Original investigation

Common and Rare Variants Genetic Association Analysis of Cigarettes per Day Among Ever-Smokers in Chronic Obstructive Pulmonary Disease Cases and Controls

Sharon M. Lutz PhD¹, Brittnei Frederiksen PhD², Ferdouse Begum PhD³, Merry-Lynn N. McDonald PhD⁴, Michael H. Cho MD^{4,5}, Brian D. Hobbs MD^{4,5}, Margaret M. Parker PhD⁴, Dawn L. DeMeo MD^{4,5}, Craig P. Hersh MD^{4,5}, Marissa A. Ehringer PhD⁶, Kendra Young PhD², Lai Jiang PhD³, Marilyn G. Foreman MD⁷, Greg L. Kinney PhD², Barry J. Make MD⁸, David A. Lomas MD⁹, Per Bakke PhD¹⁰, Amund Gulsvik PhD¹⁰, James D. Crapo MD⁸, Edwin K. Silverman MD^{4,5}, Terri H. Beaty PhD³, John E. Hokanson PhD²; the ECLIPSE and COPDGene Investigators

¹Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO; ²Department of Epidemiology, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO; ³Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; ⁴Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; ⁵Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; ⁶Institute for Behavioral Genetics, University of Colorado at Boulder, Boulder, CO; ⁷Morehouse School of Medicine, Atlanta, GA; ⁸Division of Pulmonary, Critical Care and Sleep Medicine, National Jewish Health, Denver, CO; ⁹UCL Respiratory, University College London, London, UK; ¹⁰Department of Clinical Science, Faculty of Medicine, University of Bergen, Bergen, Norway

Corresponding Author: Sharon M. Lutz, PhD, Department of Biostatistics, University of Colorado Anschutz Medical Campus, 13001 E. 17th Place, B119 Bldg. 500, Room W3128, Aurora, CO 80045, USA; E-mail: sharon.lutz@ucdenver.edu

Abstract

Introduction: Cigarette smoking is a major environmental risk factor for many diseases, including chronic obstructive pulmonary disease (COPD). There are shared genetic influences on cigarette smoking and COPD. Genetic risk factors for cigarette smoking in cohorts enriched for COPD are largely unknown.

Methods: We performed genome-wide association analyses for average cigarettes per day (CPD) across the Genetic Epidemiology of COPD (COPDGene) non-Hispanic white (NHW) ($n = 6659$) and African American (AA) ($n = 3260$), GenKOLS (the Genetics of Chronic Obstructive Lung Disease) ($n = 1671$), and ECLIPSE (the Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints) ($n = 1942$) cohorts. In addition, we performed exome array association analyses across the COPDGene NHW and AA cohorts. We considered analyses across the entire cohort and stratified by COPD case-control status.

Results: We identified genome-wide significant associations for CPD on chromosome 15q25 across all cohorts (lowest $p = 1.78 \times 10^{-15}$), except in the COPDGene AA cohort alone. Previously reported associations on chromosome 19 had suggestive and directionally consistent associations (*RAB4*, $p = 1.95 \times 10^{-6}$; *CYP2A7*, $p = 7.50 \times 10^{-5}$; *CYP2B6*, $p = 4.04 \times 10^{-4}$). When we stratified by COPD case-control status, single nucleotide polymorphisms on chromosome 15q25 were nominally associated with both NHW COPD cases ($\beta = 0.11$, $p = 5.58 \times 10^{-4}$) and controls ($\beta = 0.12$, $p = 3.86 \times 10^{-5}$). For the gene-based exome array association analysis of rare variants, there were no exome-wide significant associations. For these previously replicated associations, the most significant results were among COPDGene NHW subjects for *CYP2A7* ($p = 5.2 \times 10^{-4}$).

Conclusions: In a large genome-wide association study of both common variants and a gene-based association of rare coding variants in ever-smokers, we found genome-wide significant associations on chromosome 15q25 with CPD for common variants, but not for rare coding variants. These results were directionally consistent among COPD cases and controls.

Implications: We examined both common and rare coding variants associated with CPD in a large population of heavy smokers with and without COPD of NHW and AA descent. We replicated genome-wide significant associations on chromosome 15q25 with CPD for common variants among NHW subjects, but not for rare variants. We demonstrated for the first time that common variants on chromosome 15q25 associated with CPD are similar among COPD cases and controls. Previously reported associations on chromosome 19 showed suggestive and directionally consistent associations among common variants (*RAB4*, *CYP2A7*, and *CYP2B6*) and for rare variants (*CYP2A7*) among COPDGene NHW subjects. Although the genetic effect sizes for these single nucleotide polymorphisms on chromosome 15q25 are modest, we show that this creates a substantial smoking burden over the lifetime of a smoker.

Introduction

The most recent Surgeon General's report "The Health Consequences of Smoking—50 Years of Progress" notes tobacco continues to be an immense public health burden in the United States, and tobacco smoking has killed more than 20 million people prematurely since the first Surgeon General's report on smoking in 1964.¹ The prevalence of current cigarette smoking among adults in the United States has declined from 42% in 1965 to 18% in 2012; however, more than 42 million Americans still smoke. The findings in this recent Surgeon General's report show the decline in the prevalence of smoking has slowed in recent years, but the burden of mortality related to smoking is expected to remain at high levels for decades to come.¹

Smoking is known to have both environmental and genetic influences. The heritability of smoking behaviors has been estimated between 37% and 59%.² Genetic studies of smoking have consistently identified an association with chromosome 15q25, which includes a cluster of genes coding for nicotinic acetylcholine receptor subunits, *CHRNA5-CHRNA3-CHRNA4*. Single nucleotide polymorphisms (SNPs) on chromosome 15q25 have been associated with smoking quantity and nicotine dependence among subjects of European ancestry,^{3–10} and African Americans (AAs).¹¹ Chromosome 15q25 has also been associated with spirometric measures of pulmonary function,¹² and respiratory diseases such as chronic obstructive pulmonary disorder (COPD)^{13,14} lung cancer,^{15–18} and time to lung cancer diagnosis¹⁹ where smoking may mediate the relationship between these genes and clinically relevant outcomes.^{20,21} Recently, rare coding variants in *CHRNA5* were marginally associated with increased risk of nicotine dependence among European Americans (odds ratio [OR] = 12.9, $p = .01$) and in the same risk direction among AAs (OR = 1.5, $p = .37$).²² In addition to chromosome 15q25, common variants in several other regions have been associated with smoking-related traits such as chromosome 1 (*LPPR5*), chromosome

