Genome-wide association study of smoking behaviours in patients with COPD

Mateusz Siedlinski,[1] Michael H Cho,[1,2] Per Bakke,[3] Amund Gulsvik,[3] David A Lomas,[4] Wayne Anderson,[5] Xiangyang Kong,[6] Stephen I Rennard,[7] Terri H Beaty,[8] John E Hokanson,[9] James D Crapo,[10] Edwin K Silverman,[1,2] the COPDGene Investigators and ECLIPSE Investigators

► Additional materials are published online only. To view these files please visit the journal online (http://thorax.bmj.com).

For numbered affiliations see end of article.

Correspondence to
Edwin K Silverman, 181 Longwood Avenue, Boston, MA 02115, USA; ed.silverman@channing.harvard.edu

This article is based on research that is funded in part by grants from the National Institute of Health (NIH) and is therefore subject to the mandatory NIH Public Access Policy. The final, peer-reviewed manuscript must be deposited with the PubMed Central (PMC) database upon acceptance for publication and be made publicly accessible no later than 12 months after publication.

Received 8 March 2011
Accepted 16 May 2011
Published Online First
16 June 2011

ABSTRACT

Background Cigarette smoking is a major risk factor for chronic obstructive pulmonary disease (COPD) and COPD severity. Previous genome-wide association studies (GWAS) have identified numerous single nucleotide polymorphisms (SNPs) associated with the number of cigarettes smoked per day (CPD) and a dopamine beta-hydroxylase (DBH) locus associated with smoking cessation in multiple populations.

Objective To identify SNPs associated with lifetime average and current CPD, age at smoking initiation, and smoking cessation in patients with COPD.

Methods GWAS were conducted in four independent cohorts encompassing 3441 ever-smoking patients with COPD (Global Initiative for Obstructive Lung Disease stage II or higher). Untyped SNPs were imputed using the HapMap (phase II) panel. Results from all cohorts were meta-analysed.

Results Several SNPs near the HLA region on chromosome 6p21 and in an intergenic region on chromosome 2q21 showed associations with age at smoking initiation, both with the lowest $p=2\times10^{-7}$. No SNPs were associated with lifetime average CPD, current CPD or smoking cessation with $p<10^{-6}$. Nominally significant associations with candidate SNPs within cholinergic receptors, nicotinic, alpha 3/5 (*CHRNA3/CHRNA5*; eg, p=0.00011 for SNP rs1051730) and cytochrome P450, family 2, subfamily A, polypeptide 6 (*CYP2A6*; eg, $p=2.78\times10^{-5}$ for a non-synonymous SNP rs1801272) regions were observed for lifetime average CPD, however only *CYP2A6* showed evidence of significant association with current CPD. A candidate SNP (rs3025343) in *DBH* was significantly (p=0.015) associated with smoking cessation.

Conclusion The authors identified two candidate regions associated with age at smoking initiation in patients with COPD. Associations of *CHRNA3/CHRNA5* and *CYP2A6* loci with CPD and *DBH* with smoking cessation are also likely of importance in the smoking behaviours of patients with COPD.

Key messages

What is the key question?

► Identification of genetic markers associated with smoking behaviours in patients with chronic obstructive pulmonary disease (COPD).

What is the bottom line?

► This case-only genome-wide association study identifies two candidate loci for age at smoking onset.

Why read on?

► It additionally confirms associations of the *CYP2A6* and 15q25 loci with smoking intensity and of the *DBH* locus with smoking cessation in patients with COPD.

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a common, genetically complex disease caused, and accelerated in its progression, predominantly by tobacco smoking. High smoking intensity, likely related at least in part to nicotine addiction, increases the risk of developing COPD. Although many patients quit smoking after they are diagnosed with COPD, some continue to smoke, placing them at high risk for continued disease progression.

Smoking behaviours, such as age at smoking initiation, smoking cessation, and number of cigarettes smoked per day (CPD), are partially genetically determined and have substantial heritability.[1–4] Numerous loci and candidate genes have been suggested to contain genetic markers affecting smoking behaviours using genome-wide linkage scans[5] and genome-wide association study (GWAS) approaches.[6 7] Recent GWAS have identified loci associated with smoking cessation (dopamine beta-hydroxylase (DBH) on chromosome 9q34) and CPD (eg, cholinergic nicotinic receptors locus on chromosome 15q25 and cytochrome P450, family 2, subfamily A, polypeptide 6 (*CYP2A6*) locus on chromosome 19q13) in multiple populations,[8–11] although GWAS of smoking behaviours specifically of patients with COPD have not been reported. Interestingly, the same single nucleotide polymorphisms (SNPs) in the 15q25 locus were previously associated with development of COPD,[12] but the role of this locus in smoking behaviours in patients with COPD and whether the sole effect of this locus on COPD susceptibility relates to smoking behaviour remain unclear.

Because of their typically heavy lifetime smoking exposures, potentially related to (at least in part) enrichment in genetic variants responsible for nicotine addiction, patients with COPD can be considered as a unique population for studying smoking behaviours. However, diagnosis and

further progression of COPD are likely to modify smoking status (ie, increased efforts to quit smoking) and smoking intensity (eg, reduction of CPD). Identifying genetic factors involved in smoking cessation is of special importance in clinical practice, since quitting smoking may reduce subsequent loss of lung function in patients with COPD.[13 14] Smoking cessation results in an improvement of respiratory symptoms in patients with COPD, and is associated with reduced mortality due to COPD.[13–16] Another smoking-related phenotype, age at smoking onset, correlates with nicotine dependence in adulthood[17] and mortality due to COPD.[15] Taken together, it is of special interest to search for markers associated with smoking behaviours uniquely among patients with COPD. Likewise, it is of importance to assess whether SNPs, regarded as established genetic determinants of smoking cessation and CPD in other populations, associate with these traits in patients with COPD as well. Identification and description of such genetic variations may have significant consequences on future, targeted therapy aiming to reduce smoking among patients with COPD.

The aim of the current study was to identify SNPs associated with age at smoking initiation, smoking cessation, current and lifetime average CPD among patients with COPD, using GWAS in four independent cohorts: National Emphysema Treatment Trial (NETT), Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE), GenKOLS cohort from Bergen, Norway, and COPDGene. The authors additionally hypothesised that a subset of smoking behaviour genetic determinants in other population samples would influence smoking behaviour in patients with COPD.

METHODS

Subjects and phenotypes

Ever-smoking, Caucasian patients with at least moderate COPD (Global Initiative for Obstructive Lung Disease stage II or higher) from four independent cohorts (NETT,[18] ECLIPSE,[19] GenKOLS cohort from Bergen, Norway,[20] and COPDGene (first 1000 subjects)[21]) were studied (table 1). In total 3441 patients had at least one (out of four) non-missing, smoking-related phenotype (table 1). All phenotypes were self-reported using either a Case Report Form (smoking status and the lifetime average CPD in the ECLIPSE cohort) or modified versions of the American Thoracic Society/Division of Lung Diseases Respiratory Disease Questionnaire.[22] Because current smokers were not eligible for the NETT study, this cohort did not contribute to the analyses on smoking cessation and current CPD.

Genotyping and quality control

Different genotyping chips from Illumina (San Diego, California, USA) were used (table 1). Quality control (QC) steps were previously described in detail for the NETT, Norwegian and ECLIPSE cohorts,[23] and were applied to the COPDGene study as well. Briefly, QC steps consisted of filtering SNPs based on missing call rates (>5%) and Hardy-Weinberg Equilibrium deviation ($p<10^{-8}$), and filtering patients based on genotyping call rate (<95%), sex discrepancy, unexpected relatedness (PLINK pi-hat cut-off of 0.125), and ethnicity. Removal of cases that were outliers for genetic ancestry was performed based on principal components analysis in cases only. In the primary analysis, untyped markers were imputed using 120 founder Caucasian

Table 1 Characteristics of patients with chronic obstructive pulmonary disease (COPD) and smoking-related phenotypes studied

	Cohort			
	NETT n=362*	Norway n=851*	ECLIPSE n=1734*	COPDGene n=494*
Characteristics				
Men*, n (%)	234 (64.6)	511 (60.0)	1160 (66.9)	242 (49.0)
Age in years*, Mean (SD)	67.4 (5.8)	65.5 (10.1)	63.7 (7.1)	64.7 (8.1)
Pack-years smoked*, Mean (SD)	66.1 (30.9)	32.1 (18.6)	50.4 (27.4)	54.8 (26.8)
Post-bronchodilator FEV_1 (% pred.)*, Mean (SD)	29.1 (7.8)	50.7 (17.5)	44.8 (14.7)	48.7 (18.4)
Post-bronchodilator FEV_1/FVC*, Mean (SD)	0.32 (0.06)	0.51 (0.13)	0.45 (0.12)	0.48 (0.13)
Enrolment, Years	1998–2002	2003–2005	2005–2007	2008–2009
Genotyping technology, Chip	Illumina Quad 610	Illumina HumanHap 550 V1, V3, and Duo	Illumina HumanHap 550 V3	Illumina Human Omni1-Quad
Phenotype studied				
Age at smoking initiation				
Mean (SD)	16.6 (3.6)	18.7 (5.1)	16.9 (4.4)	16.8 (4.4)
Number of patients with non-missing phenotype	362	851	1690	494
Lambda inflation factor	0.986	0.989	1.019	0.997
Lifetime average CPD				
Mean (SD)	32.4 (13.5)	15.7 (7.8)	25.5 (12.4)	27.6 (11.8)
Number of patients with non-missing phenotype	361	851	1734	494
Lambda inflation factor	1.002	0.996	1.018	0.996
Smoking cessation				
Current smokers, n (%)	0 (0)	404 (47.5)	610 (35.3)	150 (30.6)
Former smokers, n (%)	362 (100)	447 (52.5)	1120 (64.7)	340 (69.4)
Lambda inflation factor	–	1.000	0.998	0.995
Current CPD†				
Mean (SD)	–	13.1 (6.9)	15.6 (10.8)	18.4 (12.4)
Number of patients with non-missing phenotype	–	398	565	150
Lambda inflation factor	–	0.995	1.005	0.997

*Calculated for patients with at least one non-missing phenotype.
†For patients with reported current CPD >0.
CPD, number of cigarettes smoked per day; ECLIPSE, Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points; FEV_1, forced expiratory volume in 1 s; FVC, forced vital capacity; NETT, National Emphysema Treatment Trial.

(Centre d'Etude du Polymorphisme Humain (Utah residents with ancestry from northern and western Europe) (CEU)) haplotypes from HapMap reference panel (phase II) and, as a secondary analysis, using 1000 Genomes Project data.[24 25] We limited our analysis to SNPs genotyped/imputed in at least two cohorts, with imputation r^2 coefficient ≥0.3 (for imputed SNPs only). Overall, approximately 2.5 and 6.3 million SNPs per phenotype were analysed using the reference HapMap Project and 1000 Genomes Project panels, respectively.

Association analysis of candidate SNPs

We extracted candidate SNPs achieving genome-wide significance in the previous studies on smoking cessation and CPD.[8–11 26] Following previous GWAS,[8 9] we additionally extracted two SNPs from *CYP2A6* (rs1801272 (Leu160His)) and cholinergic receptor, nicotinic, alpha 5 (CHRNA5, rs588765) based on their biological function. Since a candidate SNP from the *DBH* locus and most of the candidate SNPs from the *CYP2A6* locus were imputed in all cohorts, we searched for the best proxy SNP genotyped in at least three of the cohorts.

