

toOnline Supplementary Material

METHODS

Statistics

Values of percentage predicted post-bronchodilator Forced Expiratory Volume in 1 second (FEV₁ % pred.) were calculated using equations developed by Hankinson and colleagues for Caucasian individuals (ECLIPSE, NETT, and COPDGene cohorts).¹ In the Norwegian cohort FEV₁ % 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, FEV₁ % 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 FEV₁ % 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 FEV₁ % 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, FEV₁ % pred., and PCs (current CPD),
- age, FEV₁ % 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 (log₂-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	<i>CHRNA6</i> (5' region)	G/A	0.016	0.380	0	0.69	4/0	0.80	Yes
10	rs1329650	<i>LOC100188947</i> (intron)	G/T	0.010	0.542	28	0.25	4/3	0.72	Yes
10	rs1028936	<i>LOC100188947</i> (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; I²=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*; *CHRNA3*=*alpha-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	<i>CHRNA6</i> (5' region)	A/G	0.033	0.598	0	0.67	3/0	0.20	No
10	rs1329650	<i>LOC100188947</i> (intron)	G/T	0.004	0.942	76	0.02	3/2	0.72	Yes
10	rs1028936	<i>LOC100188947</i> (intron)	C/A	0.012	0.866	47	0.15	3/2	0.18	No

Analyses were adjusted for sex, age, % of predicted FEV₁ 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; 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*; *CHRNA3*= *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 $p < 5 \times 10^{-7}$)

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	6.55×10^{-8}	0	0.554	0.96-0.97	0.59
	rs34838160	6	33569762	ZBTB9 (3' region)	A/G	0.046	1.14×10^{-7}	0	0.468	0.88-0.92	0.59
	rs35881303	6	33569757	ZBTB9 (3' region)	G/A	0.046	1.29×10^{-7}	0	0.455	0.89-0.92	0.60
	rs4713628	6	33572760	ZBTB9 (3' region)	T/C	0.044	1.54×10^{-7}	0	0.523	0.97-0.99	0.60
	rs4713632	6	33580968	ZBTB9 (3' region)	T/G	0.044	1.59×10^{-7}	0	0.601	1.00-1.00	0.64
	rs4428487	6	33576372	ZBTB9 (3' region)	A/G	0.044	1.62×10^{-7}	0	0.612	0.99-1.00	0.64
	rs9366823	6	33579037	ZBTB9 (3' region)	T/C	0.044	1.63×10^{-7}	0	0.601	1.00-1.00	0.64
	rs9394153	6	33579608	ZBTB9 (3' region)	C/A	0.044	1.63×10^{-7}	0	0.601	1.00-1.00	0.64
	rs9380367	6	33582953	ZBTB9 (3' region)	A/G	0.044	1.71×10^{-7}	0	0.602	1.00-1.00	0.64
	rs10928927	2	130184836	-	T/C	0.057	2.7×10^{-7}	0	0.432	0.97-0.98	0.84
	rs73717741	6	4819758	CDYL (intron)	C/G	0.084	3.37×10^{-7}	0	0.705	0.91-0.95	0.93
	rs9380364	6	33579699	ZBTB9 (3' region)	C/T	0.043	4.26×10^{-7}	0	0.519	0.97-0.98	0.62
rs55645543	4	122826628	ANXA5 (intron)	C/T	0.228	4.79×10^{-7}	0	0.789	0.50-0.60	0.02	
Lifetime average CPD	rs28675338	15	76614686	AGPHD1 (intron)	C/T	0.192	1.23×10^{-7}	0	0.978	0.35-0.47	0.89
	rs77155169	2	47865120	MSH6 (intron)	C/T	0.214	2.00×10^{-7}	43	0.154	0.52-0.56	0.94
	rs117607728	10	96048626	PLCE1 (intron)	T/G	0.333	3.66×10^{-7}	0	0.708	0.51-0.64	0.02
Smoking cessation	rs114216682	1	53643673	FLJ40434 (3' region)	G/C	4.49	6.78×10^{-8}	0	0.766	0.36-0.43	0.97
Current CPD	rs56238310	3	112715929	CD96 (5' region)	C/A	0.474	1.20×10^{-7}	33	0.224	0.87-0.96	0.92
	rs76884941	3	112684475	-	T/G	0.475	1.32×10^{-7}	22	0.278	0.80-0.90	0.92
	rs76351433	2	46876047	SOCS5 (3' region)	C/A	0.692	2.15×10^{-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 $p = 5.77 \times 10^{-7}$ ($p = 2.84 \times 10^{-7}$ 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