2 (*TEX41/PABPC1P2*), chromosome 6 (*DNAH8*), chromosome 7 (*PDE1C*), chromosome 8 (*CHRNA3/CHRNA6*), chromosome 9 (*DBH*), chromosome 10 (*LOC1001889*), chromosome 11 (*BDNF*, *NCAM1*), chromosome 19 (*RAB4B*, *EGLN2*, *CYP2A7*, *CYP2B6*), and chromosome 20 (*NOL4L*).²³

The associations between SNPs on chromosome 15q25 and smoking-related diseases such as COPD and lung cancer are partially mediated through smoking.^{24,25} It is not clear if the impact of chromosome 15q25 SNPs on cigarettes per day (CPD) differs among individuals susceptible to the adverse clinical effects of cigarette smoking. It is possible that the genetic susceptibility to smoking-related diseases may be due to a stronger genetic influence on CPD that subsequently leads to increased disease risk. Given the observed relationship between chromosome 15q25 with COPD and CPD, we performed a genome-wide association (GWA) analysis of CPD in the Genetic Epidemiology of COPD (COPDGene) study, a large population of current and former smokers and a meta-analysis across the COPDGene, the Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE), and the Genetics of Chronic Obstructive Lung Disease (GenKOLS) studies, which were all enriched for COPD cases. To examine potential differences in the genetic susceptibility to CPD among COPD cases and controls, we examined the genome-wide significant SNPs for CPD stratified by COPD case-control status and formally tested for interactions between SNPs and COPD on CPD.

Methods

COPDGene Study

The COPDGene study is a multicenter observational study designed to identify genetic factors associated with COPD and to characterize COPD-related phenotypes.²⁶ This study recruited 10 192 adult

Table 1. Characteristics of COPDGene, ECLIPSE, and GenKOLS Subjects Included in Genome-Wide Association Analysis

	COPDGene NHW	COPDGene AA	ECLIPSE	GenKOLS
Sample size (no. of COPD cases)	6658 (2812)	3280 (821)	1942 (1764)	1671 (863)
Sex (%male)	47.73	44.79	66.17	58.71
Age (y)	62.0 (8.9)	54.6 (7.2)	63.1 (7.6)	60.7 (11.1)
Pack-years	47.2 (26.0)	38.3 (21.6)	48.6 (27.7)	26.0 (17.4)
Cigarettes per day	25.8 (11.4)	21.3 (10.4)	25.3 (12.7)	14.7 (7.6)
Smoking initiation age (years)	16.9 (4.3)	17.0 (5.3)	16.9 (4.4)	18.5 (5.0)
Current smoking (%)	38.9	80.0	34.8	43.8

COPD = chronic obstructive pulmonary disorder, COPDGene = Genetic Epidemiology of COPD, ECLIPSE = Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints, NHW = non-Hispanic white. For continuous variables, the mean is given first followed by the standard deviation.

smokers (current and former) who were non-Hispanic whites (NHWs) or AAs aged 44–81 years with at least 10 pack-years of smoking history. Table 1 details characteristics of the COPDGene participants included in the GWA analysis. We excluded subjects with severe α -1 antitrypsin deficiency or genotyping failure, which resulted in 9978 unrelated subjects. There were a total of 6659 NHWs and 3260 AAs with complete genotype and phenotype data available. There were 2819 NHW COPD cases (GOLD stage 2–4), 2543 NHW COPD controls (GOLD stage 0), 821 AA cases (GOLD stage 2–4), and 1749 AA controls (GOLD stage 0).

Meta-Analysis Study Populations: ECLIPSE and GenKOLS

The ECLIPSE study was a longitudinal, observational study conducted at 46 clinical centers in 12 countries.²⁷ For this study, we included 1764 COPD cases and 178 current or former smoking controls from the ECLIPSE study; we also included 863 COPD cases and 808 controls from the GenKOLS study in Bergen, Norway. For both cohorts, the COPD controls were chosen from the cohort as GOLD stage 0 subjects who were ever-smokers. In all cohorts, COPD cases were chosen as GOLD stage 2–4 COPD subjects (moderate, severe, and very severe COPD; post-bronchodilator $FEV_1 < 80\%$ predicted with $FEV_1/FVC < 0.7$). Individuals with severe α -1 antitrypsin deficiency were excluded. Both the ECLIPSE and GenKOLS cohorts were of NHW ancestry. Genotyping methods for the genome-wide SNP panels and study descriptions for the GenKOLS and ECLIPSE cohorts have been described previously.^{14,28}

Comparison to Previously Published GWA Studies: UK BiLEVE Study

We considered genome-wide significant results from a previously published genome-wide analysis of smoking-related traits in the UK Biobank Lung Exome Variant Evaluation (BiLEVE) study.²³ We chose this UK BiLEVE analysis because the study used similar enrollment criteria. This study focused on individuals of European ancestry from the UK Biobank with the middle and extremes of the FEV_1 distribution among heavy smokers (mean 35 pack-years) and never-smokers. Heavy smokers were defined as individuals with a percentage of life span smoking of at least 42% (equivalent to a minimum pack-years of 10 in the youngest participants). The study had 50 008 subjects: 10 002 individuals with low FEV_1 , 10 000 with average FEV_1 , and 5 002 with high FEV_1 from each of the heavy smoker and never-smoker groups. Genotyping was done on a custom Affymetrix Axiom array (UK BiLEVE array; Santa Clara, CA) and non-genotyped variants were imputed using a combined 1000 Genomes Project Phase and UK10K Project reference panel. Further

details regarding the genotyping methods and study design have been described previously.²⁹

Phenotypes of Smoking-Related Traits

Due to non-normality, we categorized CPD into seven categories (eg, 1 if CPD is ≤ 10 , up to 7 if the CPD is ≥ 60 , and intermediate numbers 2–6 for each increment of 10). Supplementary Figure 2a–c shows histograms of CPD and CPD binned to show that while right skewed the distribution of CPD binned has more of a bell curve than CPD. This is due to the way CPD is reported as subjects tend to report CPD values that are divisible by 5 or 10. We also considered pack-years of cigarettes smoked, but due to the similarity and high correlation ($r = 0.92$) with CPD, we did not present these results here. We also considered age of smoking initiation and current smoking status but did not find any significant results for these phenotypes. As a result, we did not present these results here.