Statistics

According to Box-Cox transformation[27] that identified approximately the best transformation of dependent variables, which could be applied to all cohorts, we studied $\log_2$-transformed age at smoking initiation and lifetime average CPD, and the square root of the current CPD. Regression models were run under an additive model for SNPs, while adjusting for potential confounders (see online supplementary material for details). A fixed effect model was used for all meta-analyses. Effect allele was defined as the one associated with later age at smoking initiation, higher CPD or higher odds for smoking cessation. Meta-analytic $p<5\times10^{-8}$ was considered as genome-wide significant[28]; $5\times10^{-8}\leq p<5\times10^{-7}$ was interpreted as a suggestive association in the genome-wide panel with $p<0.05$ as a suggestive association for candidate SNPs.

Software

Box-Cox transformations were performed with MASS package,[27] while lambda inflation factors were calculated with GenABEL package[29] in R (V.2.10.1).[30] SNP imputation was performed with MACH (V.1.0).[31] Principal components, reflecting genetic structure of each study population, were calculated and analysed with the EIGENSOFT package (V.2.0).[32] Genetic association analyses and meta-analyses were run with PLINK (V.1.0.7).[33 34] SNAP (V.2.2)[35] was used to search for proxy SNPs ($r^2\geq0.8$) using the CEU HapMap phase II panel and to assess linkage disequilibrium (LD) coefficients in 1000 Genomes Project data. SNP/gene annotations and regional association plots were created with LocusZoom (V.1.1; human genome build hg18).[36]

RESULTS

Age at smoking initiation

Analyses in the individual cohorts were adjusted for sex and principal components for genetic ancestry. In total, the meta-analysis included 3397 patients for the age at smoking initiation phenotype. Lambda inflation factors were between 0.986 and 1.019 for individual cohorts (table 1), and 1.002 for meta-analytic p values (see online figure S1 for a Q–Q plot of meta-analysis). No SNPs showed meta-analytic association p values below the genome-wide significance level. Numerous SNPs in the intergenic region on chromosome 2q21 and in the region between BCL2-antagonist/killer 1 (*BAK1*) and zinc finger and

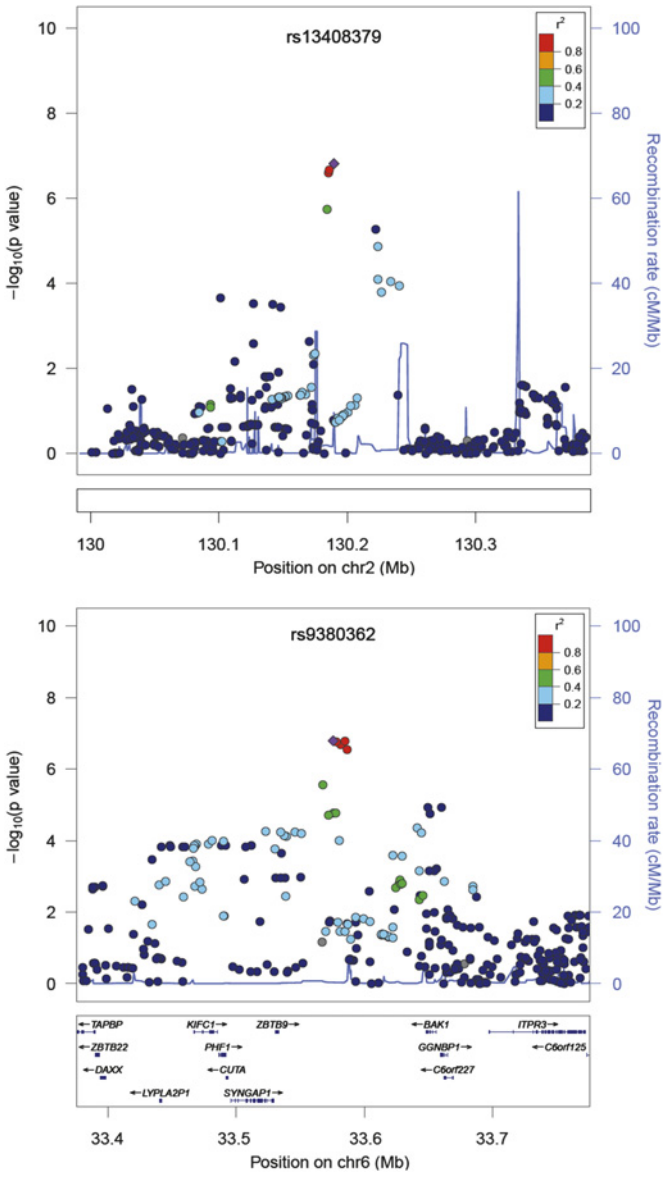


Figure 1 Regional association plots of two loci including single nucleotide polymorphisms (SNPs) associated with age at smoking initiation with meta-analytic $p<10^{-6}$. Dots correspond to meta-analysed SNPs. Colour of dots corresponds to r^2 (linkage disequilibrium (LD) coefficient in the HapMap II CEU population; grey dots correspond to SNPs with missing LD information) relative to the most significantly associated SNP (depicted with diamond) in the region that is, rs13408379 (top figure) and rs9380362 (bottom figure).

BTB domain containing 9 (*ZBTB9*) on chromosome 6p21 were associated with age at smoking initiation at a suggestive significance level (figure 1 and online table S1). In total, 24 SNPs were associated with age at smoking initiation with $p<10^{-5}$ in the meta-analysis (see online table S1).

Lifetime average CPD

Analyses in single cohorts were adjusted for sex, age and principal components for genetic ancestry. In total, the meta-analysis included 3440 patients for the lifetime average CPD phenotype. Lambda inflation factors were between 0.996 and 1.018 for individual cohorts (table 1), and 1.019 for meta-analytic p values (see online figure S2 for a Q–Q plot of meta-analysis). Thirteen SNPs were associated with lifetime average

Table 2 Single nucleotide polymorphisms (SNPs) associated with the lifetime average number of cigarettes smoked per day (CPD) (log_2-transformed) with meta-analytic $p<10^{-5}$, and candidate SNPs from *CHRNA3/CHRNA5* and *CYP2A6* loci identified by previous studies

	Chr.	Top SNP	Nearest gene within 50 kb (location)	Effect/ non-effect allele	B	p	I^2	Q stat.	N/N_{imp}	Freq.	Effect direction consistent with previous studies
Top SNPs	4	rs13104971	*SCFD2* (intron)	G/A	0.113	**1.18×10^{-6}**	0	0.92	4/3	0.13	–
	9	rs943306	*ASTN2* (intron)	T/C	0.073	**1.25×10^{-6}**	25	0.26	4/0	0.40	–
	7	rs10237067	–	A/G	0.072	**2.98×10^{-6}**	0	0.63	4/3	0.63	–
	7	rs12699125	*CALN1* (intron)	A/G	0.072	**3.17×10^{-6}**	42	0.16	4/3	0.66	–
	7	rs4718827	–	T/C	0.072	**3.36×10^{-6}**	0	0.61	4/4	0.63	–
	7	rs4717631	*CALN1* (intron)	A/G	0.072	**3.60×10^{-6}**	43	0.15	4/4	0.66	–
	7	rs17170849	*ELMO1* (intron)	C/T	0.265	**4.07×10^{-6}**	44	0.15	4/3	0.02	–
	7	rs4577845	*CALN1* (intron)	G/A	0.071	**4.57×10^{-6}**	44	0.15	4/1	0.66	–
	7	rs1859485	*CALN1* (intron)	T/C	0.069	**6.95×10^{-6}**	40	0.17	4/1	0.65	–
	12	rs11044734	–	G/C	0.112	**7.32×10^{-6}**	0	0.48	4/4	0.90	–
	12	rs11044737	–	G/A	0.112	**7.70×10^{-6}**	0	0.47	4/0	0.90	–
	16	rs190369	–	C/T	0.161	**8.89×10^{-6}**	47	0.13	4/1	0.05	–
	20	rs2869961	*CEBPB* (5′ region)	A/G	0.104	**9.65×10^{-6}**	0	0.59	4/1	0.12	–
Candidate SNPs	15	rs1051730	*CHRNA3* (exon, Tyr215Tyr)	A/G	0.060	**0.00011**	37	0.19	4/0	0.41	Yes
	15	rs16969968	*CHRNA5* (exon, Asp398Asn)	A/G	0.059	**0.00015**	30	0.23	4/3	0.41	Yes
	15	rs8034191	*AGPHD1* (intron)	C/T	0.055	**0.00036**	40	0.17	4/0	0.41	Yes
	15	rs684513	*CHRNA5* (intron)	C/G	0.029	0.163	0	0.98	4/4	0.80	Yes
	15	rs578776	*CHRNA3* (3′ UTR)	G/A	0.041	**0.020**	0	0.94	4/0	0.76	Yes
	15	rs588765	*CHRNA5* (intron)	C/T	0.031	**0.046**	47	0.13	4/4	0.60	Yes
	19	rs3733829	*EGLN2* (intron)	G/A	0.019	0.223	0	0.48	4/3	0.39	Yes
	19	rs7937	*RAB4B* (3′ UTR)	T/C	0.022	0.153	0	0.96	4/0	0.60	Yes
	19	rs1801272	*CYP2A6* (exon, Leu160His)	A/T	0.266	**2.78×10^{-5}**	0	0.72	4/4	0.96	Yes
	19	rs4105144	*CYP2A6* (5′ region)	C/T	0.073	**3.92×10^{-5}**	0	0.44	4/4	0.73	Yes
	19	rs7260329	*CYP2B6* (intron)	G/A	0.014	0.404	0	0.55	4/1	0.68	Yes
	19	rs7251570	*CYP2A6* (3′ region)	G/A	0.057	**0.00073**	0	0.84	4/4	0.71	Yes
	19	rs12461383	*CYP2A7* (3′ region)	G/C	0.063	**0.00019**	0	0.61	4/4	0.60	Yes

Analyses were adjusted for sex, age and principal components for genetic ancestry.
P values < 0.05 are depicted in bold.
N/N_{imp}, number of studies contributing to the meta-analysis/number of studies in which SNP was imputed; I^2, heterogeneity index; Q stat., p value for Q statistic; p, p value from the fixed effect meta-analysis; Freq., effect allele frequency in 3441 patients with at least one non-missing phenotype from four cohorts studied; Chr., chromosome; B, regression coefficient; UTR, untranslated region; *CHRNA3/CHRNA5*, cholinergic receptors, nicotinic, alpha 3/5; *CYP2A6*, cytochrome P450, family 2, subfamily A, polypeptide 6.

CPD with meta-analytic $p<10^{-5}$, however, no SNPs showed association with meta-analytic $p<10^{-6}$ (table 2).

Candidate SNPs from the cholinergic nicotinic receptors locus on chromosome 15q25 (rs578776, rs588765, rs8034191, rs1051730 and rs16969968) were significantly ($p<0.05$) associated with lifetime average CPD with the same direction of effect as seen in previous GWAS on CPD (table 2). The synonymous rs1051730 SNP was the top associated SNP in this locus (figure 2).