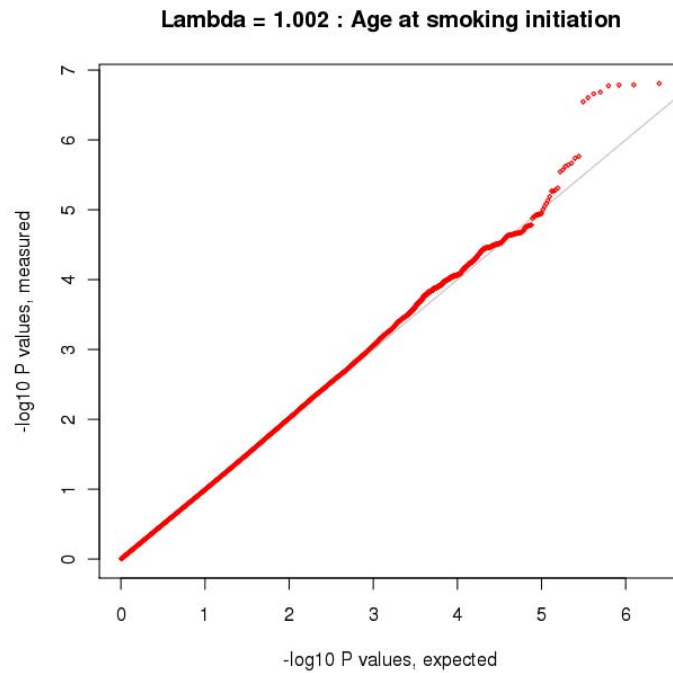


Figure 2: Q-Q plot for the meta-analysis of the lifetime average number of cigarettes smoked per day (CPD) phenotype

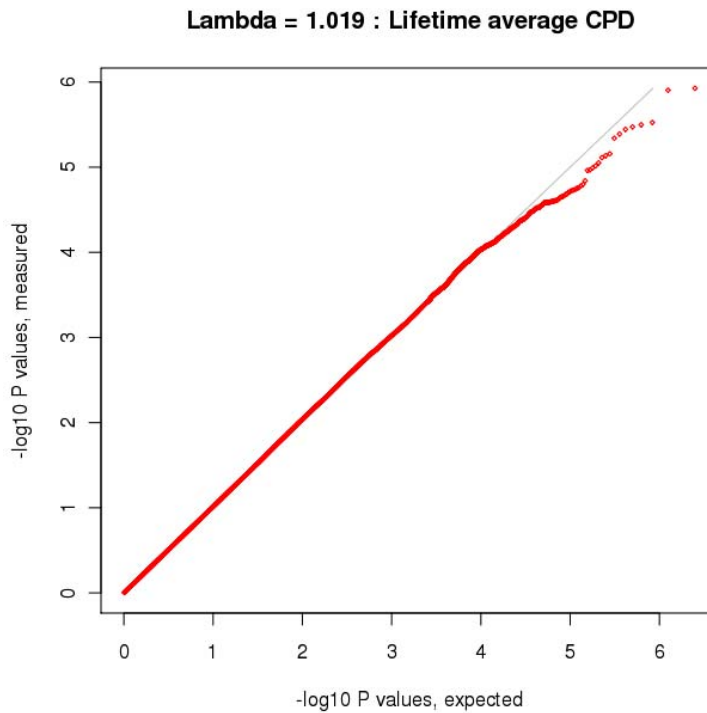
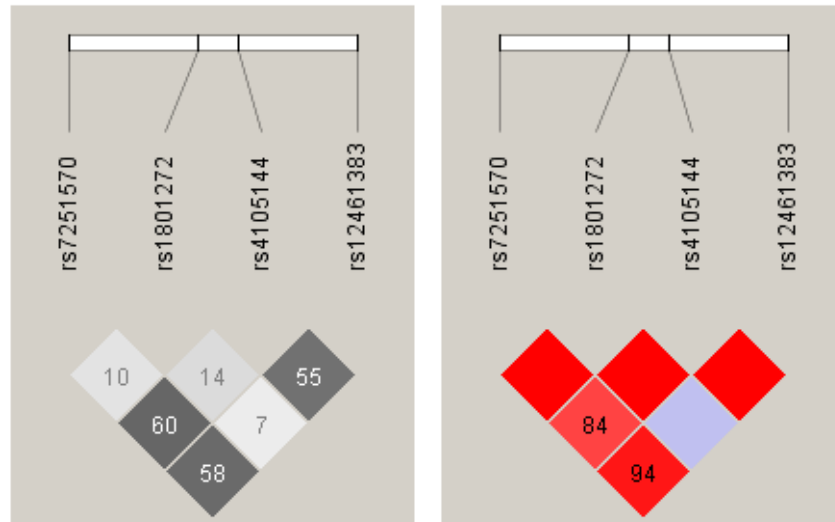


