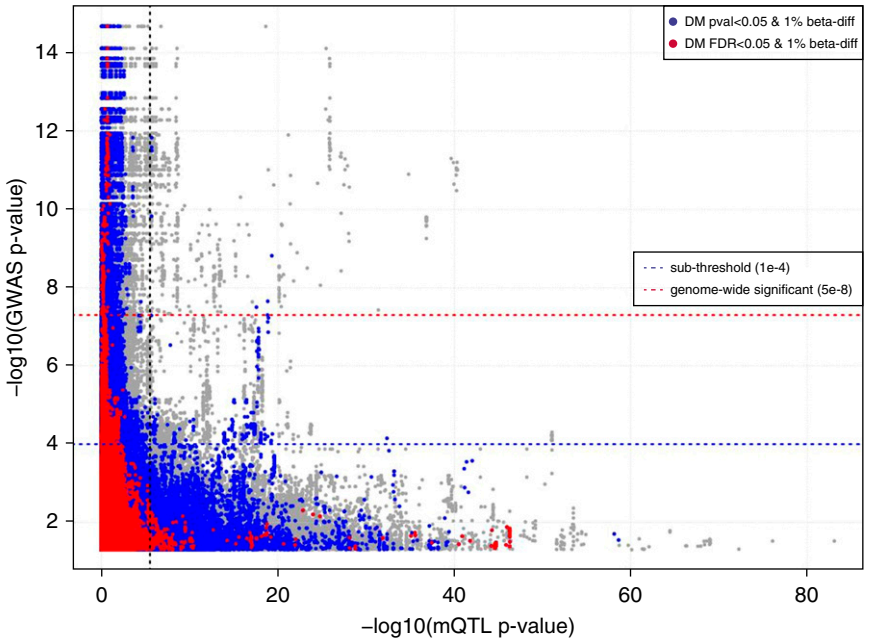

Figure 2. Scatter plot for the *cis*–methylation quantitative trait loci (mQTL) and chronic obstructive pulmonary disease genome-wide association study (GWAS) with an overlay of epigenome-wide association study (EWAS) results for the CpG mSites represented. Each point represents a SNP/QTL. GWAS *P* values (*y*-axis) are plotted against the methylation QTL *P* values (*x*-axis). A vertical dotted line indicates the threshold of significance (false discovery rate [FDR] < 5%) for the mQTL. Horizontal lines delineate genome-wide significant (red) and sub–genome-wide significant (blue) GWAS *P*-value thresholds. The significant (red; FDR < 5% and β-diff > 1%) and nominally significant (blue; $P < 0.05$ and β-diff > 1%) CpG sites from EWAS are highlighted for the CpG mSites associated with the mQTLs represented. DM = differential methylation; pval = *P* value.

devoid of significant (FDR < 5%) EWAS CpG sites.

We performed a two-way intersection of the significant mQTL results (FDR < 5%) and COPD GWAS (27) ($P < 10^{-4}$) using SNP/mQTL rsIDs. Of the 2,252 subthreshold GWAS SNPs, a total of 1,847 were found in the lung imputed genotyping data. The mQTL-GWAS overlapping SNP count was 753 of the 1,847; these GWAS results were enriched in mQTL (hypergeometric test *P* value < 0.00001, fold enrichment = 2.5). This intersection sorted by mQTL *P* value, selecting the most significant mQTL for each CpG mSite, is shown in Table E4; 34 of these 73 mSites were significant in the Shi and colleagues study (25). With respect to CpG island location annotation, there was suggestive evidence of north shore enrichment ($P < 0.15$; Figure E2) among the mSites from this intersection with GWAS. In addition, CpG mSites annotated to the genes *NUPR1* and *CYP2A7* were present in this intersection. These two genes were also significantly enriched for mSites (Table E2). For the intersection with only genome-wide significant ($P < 5.0 \times 10^{-8}$) GWAS (Table E5), 155 of the 297 significant GWAS SNPs found in the imputed lung genotyping data were in the intersecting set. This intersection highlights several COPD genes previously identified based on their proximity to the significant GWAS SNPs. In addition, 7 of the 12 CpG mSites in Table E5 were significant in the Shi and colleagues study (25). These overlaps with GWAS motivated further investigation of colocalization using Bayesian methods to hone in on the more robust COPD loci.

Colocalization of the mQTL and GWAS Associations

We tested whether the mQTL and GWAS associations were colocalized (*see* Methods). Colocalization of mQTL with COPD GWAS loci would suggest a genetic control of methylation with disease relevance and would highlight a possible regulatory role in COPD GWAS loci. All 73 significant (FDR < 5%) CpG mSites with mQTL intersecting subthreshold ($P < 10^{-4}$) GWAS (27) were assessed for colocalization. PP4 values of 0.9 and greater that have PP4/PP3 greater than 5 are considered to be significant, as those values support a single shared causal variant (Table 2). Nine of the 20 mSites with suggestive evidence of colocalization were significant in the Shi and colleagues study (25).

