Sequence variants at CHRNβ3–CHRNA6 and CYP2A6 affect smoking behavior


Smoking is a common risk factor for many diseases1. We conducted genome-wide association meta-analyses for the number of cigarettes smoked per day (CPD) in smokers (n = 31,266) and smoking initiation (n = 46,481) using samples from the ENGAGE Consortium. In a second stage, we tested selected SNPs with in silico replication in the Tobacco and Genetics (TAG) and Glaxo Smith Kline (Ox-GSK) consortia cohorts (n = 45,691 smokers) and assessed some of those in a third sample of European ancestry (n = 9,040). Variants in three genomic regions associated with CPD (P < 5 × 10−8), including previously identified loci at 15q25 represented by rs1051730[A] (effect size = 0.80 CPD, P = 2.4 × 10−69), and SNPs at 19q13 and 8p11, represented by rs4105144[C] (effect size = 0.39 CPD, P = 2.2 × 10−12) and rs6474412-T (effect size = 0.29 CPD, P = 1.4 × 10−8), respectively. Among the genes at the two newly associated loci are genes encoding nicotine-metabolizing enzymes (CYP2A6 and CYP2B6) and nicotinic acetylcholine receptor subunits (CHRNβ3 and CHRNA6), all of which have been highlighted in previous studies of smoking and nicotine dependence3–4. Nominal associations with lung cancer were observed at both 8p11 (rs6474412[T], odds ratio (OR) = 1.09, P = 0.04) and 19q13 (rs4105144[C], OR = 1.12, P = 0.0006).

Smoking behavior and nicotine dependence are considered to be influenced by genetics5. Although environmental influences play a strong role in the initiation of smoking6, the heritability of smoking persistence, smoking quantity and nicotine dependence has been high in most twin studies6,7. Sequence variants within a cluster of genes on chromosome 15q25 that encode nicotinic acetylcholine receptors (nAChRs) have recently been shown to associate with CPD6,9, nicotine dependence3,4 and smoking-related diseases such as lung cancer8,10,11, peripheral arterial disease (PAD)8 and chronic obstructive pulmonary disease (COPD)12.

To search for additional common variants associated with smoking behavior, we performed meta-analyses of genome-wide association (GWA) studies, mainly using samples of European ancestry from the ENGAGE consortium (see URLs) and focusing on two smoking phenotypes: CPD and smoking initiation. The smoking initiation analysis was performed with a total of 30,431 ever-smokers and 16,050 never-smokers, using data from 12 GWA studies: Corogene, deCODE,
The European Genome-Phenome Archive (EGA), the Erasmus Rucphen Family study (ERF), the Northern Finland Birth Cohorts (NFBC), the KORA study (Kooperative Gesundheitsforschung in der Region Augsburg), the Netherlands Twin Registry and Netherlands Study of Depression and Anxiety (NTR-NESDA), the Rotterdam study, the Sorbs study, the United Kingdom Twin Study (TwinUK) and the Wellcome Trust Case Control Consortium Study of Coronary Heart Disease (WTCCC-CAD). For CPD, we combined data from these same 12 GWA studies, plus subjects from the Nijmegen Lung and Bladder Cancer sample (NL-BLC) study, for a total of 31,266 subjects. Information on the meta-analysis studies for CPD and smoking initiation is provided in Supplementary Table 1, the Supplementary Note and in the Online Methods. After genomic control correction of each component study, we combined association data for ~2,500,000 imputed and genotyped autosomal SNPs with a fixed-effects additive meta-analysis using the inverse-variance method for CPD and smoking initiation. Quantile-quantile plots for CPD, excluding markers in the 15q25 region, displayed only modest inflation of the $\chi^2$-test statistic (genomic control inflation factor ($\lambda_{G C}$) = 1.02) (Supplementary Fig. 1a). In addition to the 15q25 locus, SNPs at two loci, 19q13 and 7p14, were genome-wide significant (GWS) for CPD ($P < 5 \times 10^{-8}$) in the meta-analysis data. The quantile-quantile plot for smoking initiation displayed weak inflation of the $\chi^2$-test statistic ($\lambda_{G C} = 1.03$) and no GWS associations (Supplementary Fig. 1b).