Figure 3: Linkage disequilibrium plots ($100 \cdot r^2$ – left, $100 \cdot D'$ – right) containing 4 Single Nucleotide Polymorphisms from the *Cytochrome P450 2A6* locus that were significantly associated in previous GWAS with the CPD and replicated in the current study with respect to the lifetime average 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 $\text{LOD} \geq 2$ while blue color indicates $\text{LOD} < 2$).

CPD=number of cigarettes smoked per day

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

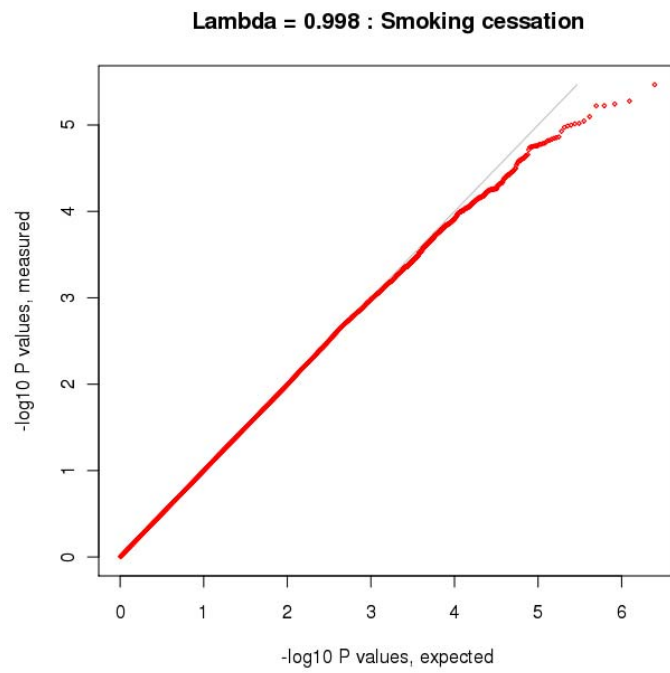
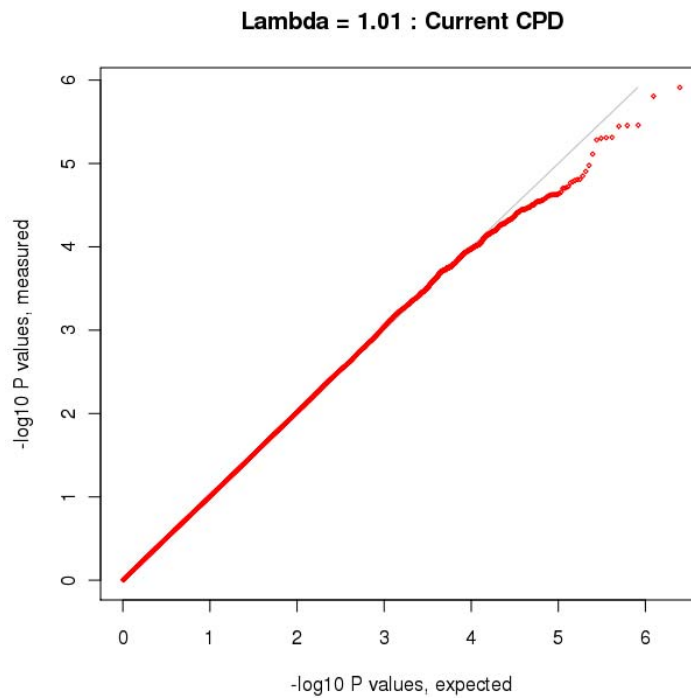


Figure 5: Q-Q plot for the meta-analysis of the current average number of cigarettes smoked per day (CPD) phenotype



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.