A CpG mSite in DNase hypersensitivity (DHS) and enhancer regions annotated to *KCNK3* (cg11273176) has significant evidence of colocalization with subthreshold GWAS (PP4 = 0.959, and PP4/PP3 > 10; Tables 2 and E6 and Figure 3). The LocusZoom plots (Figure 4) provide convincing support for a shared genetic association profile. Several findings in Table 2 were not previously annotated to a specific gene, although they highlight genes and regions of interest. For example, a CpG mSite upstream of *TCERG1L* (cg19961228) on chromosome 10 has strong colocalization evidence (Figure E3). Two mSites (cg27091865 and cg17052675) on chromosome 2 with significant colocalization are located in the gene *DCDC2C* and in a region of putative CTCF Binding (regional summary, Figures E4 and E5, Table E6). Although the effects of DNA methylation on CTCF binding are unclear, the potential impact on chromatin architecture, DNA folding, and gene regulation (4, 38) could have significant implications in disease. On chromosome 12, the CpG mSite cg13890972 demonstrates colocalization with prior GWAS (Figure E6) and is located in the gene *CCDC91*, and is also in a DHS region characterized as having open chromatin (Table E6). The mSite cg05515099 on chromosome 3 is located in the gene *FRMD4B* and is also located in a DHS region (Figure E7, Table E6). A significant mSite on chromosome 17 (cg13527508) is located in DHS and promoter flanking regions and is downstream of the gene *RAP1GAP2* (Figure E8, Table E6).

One mQTL (Table 2) is a genome-wide significant SNP from the prior COPD GWAS. This SNP is annotated by proximity to the previously identified GWAS gene *CHRNA5* (rs11633958, chr15). However, the colocalization in this region highlights a CpG mSite (cg04882995) annotated to the gene *CHRNA3* (mSite-SNP distance 50,594 bases; Figure E9). The mSite is located in an open chromatin and DHS region with putative transcription factor binding (Table E6).

The colocalization results for *EEFSEC* are an example of a subthreshold GWAS

Table 2. Colocalization Results for the Methylation Quantitative Trait Loci and Genome-Wide Association Study Data Sorted by Posterior Probability 4

mSite (CpG Site ID)	mQTL (SNP rsID)	mQTL *P* Value	GWAS *P* Value for mQTL	coloc SNP Count	PP3	PP4	PP4/PP3 Ratio	Chr	mSite Gene Symbol	mSite Gene* Location	mSite Island Location
cg11273176†	rs12468863	3.84×10^{-8}	1.47×10^{-5}	396	0.027	0.959	35.5	2	*KCNK3*	Body	N_Shelf
cg17052675†	rs12711974	8.62×10^{-52}	5.83×10^{-5}	470	0.0241	0.948	39.3	2			
cg27091865	rs13000873	1.44×10^{-9}	8.05×10^{-5}	415	0.0253	0.946	37.4	2			
cg04882995†	rs11633958	2.00×10^{-9}	7.99×10^{-12}	711	0.0804	0.92	11.4	15	*CHRNA3*	Body	Island
cg07832442†	rs34830546	1.97×10^{-8}	9.56×10^{-5}	369	0.0235	0.919	39.1	16			
cg00204512	rs149301	4.25×10^{-18}	3.20×10^{-5}	350	0.0726	0.913	12.6	16			N_Shelf
cg07451762	rs151174	1.28×10^{-10}	6.34×10^{-5}	259	0.0742	0.911	12.3	16			S_Shelf
cg13527508†	rs11078430	8.54×10^{-18}	9.19×10^{-5}	245	0.0205	0.91	44.3	17			Island
cg08044714	rs2955133	5.60×10^{-19}	4.15×10^{-6}	614	0.108	0.891	8.25	3	*EEFSEC*	Body	S_Shelf
cg26573321	rs6540991	1.53×10^{-13}	3.99×10^{-5}	379	0.0317	0.867	27.4	1	*PIK3CD*	TSS200	N_Shore
cg19961228	rs10732443	1.44×10^{-14}	5.17×10^{-5}	259	0.0164	0.862	52.5	10			
cg05515099†	rs62251459	7.45×10^{-8}	7.86×10^{-5}	308	0.0756	0.861	11.4	3			
cg08947774	rs7438245	2.31×10^{-6}	1.94×10^{-5}	252	0.0251	0.859	34.2	4	*PALLD*	Body	
cg01621080	rs231976	7.45×10^{-7}	3.50×10^{-5}	256	0.0795	0.852	10.7	16	*SULT1A2*	5′UTR	N_Shelf
cg12455300	rs4957342	1.73×10^{-6}	6.66×10^{-5}	365	0.0369	0.85	23	5	*TTC33*	3′UTR	
cg13890972‡	rs2203088	6.50×10^{-7}	7.31×10^{-5}	73	0.0119	0.835	70.5	12			
cg05377949	rs2811478	2.76×10^{-13}	1.63×10^{-5}	656	0.188	0.81	4.31	3	*EEFSEC*	Body	
cg15149645	rs17707300	2.10×10^{-12}	5.16×10^{-5}	302	0.16	0.808	5.06	16	*NUPR1*	TSS200	
cg01542023†	rs189278832	1.41×10^{-9}	4.12×10^{-5}	282	0.164	0.803	4.89	16	*NUPR1*	TSS200	
cg16576597†	rs62034325	2.76×10^{-9}	8.94×10^{-5}	300	0.18	0.783	4.34	16	*NUPR1*	TSS1500	

