nature genetics

Sequence variants at *CHRNB3–CHRNA6* and *CYP2A6* affect smoking behavior

Thorgeir E Thorgeirsson[1,2*], Daniel F Gudbjartsson[1], Ida Surakka[3,4], Jacqueline M Vink[5], Najaf Amin[6], Frank Geller[1], Patrick Sulem[1], Thorunn Rafnar[1], Tõnu Esko[7,8], Stefan Walter[6], Christian Gieger[9], Rajesh Rawal[9], Massimo Mangino[10], Inga Prokopenko[11,12], Reedik Mägi[11,12], Kaisu Keskitalo[13], Iris H Gudjonsdottir[1], Solveig Gretarsdottir[1], Hreinn Stefansson[1], John R Thompson[14], Yurii S Aulchenko[6], Mari Nelis[7,8], Katja K Aben[15,16], Martin den Heijer[15,17], Asger Dirksen[18], Haseem Ashraf[18], Nicole Soranzo[10,19], Ana M Valdes[10], Claire Steves[10], André G Uitterlinden[6,20], Albert Hofman[6], Anke Tönjes[21,22], Peter Kovacs[23], Jouke Jan Hottenga[5], Gonneke Willemsen[5], Nicole Vogelzangs[24], Angela Döring[9], Norbert Dahmen[25], Barbara Nitz[9], Michele L Pergadia[26], Berta Saez[27], Veronica De Diego[28], Victoria Lezcano[28], Maria D Garcia-Prats[29], Samuli Ripatti[3,4], Markus Perola[4], Johannes Kettunen[19], Anna-Liisa Hartikainen[30], Anneli Pouta[31], Jaana Laitinen[32], Matti Isohanni[30], Shen Huei-Yi[3,4], Maxine Allen[11], Maria Krestyaninova[33], Alistair S Hall[34], Gregory T Jones[35], Andre M van Rij[35], Thomas Mueller[36], Benjamin Dieplinger[36], Meinhard Haltmayer[36], Steinn Jonsson[37], Stefan E Matthiasson[38], Hogni Oskarsson[39], Thorarinn Tyrfingsson[40], Lambertus A Kiemeney[15,16,41], Jose I Mayordomo[42], Jes S Lindholt[43], Jesper Holst Pedersen[44], Wilbur A Franklin[45], Holly Wolf[46], Grant W Montgomery[47], Andrew C Heath[26], Nicholas G Martin[47], Pamela A F Madden[26], Ina Giegling[48], Dan Rujescu[48], Marjo-Riitta Järvelin[31,49–51], Veikko Salomaa[4], Michael Stumvoll[21], Tim D Spector[10], H-Erich Wichmann[9,52,53], Andres Metspalu[7,8], Nilesh J Samani[54], Brenda W Penninx[24], Ben A Oostra[55], Dorret I Boomsma[5], Henning Tiemeier[6], Cornelia M van Duijn[6], Jaakko Kaprio[3,13,56], Jeffrey R Gulcher[1], The ENGAGE Consortium[57], Mark I McCarthy[11,12], Leena Peltonen[3,19], Unnur Thorsteinsdottir[1,38] & Kari Stefansson[1,38]

Smoking is a common risk factor for many diseases[1]. 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 ($P < 5 \times 10^{-8}$), including previously identified SNPs at 15q25 represented by rs1051730[A] (effect size = 0.80 CPD, $P = 2.4 \times 10^{-69}$), and SNPs at 19q13 and 8p11, represented by rs4105144[C] (effect size = 0.39 CPD, $P = 2.2 \times 10^{-12}$) and rs6474412-T (effect size = 0.29 CPD, $P = 1.4 \times 10^{-8}$), respectively. Among the genes at the two newly associated loci are genes encoding nicotine-metabolizing enzymes (*CYP2A6* and *CYP2B6*) and nicotinic acetylcholine receptor subunits (*CHRNB3* and *CHRNA6*), all of which have been highlighted in previous studies of smoking and nicotine dependence[2–4]. 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 genetics[5]. Although environmental influences play a strong role in the initiation of smoking[6], the heritability of smoking persistence, smoking quantity and nicotine dependence has been high in most twin studies[6,7]. Sequence variants within a cluster of genes on chromosome 15q25 that encode nicotinic acetylcholine receptors (nAChRs) have recently been shown to associate with CPD[8,9], nicotine dependence[3,8] and smoking-related diseases such as lung cancer[8,10,11], 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,

*Complete list of author affiliations appears at the end of this paper.

Received 21 October 2009; accepted 18 March 2010; published online 25 April 2010; doi:10.1038/ng.573