Four candidate SNPs (rs7251570, rs4105144, rs1801272 and rs12461383) in the *CYP2A6* locus on chromosome 19q13 were significantly ($p<0.05$) associated with lifetime average CPD with the same direction of effect as seen in previous GWAS on CPD (table 2). A non-synonymous SNP (rs1801272) was the second most significantly associated SNP in this locus (figure 2; table 2). There is, at most, a moderate level of LD between these four SNPs according to the HapMap phase II panel ($r^2\leq0.6$ and $D'\geq0.84$; see online figure S3). Analysis of proxy SNPs that were genotyped in at least three cohorts was performed. Analysis of rs8102683 (genotyped in the NETT, ECLIPSE and Norwegian cohorts, and imputed in COPDGene) that is a proxy for rs4105144 ($r^2=0.87$), and rs7251418 (genotyped in all cohorts) that is a proxy for rs7251570 ($r^2=0.81$) confirmed the associations observed ($p=4.65\times10^{-5}$ (B=0.074 for the 'C' effect allele; $I^2=0$) and $p=0.0024$ (B=0.054 for the 'G' effect allele; $I^2=0$), respectively). No genotyped proxy SNPs could be found for rs12461383 and rs1801272 SNPs. We did not replicate, with a nominal $p<0.05$, associations between SNPs in 7p14, 8p11 and 10q23 loci, reported in previous GWAS on CPD in our analysis on lifetime average CPD (see online table S2).

Smoking cessation

Analyses in single cohorts were adjusted for age, percentage of predicted forced expiratory volume in 1 s (FEV_1) and principal components summarising genetic ancestry. In total, the meta-analysis included 1164 patients who were current smokers (defined as 'no' for smoking cessation) and 1907 patients who were former smokers (defined as 'yes' for smoking cessation). Lambda inflation factors were between 0.995 and 1.000 for individual cohorts (table 1), and 0.998 for meta-analytic p values (see online figure S4 for a Q–Q plot of meta-analysis). No SNPs showed an association with meta-analytic $p<10^{-6}$ (table 3). Candidate SNP rs3025343 in the *DBH* locus showed nominally significant association with smoking cessation with the same direction of effect as seen in recent GWAS, yet with substantial heterogeneity between studies (table 3). SNP rs3025316, a proxy for rs3025343 ($r^2=0.94$), was genotyped in all cohorts and confirmed this association ($p=0.002$, OR=1.32 for the 'T' effect allele; $I^2=18$).

Current CPD

Analyses were adjusted for sex, age, percentage of predicted FEV_1, and principal components for genetic ancestry. In total, the meta-analysis included 1113 patients who were current smokers for the current CPD phenotype. Lambda inflation factors were between 0.995 and 1.005 for individual cohorts (table 1), and 1.010 for meta-analytic p values (see online figure S5 for a Q–Q plot of meta-analysis). No SNPs showed a meta-analytic association with $p<10^{-6}$ (table 4). Among candidate SNPs, rs12461383 from the *CYP2A6* locus was significantly

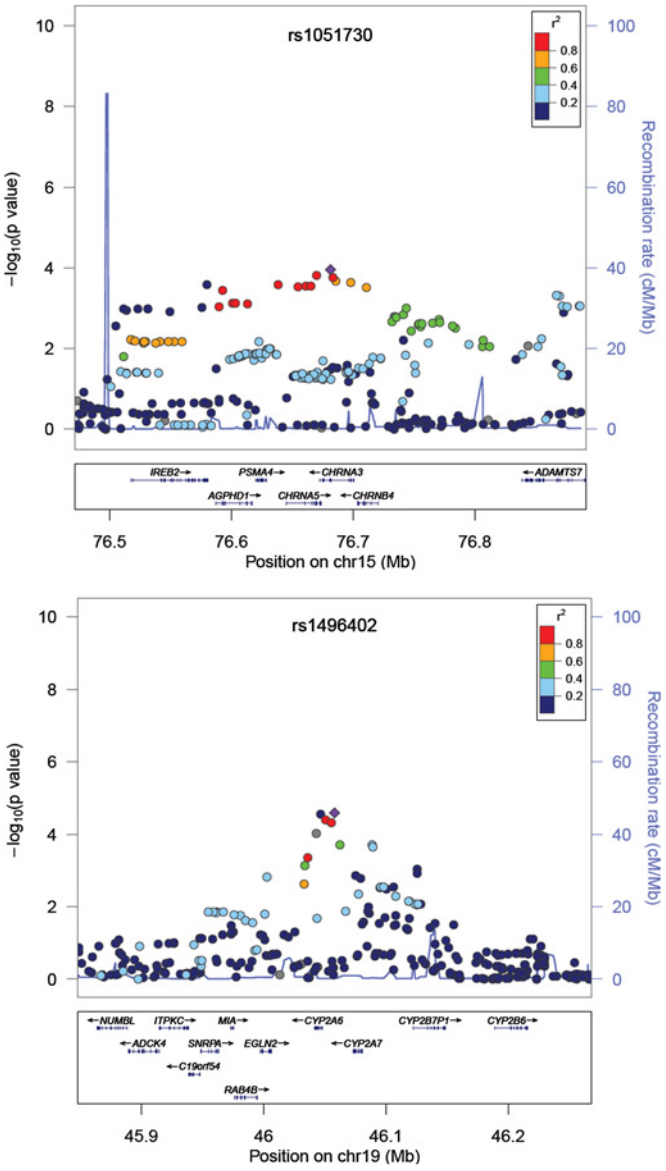


Figure 2 Regional association plots of the candidate loci centered on *CHRNA3* (top) and *CYP2A6* (bottom) and their associations with lifetime average number of cigarettes smoked per day. Dots correspond to meta-analysed SNPs. Colour of dots corresponds to r^2 (linkage disequilibrium (LD) coefficient in the HapMap II CEU population; grey dots correspond to SNPs with missing LD information) relative to the most significantly associated SNP in the region (depicted with diamond) that is, rs1051730 in the *CHRNA3* (top figure) and rs1496402 (bottom figure; the second top SNP in this region is a non-synonymous (Leu160His) rs1801272 SNP in the *CYP2A6*). Chr, chromosome; *CHRNA3*, cholinergic receptor, nicotinic, alpha 3; *CYP2A6*, cytochrome P450, family 2, subfamily A, polypeptide 6.

associated with current CPD with the same effect direction compared with previous GWAS on CPD (table 4). Candidate SNPs rs1801272 and rs7251570 showed a borderline significant association with the same effect direction compared with previous GWAS on CPD (table 4). We did not replicate, at a nominal $p<0.05$, associations between SNPs in 7p14, 8p11 and 10q23 loci reported in previous GWAS on CPD in our analysis on current CPD (online table S3).

Analyses of SNPs imputed with 1000 Genomes Project data

Analysis of all phenotypes using SNP genotypes imputed using the 1000 Genomes Project revealed 20 additional associations below the suggestive genome-wide significance level (see online table S4). SNP rs9394152 in the 6p21 locus was the most significantly associated SNP with the age at smoking initiation (meta-analytic $p=6.55\times10^{-8}$) and, similarly to the majority of SNPs associated with this phenotype below the suggestive genome-wide significance level, was characterised by high (≥0.96), cohort-specific imputation r^2 coefficients (see online table S4). We observed novel associations of three SNPs with the lifetime average CPD; the most significantly associated SNP rs28675338 ($r^2=0.07$ and $D'=1.00$ to rs1051730 according to 1000 Genomes Project Data) maps to the 15q25 locus. We also found one SNP associated with smoking cessation, with meta-analytic $p<5\times10^{-7}$ (see online table S4). Because cohort-specific, imputation r^2 coefficients of the four SNPs, associated with lifetime average CPD and smoking cessation below the suggestive genome-wide significance level, were rather modest (range 0.35–0.64; see online table S4), the observed associations certainly require additional confirmation and should be interpreted with caution. The three SNPs associated with current CPD with meta-analytic $p<5\times10^{-7}$ map to regions on chromosomes two and three (see online table S4), which were already ranked as top loci using the HapMap II reference panel (table 4).

DISCUSSION

Cigarette smoking is the most important environmental risk factor for COPD, and smoking cessation is the most important therapeutic intervention to prevent its progression. Our current study identified two loci on chromosomes 2q21 and 6p21 as candidates for containing genes influencing age at smoking initiation in patients with COPD, both showing suggestive levels of genome-wide significance. Furthermore, this study confirmed that certain SNPs in the cholinergic nicotinic receptors locus on chromosome 15q25 and SNPs in the *CYP2A6* locus on chromosome 19q13 were associated with CPD in patients with COPD. Of importance, we additionally confirmed in patients with COPD the previously reported association of a marker (rs3025343) in the *DBH* locus with smoking cessation.

The two loci mapping to chromosomes 2q21 and 6p21showed evidence of association with age at smoking initiation at a suggestive genome-wide significance level, using both HapMap II and 1000 Genomes projects reference imputation panels. The association peaks found here do not localise within any known genes, suggesting that distant regulation of yet unidentified target genes may be involved. SNPs from chromosome 6p21 are located close to the human leukocyte antigen locus, and *ZBTB9* and *BAK1* are the closest genes to the association peak. The *BAK1* gene, encoding a proapoptotic protein, has been previously associated by GWAS with testicular germ cell tumours[37] and platelet number.[38] Yet SNPs identified in these studies do not correlate, in terms of LD, with top SNPs from our age at smoking initiation GWAS.

Analysis of candidate SNPs, with respect to the lifetime average CPD, revealed that the majority of those SNPs in the *CHRNA3/CHRNA5* (15q25) and *CYP2A6* (19q13) regions significantly associated with this trait showed the same direction of effect described previously.[8–11 26] However, SNPs from the 7p14, 8p11 (*CHRNB3/CHRNA6* locus) and 10q23, as well as some SNPs from the *CYP2A6* and *CHRNA3/CHRNA5* loci, were not replicated in the current study with a nominal significance threshold. This may be caused by effect sizes that are too small to be detected in our populations or potentially to different genetic determinants of smoking behaviours within patients with COPD.