Definition of abbreviations: Chr = chromosome; GWAS = genome-wide association study; mQTL = methylation quantitative trait loci; PP = posterior probability; rsID = reference SNP ID number; UTR = untranslated region.
The mQTL with the most significant CpG mSite association is shown.
*Location relative to the first listed transcript in the complete annotation; annotations provided by assay manufacturer (36).
†mSite associated with a different mQTL in both our analysis and previous normal lung tissue *cis*-mQTL analysis (false discovery rate < 5%) (25).
‡mSite identified in previous normal lung tissue *cis*-mQTL analysis (false discovery rate < 5%) (25).

association with additional supportive evidence. The results for two mSites annotated to *EEFSEC* (cg08044714 and cg05377949) have suggestive colocalization (PP4 > 0.8 and PP4/PP3 > 4; Table 2 and Figure E10), and cg05377949 is located at a DHS region (Table E6). LocusZoom plots of association results from the mQTL analysis and prior GWAS (Figure E11) illustrate similar *P*-value profiles, as expected with suggestive colocalization. A CpG mSite (cg26573321) annotated to *PIK3CD* on chromosome 1 demonstrates strong colocalization (Table 2 and Figure E12). This mSite is in a promoter and DHS region (Table E6). Last, the complex region on chromosome 16 (Figure E13) near the genes *SULT1A2* and *NUPR1* has several CpG mSites with strong colocalization that are also located in transcription factor binding and DHS regions.

Expression Profiling for Identified Genes

We examined our previous gene expression profiling results in lung tissue samples from this study cohort (31). For the CpG mSite gene annotations listed in Table 2, we observed differential expression for *SULT1A2* ($P = 0.018$) by COPD status, providing additional regulatory evidence for the complex region on chromosome 16. When we examine the additional genes identified in the colocalization regions of Table 2, we observed differential expression of *CCDC91* ($P = 0.004$) and *FRMD4B* ($P = 0.014$) by COPD status. In addition, we examined the association between gene expression and DNA methylation for the genes in Table 2. We used data from the 86 cases and 34 control subjects for whom we have both DNA methylation and gene expression data and corrected for batch effects independently in each set. Associations were found for *KCNK3* (cg11273176: $P = 0.004$; Figure E14) and *SULT1A2* (cg01621080: $P = 7.5 \times 10^{-7}$; cg01621080: $P = 0.00016$; cg01621080 shown in Figure E14). We also observed associations for all three *NUPR1* mSites (cg15149645: $P = 0.0013$; cg01542023: $P = 0.007$; and cg16576597: $P = 0.0006$; cg15149645 shown in Figure E14). These methylation-expression findings support a possible gene regulatory role for the genetic control of DNA methylation in lung tissue.

Discussion

We examined the genetic control of DNA methylation to provide regulatory annotation and prioritization of COPD GWAS loci, through integration of genome-wide DNA methylation and genotyping data, and subsequent intersection with GWAS results. We observed that mQTLs are abundant in lung tissue and that many genome-wide significant loci from COPD GWAS are not mQTL. However, delving into the subthreshold GWAS findings, we found noncoding variants significantly enriched with mQTL.

Colocalization provided insight into the overlapping *cis*-mQTL and GWAS association profiles. The gene *KCNK3* (potassium two pore domain channel subfamily K member 3) was previously suggested as a COPD gene using protein–protein interaction data (39). The CpG mSite annotated to *KCNK3* was also found in a previous normal tissue mQTL study (25) and is located in a DHS region with putative enhancer binding within the gene. The regulatory implications and genomic reach of this CpG mSite are not

