Chronic obstructive pulmonary disease

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 Xiayang 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

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 = 2×10⁻⁷. 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 = 2.78×10⁻³ 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.

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 scans5 and genome-wide association study (GWAS) approaches.5,7 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.\(^\text{13,14}\) Smoking cessation results in an improvement of respiratory symptoms in patients with COPD, and is associated with reduced mortality due to COPD.\(^\text{13,14}\) Another smoking-related phenotype, age at smoking onset, correlates with nicotine dependence in adulthood\(^\text{17}\) and mortality due to COPD.\(^\text{15}\) Taken together, it is of special interest to search for markers associated with smoking behaviour 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,\(^\text{18}\) ECLIPSE,\(^\text{19}\) GenKOLS cohort from Bergen, Norway,\(^\text{20}\) and COPDGene (first 1000 subjects))\(^\text{21}\) 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.\(^\text{22}\) 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,\(^\text{23}\) 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\(^{-7}\)), and filtering patients based on genotyping call rate (<95%), sex discrepancy, unexpected relatedness (FLINK phat 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>
<thead>
<tr>
<th>Table 1 Characteristics of patients with chronic obstructive pulmonary disease (COPD) and smoking-related phenotypes studied</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Characteristics</strong></td>
</tr>
<tr>
<td><strong>NETT (n=362^*)</strong></td>
</tr>
<tr>
<td><strong>Norway (n=851^*)</strong></td>
</tr>
<tr>
<td><strong>ECLIPSE (n=1734^*)</strong></td>
</tr>
<tr>
<td><strong>COPDGene (n=494^*)</strong></td>
</tr>
<tr>
<td><strong>Phenotype studied</strong></td>
</tr>
<tr>
<td><strong>Age at smoking initiation</strong></td>
</tr>
<tr>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Number of patients with non-missing phenotype</td>
</tr>
<tr>
<td>Lambda inflation factor</td>
</tr>
<tr>
<td><strong>Lifetime average CPD</strong></td>
</tr>
<tr>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Number of patients with non-missing phenotype</td>
</tr>
<tr>
<td>Lambda inflation factor</td>
</tr>
<tr>
<td><strong>Smoking cessation</strong></td>
</tr>
<tr>
<td>Current smokers, (n) (%)</td>
</tr>
<tr>
<td>Former smokers, (n) (%)</td>
</tr>
<tr>
<td>Lambda inflation factor</td>
</tr>
<tr>
<td><strong>Current CPD(^\dagger)</strong></td>
</tr>
<tr>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Number of patients with non-missing phenotype</td>
</tr>
<tr>
<td>Lambda inflation factor</td>
</tr>
</tbody>
</table>

*Calculated for patients with at least one non-missing phenotype.

\(^\dagger\) 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.
Cohorts, with imputation $r^2$ coefficient limited our analysis to SNPs genotyped/imputed in at least two cohorts, calculated and analysed with the EIGENSOFT package (V.2.0).32 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 (V1.0.7).33 SNAP (V.2.2)35 was used to search for proxy SNPs that could be applied to all cohorts, we studied log2-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 $p<5\times10^{-5}$ was considered as genome-wide significant; $5\times10^{-8}<p<5\times10^{-7}$ was interpreted as a suggestive association in the genome-wide panel with $p<0.05$ as a suggestive association for candidate SNPs.

Statistics

According to Box-Cox transformation27 that identified approximately the best transformation of dependent variables, which could be applied to all cohorts, we studied log2-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 $p<5\times10^{-5}$ was considered as genome-wide significant; $5\times10^{-8}<p<5\times10^{-7}$ 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 package27 in R (V2.10.1).30 SNP imputation was performed with MACH (V1.1).31 Principal components, reflecting genetic structure of each study population, were calculated and analysed with the EIGENSOFT package (V2.0).32 Genetic association analyses and meta-analyses were run with PLINK (V1.0.7).33 SNAP (V.2.2)35 was used to search for proxy SNPs ($r^2=0.8$) 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 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
Candidate SNPs from the cholinergic nicotinic receptors locus on chromosome 15q25 (rs578776, rs588765, rs8034191, rs1051730 (exon, Tyr215Tyr) A/G 0.060 for the meta-analysis included 1113 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 rs3025343 (r²=0.94), was genotyped in all cohorts and confirmed this association (p<0.002, OR=1.32 for the ‘T’ effect allele; I²=18).

Current CPD Analyses were adjusted for sex, age, percentage of predicted forced expiratory volume in 1 s (FEV₁) and principal components summarising genetic ancestry. In total, the meta-analysis included 1116 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 1.005 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 associated with lifetime average CPD (p=4.46×10^-6, OR=1.32, I²=18).