We selected 15 regions for smoking initiation totaling 277 SNPs in 14 regions for CPD totaling 443 SNPs for in silico replication in samples from the TAG and the Ox-GSK consortia (see accompanying papers published in this issue13,14) (Supplementary Table 2).

For CPD, we included a region on chromosome 8p11 on the basis of (i) its large number of SNPs showing suggestive associations with CPD, (ii) the strong candidacy of genes in the region (CHRNA6 and CHRNA8, encoding nAChR subunits $\alpha$6 and $\beta$3) and (iii) previous suggestive evidence for association between SNPs within this region and nicotine dependence15,16.

In addition to the 15q25 locus, three new loci, 7p14, 8p11 and 19q13, were found to be GWS for CPD after combining the results from the ENGAGE meta-analysis set with those of TAG and Ox-GSK (Fig. 1, Table 1 and Supplementary Table 2). No GWS associations for the selected smoking initiation regions were observed in the combined analysis of the meta-analysis and the in silico data (Supplementary Table 2).

For further confirmation of the CPD association signals at the 7p14, 8p11 and 19q13 loci, selected markers from these regions were genotyped in additional samples ($n = 9,040$) from Iceland, Australia, Denmark, Germany and Spain (ISL-AUS-DEN-GER-SPA). Samples that were both in ENGAGE and either TAG or Ox-GSK were removed before obtaining the combined in silico results. Shown are the number of smokers (n), the allele frequency (freq.), the chromosome number (chr.) and position, the estimated allelic effects on CPD and their standard errors (P(effect) and s.e.m.).