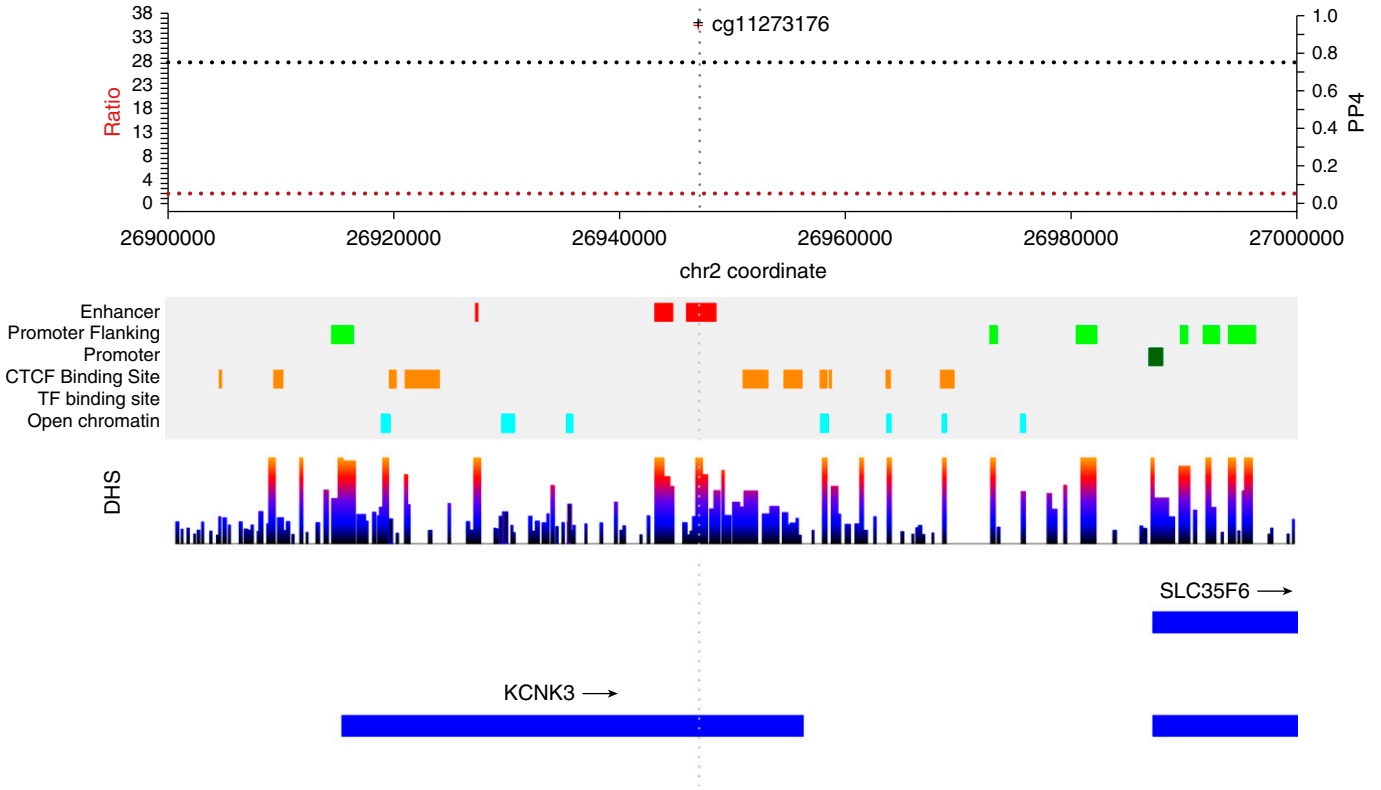

Figure 3. Annotated summary of *KCNK3* colocalization (cg11273176) on chromosome 2 (chr2). The colocalization posterior probability (PP) 4 (PP4) and PP4/PP3 ratio are shown in the top plot for the CpG mSite (cg11273176) in *KCNK3*. Regulatory information related to promoter and transcription factor (TF) binding for this genomic region is shown below, along with boundaries for genes in the region. The vertical gray dotted line passes through the genomic coordinate of the CpG mSite to align with the regulatory and gene annotations. The DNase hypersensitivity (DHS) track highlights regions of DNase hypersensitivity. The CTCF Binding Site track represents regions of *CTCF* (CCCTC-binding factor) binding.

currently known. However, our methylation-expression findings highlight a possible role in the regulation of *KCNK3* expression. Strong colocalization evidence was observed for *EEFSEC* (eukaryotic elongation factor, selenocysteine-tRNA specific) at two CpG mSites. A recent COPD genome-wide association meta-analysis in more than 60,000 subjects identified genome-wide significant loci in *EEFSEC* (40, 41). The statistical association of *EEFSEC* in this higher-powered study demonstrates the ability of our approach to elevate subthreshold results with potential relevance to COPD, without the demands of increasing sample sizes by tens of thousands.