Table 2 Single nucleotide polymorphisms (SNPs) associated with the lifetime average number of cigarettes smoked per day (CPD) (log2-transformed) with meta-analytic p<10^-5, and candidate SNPs from CHRNA3/CHRNA5 and CYP2A6 loci identified by previous studies.

<table>
<thead>
<tr>
<th>Chr.</th>
<th>Top SNP</th>
<th>Nearest gene within 50 kb (location)</th>
<th>Effect/ non-effect allele</th>
<th>B</th>
<th>p</th>
<th>I²</th>
<th>Q stat.</th>
<th>N/Nimp</th>
<th>Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top SNPs</td>
<td>4</td>
<td>rs13104971</td>
<td>SCD2 (intron)</td>
<td>G/A</td>
<td>0.113</td>
<td>1.18×10^-6</td>
<td>0.92</td>
<td>4/3</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>rs943306</td>
<td>ASTN2 (intron)</td>
<td>T/C</td>
<td>0.073</td>
<td>1.25×10^-6</td>
<td>0.26</td>
<td>4/0</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>rs10237067</td>
<td>-</td>
<td>T/C</td>
<td>0.072</td>
<td>2.98×10^-6</td>
<td>0.63</td>
<td>4/3</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>rs12691125</td>
<td>CALN1 (intron)</td>
<td>A/G</td>
<td>0.072</td>
<td>3.17×10^-6</td>
<td>0.42</td>
<td>4/3</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>rs4718827</td>
<td>-</td>
<td>T/C</td>
<td>0.072</td>
<td>3.36×10^-6</td>
<td>0.61</td>
<td>4/4</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>rs4717631</td>
<td>CALN1 (intron)</td>
<td>A/G</td>
<td>0.072</td>
<td>3.60×10^-6</td>
<td>0.43</td>
<td>4/4</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>rs17107049</td>
<td>ELMO1 (intron)</td>
<td>C/T</td>
<td>0.265</td>
<td>4.07×10^-6</td>
<td>0.44</td>
<td>4/3</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>rs4577845</td>
<td>CALN1 (intron)</td>
<td>G/A</td>
<td>0.071</td>
<td>4.57×10^-6</td>
<td>0.44</td>
<td>4/4</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>rs1859485</td>
<td>CALN1 (intron)</td>
<td>T/C</td>
<td>0.069</td>
<td>6.95×10^-6</td>
<td>0.40</td>
<td>4/4</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>rs11044734</td>
<td>-</td>
<td>G/C</td>
<td>0.112</td>
<td>7.32×10^-6</td>
<td>0.48</td>
<td>4/4</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>rs11044737</td>
<td>-</td>
<td>C/T</td>
<td>0.112</td>
<td>7.70×10^-6</td>
<td>0.47</td>
<td>4/0</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>rs190369</td>
<td>-</td>
<td>-</td>
<td>0.161</td>
<td>8.89×10^-6</td>
<td>0.47</td>
<td>4/0</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>rs2869561</td>
<td>CEBPB (5’ region)</td>
<td>A/G</td>
<td>0.112</td>
<td>9.65×10^-6</td>
<td>0.59</td>
<td>4/1</td>
<td>0.12</td>
</tr>
<tr>
<td>Candidate SNPs</td>
<td>15</td>
<td>rs1051730</td>
<td>CHRNA3 (exon, Tyr215Tyr)</td>
<td>A/G</td>
<td>0.060</td>
<td>0.00011</td>
<td>0.37</td>
<td>4/0</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>rs16969968</td>
<td>CHRNA5 (exon, Asp398Asn)</td>
<td>A/G</td>
<td>0.059</td>
<td>0.00015</td>
<td>0.30</td>
<td>4/3</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>rs8034191</td>
<td>AGPDH1 (intron)</td>
<td>C/T</td>
<td>0.055</td>
<td>0.00036</td>
<td>0.40</td>
<td>4/0</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>rs864513</td>
<td>CHRNA5 (intron)</td>
<td>C/G</td>
<td>0.029</td>
<td>0.163</td>
<td>0.98</td>
<td>4/4</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>rs578776</td>
<td>CHRNA3 (3’ UTR)</td>
<td>G/A</td>
<td>0.041</td>
<td>0.020</td>
<td>0.94</td>
<td>4/0</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>rs588765</td>
<td>CHRNA5 (intron)</td>
<td>C/T</td>
<td>0.031</td>
<td>0.046</td>
<td>0.47</td>
<td>4/4</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>rs3733829</td>
<td>EGLN2 (intron)</td>
<td>G/A</td>
<td>0.019</td>
<td>0.223</td>
<td>0.48</td>
<td>4/3</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>rs7937</td>
<td>RAB4B (3’ UTR)</td>
<td>T/C</td>
<td>0.022</td>
<td>0.153</td>
<td>0.96</td>
<td>4/0</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>rs1801272</td>
<td>CYP2A6 (5 region)</td>
<td>A/T</td>
<td>0.266</td>
<td>2.78×10^-6</td>
<td>0.72</td>
<td>4/4</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>rs4105144</td>
<td>CYP2A6 (5 region)</td>
<td>C/T</td>
<td>0.073</td>
<td>3.92×10^-6</td>
<td>0.44</td>
<td>4/4</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>rs260329</td>
<td>CYP2B6 (intron)</td>
<td>G/A</td>
<td>0.014</td>
<td>0.404</td>
<td>0.55</td>
<td>4/1</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>rs7251570</td>
<td>CYP2A6 (3 region)</td>
<td>G/A</td>
<td>0.057</td>
<td>0.00073</td>
<td>0.64</td>
<td>4/4</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>rs12461383</td>
<td>CYP2A7 (3 region)</td>
<td>G/C</td>
<td>0.063</td>
<td>0.00019</td>
<td>0.61</td>
<td>4/4</td>
<td>0.60</td>
</tr>
</tbody>
</table>
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=6.55×10^{-6}) 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 rs28675538 ($r^2=0.07$ 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<5×10^{-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<5×10^{-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).
from the dopamine beta-hydroxylase (DBH) locus identified by previous genome-wide association studies (GWAS) on smoking cessation.