Genotyping, Quality Control, and Imputation for Common Variants in the COPDGene Study

All COPDGene subjects were genotyped using the Illumina HumanOmniExpress-12 v1 by Illumina (San Diego, CA). Details of genotyping quality control have been previously described.¹³ Imputation on the COPDGene cohorts was performed using MaCH and minimac.^{30,31} Prephasing and imputation were both performed using 30 rounds and 200 states, with regions divided into 1 Mb segments with 500-kb flanks. Reference panels for the NHW and AA subjects were the 1000 Genomes Phase I (v3) European (EUR) and cosmopolitan reference panels, respectively.³² Imputed variants with an R -squared value of less than 0.3 were dropped from further analysis. SNPs with minor allele frequency less than 1% were excluded. Further details concerning genotyping, quality control, and imputation are posted on the COPDGene Web site (www.copdgene.org). All SNP locations are based on the NCBI37/hg19 assembly.

Genotyping and Quality Control for Exome Chip Variants in the COPDGene Study

Genotyping with the exome chip was done for this same COPDGene study cohort on Illumina Human exome array (v1.1 and v1.2). Standard QC protocols were followed to clean called variants on this exome array data as detailed previously.²⁹

Statistical Analyses for Common Variants

GWA tests were performed in Plink (v.1.07, Boston, MA) stratified by race.³³ Linear regression analyses of CPD were adjusted for age, sex, and genetic ancestry (as summarized by principal components computed within racial group) by including these covariates in the model.³⁴

We chose this set of precision variables and confounders to match other GWA studies. We assumed additive model for each SNP (ie, 0,1,2 for 0,1,2 copies of the minor allele, respectively). For the primary analysis, we did not adjust for current smoking status due to the correlation with CPD ($r = -0.19$). For [Supplementary Tables 1–3](#), we re-ran all analyses adjusting for current smoking status in addition to age, sex, and genetic ancestry (as summarized by principal components computed within racial group) by including these covariates in the model. The results summarized in [Supplementary Tables 1–3](#) are very similar to those shown in [Tables 2–5](#). A fixed-effects meta-analysis was performed using METAL (v 2010-08-01, Ann Arbor, MI)³⁵ for CPD adjusting for the same covariates mentioned earlier (age, sex, and genetic ancestry as summarized by principal components) for the COPDGene, ECLIPSE, and GenKOLS cohorts. We used a fixed-effects meta-analysis because in the presence of between-study heterogeneity, fixed-effects models are known to produce tighter confidence intervals than random-effects models.³⁶ Genome-wide significant associations were defined by $p < 5 \times 10^{-8}$. The power analysis for these linear regressions that examines the common variants associated with CPD in the NHW COPDGene, AA COPDGene, and meta-analysis cohorts is included in [Supplementary Table 3a](#).

In addition, using SNPs from [Table 2](#), we stratified by COPD case–control status (GOLD spirometry stage 2–4 for cases and GOLD spirometry stage 0 for controls). GOLD stage 1 subjects are an indeterminate group regarding COPD status, meeting spirometric criteria for mild COPD but likely also including normal

older subjects, because FEV₁/FVC typically falls with age. Because we wanted to focus on subjects with definite COPD or normal spirometry, we excluded GOLD stage 0 and PRISM subjects.³⁷ The power analysis for these methods that examine the common variants associated with CPD stratified by COPD case–control status are in [Supplementary Table 3b](#). In addition, we tested for an SNP by COPD interaction on CPD. The power analysis for these linear regressions that examines the SNP by COPD interaction on CPD is given in [Supplementary Table 3c](#). Using the SNPs from [Table 2](#), we also stratified by sex, and in addition, tested for an SNP by sex interaction on CPD. We did not find any significant sex differences of SNPs for CPD, and the results of these analyses are in [Supplementary Tables 4a and b](#).

Statistical Analyses for Exome Array

Using data from the exome array, variants with minor allele frequency value of less than 5% and minimum minor allele count of 5 were collapsed into gene sets and analyzed using SKAT-O³⁸ separately for NHW and AA COPDGene subjects. These analyses were restricted to putatively functional variants based on GENCODE v14 annotations via EPACTS where the annotations were nonsynonymous, splice site, or stop gain or loss.²⁹ The same outcomes and covariates were used for both the rare and common variant analyses with the addition of a covariate for exome array platform (v1.1 vs. v1.2) in the NHW analyses and the inclusion of exome array-specific ancestry-based principal components. Among the AA and

Table 2. Genome-Wide Significant Results for CPD and Comparison of Previously Replicated Regions From the UK BiLEVE Study²³

												Meta-analysis of COPDGene NHW, COPDGene AA, ECLIPSE, and GenKOLS (<i>n</i> = 13551)		
					COPDGene NHW (<i>n</i> = 6658)			COPDGene AA (<i>n</i> = 3260)						
	SNP	Position (bp)	CHR	Nearest gene	Coded allele	Allele freq	β	<i>p</i>	Allele freq	β	<i>p</i>	Allele freq	β	<i>p</i>
Chromosome 15q25	rs8192482	78886198	15	<i>CHRNA3</i>	T/T/T	0.37	0.13	5.37E-11	0.06	0.05	.41	0.37	0.12	1.78E-15
	rs11633958	78866445	15	<i>CHRNA5</i>	T/T/T	0.37	0.13	9.41E-11	0.06	0.04	.46	0.37	0.12	2.18E-15
	rs72738786	78828086	15	<i>AGPHD1</i>	T/T/T	0.37	0.13	3.41E-11	0.14	0.02	.61	0.37	0.12	5.78E-15
	rs2869548	78922638	15	<i>CHRNA4</i>	A/A/A	0.38	0.12	2.17E-10	0.06	0.05	.31	0.39	0.12	2.82E-14
	rs11858836	78783277	15	<i>IREB2</i>	A/A/A	0.36	0.11	1.42E-8	0.07	0.01	.83	0.36	0.10	1.26E-10
Replicated regions from the UK BiLEVE study	rs61784651	99445471	1	<i>LPPR5</i>	T/T/T	0.16	-0.03	0.27	0.06	-0.08	.15	0.16	0.02	0.32
	rs10193706	146316319	2	<i>TEX41/ PABPC1P2</i>	A/C/A	0.46	0.01	0.46	0.19	0.01	.96	0.51	0.01	0.4
	rs10807199	38901867	6	<i>DNAH8</i>	T/T/T	0.46	0.01	0.86	0.15	0.02	.65	0.47	-0.01	0.56
	rs215605	32336965	7	<i>PDE1C</i>	G/T/T	0.38	0.02	0.21	0.28	0.01	.83	0.37	-0.01	0.43
	rs13280604	42559586	8	<i>CHRNA3</i>	G/A/A	0.22	0.04	0.27	0.45	-0.06	.03	0.78	-0.02	0.24
	rs3025343	136478355	9	<i>DBH</i>	A/A/A	0.12	-0.02	0.52	0.02	0.07	.40	0.12	0.001	0.99
	rs1329650	93348120	10	<i>LOC100188947</i>	T/T/T	0.27	-0.01	0.58	0.10	0.02	.63	0.28	-0.03	0.12
	rs6265	27679916	11	<i>BDNF</i>	T/T/T	0.18	-0.02	0.48	0.04	-0.02	.75	0.18	-0.02	0.39
	rs4466874	112861434	11	<i>NCAM1</i>	C/C/T	0.41	-0.06	9.34E-4	0.41	0.05	.06	0.58	0.03	0.02
	rs7937	41302706	19	<i>RAB4B</i>	C/T/T	0.43	-0.09	1.95E-6	0.30	0.03	.34	0.42	0.07	1.05E-5
	rs3733829	41310571	19	<i>EGLN2</i>	G/G/A	0.37	0.06	3.86E-3	0.08	0.03	.49	0.47	-0.04	0.01
	rs12461383	41370338	19	<i>CYP2A7</i>	C/G/C	0.47	-0.09	2.47E-4	0.23	0.04	.26	0.47	-0.09	7.50E-5
	rs7260329	41521638	19	<i>CYP2B6</i>	A/A/A	0.31	-0.07	4.04E-4	0.15	-0.04	.23	0.34	-0.05	8.20E-4
	rs4911243	31162568	20	<i>NOL4L</i>	A/G/A	0.33	0.04	0.03	0.17	-0.03	.46	0.35	0.02	0.18