Table 1 Association of markers in four chromosomal regions with CPD

<table>
<thead>
<tr>
<th>SNP</th>
<th>Effect</th>
<th>Other</th>
<th>Freq.</th>
<th>Chr.</th>
<th>Position</th>
<th>ENGAGE*</th>
<th>TAG and Ox-GSK*</th>
<th>ISL-AUS-DEN-GER-SPA*</th>
<th>Combined*</th>
<th>n</th>
<th>Effect (s.e.m.)</th>
<th>P</th>
<th>$I^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1051730</td>
<td>A</td>
<td>G</td>
<td>0.339</td>
<td>15q25</td>
<td>76,681,394</td>
<td>0.84 (0.07)</td>
<td>2.1 × 10$^{-13}$</td>
<td>0.78 (0.06)</td>
<td>5.6 × 10$^{-38}$</td>
<td>76,972</td>
<td>0.80 (0.05)</td>
<td>2.4 × 10$^{-69}$</td>
<td>0.035</td>
</tr>
<tr>
<td>rs4674412</td>
<td>T</td>
<td>C</td>
<td>0.784</td>
<td>8p11</td>
<td>42,669,655</td>
<td>0.31 (0.08)</td>
<td>1.7 × 10$^{-6}$</td>
<td>0.30 (0.07)</td>
<td>2.6 × 10$^{-5}$</td>
<td>84,956</td>
<td>0.29 (0.05)</td>
<td>1.4 × 10$^{-8}$</td>
<td>0.24</td>
</tr>
<tr>
<td>rs13290604</td>
<td>A</td>
<td>G</td>
<td>0.84</td>
<td>8p11</td>
<td>46,674,743</td>
<td>0.31 (0.08)</td>
<td>1.2 × 10$^{-7}$</td>
<td>0.30 (0.07)</td>
<td>2.7 × 10$^{-5}$</td>
<td>76,670</td>
<td>0.31 (0.05)</td>
<td>1.3 × 10$^{-8}$</td>
<td>0.24</td>
</tr>
<tr>
<td>rs215614</td>
<td>G</td>
<td>A</td>
<td>0.356</td>
<td>7p14</td>
<td>32,313,860</td>
<td>0.38 (0.07)</td>
<td>2.4 × 10$^{-8}$</td>
<td>0.17 (0.06)</td>
<td>3.6 × 10$^{-3}$</td>
<td>86,259</td>
<td>0.22 (0.04)</td>
<td>2.1 × 10$^{-7}$</td>
<td>0.018</td>
</tr>
<tr>
<td>rs215605</td>
<td>T</td>
<td>G</td>
<td>0.357</td>
<td>7p14</td>
<td>32,303,490</td>
<td>0.39 (0.07)</td>
<td>1.7 × 10$^{-7}$</td>
<td>0.17 (0.06)</td>
<td>3.5 × 10$^{-3}$</td>
<td>77,012</td>
<td>0.26 (0.04)</td>
<td>5.4 × 10$^{-9}$</td>
<td>0.12</td>
</tr>
<tr>
<td>rs7937</td>
<td>T</td>
<td>C</td>
<td>0.560</td>
<td>19q13</td>
<td>45,994,546</td>
<td>0.34 (0.07)</td>
<td>2.2 × 10$^{-7}$</td>
<td>0.19 (0.06)</td>
<td>1.1 × 10$^{-3}$</td>
<td>86,319</td>
<td>0.24 (0.04)</td>
<td>2.4 × 10$^{-9}$</td>
<td>0.45</td>
</tr>
<tr>
<td>rs1801272</td>
<td>A</td>
<td>T</td>
<td>0.961</td>
<td>19q13</td>
<td>46,466,373</td>
<td>1.08 (0.27)</td>
<td>7.0 × 10$^{-6}$</td>
<td>0.41 (0.24)</td>
<td>8.4 × 10$^{-2}$</td>
<td>66,380</td>
<td>0.68 (0.18)</td>
<td>1.1 × 10$^{-4}$</td>
<td>0.50</td>
</tr>
<tr>
<td>rs4105144</td>
<td>C</td>
<td>T</td>
<td>0.704</td>
<td>19q13</td>
<td>46,050,464</td>
<td>0.59 (0.10)</td>
<td>1.2 × 10$^{-9}$</td>
<td>0.31 (0.08)</td>
<td>5.8 × 10$^{-5}$</td>
<td>83,317</td>
<td>0.39 (0.06)</td>
<td>2.2 × 10$^{-12}$</td>
<td>0.51</td>
</tr>
<tr>
<td>rs7260329</td>
<td>G</td>
<td>A</td>
<td>0.687</td>
<td>19q13</td>
<td>46,213,748</td>
<td>0.43 (0.07)</td>
<td>1.1 × 10$^{-9}$</td>
<td>0.06 (0.06)</td>
<td>0.36</td>
<td>86,092</td>
<td>0.20 (0.04)</td>
<td>5.5 × 10$^{-6}$</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Results are given for the ENGAGE analysis, the in silico replication obtained by combining results from TAG and Ox-GSK and the results of single-SNP assay replications in samples from Iceland, Australia, Denmark, Germany and Spain (ISL-AUS-DEN-GER-SPA). Samples that were both in ENGAGE and either TAG or Ox-GSK were removed before obtaining the combined in silico results. Shown are the number of smokers (n), the allele frequency (freq.), the chromosome number (chr.) and position, the estimated allelic effects on CPD and their standard errors (P(effect) and s.e.m.). $I^2$ is an estimate of the proportion of total variation in study estimates that is due to heterogeneity ($P^2$).

*Meta-analysis, n = 31,266. **In silico replication, n = 45,691. *Direct genotyping, n = 9,040. *$I^2$ (rs1051730) = 85,997. **$I^2$ (rs1051730) = 9,040.

© 2010 Nature America, Inc. All rights reserved.
HapMap CEU samples) was previously reported to associate with nicotine dependence to an extent that could not be accounted for by rs1051730. As with rs2036534, rs578776 is in weak LD with rs1051730 (\(r^2 = 0.21, D' = 1.0\) in the HapMap CEU samples). However, rs578776 is not correlated with rs2869046 (\(r^2 = 0.038, D' = 0.41\) in the HapMap CEU samples) and thus does not explain the signal described here for rs2869046 at the 15q25 locus (Supplementary Fig. 3).