Table 3 Single nucleotide polymorphisms (SNPs) associated with smoking cessation with meta-analytic $p<10^{-5}$, and a candidate SNP (last row) from the dopamine beta-hydroxylase (DBH) locus identified by previous genome-wide association studies (GWAS) on smoking cessation

Chr.	Top SNP	Nearest gene within 50 kb	Effect/ non-effect allele	OR	p	I^2	Q stat.	N/N_{imp}	Freq.	Effect direction consistent with previous GWAS
10	rs10794613	*FLJ46361* (5′ region)	G/C	1.43	**3.41×10^{-6}**	0	0.51	3/3	0.21	–
3	rs13064954	*CCNL1* (3′ region)	A/G	1.94	**5.28×10^{-6}**	0	0.60	3/2	0.05	–
10	rs1896376	*CPMX2* (intron)	C/T	1.83	**5.71×10^{-6}**	57	0.10	3/2	0.95	–
13	rs9506942	–	C/G	1.29	**5.96×10^{-6}**	0	0.72	3/3	0.57	–
13	rs9552733	–	G/A	1.29	**5.99×10^{-6}**	0	0.69	3/0	0.57	–
3	rs9866141	*VEPH1* (3′ region)	T/C	1.88	**7.99×10^{-6}**	0	0.66	3/3	0.06	–
3	rs1165640	–	C/T	1.56	**8.98×10^{-6}**	0	0.57	3/0	0.10	–
12	rs10861185	*TXNRD1* (intron)	C/A	1.29	**9.57×10^{-6}**	16	0.31	3/2	0.57	–
10	rs727417	*CPXM2* (intron)	C/G	1.71	**9.67×10^{-6}**	38	0.20	3/3	0.94	–
9	rs3025343	*DBH* (5′ region)	G/A	1.24	**0.015**	46	0.16	3/3	0.87	Yes

Analyses were adjusted for age, % of predicted forced expiratory volume in 1 s (FEV_1) and principal components for genetic ancestry. Patients who were current smokers were considered as 'controls', while patients who were former smokers were considered as 'cases'.
P values < 0.05 are depicted in bold.
N/N_{imp}, number of studies contributing to the meta-analysis/number of studies in which SNP was imputed; I^2, heterogeneity index; Q stat., p value for Q statistic; p, p value from the fixed effect meta-analysis; Freq., effect allele frequency in 3441 patients with at least one non-missing phenotype from four cohorts studied; Chr., chromosome.

Besides the 'neutral' or reference haplotype, it has been suggested there are at least two other haplotypes with independent effects on CPD in this 15q25 region.[26 39] Our study confirms previous observations that rs1051730, and tagging SNPs such as rs16969968, exhibit relatively strong effects in this region, and SNPs rs588765 and rs578776, determine independent haplotypes associated with lifetime CPD in patients with COPD as well. However, we could not replicate the association between lifetime average CPD and rs684513, that is, the top SNP in the analysis on CPD when conditioning for rs1051730 in the Tobacco Genetics Consortium.[10] Importantly, the effect directions of all candidate SNPs in this 15q25 locus agree with those previously reported, suggesting larger COPD cohorts are required to establish significance of independent SNPs/haplotypes in this region.

Interestingly, effect sizes of replicated SNPs from the *CHRNA3/CHRNA5* locus are similar to most of those from the *CYP2A6* locus in our study,which contrasts with recent GWAS where bigger effect sizes were seen for the former locus.[8] Additionally, variations in the *CHRNA3/CHRNA5*, but not *CYP2A6*, locus showed substantial heterogeneity in genetic effects in our study, and as observed in previous GWAS as well.[8 10] We speculate that this might be due to between-study differences in general (eg, educational status or peer smoking[40]) or

Table 4 Single nucleotide polymorphisms (SNPs) associated with the current number of cigarettes smoked per day (CPD) (square root transformed) with meta-analytic $p<10^{-5}$, and candidate SNPs from *CHRNA3/CHRNA5* and *CYP2A6* loci identified by previous studies

	Chr.	Top SNP	Nearest gene within 50 kb	Effect/ non-effect allele	B	p	I^2	Q stat.	N/N_{imp}	Freq.	Effect direction consistent with previous studies
Top SNPs	3	rs1881681	–	C/A	0.363	**1.22×10^{-6}**	33	0.22	3/0	0.88	–
	8	rs11984631	–	T/G	1.351	**1.56×10^{-6}**	37	0.21	2/0	0.99	–
	2	rs12615264	*SOCS5* (3′ region)	T/G	0.644	**3.46×10^{-6}**	0	0.38	3/3	0.03	–
	2	rs11682595	*SOCS5* (3′ region)	T/G	0.642	**3.49×10^{-6}**	0	0.41	3/3	0.03	–
	2	rs11125090	*SOCS5* (3′ region)	A/G	0.641	**3.58×10^{-6}**	0	0.42	3/2	0.03	–
	20	rs297755	–	G/A	0.291	**4.86×10^{-6}**	49	0.14	3/3	0.17	–
	4	rs3893377	*BST1* (3′ region)	C/T	0.289	**4.90×10^{-6}**	51	0.13	3/3	0.80	–
	4	rs10019008	*BST1* (3′ region)	C/T	0.288	**4.99×10^{-6}**	53	0.12	3/0	0.80	–
	4	rs11947310	*BST1* (3′ region)	A/C	0.289	**5.21×10^{-6}**	53	0.12	3/3	0.80	–
	4	rs10018756	*BST1* (3′ region)	A/T	0.275	**7.71×10^{-6}**	48	0.15	3/3	0.77	–
Candidate SNPs	15	rs1051730	*CHRNA3* (exon, Tyr215Tyr)	G/A	0.010	0.851	32	0.23	3/0	0.59	No
	15	rs16969968	*CHRNA5* (exon, Asp398Asn)	G/A	0.009	0.863	30	0.24	3/2	0.59	No
	15	rs8034191	*AGPHD1* (intron)	T/C	0.006	0.914	5	0.35	3/0	0.59	No
	15	rs684513	*CHRNA5* (intron)	C/G	0.100	0.153	0	0.78	3/3	0.80	Yes
	15	rs578776	*CHRNA3* (3′ UTR)	G/A	0.052	0.388	0	0.98	3/0	0.76	Yes
	15	rs588765	*CHRNA5* (intron)	T/C	0.048	0.352	23	0.27	3/3	0.40	No
	19	rs3733829	*EGLN2* (intron)	G/A	0.005	0.930	40	0.19	3/2	0.39	Yes
	19	rs7937	*RAB4B* (3′ UTR)	T/C	0.027	0.590	0	0.37	3/0	0.60	Yes
	19	rs1801272	*CYP2A6* (exon, Leu160His)	A/T	0.362	0.063	0	1.00	3/3	0.96	Yes
	19	rs4105144	*CYP2A6* (5′ region)	C/T	0.069	0.217	0	0.55	3/3	0.73	Yes
	19	rs7260329	*CYP2B6* (intron)	G/A	0.005	0.922	0	0.76	3/0	0.68	Yes
	19	rs7251570	*CYP2A6* (3′ region)	G/A	0.105	0.052	0	0.87	3/3	0.71	Yes
	19	rs12461383	*CYP2A7* (3′ region)	G/C	0.143	**0.008**	0	0.62	3/3	0.60	Yes

Analyses were adjusted for sex, age, % of predicted forced expiratory volume in 1 s (FEV_1) and principal components for genetic ancestry.
P values < 0.05 are depicted in bold.
N/N_{imp}, number of studies contributing to the meta-analysis/number of studies in which SNP was imputed; I^2, heterogeneity index; Q stat., p value for Q statistic; p, p value from the fixed effect meta-analysis; Freq., effect allele frequency in 3441 patients with at least one non-missing phenotype from four cohorts studied; Chr., chromosome; B, regression coefficient; UTR, untranslated region; *CHRNA3/CHRNA5*, cholinergic receptors, nicotinic, alpha 3/5; *CYP2A6*, cytochrome P450, family 2, subfamily A, polypeptide 6.

COPD-specific (eg, reporting lifetime average CPD which may be influenced by severity of disease) characteristics.

Analysis of current CPD, which was much less powered than lifetime CPD because of a lower number of patients, was still able to detect some evidence of association with markers near *CYP2A6*, yet not for those located in the *CHRNA3/CHRNA5* region. CYP2A6 is an enzyme primarily responsible for conversion of nicotine to cotinine in the liver, and rs1801272 (Leu160His) codes for the CYP2A6*2 allele, which inactivates the enzyme.[41] This SNP showed the largest effect size on both current and lifetime average CPD among all analysed candidate SNPs in the region. Leu160His is in LD (D'=1.0) with other SNPs that were replicated in the *CYP2A6* locus. This agrees with previous GWAS and suggests that the rs1801272 SNP may be a true causative variant, while the other associations observed may be due to partial tagging by this SNP.[8] Importantly, we show that the genotyped proxy SNPs in the *CYP2A6* locus confirmed our analysis on imputed SNPs with respect to lifetime average CPD. The lack of convincing association for the proxy SNP rs7251418 with current CPD may be explained by the relatively smaller sample size in this analysis, and incomplete LD between the target and the proxy SNPs. Previous GWAS suggested that other genes may be associated with CPD in the 19q13 region, yet we observed no nominally significant associations for *EGLN2*, *RAB4B* and *CY2B6*.

Analysis of smoking cessation did not reveal any loci associated below the suggestive significance level, and it showed that the previously reported[10] association between rs3025343 in the *DBH* locus can be replicated in patients with COPD with a nominal significance threshold. *DBH* is a plausible candidate gene for smoking cessation because it participates in the metabolism of dopamine. This neurotransmitter is released from neurons in response to nicotine,[3] and is an important mediator of addiction behaviours such as smoking. The rs3025343 SNP that we extracted showed a substantial heterogeneity in this effect, which may reflect the between-study differences in factors such as use of nicotine replacement therapy or socioeconomic status. Interestingly, the best proxy SNP rs3025316 (genotyped in all cohorts) showed a more pronounced effect with a somewhat smaller heterogeneity index compared with the rs3025343.

Our study possesses several limitations, and some of them are typical for many GWAS. For example, the size of the current study may have been too small to detect associations at a genome-wide significance level. Phenotypes studied are genetically complex and are likely determined by many genes of modest effects, which makes them difficult to detect with genome-wide significance for the currently studied sample size. It is worth noting that the much larger Tobacco and Genetics consortium did not identify any genome-wide significant associations for age at smoking initiation in a meta-analysis encompassing over 20 000 subjects.[10] This emphasises the need for even larger studies to study smoking behaviours in order to detect variants with presumably low effect sizes. Despite the modest effect sizes of the genetic variants implicated by GWAS, key biological pathways may be identified using such approaches. Second, the genotype imputation accuracy could have had an impact on our results. Many top SNPs from the analyses of age at smoking initiation were imputed in all four cohorts; however, the association peaks of these regions also contained SNPs genotyped in the majority (or even all) of the cohorts, which likely makes these findings more reliable. Given the fact that different, and not fully overlapping, genotyping chips were used, this imputation was crucial to obtain a comprehensive overview of many genetic associations. In our study, this is of special importance for the non-synonymous SNP rs1801272 in *CYP2A6*, which had to be imputed in all cohorts and has no known proxy SNPs. Assuming that the association of this SNP with CPD is a true positive and possibly causal, imputation was the only one way to detect it. Lastly, we must acknowledge that COPD diagnosis has a significant impact on smoking behaviours studied, and especially on current CPD and smoking cessation. We took into account the severity of COPD, reflected in the level of lung function, as a potential confounder when analysing these two phenotypes. However, we hypothesise that additional factors such as frequency of exacerbations may also affect smoking behaviours in the patients with COPD who were studied here. Additionally, it is plausible that social factors, such as smoking trends changing over time, are important environmental determinants of smoking initiation and intensity, and they might have potentially confounded the genetic associations found.

In summary we identified two candidate loci associated with age at smoking initiation in patients with COPD. We showed that variation in the *CHRNA3/CHRNA5* locus on chromosome 15q25 locus and the *CYP2A6* locus on chromosome 19q13 were associated with lifetime average CPD among patients with COPD. The latter gene may play a significant role in the current smoking intensity among patients with COPD. Future studies in larger populations will be required to determine the overlap between genetic determinants of smoking behaviour in the general population and in patients with COPD.