AA = African American, CPD = cigarettes per day, COPD = chronic obstructive pulmonary disorder, COPDGene = Genetic Epidemiology of COPD, ECLIPSE = Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints, NHW = non-Hispanic white, SNP = single nucleotide polymorphism. The “Coded allele” column shows alleles for COPDGene NHW cohort, COPDGene AA cohort, the meta-analysis of COPDGene NHW, COPDGene AA, ECLIPSE and GenKOLS. Correcting for the 12 regions listed, all p values less than 4.17E-3 (0.05/12) are in bold. Note that due to the categorization of CPD, the β coefficients listed are for an average increase of 10 CPD.

NHW in COPDGene, 14 559 and 14 677 genes were included in the analysis, respectively, which generated a Bonferroni critical value of p equal to 3.4×10^{-6} for each cohort.

Results

Common Variants: NHWs and AAs in the COPDGene Study and a Meta-Analysis of COPDGene, ECLIPSE, and GenKOLS Cohorts

Among all NHW COPDGene participants, multiple SNPs within chromosome 15q25 reached genome-wide significance ($p < 5.0E-8$) for CPD as shown in Table 2. There were no genome-wide significant results for CPD among AA in the COPDGene cohort. For the meta-analysis of COPDGene, ECLIPSE, and GenKOLS cohorts, all genome-wide significant results were again within chromosome 15q25. Supplementary Figure 1 includes regional plots for CPD for this chromosome 15 region from the meta-analysis, among NHW and AA subjects in the COPDGene study. Note that there is a similar association within chromosome 15q25 for the meta-analysis and among NHW subjects in the COPDGene study, but there was no significant GWA between any SNPs within chromosome 15q25 and CPD among AA subjects in the COPDGene study.

Comparison to Previously Published GWA Studies

We considered genome-wide significant results from a previously published genome-wide analysis of smoking-related traits in the UK BiLEVE study.²³ Table 2 shows p values of these SNPs from the UK BiLEVE analyses in the COPDGene subpopulations analyzed above. Previously reported associations on chromosome 19 (*RAB4*, $p = 1.95 \times 10^{-6}$; *CYP2A7*, $p = 7.50 \times 10^{-5}$; *CYP2B6*, 4.04×10^{-4})

had suggestive and directionally consistent associations. For the exome array association analysis of rare variants in these previously replicated associations, the most significant results were among the COPDGene NHW subjects for *CYP2A7* ($p = 5.2 \times 10^{-4}$), as shown in Table 5. In particular, chromosome 15q25 did not reach exome chip significance with CPD among the NHW cohort (*CHRNA3*, $p = .17$; *CHRNA5*, $p = .06$; *AGPHD1*, $p = .27$; *CHRNA4*, $p = .87$; *IREB2*, $p = .27$) and the AA cohort (*CHRNA3*, $p = .52$; *CHRNA5*, $p = .30$; *AGPHD1*, $p = .02$; *CHRNA4*, $p = .87$; *IREB2*, $p = .02$).

CPD Genetic Associations by COPD

Case–Control Status

When we examined the SNPs listed in Table 2 stratified by COPD case–control status, SNPs on chromosome 15q25 were nominally associated in both NHW COPD cases ($\beta = 0.11$, $p = 5.58 \times 10^{-4}$) and controls ($\beta = 0.12$, $p = 3.86 \times 10^{-5}$) (Tables 3 and 4). However, on chromosome 19, there was a stronger but directionally consistent signal with CPD in COPD cases than controls for *RAB4* ($\beta = -0.13$ vs. $\beta = -0.05$) and there was a stronger but directionally consistent signal with CPD in COPD controls than cases for *CYP2B6* ($\beta = -0.10$ vs. $\beta = -0.03$). When we tested for an SNP by COPD interaction on CPD, none of these interactions in Tables 3 and 4 were significant. The SNP by COPD interaction on CPD for *RAB4* was marginally significant ($p = .07$), but the SNP by COPD interaction on CPD for *CYP2B6* was not significant ($p = .12$).

Rare Variants: Exome Genotyping in COPDGene

NHWs and AAs

Using a Bonferroni correction for the number of regions tested in the exome array, we considered a significance level of 3.4×10^{-6}

Table 3. SNPs From Table 2 With the Analysis Stratified by COPD Case–Control Status Among COPDGene NHW