The SNPs on chromosome 8p11 that reached genome-wide significance in the HapMap CEU samples, rs6474412 (effect size = 0.29 CPD, \(P = 1.4 \times 10^{-5}\)) and rs13280604 (effect size = 0.31 CPD, \(P = 1.3 \times 10^{-5}\)), are in perfect LD with a variant (rs13277254[T]) that was previously highlighted as being suggestively associated in a GWA study of nicotine dependence using controls who smoked but had not developed nicotine dependence (rs13277254[T], \(OR = 1.19, P = 6 \times 10^{-5}\)). The other SNPs in the region showing association with CPD (Fig. 1) are all in strong LD with these two SNPs. rs6474412 is located about 2.1 kb from the 5′ end of CHRNA6, the gene encoding the β3 nAChR subunit, and belongs to a group of highly correlated SNPs that includes two SNPs in exons of CHRNA3: a synonymous SNP (rs4593) and a nonsynonymous SNP (rs4952). Although CHRNA3 is implicated by its proximity to the location of the associating SNPs, these markers could be tagging variation elsewhere within the LD block that also contains CHRNA6 (encoding the α6 nAChR subunit) (Fig. 1).

Nine different nicotinic cholinergic receptor subunits (α2–α7 and β2–β4) are expressed in the human brain and they combine with each other in diverse patterns to form various types of functional pentameric receptors distinguished by subunit composition and sensitivity to nicotine. Rodent studies have implicated CHRNA6 and CHRN3B receptor subunits in nicotine-induced dopamine release. Neither CHRNA6 nor CHRN3B are expressed in lung tissue.

The CPD-associated markers on chromosome 19q13 are located in a region harboring CYP2A6, which encodes CYP2A6, an enzyme that plays a major role in the oxidation of nicotine in human liver microsomes; this region also harbors several other genes and pseudogenes belonging to the CYP gene family (Fig. 1). Several sequence variants in or near CYP2A6 that reduce CYP2A6’s enzymatic activity have been identified. For some of these variants, effects on smoking behavior have been suggested. In the present study, the most significant association in the region was observed with rs4105144. This SNP is in LD with CYP2A6*2 (rs1801272) (\(r^2 = 0.13\) and \(D' = 1.0\) in the HapMap CEU samples) and the CYP2A6*2 reduced-function allele is only found on the background of rs4105144[C], which associates with reduced smoking quantity. Although the effect of rs4105144 (effect size \(±\) s.e.m.) (0.39 ± 0.06 CPD) is smaller than that of rs1801272 (0.68 ± 0.18 CPD) (Table 1), its association is more significant (that is, a lower \(P\) value) because of higher minor allele frequency. This suggests that rs4105144[C] may be tagging many reduced-function variants. The second most significant association in the region was with rs7937, in the untranslated 3′ end of RAB4B, which is in LD with rs4105144 (\(r^2 = 0.32, D' = 0.82\) in the HapMap CEU samples). The third most significant association in the region was with rs7260329, which is almost independent of rs4105144 (\(r^2 = 0.0064, D' = 0.091\) in the HapMap CEU samples). rs7260329 is an intronic SNP in CYP2B6, but the product of this gene converts nicotine to cotinine with about 10% of the catalytic activity of the CYP2A6 enzyme; this product also metabolizes several drugs of abuse, as well as bupropion, an atypical antidepressant also used as a smoking cessation aid. The CYP2B6 levels in the human brain are higher than those of CYP2A6 and are altered in smokers and alcoholics.

One of the ENGAGE studies (GenMetS) contained information on immune-reactive serum cotinine levels for a set of samples \((n = 485)\), which reflect the catalytic activity of the CYP2A and CYP2B gene products. One of the SNPs associated with CPD at 19q13 (rs7937[G]) showed nominally significant association with cotinine levels (effect size = 1.16, \(P = 0.0031\)), whereas the markers at 8p11 and 7p14 did not associate with cotinine levels in this sample \((P > 0.26)\). Two SNPs at 19q13, rs2233152[A] and rs2287692[A], showed a stronger association with cotinine levels than rs7937 (effect size = 1.92, \(P = 0.00021\)); however, these two SNPs did not associate strongly with CPD in the ENGAGE samples \((P = 0.013)\) and most of the markers showing the strongest association with CPD did not associate with cotinine levels.