Author affiliations
[1]Channing Laboratory, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA
[2]Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA
[3]Department of Thoracic Medicine, Haukeland University Hospital and Institute of Medicine, University of Bergen, Bergen, Norway
[4]Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK
[5]GlaxoSmithKline Research and Development, Research Triangle Park, North Carolina, USA
[6]GlaxoSmithKline Research and Development, King of Prussia, Pennsylvania, USA
[7]Department of Pulmonary and Critical Care Medicine, University of Nebraska Medical Center, Omaha, Nebraska, USA
[8]Johns Hopkins School of Public Health, Baltimore, Maryland, USA
[9]Department of Epidemiology, Colorado School of Public Health, University of Colorado Denver, Aurora, Colorado, USA
[10]Department of Medicine, National Jewish Health, Denver, Colorado, USA

Acknowledgements ECLIPSE Steering Committee: Harvey Coxson (Canada), Lisa Edwards (GlaxoSmithKline, USA), Katharine Knobil (Co-chair, GlaxoSmithKline, UK), David Lomas (UK), William MacNee (UK), Edwin Silverman (USA), Ruth Tal-Singer (GlaxoSmithKline, USA), Jørgen Vestbo (Co-chair, Denmark), Julie Yates (GlaxoSmithKline, USA).
ECLIPSE Scientific Committee: Alvar Agusti (Spain), Peter Calverley (UK), Bartolome Celli (USA), Courtney Crim (GlaxoSmithKline, USA), Bruce Miller (GlaxoSmithKline, UK), William MacNee (Chair, UK), Stephen Rennard (USA), Ruth Tal-Singer (GlaxoSmithKline, USA), Emiel Wouters (The Netherlands), Julie Yates (GlaxoSmithKline, USA).
ECLIPSE Investigators: Bulgaria: Yavor Ivanov, Pleven; Kosta Kostov, Sofia. Canada: Jean Bourbeau, Montreal, Que; Mark Fitzgerald, Vancouver, BC; Paul Hernandez, Halifax, NS; Kieran Killian, Hamilton, On; Robert Levy, Vancouver, BC; Francois Maltais, Montreal, Que; Denis O'Donnell, Kingston, On. Czech Republic: Jan Krepelka, Praha. Denmark: Jørgen Vestbo, Hvidovre. The Netherlands: Emiel Wouters, Horn-Maastricht. New Zealand: Dean Quinn, Wellington. Norway: Per Bakke, Bergen. Slovenia: Mitja Kosnik, Golnik. Spain: Alvar Agusti, Jaume Sauleda, Palma de Mallorca. Ukraine: Yuri Feschenko, Kiev; Vladamir Gavrisyuk, Kiev; Lyudmila Yashina, Kiev; Nadezhda Monogarova, Donetsk. United Kingdom: Peter Calverley, Liverpool; David Lomas, Cambridge; William MacNee, Edinburgh; David Singh, Manchester; Jadwiga Wedzicha, London. United States of America: Antonio Anzueto, San Antonio, Texas; Sidney Braman, Providence, Rhode Island; Richard Casaburi, Torrance, California; Bart Celli, Boston, Massachusetts; Glenn Giessel, Richmond, Virginia; Mark Gotfried, Phoenix, Arizona; Gary Greenwald, Rancho Mirage, California; Nicola Hanania, Houston, Texas; Don Mahler, Lebanon, New

Hampshire; Barry Make, Denver, Colorado; Stephen Rennard, Omaha, Nebraska; Carolyn Rochester, New Haven, Connecticut; Paul Scanlon, Rochester, Minnesota; Dan Schuller, Omaha, Nebraska; Frank Sciurba, Pittsburgh, Pennsylvania; Amir Sharafkhaneh, Houston, Texas; Thomas Siler, St. Charles, Missouri; Edwin Silverman, Boston, Massachusetts; Adam Wanner, Miami, Florida; Robert Wise, Baltimore, Maryland; Richard ZuWallack, Hartford, Connecticut.
We acknowledge the co-investigators in the NETT Genetics Ancillary Study including Joshua Benditt, Gerard Criner, Malcolm DeCamp, Philip Diaz, Mark Ginsburg, Larry Kaiser, Marcia Katz, Mark Krasna, Neil MacIntyre, Barry Make, Rob McKenna, Fernando Martinez, Zab Mosenifar, John Reilly, Andrew Ries, Paul Scanlon, Frank Sciurba, and James Utz.
COPDGene Investigators
The members of the COPDGene study group as of June 2010:
Ann Arbour VA: Jeffrey Curtis, MD (PI), Ella Kazerooni, MD (RAD)
Baylor College of Medicine, Houston, Texas: Nicola Hanania, MD, MS (PI), Philip Alapat, MD, Venkata Bandi, MD, Kalpalatha Guntupalli, MD, Elizabeth Guy, MD, Antara Mallampalli, MD, Charles Trinh, MD (RAD), Mustafa Atik, MD
Brigham and Women's Hospital, Boston, Massachusetts: Dawn DeMeo, MD, MPH (Co-PI), Craig Hersh, MD, MPH (Co-PI), George Washko, MD, Francine Jacobson, MD, MPH (RAD)
Columbia University, New York, New York: R. Graham Barr, MD, DrPH (PI), Byron Thomashow, MD, John Austin, MD (RAD)
Duke University Medical Center, Durham, North Carolina: Neil MacIntyre, Jr., MD (PI), Lacey Washington, MD (RAD), H Page McAdams, MD (RAD)
Fallon Clinic, Worcester, Massachusetts: Richard Rosiello, MD (PI), Timothy Bresnahan, MD (RAD)
Health Partners Research Foundation, Minneapolis, Minnesota: Charlene McEvoy, MD, MPH (PI), Joseph Tashjian, MD (RAD)
Johns Hopkins University, Baltimore, Maryland: Robert Wise, MD (PI), Nadia Hansel, MD, MPH, Robert Brown, MD (RAD), Gregory Diette, MD
Los Angeles Biomedical Research Institute at Harbour UCLA Medical Center, Los Angeles, California: Richard Casaburi, MD (PI), Janos Porszasz, MD, PhD, Hans Fischer, MD, PhD (RAD), Matt Budoff, MD
Michael E. DeBakey VAMC, Houston, Texas: Amir Sharafkhaneh, MD (PI), Charles Trinh, MD (RAD), Hirani Kamal, MD, Roham Darvishi, MD
Minneapolis VA: Dennis Niewoehner, MD (PI), Tadashi Allen, MD (RAD), Quentin Anderson, MD (RAD), Kathryn Rice, MD
Morehouse School of Medicine, Atlanta, Georgia: Marilyn Foreman, MD, MS (PI), Gloria Westney, MD, MS, Eugene Berkowitz, MD, PhD (RAD)
National Jewish Health, Denver, Colorado: Russell Bowler, MD, PhD (PI), Adam Friedlander, MD, David Lynch, MB (RAD), Joyce Schroeder, MD (RAD), John Newell, Jr., MD (RAD)
Temple University, Philadelphia, Pennsylvania: Gerard Criner, MD (PI), Victor Kim, MD, Nathaniel Marchetti, DO, Aditi Satti, MD, A. James Mamary, MD, Robert Steiner, MD (RAD), Chandra Dass, MD (RAD)
University of Alabama, Birmingham, Alabama: William Bailey, MD (PI), Mark Dransfield, MD (Co-PI), Hrudaya Nath, MD (RAD)
University of California, San Diego, California: Joe Ramsdell, MD (PI), Paul Friedman, MD (RAD)
University of Iowa, Iowa City, Iowa: Geoffrey McLennan, MD, PhD (PI), Edwin JR van Beek, MD, PhD (RAD), Brad Thompson, MD (RAD), Dwight Look, MD
University of Michigan, Ann Arbor, Michigan: Fernando Martinez, MD (PI), MeiLan Han, MD, Ella Kazerooni, MD (RAD)
University of Minnesota, Minneapolis, Minnesota: Christine Wendt, MD (PI), Tadashi Allen, MD (RAD)
University of Pittsburgh, Pittsburgh, Pennsylvania: Frank Sciurba, MD (PI), Joel Weissfeld, MD, MPH, Carl Fuhrman, MD (RAD), Jessica Bon, MD
University of Texas Health Science Center at San Antonio, San Antonio, Texas: Antonio Anzueto, MD (PI), Sandra Adams, MD, Carlos Orozco, MD, Mario Ruiz, MD (RAD)
Administrative Core: James Crapo, MD (PI), Edwin Silverman, MD, PhD (PI), Barry Make, MD, Elizabeth Regan, MD, Sarah Moyle, MS, Douglas Stinson
Genetic Analysis Core: Terri Beaty, PhD, Barbara Klanderman, PhD, Nan Laird, PhD, Christoph Lange, PhD, Michael Cho, MD, Stephanie Santorico, PhD, John Hokanson, MPH, PhD, Dawn DeMeo, MD, MPH, Nadia Hansel, MD, MPH, Craig Hersh, MD, MPH, Jacqueline Hetmanski, MS, Tanda Murray
Imaging Core: David Lynch, MB, Joyce Schroeder, MD, John Newell, Jr., MD, John Reilly, MD, Harvey Coxson, PhD, Philip Judy, PhD, Eric Hoffman, PhD, George Washko, MD, Raul San Jose Estepar, PhD, James Ross, MSc, Rebecca Leek, Jordan Zach, Alex Kluiber, Jered Sieren, Heather Baumhauer, Verity McArthur, Dzimitry Kazlouski, Andrew Allen, Tanya Mann, Anastasia Rodionova
PFT QA Core, LDS Hospital, Salt Lake City, UT: Robert Jensen, PhD
Biological Repository, *Johns Hopkins University, Baltimore, Maryland*: Homayoon Farzadegan, PhD, Stacey Meyerer, Shivam Chandan, Samantha Bragan
Data Coordinating Center and Biostatistics, *National Jewish Health, Denver, Colorado*: James Murphy, PhD, Douglas Everett, PhD, Carla Wilson, MS, Ruthie Knowles, Amber Powell, Joe Piccoli, Maura Robinson, Margaret Forbes, Martina Wamboldt
Epidemiology Core, *University of Colorado School of Public Health, Denver, Colorado*: John Hokanson, MPH, PhD, Marci Sontag, PhD, Jennifer Black-Shinn, MPH, Gregory Kinney, MPH

Funding This work was supported by US National Institutes of Health (NIH) grants R01 HL075478, R01 HL084323, P01 HL083069, U01 HL089856 (Silverman), K12HL089990 (Cho), and U01 HL089897 (Crapo). The National Emphysema Treatment Trial was supported by the National Heart, Lung, and Blood Institute contracts (N01HR76101, N01HR76102, N01HR76103, N01HR76104, N01HR76105, N01HR76106, N01HR76107, N01HR76108, N01HR76109, N01HR76110, N01HR76111, N01HR76112, N01HR76113, N01HR76114, N01HR76115, N01HR76116, N01HR76118, N01HR76119), the Centers for Medicare and Medicaid Services, and the Agency for Healthcare Research and Quality. The Norway cohort and the ECLIPSE study (http://www.Clinicaltrials.gov identifier NCT00292552; GSK Code SCO104960) are funded by GlaxoSmithKline. The COPDGene project was supported by Award Number U01HL089897 and Award Number U01HL089856 from the National Heart, Lung, and Blood Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, and Blood Institute or the National Institutes of Health. The COPDGene project is also supported by the COPD Foundation through contributions made to an Industry Advisory Board comprised of AstraZeneca, Pfizer, Boehringer Ingelheim, Novartis and Sepracor. M. Siedlinski is a recipient of a post-doctoral fellowship from the Niels Stensen Foundation.