	SNP	CHR	Nearest gene	Coded allele	COPDGene NHW COPD cases ($n = 2819$)			COPDGene NHW COPD controls ($n = 2534$)			Interaction of COPD and SNP on CPD ($n = 5353$)		
					Allele freq	β	p	Allele freq	β	p	Allele freq	β	p
Chromosome 15q25	rs8192482	15	<i>CHRNA3</i>	T/T/T/T	0.39	0.11	5.58E-4	0.34	0.12	3.86E-5	0.37	-0.02	.61
	rs11633958	15	<i>CHRNA5</i>	T/T/T/T	0.39	0.11	5.06E-4	0.34	0.12	5.34E-5	0.37	-0.02	.66
	rs72738786	15	<i>AGPHD1</i>	T/T/T/T	0.39	0.11	3.33E-4	0.34	0.12	5.81E-5	0.37	-0.01	.74
	rs2869548	15	<i>CHRNA4</i>	A/A/A/A	0.40	0.10	7.94E-4	0.36	0.11	2.81E-4	0.38	-0.01	.81
	rs11858836	15	<i>IREB2</i>	A/A/A/A	0.38	0.08	7.17E-4	0.33	0.11	2.36E-4	0.36	-0.03	.47
Replicated regions from the	rs61784651	1	<i>LPPR5</i>	T/T/T/T	0.15	-0.01	0.90	0.16	-0.05	0.24	0.15	0.04	.48
	rs10193706	2	<i>TEX41/</i> <i>PABPC1P2</i>	A/A/C/C	0.46	0.01	0.83	0.47	0.04	0.12	0.47	-0.04	.37
UK BiLEVE study	rs10807199	6	<i>DNAH8</i>	T/T/T/T	0.48	0.02	0.43	0.45	-0.04	0.14	0.46	0.07	.11
	rs215605	7	<i>PDE1C</i>	G/G/T/T	0.39	0.02	0.57	0.37	0.02	0.56	0.38	0.001	.98
	rs13280604	8	<i>CHRNA3</i>	G/G/A/A	0.21	0.05	0.15	0.22	0.02	0.53	0.22	0.03	.51
	rs3025343	9	<i>DBH</i>	A/A/A/A	0.12	-0.02	0.64	0.12	0.01	0.82	0.12	-0.03	.65
	rs1329650	10	<i>LOC100188947</i>	T/T/T/T	0.28	-0.01	0.79	0.27	-0.03	0.39	0.28	0.02	.72
	rs6265	11	<i>BDNF</i>	T/T/T/T	0.18	-0.02	0.64	0.19	-0.01	0.76	0.19	0.001	.93
	rs4466874	11	<i>NCAM1</i>	C/C/C/C	0.41	-0.07	0.02	0.40	-0.07	0.02	0.41	0.001	.94
	rs7937	19	<i>RAB4B</i>	C/C/T/T	0.40	-0.13	2.49E-5	0.45	-0.05	0.05	0.42	-0.08	.07
	rs3733829	19	<i>EGLN2</i>	G/G/G/G	0.38	0.08	0.02	0.36	0.06	0.04	0.37	0.02	.72
	rs12461383	19	<i>CYP2A7</i>	C/C/G/G	0.46	-0.1	9.63E-03	0.48	-0.06	0.09	0.47	-0.04	.46
	rs7260329	19	<i>CYP2B6</i>	A/A/A/A	0.30	-0.03	0.44	0.32	-0.1	1.19E-3	0.31	0.07	.12
	rs4911243	20	<i>NOL4L</i>	A/A/G/G	0.34	0.04	0.24	0.33	0.05	0.12	0.34	-0.01	.90

CHR = chromosome, CPD = cigarettes per day, COPD = chronic obstructive pulmonary disorder, COPDGene = Genetic Epidemiology of COPD, NHW = non-Hispanic whites, SNPs = single nucleotide polymorphisms. Correcting for the 12 regions listed, all p values less than 4.17E-3 (0.05/12) are in bold.

Table 4. SNPs From Table 2 With the Analysis Stratified by COPD Case–Control Status Among COPDGene AA

					COPDGene AA COPD cases (<i>n</i> = 821)			COPDGene AA COPD controls (<i>n</i> = 1749)			Interaction of COPD and SNP on CPD (<i>n</i> = 2570)		
	SNP	CHR	Nearest gene	Coded allele	Allele freq	β	Allele freq		Allele freq	β	Allele Freq	β	<i>p</i>
Chromosome 15q25	rs8192482	15	<i>CHRNA3</i>	T/T/T/T	0.06	0.08	0.06	0.08	0.06	0.08	0.06	0.06	.65
	rs11633958	15	<i>CHRNA5</i>	T/T/T/T	0.07	0.13	0.07	0.13	0.07	0.13	0.06	0.14	.26
	rs72738786	15	<i>AGPHD1</i>	T/T/T/T	0.14	0.07	0.14	0.07	0.14	0.07	0.14	0.05	.56
	rs2869548	15	<i>CHRNA4</i>	A/A/A/A	0.07	0.07	0.07	0.07	0.07	0.07	0.06	0.02	.86
Replicated regions from the UK BiLEVE study	rs11858836	15	<i>IREB2</i>	A/A/A/A	0.07	0.09	0.07	0.09	0.07	0.09	0.07	0.13	.29
	rs61784651	1	<i>LPPR5</i>	T/T/T/T	0.05	-0.16	0.05	-0.16	0.05	-0.16	0.06	-0.10	.46
	rs10193706	2	<i>TEX41/PABPC1P2</i>	A/A/C/C	0.19	-0.02	0.19	-0.02	0.19	-0.02	0.19	-0.01	.93
	rs10807199	6	<i>DNAH8</i>	T/T/T/T	0.15	-0.01	0.15	-0.01	0.15	-0.01	0.15	-0.03	.71
	rs215605	7	<i>PDE1C</i>	G/G/T/T	0.28	0.03	0.28	0.03	0.28	0.03	0.28	-0.02	.79
	rs13280604	8	<i>CHRNA3</i>	G/G/A/A	0.27	-0.08	0.27	-0.08	0.27	-0.08	0.28	0.05	.47
	rs3025343	9	<i>DBH</i>	A/A/A/A	0.03	0.06	0.03	0.06	0.03	0.06	0.02	-0.05	.81
	rs1329650	10	<i>LOC100188947</i>	T/T/T/T	0.11	0.01	0.11	0.01	0.11	0.01	0.10	-0.02	.82
	rs6265	11	<i>BDNF</i>	T/T/T/T	0.04	-0.08	0.04	-0.08	0.04	-0.08	0.04	-0.01	.94
	rs4466874	11	<i>NCAM1</i>	C/C/C/C	0.39	0.11	0.39	0.11	0.39	0.11	0.41	0.08	.22
	rs7937	19	<i>RAB4B</i>	C/C/T/T	0.31	0.01	0.31	0.01	0.31	0.01	0.31	0.001	1.00
	rs3733829	19	<i>EGLN2</i>	G/G/G/G	0.09	0.03	0.09	0.03	0.09	0.03	0.08	-0.01	.94
rs12461383	19	<i>CYP2A7</i>	C/C/G/G	0.23	0.07	0.23	0.07	0.23	0.07	0.23	0.07	.34	
rs7260329	19	<i>CYP2B6</i>	A/A/A/A	0.16	-0.07	0.16	-0.07	0.16	-0.07	0.15	-0.03	.70	
rs4911243	20	<i>NOL4L</i>	A/A/G/G	0.16	0.02	0.16	0.02	0.16	0.02	0.17	0.13	.12	

AA = African American, CHR = chromosome, CPD = cigarettes per day, COPD = chronic obstructive pulmonary disorder, COPDGene = Genetic Epidemiology of COPD, SNPs = single nucleotide polymorphisms. Correcting for the 12 regions listed, all p values less than 4.17E-3 (0.05/12) are in bold.