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. 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.

We speculate that this might be due to between-study differences in general (e.g. educational status or peer smoking) or those previously reported.

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

<table>
<thead>
<tr>
<th>Chr.</th>
<th>Top SNP</th>
<th>Nearest gene within 50 kb</th>
<th>Effect/ non-effect allele</th>
<th>OR</th>
<th>p</th>
<th>I^2</th>
<th>Q stat.</th>
<th>N/Nimp</th>
<th>Freq.</th>
<th>Effect direction consistent with previous GWAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>rs10794613</td>
<td>FLJ46381 (5' region)</td>
<td>G/C</td>
<td>1.43</td>
<td>3.41 × 10^{-6}</td>
<td>0</td>
<td>0.51</td>
<td>3/3</td>
<td>0.21</td>
<td>--</td>
</tr>
<tr>
<td>3</td>
<td>rs13064954</td>
<td>CCN1 (3' region)</td>
<td>A/G</td>
<td>1.94</td>
<td>5.28 × 10^{-6}</td>
<td>0</td>
<td>0.60</td>
<td>3/2</td>
<td>0.05</td>
<td>--</td>
</tr>
<tr>
<td>10</td>
<td>rs1896376</td>
<td>CPMX2 (intron)</td>
<td>C/T</td>
<td>1.83</td>
<td>5.71 × 10^{-6}</td>
<td>57</td>
<td>0.10</td>
<td>3/2</td>
<td>0.95</td>
<td>--</td>
</tr>
<tr>
<td>13</td>
<td>rs9506942</td>
<td>--</td>
<td>C/G</td>
<td>1.29</td>
<td>5.96 × 10^{-6}</td>
<td>0</td>
<td>0.72</td>
<td>3/3</td>
<td>0.57</td>
<td>--</td>
</tr>
<tr>
<td>13</td>
<td>rs9525733</td>
<td>--</td>
<td>G/A</td>
<td>1.29</td>
<td>5.96 × 10^{-6}</td>
<td>0</td>
<td>0.69</td>
<td>3/0</td>
<td>0.57</td>
<td>--</td>
</tr>
<tr>
<td>9</td>
<td>rs9866114</td>
<td>VEPH1 (3' region)</td>
<td>T/C</td>
<td>1.88</td>
<td>7.99 × 10^{-6}</td>
<td>0</td>
<td>0.66</td>
<td>3/3</td>
<td>0.06</td>
<td>--</td>
</tr>
<tr>
<td>3</td>
<td>rs1165640</td>
<td>--</td>
<td>C/T</td>
<td>1.56</td>
<td>8.98 × 10^{-6}</td>
<td>0</td>
<td>0.57</td>
<td>3/0</td>
<td>0.10</td>
<td>--</td>
</tr>
<tr>
<td>12</td>
<td>rs10861185</td>
<td>TXNRD1 (intron)</td>
<td>C/A</td>
<td>1.29</td>
<td>9.57 × 10^{-6}</td>
<td>16</td>
<td>0.31</td>
<td>3/2</td>
<td>0.57</td>
<td>--</td>
</tr>
<tr>
<td>10</td>
<td>rs7274171</td>
<td>CPXM2 (intron)</td>
<td>C/G</td>
<td>1.71</td>
<td>9.67 × 10^{-6}</td>
<td>38</td>
<td>0.20</td>
<td>3/3</td>
<td>0.94</td>
<td>--</td>
</tr>
<tr>
<td>9</td>
<td>rs3025343</td>
<td>DBH (5' region)</td>
<td>G/A</td>
<td>1.24</td>
<td>0.015</td>
<td>46</td>
<td>0.16</td>
<td>3/3</td>
<td>0.87</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Analyses were adjusted for age, % of predicted forced expiratory volume in 1 s (FEV₁) 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/Nimp, 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. Our study confirms previous observations that rs1051750, and tagging SNPs such as rs16969968, exhibit relatively strong effects in this region, and SNPs rs587765 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 rs1051750 in the Tobacco Genetics Consortium. Importantly, the effect directions of all candidate SNPs in this 15q25 locus agree with...
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 in is LD (D2=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.35 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 ECLN2, RAB4B and CY2B6.