Competing interests No competing interests: MS, AG, GlaxoSmithKline employee: WA, XK, NIH grant: JEH, THB, EKS, JDC, GlaxoSmithKline grant: EKS, DAL, PB, COPD Foundation grant: EKS, JDC, GlaxoSmithKline consulting fee or honorarium and support for travel to meetings for the study or other purposes: EKS, DAL, COPD Foundation support for travel to meetings for the study or other purposes: EKS, JDC, AstraZeneca and GlaxoSmithKline—payment for lectures including service on speakers bureaus and consultancy (outside the submitted work): EKS, PB, Pfizer—payment for lectures including service on speakers bureaus and consultancy (outside the submitted work): PB, GlaxoSmithKline—board membership, consultancy, payment for lectures including service on speakers bureaus and consultancy (outside the submitted work): DAL, Boehringer Ingelheim consultancy (outside the submitted work): DAL, GlaxoSmithKline fees for participation in review activities such as data monitoring boards, statistical analysis, end point committees, and the like: DAL, Consulting fee or honorarium from AstraZeneca, Novartis, Otsuka: SIR, Grant from AstraZeneca, Biomarck, Centocor, Mpex, Nabi, Novartis, Otsuka: SIR, Almirall, Novartis, Nycomed, Pfizer board membership (outside the submitted work): SIR, Almirall, AstraZeneca, Boehringer Ingelheim, California Allergy Society, Creative Educational Concept, France Foundation, GlaxoSmithKline, Information TV, Network for Continuing Education, Novartis, Nycomed, Pfizer—travel/accommodations expenses covered or reimbursed (outside the submitted work): SIR, Able Associates, Adelphi Research, APT Pharma/Britnall, Aradigm, AstraZeneca, Boehringer Ingelheim, Chiesi, Common Health, Consult Complete, COPDForum, Data Monitor, Decision Resource, Defined Health, Dey, Dunn Group, Easton Associates, Equinox, Gerson, GlaxoSmithKline, Infomed, KOL Connection, M. Pankove, MedaCorp, MDRx Financial, Mpex, Oriel Therapeutics, Otsuka, Pennside, PharmaVentures, Pharmaxis, Price Waterhouse, Propagate, Pulmatrix, Reckner Associates, Recruiting Resources, Roche, Schlesinger Medical, Scimed, Sudler and Hennessey, TargeGen, Theravance, UBC, Uptake Medical, VantagePoint Mangement—consultancy (outside the submitted work): SIR.

Ethics approval Partners Human Research Committee and IRBs at each clinical center.

Contributors funding WA, JDC, EKS; concept and design: MS, EKS; data collection: PB, AG, DAL, WA, SIR, JDC, EKS; data analysis: MS, MHC, EKS; statistical support: MS, MHC, XK, THB, JEH, EKS; manuscript writing/editing: all authors.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Vink JM,** Posthuma D, Neale MC, *et al*. Genome-wide linkage scan to identify loci for age at first cigarette in Dutch sibling pairs. *Behav Genet* 2006;**36**:100—11.
2. **Li MD,** Cheng R, Ma JZ, *et al*. A meta-analysis of estimated genetic and environmental effects on smoking behavior in male and female adult twins. *Addiction* 2003;**98**:23—31.
3. **Batra V,** Patkar AA, Berrettini WH, *et al*. The genetic determinants of smoking. *Chest* 2003;**123**:1730—9.
4. **Schnoll RA,** Johnson TA, Lerman C. Genetics and smoking behavior. *Curr Psychiatry Rep* 2007;**9**:349—57.
5. **Han S,** Gelernter J, Luo X, *et al*. Meta-analysis of 15 genome-wide linkage scans of smoking behavior. *Biol Psychiatry* 2010;**67**:12—19.
6. **Caporaso N,** Gu F, Chatterjee N, *et al*. Genome-wide and candidate gene association study of cigarette smoking behaviors. *PLoS One* 2009;**4**:e4653.
7. **Liu YZ,** Pei YF, Guo YF, *et al*. Genome-wide association analyses suggested a novel mechanism for smoking behavior regulated by IL15. *Mol Psychiatry* 2009;**14**:668—80.
8. **Thorgeirsson TE,** Gudbjartsson DF, Surakka I, *et al*. Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat Genet* 2010;**42**:448—53.

9. **Liu JZ,** Tozzi F, Waterworth DM, *et al*. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet* 2010;**42**:436—40.
10. **Tobacco and Genetics Consortium.** Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* 2010;**42**:441—7.
11. **Thorgeirsson TE,** Geller F, Sulem P, *et al*. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* 2008;**452**:638—42.
12. **Pillai SG,** Ge D, Zhu G, *et al*. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet* 2009;**5**:e1000421.
13. **Godtfredsen NS,** Lam TH, Hansel TT, *et al*. COPD-related morbidity and mortality after smoking cessation: status of the evidence. *Eur Respir J* 2008;**32**:844—53.
14. **Andreas S,** Hering T, Muhlig S, *et al*. Smoking cessation in chronic obstructive pulmonary disease: an effective medical intervention. *Dtsch Arztebl Int* 2009;**106**:276—82.
15. **Pauwels RA,** Rabe KF. Burden and clinical features of chronic obstructive pulmonary disease (COPD). *Lancet* 2004;**364**:613—20.
16. **Kanner RE,** Connett JE, Williams DE, *et al*. Effects of randomized assignment to a smoking cessation intervention and changes in smoking habits on respiratory symptoms in smokers with early chronic obstructive pulmonary disease: the Lung Health Study. *Am J Med* 1999;**106**:410—16.
17. **Hu MC,** Davies M, Kandel DB. Epidemiology and correlates of daily smoking and nicotine dependence among young adults in the United States. *Am J Public Health* 2006;**96**:299—308.
18. **Fishman A,** Martinez F, Naunheim K, *et al*. A randomized trial comparing lung-volume-reduction surgery with medical therapy for severe emphysema. *N Engl J Med* 2003;**348**:2059—73.
19. **Vestbo J,** Anderson W, Coxson HO, *et al*. Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE). *Eur Respir J* 2008;**31**:869—73.
20. **Zhu G,** Warren L, Aponte J, *et al*. The SERPINE2 gene is associated with chronic obstructive pulmonary disease in two large populations. *Am J Respir Crit Care Med* 2007;**176**:167—73.
21. **Regan EA,** Hokanson JE, Murphy JR, *et al*. Genetic epidemiology of COPD (COPDGene) study design. *COPD* 2010;**7**:32—43.
22. **Ferris BG.** Epidemiology Standardization Project (American Thoracic Society). *Am Rev Respir Dis* 1978;**118**:1—120.
23. **Cho MH,** Boutaoui N, Klanderman BJ, *et al*. Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat Genet* 2010;**42**:200—2.
24. **The International HapMap Consortium.** The international HapMap project. *Nature* 2003;**426**:789—96.
25. **Durbin RM,** Abecasis GR, Altshuler DL, *et al*. A map of human genome variation from population-scale sequencing. *Nature* 2010;**467**:1061—73.
26. **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**:e1001053.
27. **Venables WN,** Ripley BD. *Modern applied statistics with S*. 4th edn. New York: Springer, 2002.
28. **Dudbridge F,** Gusnanto A. Estimation of significance thresholds for genomewide association scans. *Genet Epidemiol* 2008;**32**:227—34.
29. **Aulchenko YS,** Ripke S, Isaacs A, *et al*. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 2007;**23**:1294—6.
30. **R Development Core Team.** R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2010. http://www.R-project.org/.
31. **Li Y,** Willer C, Sanna S, *et al*. Genotype imputation. *Annu Rev Genomics Hum Genet* 2009;**10**:387—406.
32. **Price AL,** Patterson NJ, Plenge RM, *et al*. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;**38**:904—9.
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**:559—75.
34. **Purcell S.** *PLINK (Version 1.0.7)*. http://pngu.mgh.harvard.edu/~purcell/plink/.
35. **Johnson AD,** Handsaker RE, Pulit SL, *et al*. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 2008;**24**:2938—9.
36. **Pruim RJ,** Welch RP, Sanna S, *et al*. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 2010;**26**:2336—7.
37. **Rapley EA,** Turnbull C, Al Olama AA, *et al*. A genome-wide association study of testicular germ cell tumor. *Nat Genet* 2009;**41**:807—10.
38. **Soranzo N,** Spector TD, Mangino M, *et al*. A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nat Genet* 2009;**41**:1182—90.
39. **Weiss RB,** Baker TB, Cannon DS, *et al*. A candidate gene approach identifies the CHRNA5-A3-B4 region as a risk factor for age-dependent nicotine addiction. *PLoS Genet* 2008;**4**:e1000125.
40. **Johnson EO,** Chen LS, Breslau N, *et al*. Peer smoking and the nicotinic receptor genes: an examination of genetic and environmental risks for nicotine dependence. *Addiction* 2010;**105**:2014—22.
41. **Ray R,** Tyndale RF, Lerman C. Nicotine dependence pharmacogenetics: role of genetic variation in nicotine-metabolizing enzymes. *J Neurogenet* 2009;**23**:252—61.

toOnline Supplementary Material

METHODS

Statistics

Values of percentage predicted post-bronchodilator Forced Expiratory Volume in 1 second (FEV_1 % pred.) were calculated using equations developed by Hankinson and colleagues for Caucasian individuals (ECLIPSE, NETT, and COPDGene cohorts).[1] In the Norwegian cohort FEV_1 % pred. was calculated using equations developed by Gulsvik and colleagues for the Norwegian population.[2]

All of the association analyses were adjusted for principal components for genetic ancestry (PCs; $p<0.05$ according to Tracy-Widom statistics) which were generated within COPD cases for the specific study populations.

Multivariate regression analyses (1 per phenotype per cohort), with PCs included, were run to test potentially relevant confounders. We considered sex as a potential confounder in all of the analyses, and age in the analyses on the lifetime average/current number of cigarettes smoked per day (CPD) and smoking cessation phenotypes. Since we expected that severity of COPD may affect current smoking intensity and efforts to quit smoking in COPD patients, FEV_1 % pred. was additionally considered as a potential confounder of the analyses on current CPD and smoking cessation.

In the analyses concerning log_2-transformed age at smoking initiation, female sex was significantly ($p<0.05$) associated with later age at smoking initiation in all cohorts.

Female sex was significantly associated with lower lifetime average CPD (log_2-transformed) in all cohorts. Age was significantly associated with lower lifetime average CPD (log_2-transformed) in the NETT, ECLIPSE, and GenKOLS cohorts, while in the COPDGene a borderline significant ($p=0.07$) association in the opposite direction was observed.