Our study of genome-wide significant loci highlighted *CHRNA3* (cholinergic receptor nicotinic alpha 3 subunit) within a locus implicated in previous COPD GWAS (27, 42). The genes *CYP2A7* (cytochrome P450 family 2 subfamily A member 7) and *NUPR1* (nuclear protein 1, transcriptional regulator) were enriched in mSites and associated with an mQTL having subthreshold significance in prior GWAS. In a more recent GWAS of lung function of COPD, *CYP2A7* was identified in a genome-wide significant locus on chromosome 19 (43). A region of interest for future detailed study in COPD is located near *IL27* (interleukin 27), *SULT1A1*, and *SULT1A2* (sulfotransferase family 1A member 2). This region also includes *NUPR1*. The complexity and genetic relevance of this region on chromosome 16 was highlighted in an association study of exome array data in COPD (44). The recent Hobbs and colleagues COPD meta-analysis also identified a genome-wide locus in this region of chromosome 16 (40). Our approach may narrow the search for pathogenic genes within these complex regions identified through GWAS.

The mQTLs in this study were distributed throughout the genome. Within the top 20 mQTL results, 11 of the CpG mSites were found in a previous *cis*-mQTL analysis in normal lung tissue (25); all 11 associated mQTLs in normal tissue were found within our significant results, suggesting that some are lung QTLs important for normal function and development, but others may be driven by exposure or disease. Nine of the 20 CpG mSites in the colocalization results were significant (FDR $<$ 5%) in the Shi and colleagues study, including the mSite annotated to *CHRNA3* (25). This, together with the replication of top colocalization findings for *KCNK3*, was a robust recapitulation of genetic effects on DNA methylation in lung tissue, independent of disease status.

We observed CpG island shore and shelf annotations among the significant CpG mSites, but only enrichment of CpG island northern shores is suggested, when limited to a GWAS *P* value threshold of 10^{-4}; enrichment of northern shore sites has been previously observed (25). This suggests that CpG mSites related to disease susceptibility via genetic effects are a subset with more regulatory implications (25, 45, 46), and these northern-shore regions are of particular interest for functional characterization. We also observed