Analysis of smoking cessation did not reveal any loci associated below the suggestive significance level, and it showed that the previously reported10 association between rs3025545 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,9 and is an important mediator of addiction behaviours such as smoking. The rs3025545 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 rs3025516 (genotyped in all cohorts) showed a more pronounced effect with a somewhat smaller heterogeneity index compared with the rs3025543.

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 emphasizes 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 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 hypothesize 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**

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

**Acknowledgements**

ECLIPSE Steering Committee: Harvey Coxson (Canada), Lisa Edwards (GlaxoSmithKline, UK), Katharine Knoebi (Co-chair, GlaxoSmithKline, UK), David Lomas (UK), William MacNee (UK), Edwin Silverman (USA), Ruth Tal-Singer (GlaxoSmithKline, UK), Jørgen Vestbo (Co-chair, Denmark), Julie Yates (GlaxoSmithKline, UK).

ECLIPSE Scientific Committee: Alvar Agusti (Spain), Peter Calverley (UK), Bartolome Cell (USA), Courtney Crim (GlaxoSmithKline, USA), Bruce Miller (GlaxoSmithKline, UK), William MacNee (Chair, UK), Stephen Rennard (USA), Ruth Tal-Singer (GlaxoSmithKline, UK), Emiel Wouters (The Netherlands), Julie Yates (GlaxoSmithKline, UK).

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 HL038069, U01 HL088856 (Silverman), K12HL098990 (Cho), and U01 HL08987 (Capan). The National Emphysema Treatment Trial was supported by the National Heart Lung and Blood Institute (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 Norwalk cohort and the ECLIPSE study (http://www.Clinicaltrials.gov identifier NCT00292552, GSK Code SC1019480) are funded by GlaxoSmithKline. The COPDGene project was supported by Award Number U01HL08987 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 Sanofi. 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, XX, 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 grant for travel to meetings for the study or other purposes: 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): EKS, 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, Biomark, Centocor, Mpx, 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, Aradig, 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. Pankowe, MediaCorp, MDRx Financial, Mpx, Oriel Therapeutics, Otsuka, Pennside, PharmaVentures, Pharmaxis, Pice Wardenhouse, Propagata, Pulmatrix, Reckner Associates, Recruiting Resources, Roche, Schlesinger Medical, Scimed, Sudler and Hennessy, TargetGen, Theravance, UBC, Uptake Medical, VantagePoint Management—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, JDC, EKS; manuscript writing/editing: all authors.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES
Chronic obstructive pulmonary disease


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


Thorax 2011 66: 894-902 originally published online June 16, 2011
doi: 10.1136/thoraxjnl-2011-200154

Updated information and services can be found at:
http://thorax.bmj.com/content/66/10/894

These include:

Supplementary Material
Supplementary material can be found at:
http://thorax.bmj.com/content/suppl/2011/06/29/thoraxjnl-2011-200154.DC1

References
This article cites 38 articles, 5 of which you can access for free at:
http://thorax.bmj.com/content/66/10/894#BIBL

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections
Articles on similar topics can be found in the following collections

Health education (1223)
Smoking (1037)
Tobacco use (1039)
Smoking cessation (154)
Health effects of tobacco use (211)
Molecular genetics (211)
Tobacco related disease (17)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/