Table 5. p Values for the Associations of Rare Variants in Each Gene With CPD Using the Genes From Table 2

	CHR	Nearest gene	COPDGene NHW			COPDGene AA		
			All (n = 6658)	Cases (n = 2819)	Controls (n = 2534)	All (n = 3260)	Cases (n = 821)	Controls (n = 1749)
Chromosome 15q25	15	CHRNA3	0.17	0.24	0.41	0.52	0.52	0.49
	15	CHRNA5	0.06	0.15	1.00	0.30	0.33	0.45
	15	AGPHD1	0.27	0.01	0.28	0.02	0.43	0.05
	15	CHRNA4	0.87	0.82	0.84	0.87	0.19	0.50
	15	IREB2	0.27	0.77	0.82	0.02	NA	0.15
Replicated regions from the UK	1	LPPR5	0.83	0.76	0.81	0.19	0.88	0.11
BiLEVE study	2	TEX41/PABPC1P2	0.56	0.85	0.22	0.64	1.00	0.33
	6	DNAH8	0.87	0.35	0.87	0.38	0.42	0.51
	7	PDE1C	1.00	0.05	0.14	0.60	0.32	0.74
	8	CHRNA3	0.64	1.00	0.65	0.86	NA	1.00
	9	DBH	0.52	0.64	0.50	0.70	1.00	0.45
	10	LOC100188947	0.87	0.79	0.71	0.11	0.51	0.69
	11	BDNF	1.00	0.76	0.46	0.26	0.48	0.56
	11	NCAM1	0.76	0.30	0.82	0.03	0.27	0.28
	19	RAB4B-EGLN2	0.44	0.69	0.57	0.04	0.15	0.27
	19	CYP2A7	5.2E-4	0.05	0.08	0.47	0.51	0.97
	19	CYP2B6	0.56	0.45	0.03	0.48	0.51	0.21
	20	NOL4L	0.32	0.55	0.89	0.84	NA	NA

AA = African American, CHR = chromosome, CPD = cigarettes per day, COPD = chronic obstructive pulmonary disorder, COPDGene = Genetic Epidemiology of COPD, NA = not applicable, NHW = non-Hispanic white. Correcting for the 12 regions listed, all p values less than 4.17E-3 (0.05/12) are in bold. NAs are given if there were not enough rare variants in the region for the test statistic.

in the SKAT-O gene-based tests for association between rare variants and CPD. No genes reached this significance threshold among NHW in the COPDGene study. Among AA in COPDGene study, one gene reached this significance threshold (*MYLIP*, $p = 1.41 \times 10^{-6}$). However, there were only seven SNPs in this region, and only one

SNP (rs151199797, minor allele frequency = 0.005) was responsible for this signal. When this SNP (rs151199797) was excluded from this region in the rare variant analysis, the *MYLIP* gene was no longer associated with CPD among AA subjects ($p = .51$). This region on chromosome 6 (*MYLIP*) including rs151199797 was not

replicated at the significance level of 3.4×10^{-6} among NHW in the COPDGene study ($p = .034$).

Discussion

In a large GWA study of current and former heavy smokers in a population of COPD cases and controls, we have replicated the association of common variants on chromosome 15q25 with CPD. This region contains a cluster of genes coding for nicotinic acetylcholine receptors and other genes that are associated with smoking behavior,^{3–11} pulmonary function,¹² and diseases directly related to smoking including lung cancer and COPD.^{15–19} These findings were replicated in the COPDGene cohort and then confirmed in a meta-analysis of two other NHW ancestry cohorts (ie, ECLIPSE and GenKOLS). The SNPs associated with CPD on chromosome 15q25 do not represent independent associations. There is strong linkage disequilibrium (LD) between the SNPs on chromosome 15q25 associated with CPD in this study for NHW ($R^2 = 0.71$ – 1.0 as seen in [Supplementary Figure 4a](#)). Also, a known functional variant (rs16969968), which leads to an amino acid change in *CHRNA5*³⁹ and has been associated with nicotine dependence,⁴⁰ is in LD with the genome-wide significant SNPs associated with CPD in the current study for NHW ($R^2 = 0.655$ – 1.0 as shown in [Supplementary Figure 4a](#)). Thus, the genome-wide significant SNPs on chromosome 15q25 for CPD among NHW in this study may be tagging rs16969968 or these SNPs may be tagging another variant on chromosome 15q25. Of note, the genome-wide significant SNPs for CPD on chromosome 15q25 among NHW in this study are not in strong LD with two known SNPs in *CHRNA5*, rs880395 and rs588765 ($R^2 = 0.229$ – 0.332), that influence brain mRNA expression.⁴¹ In addition, rare variant gene-based analyses for chromosome 15q25 were not significantly associated with CPD.

Previously reported associations on chromosome 19 (*RAB4*, $p = 1.95\text{E-}06$; *CYP2A7*, $p = 7.50\text{E-}05$; *CYP2B6*, $p = 4.04\text{E-}04$) showed suggestive and directionally consistent associations among common variants and for rare variants among COPDGene NHW subjects for *CYP2A7* ($p = 5.2 \times 10^{-4}$). The mechanism for the associations between these SNPs on chromosome 19 and CPD is not known; however, it appears that these associations represent distinct genetic associations given the weak LD among these SNPs on chromosome 19 for NHW subjects as seen in [Supplementary Figure 4c](#). Previous studies have demonstrated the importance of the chromosome 19q13, *CYP2A7/CYP2B6* region, and nicotine metabolism.^{42,43}

To test the hypothesis that the genetic susceptibility to CPD was different among COPD cases and controls, analyses were stratified by COPD case-control status, and in addition, we tested for an SNP by COPD interaction on CPD. Although the association of SNPs on chromosome 15q25 was nominally associated with CPD in NHW COPD cases and controls, there was not a significant SNP by COPD interaction on CPD for these chromosome 15q25 SNPs. There was a nominal association between rs7937 in *RAB4B* and CPD among COPD cases. However, the SNP by COPD interaction on CPD for *RAB4B* among NHW was only marginally significant ($p = .07$). Among NHW COPD controls, rs7260329 in *CYP2B6* on chromosome 19 was nominally associated with CPD. However, the SNP by COPD interaction on CPD for *CYP2B6* was not statistically significant ($p = .12$). Although differences of the genetic susceptibility to CPD among COPD cases and controls were intriguing, convincing evidence for the SNP by COPD interaction was not found in this study.