We next assessed the SNPs from the newly identified regions associating with CPD for association with nicotine dependence, defined...
**Table 2** Association of SNPs in four chromosomal regions with lung cancer in four populations

<table>
<thead>
<tr>
<th>Population</th>
<th>rs6474412[T], chromosome 8p11</th>
<th>rs15614[G], chromosome 7p14</th>
<th>rs7937[T], chromosome 19q13</th>
<th>rs4105144[C], chromosome 19q13</th>
<th>rs7260329[G], chromosome 19q13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iceland</td>
<td>839</td>
<td>36,606</td>
<td>836</td>
<td>36,552</td>
<td>831</td>
</tr>
<tr>
<td>Denver</td>
<td>192</td>
<td>856</td>
<td>193</td>
<td>848</td>
<td>189</td>
</tr>
<tr>
<td>Spain</td>
<td>351</td>
<td>1,195</td>
<td>453</td>
<td>1,330</td>
<td>457</td>
</tr>
<tr>
<td>Netherlands</td>
<td>515</td>
<td>769</td>
<td>502</td>
<td>1,709</td>
<td>519</td>
</tr>
<tr>
<td>IARC*</td>
<td>1,914</td>
<td>2,506</td>
<td>1,917</td>
<td>2,513</td>
<td>1,899</td>
</tr>
<tr>
<td>Combined</td>
<td>4,403</td>
<td>41,340</td>
<td>4,528</td>
<td>42,296</td>
<td>4,477</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n</th>
<th>Control</th>
<th>n</th>
<th>Control</th>
<th>n</th>
<th>Control</th>
<th>n</th>
<th>Control</th>
<th>n</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case</td>
<td>36,606</td>
<td>Case</td>
<td>36,552</td>
<td>Case</td>
<td>36,606</td>
<td>Case</td>
<td>36,552</td>
<td>Case</td>
<td>36,454</td>
</tr>
<tr>
<td>Control</td>
<td>36,606</td>
<td>Control</td>
<td>36,552</td>
<td>Control</td>
<td>36,606</td>
<td>Control</td>
<td>36,552</td>
<td>Control</td>
<td>40,407</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Freq.</th>
<th>Case</th>
<th>Control</th>
<th>OR (95% c.i.)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.784</td>
<td>0.770</td>
<td>1.08 (0.96, 1.22)</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>0.805</td>
<td>0.790</td>
<td>1.09 (0.83, 1.44)</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>0.819</td>
<td>0.764</td>
<td>1.40 (1.13, 1.72)</td>
<td>0.0019</td>
<td></td>
</tr>
<tr>
<td>0.828</td>
<td>0.809</td>
<td>1.13 (0.92, 1.39)</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>0.778</td>
<td>0.763</td>
<td>1.10 (0.99, 1.22)</td>
<td>0.072</td>
<td></td>
</tr>
</tbody>
</table>

Shown are the number of cases and controls (n), the frequencies of the effect allele (Table 1) in cases and controls, the odds ratio and 95% confidence intervals and the P value for the test of association. The results for PAD are shown in Supplementary Table 4a.

*For IARC, results for rs6474412 and rs215614 were not available and here we report results for rs6474414 and rs215605, respectively, both of which are perfect surrogates in the HapMap CEU samples (r² = 1).

As we had previously found association of the 15q25 region with lung cancer and PAD, we directly genotyped selected markers from the 7p14, 8p11 and 19q13 regions for association with lung cancer (including 2,019 cases and 40,509 controls) and PAD (2,855 cases and 611 Dutch population controls. SNPs on chromosome 8p11 and chromosome 7p14 associated nominally with nicotine dependence but none of the SNPs on chromosome 19q13 did (Supplementary Table 4a).