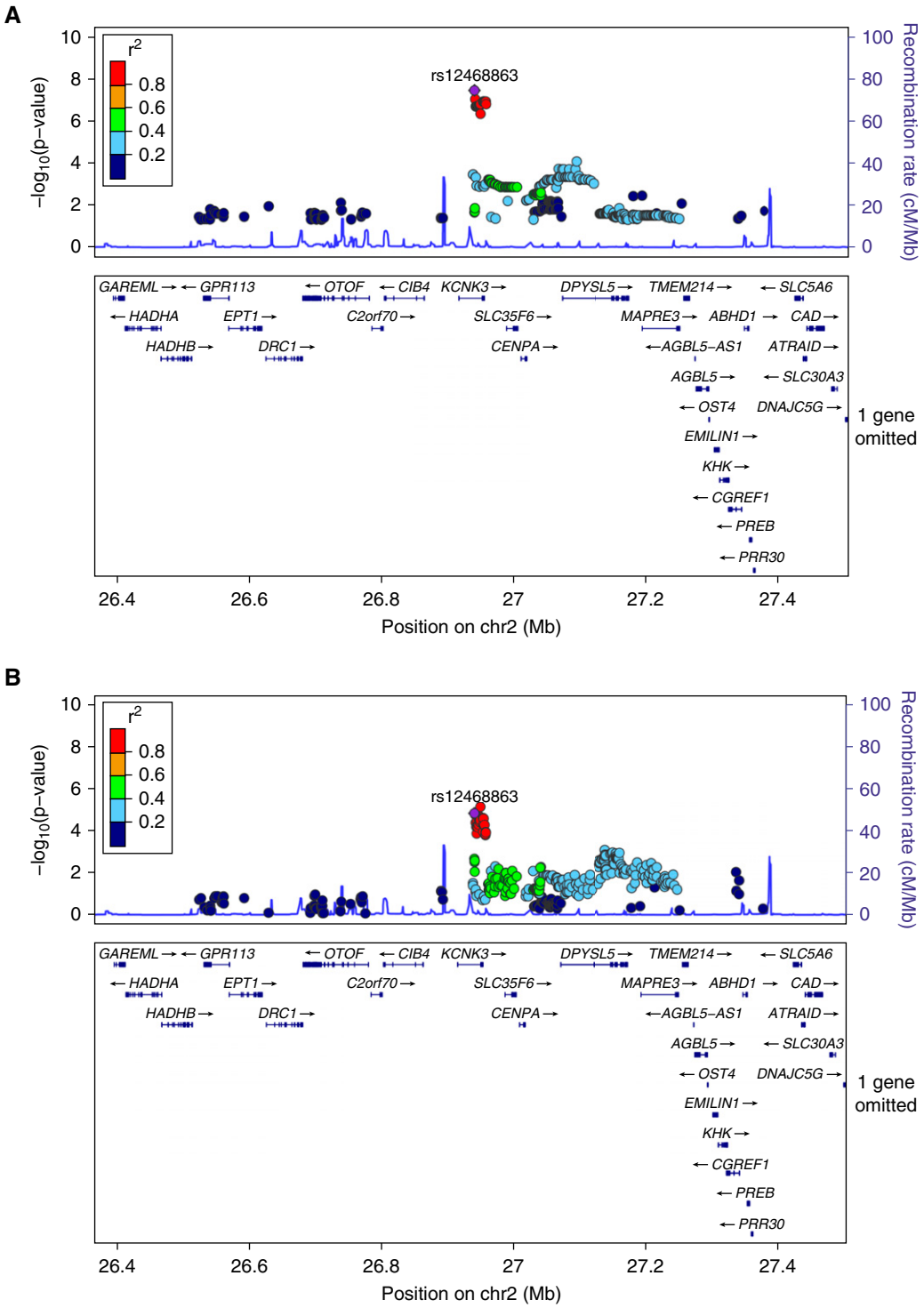

Figure 4. LocusZoom plots for the *KCNK3* colocalization analysis for cg11273176. (*A*) Methylation quantitative trait loci (mQTL) results. (*B*) Genome-wide association study (GWAS) results. Plots of the mQTL and GWAS *P* values in the region surrounding *KCNK3* are shown in LocusZoom format; the plotted regions flank *KCNK3* by 550,000 bases. The purple circle is the marker for the rsID 12468863 (rs12468863). chr = chromosome; rsID = reference SNP ID number.

enrichment of mQTL in the subthreshold GWAS results, as previously observed in other complex diseases (12, 45) and observed with expression QTLs (47). The associations between DNA methylation and gene expression for *KCNK3* (location: gene body), *SULT1A2* (location: 5′UTR), and *NUPR1* (location: transcription start site) provide a compelling demonstration of gene regulation. Although hypermethylation and concomitant reduction in gene expression for the

KCNK3 site in the gene body is perhaps the opposite of what might be predicted (3), the complex regulatory effects of DNA methylation are still not fully understood (2, 48). Together, our data support disease relevance for DNA methylation at regions of gene regulatory impact.

The limitations of our study involve the potential contribution of cellular heterogeneity that is inherent in experiments with homogenized lung tissue. Future validation of these findings in single lung cell types will address this issue. As technology advances, we will investigate DNA methylation regulatory effects *in vitro* to provide insight into causal mechanisms. In addition, our study focused on severe COPD. The study of other levels of severity or related COPD phenotypes, including emphysema and airway disease, will provide greater context for these findings. Last, distant normal tissue was sourced from lung nodule resection samples; although there is concern for a potential field effect from cancerous nodules, no enrichment for cancer-related gene sets was observed in our pathway analyses, which argues against such bias.

Epigenetic studies complement genetic association studies to identify COPD pathogenesis genes. Our study has provided regulatory characterization of loci among significant and subthreshold GWAS results. In the case of *EEFSEC*, we provided molecular justification to elevate a prior subthreshold result that has only recently been validated by virtue of its proximity to a genome-wide significant locus in a large study. Other genes identified near subthreshold loci, such as *KCNK3*, could rise to significance in larger GWAS of lung disease, as colocalization has provided compelling evidence for this gene. As demonstrated for the *CHRNA5* locus from prior COPD GWAS, the methods presented in this study may identify causal variants and reveal regulatory mechanisms in regions with loci known to be associated with COPD. Colocalization of methylation QTL and GWAS loci provides regulatory characterization of significant and subthreshold GWAS, supporting a role for genetic control of methylation in COPD pathogenesis. ■

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

Acknowledgment: The authors thank Drs. Amund Gulsvik, Per Bakke, Augusto Litonjua, Pantel Vokonas, and Ruth Tal-Singer, and the GenKOLS (Genetics of Chronic Obstructive Lung Disease), NETT (National Emphysema Treatment Trial)/NAS (Normative Aging Study), ECLIPSE (Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points), and COPDGene studies for use of their genome-wide association study meta-analysis data. The COPDGene study (R01 HL089856 and R01 HL089897) (NCT00608764) was funded by the NIH and is also supported by the COPD Foundation through contributions made to an Industry Advisory Board composed of AstraZeneca, Boehringer Ingelheim, Novartis, Pfizer, Siemens, GSK, and Sunovion. NETT was supported by grants NHLBI N01HR76101, N01HR76102, N01HR76103, N01HR76104, N01HR76105, N01HR76106, N01HR76107, N01HR76108, N01HR76109, N01HR76110, N01HR76111, N01HR76112, N01HR76113, N01HR76114, N01HR76115, N01HR76116, N01HR76118, and N01HR76119, the Centers for Medicare and Medicaid Services, and the Agency for Healthcare Research and Quality. NAS is supported by the Cooperative Studies Program/ERIC of the U.S. Department of Veterans Affairs and is a component of the Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC). The Norway GenKOLS study (GSK code RES11080), the ECLIPSE study (NCT00292552; GSK code SCO104960), and the ICGN study were funded by GlaxoSmithKline.