Female sex was significantly ($p<0.05$) associated with lower current CPD (square root-transformed) in all cohorts. Age significantly associated with lower current CPD (square root-transformed) in the ECLIPSE and GenKOLS cohorts, while in COPDGene, a non-significant trend in the same direction was observed. Higher level of FEV_1 % pred. was significantly associated with higher current CPD (square root-transformed) in the COPDGene and ECLIPSE cohorts, and a non-significant association, in the same direction, in the GenKOLS cohort was observed.

Analyses on smoking cessation revealed that sex was not significantly associated with this phenotype in any cohort ($p>0.36$; effect direction not consistent between cohorts). Age was significantly associated with higher odds for smoking cessation in all cohorts. Higher level of FEV_1 % pred. was significantly associated with lower odds for smoking cessation in the GenKOLS and COPDGene cohorts, and a non-significant association in the same direction was observed in the ECLIPSE cohort.

Similar results were obtained for models concerning smoking cessation without adjustment for sex.

In the final analyses we retained potential confounders significantly associated with the specific phenotype in the majority of cohorts analyzed. Thus, final analyses were adjusted for:

- sex and PCs (age at smoking initiation),
- sex, age, and PCs (lifetime average CPD),
- sex, age, FEV_1 % pred., and PCs (current CPD),
- age, FEV_1 % pred., and PCs (smoking cessation)

Software

Linkage disequilibrium plot was created with Haploview (ver. 4.2; default settings)[3] for Caucasian individuals from HapMap phase II.

TABLES

Table 1: SNPs associated with meta-analytic $p<10^{-5}$ with age at smoking initiation (log_2-transformed)

Chr.	SNP	Nearest gene within 50kb	Effect/Non-Effect allele	B	p	Q stat	I^2	N/N_{imp}	Freq
2	rs13408379	-	G/A	0.059	**$1.55x10^{-7}$**	0.45	0	4/1	0.85
6	rs9380362	*ZBTB9* (3'region)	C/T	0.044	**$1.62x10^{-7}$**	0.61	0	4/4	0.64
6	rs7747216	-	T/C	0.044	**$1.63x10^{-7}$**	0.60	0	4/4	0.64
6	rs7743060	*ZBTB9* (3'region)	T/C	0.044	**$1.68x10^{-7}$**	0.60	0	4/4	0.64
6	rs769051	*ZBTB9* (3'region)	G/T	0.044	**$2.07x10^{-7}$**	0.64	0	4/0	0.64
2	rs4662986	-	G/A	0.058	**$2.18x10^{-7}$**	0.45	0	4/3	0.84
2	rs4662984	-	A/G	0.057	**$2.48x10^{-7}$**	0.44	0	4/4	0.84
6	rs9296092	-	G/A	0.044	**$2.84x10^{-7}$**	0.62	0	4/3	0.64
4	rs17212303	-	C/G	0.094	**$1.72x10^{-6}$**	0.88	0	4/4	0.96
2	rs10189546	-	G/A	0.059	**$1.83x10^{-6}$**	0.60	0	4/0	0.88
4	rs17278117	-	A/G	0.093	**$2.15x10^{-6}$**	0.85	0	4/4	0.96
6	rs6926191	*CDYL* (intron)	A/G	0.080	**$2.27x10^{-6}$**	0.52	0	4/0	0.94
5	rs11242496	*NUDT12* (5' region)	A/G	0.043	**$2.39x10^{-6}$**	0.90	0	4/4	0.30
6	rs11757081	*ZBTB9* (3'region)	A/T	0.045	**$2.69x10^{-6}$**	0.82	0	4/4	0.73
20	rs6136813	*SLC24A3* (intron)	T/G	0.043	**$2.85x10^{-6}$**	0.59	0	4/4	0.26
7	rs7795747	*GIMAP2* (5' region)	A/G	0.242	**$4.87x10^{-6}$**	0.65	0	4/4	0.01
12	rs7134221	-	C/A	0.038	**$5.18x10^{-6}$**	0.93	0	4/4	0.49
18	rs17194666	*C18orf62* (3'region)	C/A	0.040	**$5.38x10^{-6}$**	0.23	31	4/4	0.33
2	rs13035304	-	G/A	0.063	**$5.40x10^{-6}$**	0.91	0	4/4	0.88
10	rs1014246	*HSPA12A* (intron)	T/C	0.040	**$6.44x10^{-6}$**	0.62	0	4/0	0.30
5	rs10045413	*CXCL14* (3'region)	T/C	0.039	**$7.35x10^{-6}$**	0.25	27	4/0	0.68
5	rs11242497	*NUDT12* (5' region)	A/G	0.039	**$8.16x10^{-6}$**	0.90	0	4/0	0.31
10	rs1867982	*CDH23* (intron)	A/G	0.058	**$8.97x10^{-6}$**	0.06	60	4/4	0.11
17	rs11871183	-	G/T	0.038	**$9.91x10^{-6}$**	0.56	0	4/0	0.35

Analyses were adjusted for sex and principal components for genetic ancestry.

N/N_{imp} = Number of studies contributing to meta-analysis / number of studies where SNP was imputed; I2=heterogeneity index; Q stat.=p value for Q statistic; p=p value from the fixed effect meta-analysis; SNP=Single Nucleotide Polymorphism; Freq. = Effect allele frequency in 3,441 subjects with at least 1 non-missing phenotype from 4 cohorts studied; Chr.=Chromosome; B=regression coefficient

Table 2: Associations of candidate SNPs mapping to loci 7p14, 8p11 and 10q23 with lifetime average CPD

Chr.	Top SNP	Nearest gene within 50kb (location)	Effect/Non-Effect allele	B	p	I^2	Q stat	N/N_{imp}	Freq.	Effect direction consistent with previous GWAS
7	rs10264177	-	A/G	0.006	0.712	50	0.11	4/4	0.66	No
7	rs9771228	-	C/T	0.007	0.678	40	0.17	4/3	0.36	Yes
7	rs7779180	-	A/G	0.014	0.445	0	0.75	4/1	0.78	No
7	rs215605	-	T/G	0.004	0.818	0	0.48	4/0	0.63	No
8	rs13273442	*CHRNB3* (5' region)	G/A	0.016	0.380	0	0.69	4/0	0.80	Yes
10	rs1329650	*LOC100188947* (intron)	G/T	0.010	0.542	28	0.25	4/3	0.72	Yes
10	rs1028936	*LOC100188947* (intron)	A/C	0.001	0.958	18	0.30	4/3	0.82	Yes

Analyses were adjusted for sex, age and principal components for genetic ancestry.

N/N_{imp} = Number of studies contributing to meta-analysis / number of studies where SNP was imputed; I^2=heterogeneity index; p=p value from the fixed effect meta-analysis; SNP=Single Nucleotide Polymorphism; CPD=number of cigarettes smoked per day; Freq. = Effect allele frequency in 3,441 subjects with at least 1 non-missing phenotype from 4 cohorts studied; Chr.=Chromosome; B=regression coefficient; *CHRNA6=alpha-nicotinic acetylcholine receptor 6*; *CHRNB3= beta-nicotinic acetylcholine receptor 3*

Table 3: Associations of candidate SNPs mapping to loci 7p14, 8p11 and 10q23 with current CPD

Chr.	Top SNP	Nearest gene within 50kb	Effect/Non-Effect allele	B	p	I^2	Q stat	N/N_{imp}	Freq.	Effect direction consistent with previous GWAS
7	rs10264177	-	A/G	0.023	0.683	0	0.68	3/3	0.66	No
7	rs9771228	-	C/T	0.005	0.916	0	0.77	3/2	0.36	Yes
7	rs7779180	-	G/A	0.010	0.866	0	0.84	3/1	0.22	Yes
7	rs215605	-	T/G	0.009	0.858	0	0.77	3/0	0.63	No
8	rs13273442	*CHRNB3* (5' region)	A/G	0.033	0.598	0	0.67	3/0	0.20	No
10	rs1329650	*LOC100188947* (intron)	G/T	0.004	0.942	76	**0.02**	3/2	0.72	Yes
10	rs1028936	*LOC100188947* (intron)	C/A	0.012	0.866	47	0.15	3/2	0.18	No

Analyses were adjusted for sex, age, % of predicted FEV_1 and principal components for genetic ancestry.

N/N_{imp} = Number of studies contributing to meta-analysis / number of studies where SNP was imputed; I^2=heterogeneity index; p=p value from the fixed effect meta-analysis; SNP=Single Nucleotide Polymorphism; CPD=number of cigarettes smoked per day; Freq. = Effect allele frequency in 3,441 subjects with at least 1 non-missing phenotype from 4 cohorts studied; Chr.=Chromosome; B=regression coefficient; *CHRNA6=alpha-nicotinic acetylcholine receptor 6*; *CHRNB3= beta-nicotinic acetylcholine receptor 3*

Table 4: Single Nucleotide Polymorphisms (SNPs) imputed using 1000 Genomes Project (pilot 1) panel and achieving a suggestive level of genome-wide significance (meta-analytic $p<5x10^{-7}$)

Phenotype	SNP id	Chr.	Position (hg18)	Nearest gene within 50kb	Effect/ Non-Effect allele	B/OR*	p	I^2	Q stat.	Imputation r^2 coefficient (range)	Freq.
Age at smoking initiation	rs9394152	6	33573460	*ZBTB9* (3'region)	T/C	0.045	$6.55x10^{-8}$	0	0.554	0.96-0.97	0.59
	rs34838160	6	33569762	*ZBTB9* (3'region)	A/G	0.046	$1.14x10^{-7}$	0	0.468	0.88-0.92	0.59
	rs35881303	6	33569757	*ZBTB9* (3'region)	G/A	0.046	$1.29x10^{-7}$	0	0.455	0.89-0.92	0.60
	rs4713628	6	33572760	*ZBTB9* (3'region)	T/C	0.044	$1.54x10^{-7}$	0	0.523	0.97-0.99	0.60
	rs4713632	6	33580968	*ZBTB9* (3'region)	T/G	0.044	$1.59x10^{-7}$	0	0.601	1.00-1.00	0.64
	rs4428487	6	33576372	*ZBTB9* (3'region)	A/G	0.044	$1.62x10^{-7}$	0	0.612	0.99-1.00	0.64
	rs9366823	6	33579037	*ZBTB9* (3'region)	T/C	0.044	$1.63x10^{-7}$	0	0.601	1.00-1.00	0.64
	rs9394153	6	33579608	*ZBTB9* (3'region)	C/A	0.044	$1.63x10^{-7}$	0	0.601	1.00-1.00	0.64
	rs9380367	6	33582953	*ZBTB9* (3'region)	A/G	0.044	$1.71x10^{-7}$	0	0.602	1.00-1.00	0.64
	rs10928927	2	130184836	-	T/C	0.057	$2.7x10^{-7}$	0	0.432	0.97-0.98	0.84
	rs73717741	6	4819758	*CDYL* (intron)	C/G	0.084	$3.37x10^{-7}$	0	0.705	0.91-0.95	0.93
	rs9380364	6	33579699	*ZBTB9* (3'region)	C/T	0.043	$4.26x10^{-7}$	0	0.519	0.97-0.98	0.62
	rs55645543	4	122826628	*ANXA5* (intron)	C/T	0.228	$4.79x10^{-7}$	0	0.789	0.50-0.60	0.02
Lifetime average CPD	rs28675338	15	76614686	*AGPHD1* (intron)	C/T	0.192	$1.23x10^{-7}$	0	0.978	0.35-0.47	0.89
	rs77155169	2	47865120	*MSH6* (intron)	C/T	0.214	$2.00x10^{-7}$	43	0.154	0.52-0.56	0.94
	rs117607728	10	96048626	*PLCE1* (intron)	T/G	0.333	$3.66x10^{-7}$	0	0.708	0.51-0.64	0.02
Smoking cessation	rs114216682	1	53643673	*FLJ40434* (3' region)	G/C	4.49	$6.78x10^{-8}$	0	0.766	0.36-0.43	0.97
Current CPD	rs56238310	3	112715929	*CD96* (5' region)	C/A	0.474	$1.20x10^{-7}$	33	0.224	0.87-0.96	0.92
	rs76884941	3	112684475	-	T/G	0.475	$1.32x10^{-7}$	22	0.278	0.80-0.90	0.92
	rs76351433	2	46876047	*SOCS5* (3' region)	C/A	0.692	$2.15x10^{-7}$	0	0.753	0.72-0.92	0.04