In conclusion, we have discovered sequence variants associated with smoking behavior within regions harboring nAChR genes (CHRN3–CHRNA6, 8p11) and nicotine-metabolizing enzyme genes (CYP2A6–CYP2B6, 19q13). The 8p11 association is reminiscent of that with chromosome 15q25; both regions contain genes encoding γ-aminobutyric acid receptor subunits, as well as PDE1C, CDH13 and A2BP1, which were all highlighted in a previous GWA study of nicotine dependence and smoking cessation. Some of these genes may play a role in smoking behavior. The 13 regions that were selected from the ENGAGE meta-analysis of CPD but did not reach GWAS suggest a number of interesting functional candidate genes, including GABRA1 and GABRG2 (genes encoding γ-aminobutyric acid receptor subunits), as well as PDE1C, CDH13 and A2BP1, which were all highlighted in a previous GWA study of nicotine dependence and smoking cessation. Some of these genes may play a role in smoking behavior.
waypath will be even more difficult in case of these new variants, as their effects on both CPD and lung cancer are smaller, and further studies are warranted.


**METHODS**

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

**Note.** Supplementary information is available on the Nature Genetics website.

**ACKNOWLEDGMENTS**

We thank all the participants in the genetic studies whose contributions made this work possible. This work was supported in part by the US National Institutes of Health (R01-DA017932) and the European Commission’s Sixth Framework Program, Integrated Project GENADDICT (LSHM-CT-2004-003166). The ENGAGE smoking consortium was formed through a component of the Integrated Project ENGAGE, supported by the European Commission’s Seventh Framework Program, grant agreement HEALTH-F4-2007-201413. ENGAGE projects have benefited from the SIMBioMS platform, which has greatly facilitated data exchange and annotation. Further acknowledgments are listed in the Supplementary Note online.

**AUTHOR CONTRIBUTIONS**


**COMPETING FINANCIAL INTERESTS**

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturegenetics/.

Published online at http://www.nature.com/naturegenetics/.

Reprints and permissions information is available online at http://rnp.nature.com/reprintsandpermissions/.

Oulu, Finland. European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK. Multidisciplinary Cardiovascular Research Centre (MCRC), Leeds Institute of Genetics, Health and Therapeutics (LIGHT), University of Leeds, Leeds, UK. Vascular Research Group, Otago Medical School, Dunedin, New Zealand. Department of Laboratory Medicine, Konventospital Bamherzige Brander Linz, Linz, Austria. Landspitali University Hospital, Department of Medicine, Reykjavik, Iceland. University of Iceland, Faculty of Medicine, Reykjavik, Iceland. Therapea, Reykjavik, Iceland. National Center of Addiction Medicine, Vogur Hospital, Reykjavik, Iceland. Radboud University Nijmegen Medical Centre, Department of Urology, Nijmegen, The Netherlands. University of Zaragoza, Zaragoza, Spain. Vascular Research Unit, Viborg Hospital, Viborg, Denmark. Department of Cardiothoracic Surgery, Rigshospitalen, University of Copenhagen, Copenhagen, Denmark. Department of Pathology, University of Colorado Denver, Aurora, Colorado, USA. Community and Behavioral Health, Colorado School of Public Health, University of Colorado Denver, Aurora, Colorado, USA. Queensland Institute of Medical Research, Queensland, Australia. Department of Psychiatry, University of Munich (LMU), Munich, Germany. Department of Epidemiology and Public Health, Imperial College, Faculty of Medicine, London, UK. Institute of Health Sciences, University of Oulu, Oulu, Finland. Biocenter Oulu, University of Oulu, Oulu, Finland. Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany. Klinikum Grosshadern, Munich, Germany. Department of Cardiovascular Sciences, University of Leicester, Clinical Sciences Wing, Glenfield Hospital, Leicester, UK. Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands. Department of Mental Health and Alcohol Abuse Services, National Institute for Health and Welfare, Helsinki, Finland. A full list of members is provided in the Supplementary Note. Correspondence should be addressed to K.S. (kstefans@decode.is) or U.T. (unnur.thorsteinsdottir@decode.is).
ONLINE METHODS

Study subjects. Written informed consent was obtained from all subjects in the study populations from 12 countries (Australia, Austria, Denmark, Estonia, Finland, Germany, Iceland, The Netherlands, New Zealand, Spain, the United States and the United Kingdom). Inclusion in the study required the availability of genotypes from either previous GWA studies or follow-up genotyping of selected SNPs in additional subjects. All subjects were of European descent. The sample sizes for each of the samples used in the study are listed in Supplementary Table 1. For the ENGAGE meta-analysis of CPD, data for 31,266 smokers were used, and for the meta-analysis of smoking initiation, the number of cases and controls were 30,431 and 16,050, respectively. A brief description of each sample is provided in the Supplementary Note.