References

1. Vestbo J, Hurd SS, Agustí AG, Jones PW, Vogelmeier C, Anzueto A, *et al*. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* 2013;187:347–365.
2. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, *et al*.; Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. *Nature* 2015;518:317–330.
3. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* 2012;13:484–492.
4. Meier K, Recillas-Targa F. New insights on the role of DNA methylation from a global view. *Front Biosci* 2017;22:644–668.
5. Qiu W, Wan E, Morrow J, Cho MH, Crapo JD, Silverman EK, *et al*. The impact of genetic variation and cigarette smoke on DNA methylation in current and former smokers from the COPDGene study. *Epigenetics* 2015;10:1064–1073.
6. Wan ES, Qiu W, Baccarelli A, Carey VJ, Bacherman H, Rennard SI, *et al*. Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome. *Hum Mol Genet* 2012;21:3073–3082.
7. Wan ES, Qiu W, Carey VJ, Morrow J, Bacherman H, Foreman MG, *et al*. Smoking-associated site-specific differential methylation in buccal mucosa in the COPDGene study. *Am J Respir Cell Mol Biol* 2015;53:246–254.
8. Wan ES, Qiu W, Baccarelli A, Carey VJ, Bacherman H, Rennard SI, *et al*. Systemic steroid exposure is associated with differential methylation in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2012;186:1248–1255.
9. Bjornsson HT, Fallin MD, Feinberg AP. An integrated epigenetic and genetic approach to common human disease. *Trends Genet* 2004;20: 350–358.
10. Feinberg AP. Epigenomics reveals a functional genome anatomy and a new approach to common disease. *Nat Biotechnol* 2010;28: 1049–1052.
11. Olsson AH, Volkov P, Bacos K, Dayeh T, Hall E, Nilsson EA, *et al*. Genome-wide associations between genetic and epigenetic variation influence mRNA expression and insulin secretion in human pancreatic islets. *PLoS Genet* 2014;10:e1004735.
12. Hannon E, Dempster E, Viana J, Burrage J, Smith AR, Macdonald R, *et al*. An integrated genetic-epigenetic analysis of schizophrenia: evidence for co-localization of genetic associations and differential DNA methylation. *Genome Biol* 2016;17:176.
13. Kato N, Loh M, Takeuchi F, Verweij N, Wang X, Zhang W, *et al*.; BIOS-consortium; CARDIo GRAMplusCD; LifeLines Cohort Study; InterAct Consortium. Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation. *Nat Genet* 2015;47:1282–1293.
14. Volkov P, Olsson AH, Gillberg L, Jørgensen SW, Brøns C, Eriksson K-F, *et al*. A genome-wide mQTL analysis in human adipose tissue identifies genetic variants associated with DNA methylation, gene expression and metabolic traits. *PLoS One* 2016;11:e0157776.
15. Bonder MJ, Luijk R, Zhernakova DV, Moed M, Deelen P, Vermaat M, *et al*.; BIOS Consortium. Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet* 2017;49: 131–138.
16. Banovich NE, Lan X, McVicker G, van de Geijn B, Degner JF, Blischak JD, *et al*. Methylation QTLs are associated with coordinated changes in transcription factor binding, histone modifications, and gene expression levels. *PLoS Genet* 2014;10:e1004663.
17. Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Pique-Regi R, Degner JF, *et al*. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol* 2011;12:R10.
18. Shakhbazov K, Powell JE, Hemani G, Henders AK, Martin NG, Visscher PM, *et al*. Shared genetic control of expression and methylation in peripheral blood. *BMC Genomics* 2016;17:278.
19. Chen L, Ge B, Casale FP, Vasquez L, Kwan T, Garrido-Martín D, *et al*. Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell* 2016;167:1398–1414.e24.