Genetic Burden of Smoking

Among all NHW COPDGene participants, multiple SNPs within chromosome 15q25 reached genome-wide significance ($p < 5.0\text{E-}8$) for average CPD; however, the genetic effect size for the most significant SNP rs8192482 [*CHRNA3*] was modest ($\beta = 0.13$ for an average increase of 10 CPD). This modest effect size creates a substantial smoking burden over the lifetime of a smoker. For instance, given the average smoking duration in the COPDGene study of 36.7 years, each copy of the risk allele at rs8192482 would contribute an average of 474.5 cigarettes more per year and an average of 17 414 more cigarettes over the lifetime of this NHW COPDGene smoker adjusting for age, sex, and genetic ancestry (via principal components).

Differences Among NHW and AA COPDGene Subjects

Within the COPDGene study, on average, NHW subjects smoked more CPD (mean = 25.80, SD = 11.42) compared to AA subjects (mean = 21.29, SD = 10.40). However, there is a larger proportion of current smokers among AA subjects (80%) compared to NHW subjects (39%). This is consistent with previously observed difference in smoking patterns between AA and NHW.⁴³ In that study, AAs made more attempts to quit smoking but overall were less successful in maintaining smoking cessation.⁴⁴ The underlying causes of these differences are complex and likely relate both to underlying susceptibility to smoking behavior and to societal factors. For example, among NHW, the LD on chromosome 15q25 is relatively strong as shown in [Supplementary Figure 4a](#) as compared to AA as shown in [Supplementary Figure 4b](#). There are also differences for allele frequencies among AA and NHW. For the most significant association (rs8192482 [*CHRNA3*]) for CPD, the allele frequency for the coded T allele is 37% for NHW COPDGene subjects but only 6% for AA subjects. There is also a significant difference in sample sizes among the NHW subjects ($n = 6659$) and AA COPDGene subjects ($n = 3260$). Given the differences in sample sizes and minor allele frequency, the power analysis in [Supplementary Figure 3a–c](#) shows that there is a large difference in power for the analyses among NHW and AA. These differences in the COPDGene AA and NHW subjects could contribute to the differing results between these the two populations.

Potential Limitations

The COPDGene cohort was ascertained based on smoking history and is deliberately enriched for COPD cases. Although this ascertainment scheme maximizes the efficiency of case-control studies from the COPDGene cohort, analyzing secondary phenotypes (eg, quantitative measure of smoking behaviors) in any case-control study can be biased due to the ascertainment scheme. This is only an issue for SNPs associated with both the ascertainment condition and the secondary phenotype.⁴⁵ Because our analysis focused on CPD, when we adjusted for GOLD stage (a measure of severity of COPD) in an additional analysis and observed a decrease in the strength of the association (as measured by β), markers on chromosome 15q25 still gave the most significant signal. This suggests that our analysis should be robust against this source of sampling bias due to ascertainment.⁴⁵

Although COPDGene included substantial numbers of both AA and NHW subjects, the sample size for AA subjects was considerably smaller and therefore had less statistical power. In addition, CPD is self-reported and may suffer from measurement error.⁴⁶ Biomarkers for smoking history such as exhaled carbon monoxide, urine and blood cotinine, and the nicotine metabolite ratio could provide better measurements of smoking exposure than the reported number of CPD.²⁰

Supplementary Material

Supplementary data can be found online at <http://www.ntr.oxford-journals.org>.

Funding

This work was supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health (K01HL125858 to SML). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This work was also supported by the National Heart, Lung, and Blood Institute (NHLBI P01 HL105339 and R01 HL089856 to EKS; K08 HL097029 and R01 HL113264 to MHC; R01 HL089897 to JDC.; R00 HL121087 to MNM; K08 HL136928 to BH), Parker B. Francis Research Opportunity Award (to MNM and BH), and a grant from the Alpha-1 Foundation (to MHC). The COPDGene study (NCT00608764) was supported by NHLBI U01 HL089897 and U01 HL089856. The COPDGene study is also supported by the COPD Foundation through contributions made to an Industry Advisory Committee comprised of AstraZeneca, Boehringer-Ingelheim, GlaxoSmithKline, Novartis, and Sunovion. The Norway GenKOLS study (Genetics of Chronic Obstructive Lung Disease, GSK code RES11080) and the ECLIPSE study (NCT00292552; GSK code SCO104960) were funded by GlaxoSmithKline. The funding sources played no role in the design of the study or the decision to submit the manuscript for publication. A full list of the COPDGene and ECLIPSE investigators are given in the supplement.

Declaration of Interests

Regarding conflicts of interest, in the past 3 years, EKS received honoraria and consulting fees from Merck; grant support and consulting fees from GlaxoSmithKline; and honoraria from Novartis. DAL is a consultant and has received grant support and honoraria from GlaxoSmithKline. He chaired the Respiratory Therapy Area Board at GlaxoSmithKline (2012–2015). AG has participated in the advisory boards of Chiesi Pharma AB, Sverige; Novartis, Norge; Takeda Nycomed, Norge; AstraZeneca, Norge; and Boehringer Ingelheim, Norge. No other authors reported conflicts of interest.