Presented SNPs were imputed in all cohorts. All cohorts contributed to the analyses of presented SNPs (except for the NETT cohort, which was not included in the analyses of smoking cessation and current CPD). Similarly to the results obtained with HapMap phase II reference imputation panel (supplementary table 1), SNPs rs13408379, rs9380362, rs7747216, rs7743060, rs769051, rs4662986 and rs4662984 were also associated with age at smoking initiation below suggestive genome-wide significance level, and are not depicted in the table. SNP rs9296092 was associated with age at smoking initiation with meta-analytic $p=5.77x10^{-7}$ ($p=2.84 x10^{-7}$ for HapMap phase II reference imputation panel)

Table 4 footnotes:

*OR is presented only for a single SNP from the analysis of smoking cessation

I^2=heterogeneity index; Q stat.=p value for Q statistic; p=p value from the fixed effect meta-analysis; SNP=Single Nucleotide Polymorphism; CPD=number of cigarettes smoked per day; Freq. = Effect allele frequency in 3,441 subjects with at least 1 non-missing phenotype from 4 cohorts studied; Chr.=Chromosome; B=regression coefficient; OR=Odds Ratio

FIGURES

Figure 1: Q-Q plot for the meta-analysis of age at smoking initiation phenotype

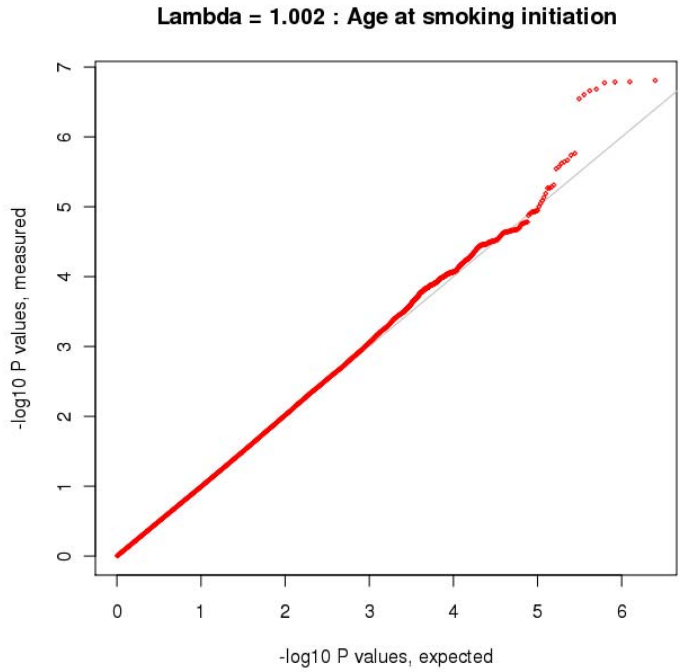


Figure 2: Q-Q plot for the meta-analysis of the lifetime average number of cigarettes smoked per day (CPD) phenotype

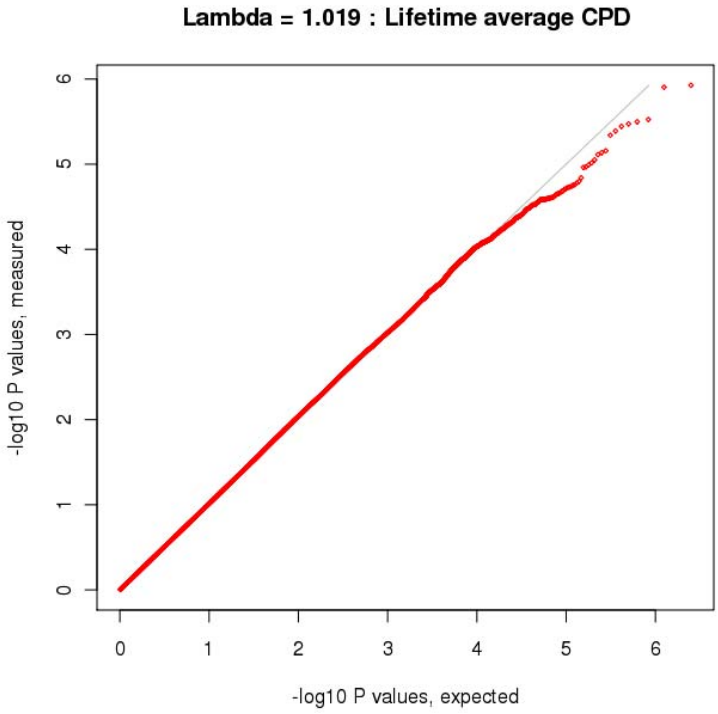
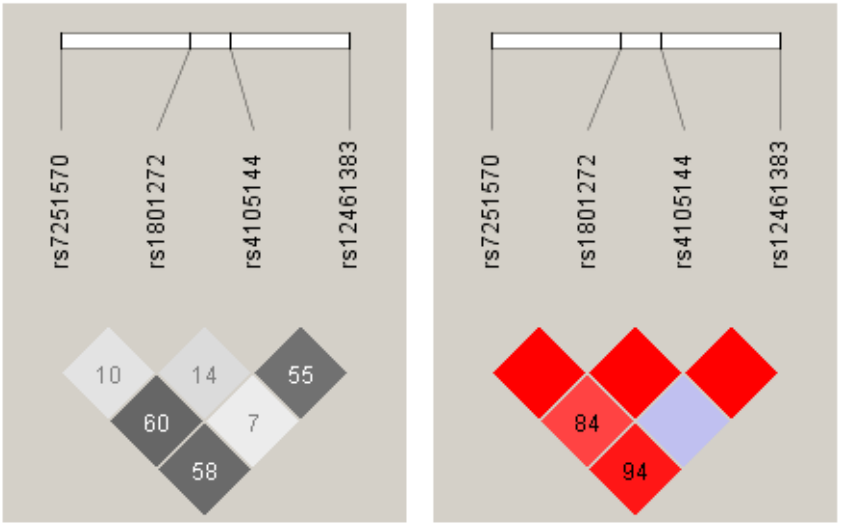


Figure 3: Linkage disequilibrium plots (100·r2 – left, 100·D' – right) containing 4 Single Nucleotide Polymorphisms from the *Cytochrome P450 2A6* locus that were significantly associated in previous GWAS with the CPD and replicated in the current study with respect to the lifetime average CPD



Plots were created for Caucasian individuals from HapMap phase II. The lack of D' values on the right plot indicates D'=1 (red color indicates LOD≥2 while blue color indicates LOD<2).

CPD=number of cigarettes smoked per day

Figure 4: Q-Q plot for the meta-analysis of the smoking cessation phenotype

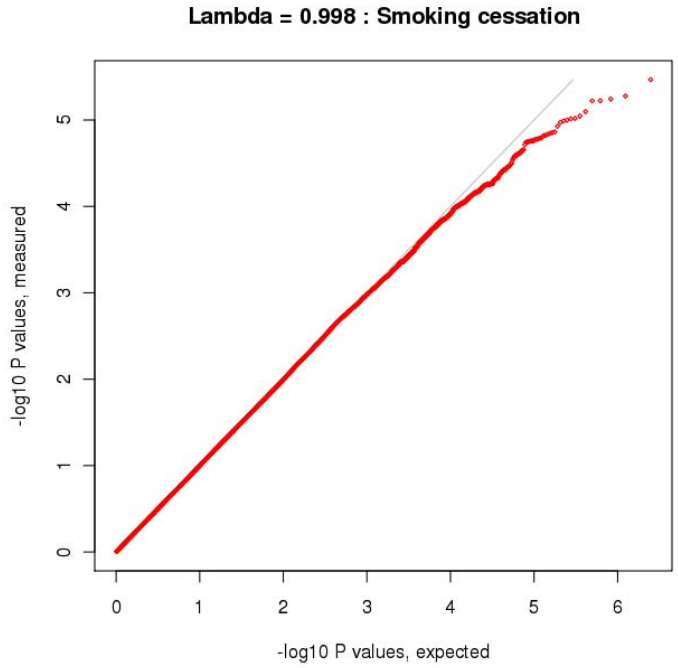
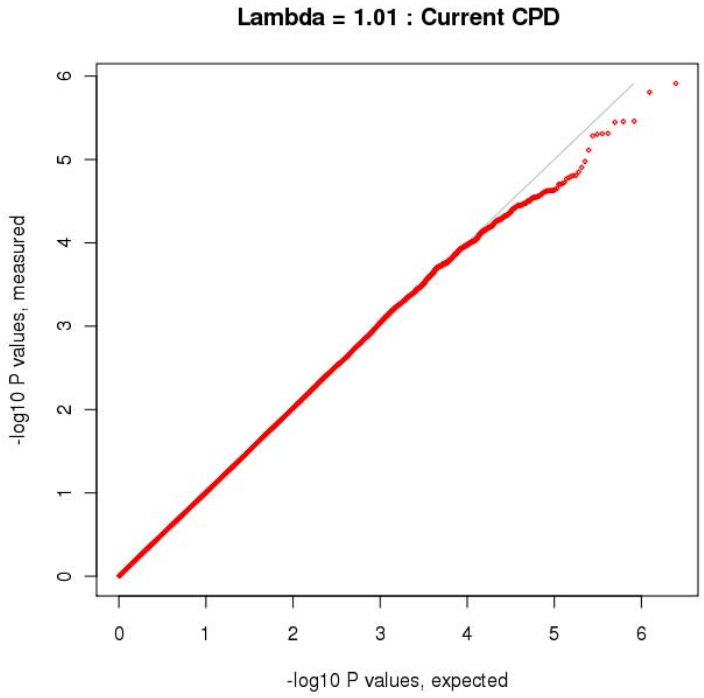


Figure 5: Q-Q plot for the meta-analysis of the current average number of cigarettes smoked per day (CPD) phenotype



REFERENCES

1. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med* 1999;**159**:179-87.

2. Gulsvik A, Tosteson T, Bakke P*, et al.* Expiratory and inspiratory forced vital capacity and one-second forced volume in asymptomatic never-smokers in Norway. *Clin Physiol* 2001;**21**:648-60.

3. Barrett JC, Fry B, Maller J*, et al.* Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;**21**:263-5.