20. Cheung WA, Shao X, Morin A, Siroux V, Kwan T, Ge B, *et al*. Functional variation in allelic methylomes underscores a strong genetic contribution and reveals novel epigenetic alterations in the human epigenome. *Genome Biol* 2017;18:50.
21. Cheng L, Liu J, Li B, Liu S, Li X, Tu H. Cigarette smoke-induced hypermethylation of the GCLC gene is associated with COPD. *Chest* 2016;149:474–482.
22. Morrow JD, Cho MH, Hersh CP, Pinto-Plata V, Celli B, Marchetti N, *et al*. DNA methylation profiling in human lung tissue identifies genes associated with COPD. *Epigenetics* 2016;11:1–10.
23. Vucic EA, Chari R, Thu KL, Wilson IM, Cotton AM, Kennett JY, *et al*. DNA methylation is globally disrupted and associated with expression changes in chronic obstructive pulmonary disease small airways. *Am J Respir Cell Mol Biol* 2014;50:912–922.
24. Buro-Auriemma LJ, Salit J, Hackett NR, Walters MS, Strulovici-Barel Y, Staudt MR, *et al*. Cigarette smoking induces small airway epithelial epigenetic changes with corresponding modulation of gene expression. *Hum Mol Genet* 2013;22:4726–4738.
25. Shi J, Marconett CN, Duan J, Hyland PL, Li P, Wang Z, *et al*. Characterizing the genetic basis of methylome diversity in histologically normal human lung tissue. *Nat Commun* 2014;5:3365.
26. Yoo S, Takikawa S, Geraghty P, Argmann C, Campbell J, Lin L, *et al*. Integrative analysis of DNA methylation and gene expression data identifies EPAS1 as a key regulator of COPD. *PLoS Genet* 2015;11: e1004898.
27. Cho MH, McDonald M-LN, Zhou X, Mattheisen M, Castaldi PJ, Hersh CP, *et al*.; NETT Genetics, ICGN, ECLIPSE and COPDGene Investigators. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *Lancet Respir Med* 2014;2: 214–225.
28. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, *et al*. Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet* 2014;10:e1004383.
29. Wang X, Tucker NR, Rizki G, Mills R, Krijger PH, de Wit E, *et al*. Discovery and validation of sub-threshold genome-wide association study loci using epigenomic signatures. *eLife* 2016;5:e10557.
30. Morrow JD, Cho MH, Chase RP, Hersh CP, Castaldi PJ, Pinto-Plata VM, *et al*. Genetic and epigenetic colocalizing associations identify potential COPD genes [abstract]. *Am J Respir Crit Care Med* 2017; 195:A2448.
31. Morrow JD, Zhou X, Lao T, Jiang Z, DeMeo DL, Cho MH, *et al*. Functional interactors of three genome-wide association study genes are differentially expressed in severe chronic obstructive pulmonary disease lung tissue. *Sci Rep* 2017;7:44232.
32. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 2012;28:1353–1358.
33. Westra H-J, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, *et al*. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* 2013;45:1238–1243.
34. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, *et al*. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* 2005;21: 3439–3440.
35. Guo H, Fortune MD, Burren OS, Schofield E, Todd JA, Wallace C. Integration of disease association and eQTL data using a Bayesian colocalisation approach highlights six candidate causal genes in immune-mediated diseases. *Hum Mol Genet* 2015;24:3305–3313.
36. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, *et al*. High density DNA methylation array with single CpG site resolution. *Genomics* 2011;98:288–295.
37. Kamburov A, Stelzl U, Lehrach H, Herwig R. The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res* 2013;41: D793–D800.
38. Maurano MT, Wang H, John S, Shafer A, Canfield T, Lee K, *et al*. Role of DNA methylation in modulating transcription factor occupancy. *Cell Reports* 2015;12:1184–1195.
39. McDonald M-LN, Mattheisen M, Cho MH, Liu Y-Y, Harshfield B, Hersh CP, *et al*.; GenKOLS, COPDGene and ECLIPSE study investigators. Beyond GWAS in COPD: probing the landscape between gene-set associations, genome-wide associations and protein-protein interaction networks. *Hum Hered* 2014;78: 131–139.
40. Hobbs BD, de Jong K, Lamontagne M, Bossé Y, Shrine N, Artigas MS, *et al*.; COPDGene Investigators; ECLIPSE Investigators; LifeLines Investigators; SPIROMICS Research Group; International COPD Genetics Network Investigators; UK BiLEVE Investigators; International COPD Genetics Consortium. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet* 2017;49: 426–432.
41. Wain LV, Shrine N, Artigas MS, Erzurumluoglu AM, Noyvert B, Bossini-Castillo L, *et al*.; Understanding Society Scientific Group; Geisinger-Regeneron DiscovEHR Collaboration. Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat Genet* 2017;49: 416–425.
42. Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, *et al*.; ICGN Investigators. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet* 2009;5:e1000421.
43. Lutz SM, Cho MH, Young K, Hersh CP, Castaldi PJ, McDonald M-L, *et al*. A genome-wide association study identifies risk loci for spirometric measures among smokers of European and African ancestry. *BMC Genet* 2015;16:138.
44. Hobbs BD, Parker MM, Chen H, Lao T, Hardin M, Qiao D, *et al*.; NETT Genetics Investigators; ECLIPSE Investigators; COPDGene Investigators; International COPD Genetics Network Investigators. Exome array analysis identifies a common variant in *IL27* associated with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2016; 194:48–57.
45. Gaunt TR, Shihab HA, Hemani G, Min JL, Woodward G, Lyttleton O, *et al*. Systematic identification of genetic influences on methylation across the human life course. *Genome Biol* 2016;17:61.
46. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, *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.
47. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet* 2010;6:e1000888.
48. van Eijk KR, de Jong S, Boks MPM, Langeveld T, Colas F, Veldink JH, *et al*. Genetic analysis of DNA methylation and gene expression levels in whole blood of healthy human subjects. *BMC Genomics* 2012;13:636.