References

- US Department of Health and Human Services. *Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General*. 2014. www.surgeongeneral.gov/library/reports/50-years-of-progress/exec-summary.pdf.
- Li MD, Cheng R, Ma JZ, Swan GE. A meta-analysis of estimated genetic and environmental effects on smoking behavior in male and female adult twins. *Addiction*. 2003;98(1):23–31.
- Bierut LJ, Madden PA, Breslau N, et al. Novel genes identified in a high-density genome wide association study for nicotine dependence. *Hum Mol Genet*. 2007;16(1):24–35.
- Thorgerirsson TE, Geller F, Sulem P, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature*. 2008;452(7187):638–642.
- Liu JZ, Tozzi F, Waterworth DM, et al.; Wellcome Trust Case Control Consortium. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet*. 2010;42(5):436–440.
- Thorgerirsson TE, Gudbjartsson DF, Surakka I, et al.; ENGAGE Consortium. Sequence variants at CHRN3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat Genet*. 2010;42(5):448–453.
- TAG Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet*. 2010;42:441–447. <https://www.nature.com/articles/ng.571>.
- Saccone NL, Culverhouse RC, Schwantes-An TH, et al. Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer and COPD. *PLoS Genet*. 2010;6(8):1–16.
- Bierut LJ, Stitzel JA, Wang JC, et al. Variants in nicotinic receptors and risk for nicotine dependence. *Am J Psychiatry*. 2008;165(9):1163–1171.
- Hancock DB, Reginsson GW, Gaddis NC, et al. Genome-wide meta-analysis reveals common splice site acceptor variant in CHRNA4 associated with nicotine dependence. *Transl Psychiatry*. 2015;5:e651. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4930126/pdf/tp2015149a.pdf>.
- David SP, Hamidovic A, Chen GK, et al. Genome-wide meta-analyses of smoking behaviors in African Americans. *Transl Psychiatry*. 2012;2:e119. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3365260/pdf/tp201241a.pdf>.
- Lutz SM, Cho MH, Young K, et al.; ECLIPSE Investigators; COPDGene Investigators. A genome-wide association study identifies risk loci for spirometric measures among smokers of European and African ancestry. *BMC Genet*. 2015;16(138):1–12.
- Cho MH, McDonald ML, Zhou X, 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(3):214–225.
- Pillai SG, Ge D, Zhu G, 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(3):e1000421.
- Wang JC, Bierut LJ, Goate AM. Variants weakly correlated with CHRNA5 D398N polymorphism should be considered in transcriptional deregulation at the 15q25 locus associated with lung cancer risk. *Clin Cancer Res*. 2009;15(17):5599; author reply 5599.
- Falvella FS, Galvan A, Frullanti E, et al. Transcription deregulation at the 15q25 locus in association with lung adenocarcinoma risk. *Clin Cancer Res*. 2009;15(5):1837–1842.
- Amos CI, Wu X, Broderick P, et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet*. 2008;40(5):616–622.
- Thorgerirsson TE, Geller F, Sulem P, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature*. 2008;452(7187):638–642.
- Chen LS, Hung RJ, Baker T, et al. CHRNA5 risk variant predicts delayed smoking cessation and earlier lung cancer diagnosis: a meta-analysis. *J Natl Cancer Inst*. 2015;107(5):1–9. <https://www.ncbi.nlm.nih.gov/pubmed/25873736>.
- Lutz SM, Hokanson JE. Genetic influences on smoking and clinical disease. Understanding behavioral and biological pathways with mediation analysis. *Ann Am Thorac Soc*. 2014;11(7):1082–1083.
- Lutz SM, Hokanson JE. Mediation analysis in genome-wide association studies: current perspectives. *Open Bioinform J*. 2015;2015(7):1–5.
- Olsson E, Saccone NL, Johnson EO, et al. Rare, low frequency and common coding variants in CHRNA5 and their contribution to nicotine dependence in European and African Americans. *Mol Psychiatry*. 2016;21(5):601–607.
- Wain LV, Shrine N, Miller S, et al. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *Lancet Respir*. 2015;3(10):769–781.
- VanderWeele TJ, Asomaning K, Tchetgen Tchetgen EJ, et al. Genetic variants on 15q25.1, smoking, and lung cancer: an assessment of mediation and interaction. *Am J Epidemiol*. 2012;175(10):1013–1020.
- Siedlinski M, Tingley D, Lipman PJ, et al.; COPDGene and ECLIPSE Investigators. Dissecting direct and indirect genetic effects on chronic obstructive pulmonary disease (COPD) susceptibility. *Hum Genet*. 2013;132(4):431–441.
- Regan EA, Hokanson JE, Murphy JR, et al. Genetic epidemiology of COPD (COPDGene) study design. *COPD*. 2010;7(1):32–43.
- Vestbo J, Anderson W, Coxson HO, et al.; ECLIPSE investigators. Evaluation of COPD longitudinally to identify predictive surrogate end-points (ECLIPSE). *Eur Respir J*. 2008;31(4):869–873.

28. Cho MH, Boutaoui N, Klanderman BJ, et al. Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat Genet.* 2010;42(3):200–202.
29. Hobbs BD, Parker MM, Chen H, 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(1):48–57.
30. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet.* 2012;44(8):955–959.
31. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol.* 2010;34(8):816–834.
32. Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature.* 2010;491:56–65. <https://www.nature.com/articles/nature11632>.
33. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559–575.
34. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006;38(8):904–909.
35. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010;26(17):2190–2191.
36. de Bakker PIW, Ferreira MAR, Jia X, et al. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet.* 2008;17(2):122–128.
37. Wan ES, Castaldi PJ, Cho MH, et al.; COPDGene Investigators. Epidemiology, genetics, and subtyping of preserved ratio impaired spirometry (PRISm) in COPDGene. *Respir Res.* 2014;15(89):1–13. <https://respiratory-research.biomedcentral.com/track/pdf/10.1186/s12931-014-0089-y>.
38. Lee S, Emond MJ, Bamshad MJ, et al.; NHLBI GO Exome Sequencing Project—ESP Lung Project Team. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet.* 2012;91(2):224–237.
39. Bierut LJ. Genetic vulnerability and susceptibility to substance dependence. *Neuron.* 2011;69(4):618–627.
40. Saccone NL, Wang JC, Breslau N, et al. The CHRNA5-CHRNA3-CHRNA4 nicotinic receptor subunit gene cluster affects risk for nicotine dependence in African-Americans and in European-Americans. *Cancer Res.* 2009;69(17):6848–6856.
41. Barrie ES, Hartmann K, Lee SH, et al. The CHRNA5/CHRNA3/CHRNA4 nicotinic receptor regulome: genomic architecture, regulatory variants, and clinical associations. *Hum Mutat.* 2017;38(1):112–119.
42. Baurley JW, Edlund CK, Pardamean CI, et al. Genome-wide association of the laboratory-based nicotine metabolite ratio in three ancestries. *Nicotine Tob Res.* 2016;18(9):1837–1844.
43. Loukola A, Buchwald J, Gupta R, et al. A genome-wide association study of a biomarker of nicotine metabolism. *PLoS Genet.* 2015;11(9):e1005498.
44. Trinidad DR, Xie B, Fagan P, et al. Disparities in the population distribution of African American and non-Hispanic white smokers along the quitting continuum. *Health Educ Behav.* 2015;42(6):742–751.
45. Lutz SM, Hokanson JE, Lange C. An alternative hypothesis testing strategy for secondary phenotype data in case-control genetic association studies. *Front Genet.* 2014;4(188):1–4.
46. Bloom AJ, Hartz SM, Baker TB, et al. Beyond cigarettes per day. A genome-wide association study of the biomarker carbon monoxide. *Ann Am Thorac Soc.* 2014;11(7):1003–1010.