Genome-wide genotyping. Samples were genotyped on various platforms according to study. Most of the ENGAGE projects used the Illumina platform, either HumanHap300–HH370 (DCGN and NLBLC), HumanHap550 (Rotterdam Study, ERF and TwinUK), or the 610-Quad (Corogene, GenMets/FTC and NFBC); other studies used Affymetrix 500k (KORA, Sorbs and WTCCC-CAD) and Perlegen 600k (NTR-NESDA). SNP imputation was based on the Phase II CEU HapMap samples29 and was done mostly using IMPUTE30; however, some studies used MACH31 (Corogene, GenMets/FTC and NFBC), yielding a total of approximately 2.5 million SNPs. SNPs were excluded if they had (i) yield lower than 95%, (ii) minor allele frequency less than 1% in the population or (iii) showed significant deviation from Hardy-Weinberg equilibrium in the controls ($P < 0.001$). Any samples with a call rate below 98% were excluded from the analysis.

Single SNP genotyping. Single-SNP genotyping for all samples was carried out at deCODE genetics in Reykjavik, Iceland, applying the same platform to all populations studied. All single-SNP genotyping was carried out using the Centaurus (Nanogen) platform32. The quality of each Centaurus SNP assay was evaluated by genotyping each assay on the CEU samples and comparing the results with the HapMap data29. All assays had mismatch rate <0.5%. Additionally, all markers were re-genotyped on more than 10% of samples typed with the Illumina platform, resulting in an observed mismatch in less than <0.5% of samples.

Association analysis. For the quantitative trait association analysis, that is, smoking quantity measured in CPD, a classical linear regression was fit to test for association using the genotype as an additive covariate (or expected allele count for imputed SNPs) and the CPD categories as the response variable. An additive model for SNP effects was assumed in all instances. The smoking categories used were: 1–10 CPD, 11–20 CPD, 21–30 CPD and 31 CPD and over, and associations with quantitative traits were performed adjusting for sex and year of birth. We converted the result to CPD by dividing the categorical effect size by 10. The association analysis was performed by most of the ENGAGE studies using SNPTEST29, but MACH31 (KORA), ProbABEL (ERF, Rotterdam and KORA) and GenABEL (TwinUK) were also used.

For case control association analysis, for example, when comparing PAD, lung cancer or nicotine-dependent cases to population controls, we used a standard likelihood ratio statistic, implemented in the NEMO software32 to calculate two-sided $P$ values for each individual allele, assuming a multiplicative model for risk, i.e. that the risk of the two alleles a person carries multiplies34. Combined significance levels were calculated by weighing $z$ scores by the inverse of the square root of each study’s effective sample size.

Heterogeneity was examined using a likelihood ratio test by comparing the null hypothesis of the effect being the same in all populations to the alternative hypothesis of each population having a different effect. $I^2$ lies between 0% and 100% and describes the proportion of total variation in study estimates that is due to heterogeneity35.

Correction for relatedness and stratification. We estimated an inflation factor for each GWA scan by calculating the average of the $\chi^2$ statistics, which is a method of genomic control36 to adjust for both relatedness and potential population stratification. The inflation factors for CPD and smoking initiation were estimated within each study by the ratio of the median of the $\chi^2$-test statistic and its expected value (0.675), or was estimated as 1 if this ratio was calculated to be less than 1, and all the results presented from association with these traits were adjusted based on these inflation factors. The inflation factors used for correction are listed for each of the studies in Supplementary Table 1 for CPD and smoking initiation.

In-silico replication studies. The TAG and Ox-GSK consortia provided results for the selected SNPs using the same methods (that is, categorical CPD corrected for age and sex) as described above and provided results from each of the participating populations. Data from samples also present in the ENGAGE analysis were excluded from the in-silico replication stage, and data derived from samples participating in both the TAG and the Ox-GSK consortia were entered only once into the analysis.