Table 1 Association of markers in four chromosomal regions with CPD

	Allele					ENGAGE[a]		TAG and Ox-GSK[b]		ISL-AUS-DEN-GER-SPA[c]		Combined[d]				
SNP	Effect	Other	Freq.	Chr.	Position	Effect (s.e.m.)	*P*	Effect (s.e.m.)	*P*	Effect (s.e.m.)	*P*	*n*	Effect (s.e.m.)	*P*	P_{het}	I^2
rs1051730	A	G	0.339	15q25	76,681,394	0.84 (0.07)	2.1×10^{-33}	0.78 (0.06)	5.6×10^{-38}			76,972	0.80 (0.05)	2.4×10^{-69}	0.035	32
rs6474412	T	C	0.784	8p11	42,669,655	0.31 (0.08)	1.7×10^{-4}	0.30 (0.07)	2.6×10^{-5}	0.19 (0.18)	0.30	84,956	0.29 (0.05)	1.4×10^{-8}	0.24	13
rs13280604	A	G	0.784	8p11	42,678,743	0.31 (0.08)	1.2×10^{-4}	0.30 (0.07)	2.7×10^{-5}			76,670	0.31 (0.05)	1.3×10^{-8}	0.24	14
rs215614	G	A	0.356	7p14	32,313,860	0.38 (0.07)	2.4×10^{-8}	0.17 (0.06)	3.6×10^{-3}	–0.15 (0.16)	0.35	86,259	0.22 (0.04)	2.1×10^{-7}	0.018	34
rs215605	G	T	0.357	7p14	32,303,490	0.39 (0.07)	1.7×10^{-8}	0.17 (0.06)	3.5×10^{-3}			77,012	0.26 (0.04)	5.4×10^{-9}	0.12	22
rs7937	T	C	0.560	19q13	45,994,546	0.34 (0.07)	2.2×10^{-7}	0.19 (0.06)	1.1×10^{-3}	0.19 (0.14)	0.17	86,319	0.24 (0.04)	2.4×10^{-9}	0.45	1
rs1801272	A	T	0.961	19q13	46,046,373	1.08 (0.27)	7.0×10^{-5}	0.41 (0.24)	8.4×10^{-2}			66,380	0.68 (0.18)	1.1×10^{-4}	0.50	0
rs4105144	C	T	0.704	19q13	46,050,464	0.59 (0.10)	1.2×10^{-9}	0.31 (0.08)	5.8×10^{-5}	0.27 (0.15)	0.069	83,317	0.39 (0.06)	2.2×10^{-12}	0.51	0
rs7260329	G	A	0.687	19q13	46,213,478	0.43 (0.07)	1.1×10^{-9}	0.06 (0.06)	0.36	0.08 (0.16)	0.65	86,092	0.20 (0.04)	5.5×10^{-6}	0.12	21

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-AUST-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 effect allele and the other allele, the allele frequencies (freq.), the chromosome number (chr.) and position, the estimated allelic effects on CPD and their standard errors in CPD (effect and s.e.m.), the *P* value for the test of association, the *P* value for the test for heterogeneity in effect size (P_{het}) and an estimate of the proportion of total variation in study estimates that is due to heterogeneity (I^2). [a]Meta-analysis, $n = 31,266$. [b]*In silico* replication, $n = 45,691$. [c]Direct genotyping, $n = 9,040$. [d]$n = 85,997$.

The Estonian Genome Project of University of Tartu (EGPUT), 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 Controls 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 χ^2-test statistic (genomic control inflation factor (λ_{GC}) = 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 χ^2-test statistic ($\lambda_{GC} = 1.03$) and no GWS associations (**Supplementary Fig. 1b**).

We selected 15 regions for smoking initiation totaling 277 SNPs and 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 issue[13,14]) (**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 *CHRNAB3*, encoding nAChR subunits α6 and β3) and (iii) previous suggestive evidence for association between SNPs within this region and nicotine dependence[2,3].

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 (**Table 1**). The markers at 8p11 and at the 19q13 loci had effects in the same direction but the marker on 7p14 did not (**Table 1**). After combining these data with ENGAGE results and the *in silico* replication, the 8p11 and the 19q13 loci remained GWS but the 7p14 locus did not (**Table 1**). The CPD association results for the SNPs in **Table 1** for each study are presented in **Supplementary Table 3** and **Supplementary Figure 2**.

Nominally significant heterogeneity in the strength of association with CPD was observed for rs1051730 at 15q25 ($P = 0.035$, fraction of variation due to heterogeneity (I^2) = 32%) and rs215614 at 7p14 ($P = 0.018$, $I^2 = 34\%$). Given previous and current evidence, it is highly likely that there is a true association between 15q25 and CPD. Therefore, its heterogeneity must be due to differences between the study populations used, such as different CPD information ascertainment, different types of cigarettes being used, different phenotypic and demographic ascertainment strategies or different genetic structures. The heterogeneity observed at 7p14 could be caused by a false positive finding or, if it is indeed a true positive, some combination of the 'winner's curse', which inflates initial effect estimates, and the same sort of differences driving the heterogeneity at 15q25.

The strongest associations observed with CPD in the combined analysis were with SNPs within the previously identified region on chromosome 15q25 (rs1051730[A], $P = 2.4 \times 10^{-69}$, effect size = 0.80 CPD) (**Fig. 1** and **Table 1**). We searched for additional association signals in the 15q25 region which were not accounted for by rs1051730 by performing linear regression using the rs1051730 allele count as a covariate in a subset of the ENGAGE samples ($n = 23,089$). The residual signals were mostly tagged by two SNPs in relatively low linkage disequilibrium (LD), rs2869046[T] ($P = 4.8 \times 10^{-5}$, effect size = 0.5) and rs2036534[T] ($P = 9.1 \times 10^{-5}$, effect size = 0.50 CPD) ($r^2 = 0.080$ and $D' = 0.65$ in the HapMap CEU samples) (**Supplementary Fig. 3**). These two SNPs are also in fairly weak LD with rs1051730 ($r^2 = 0.12$, $D' = 0.49$ and $r^2 = 0.18$, $D' = 1.05$ in the HapMap CEU samples for rs2869046 and rs2036534, respectively). These data suggest that either the three variants (rs2869046[T], rs2036534[T] and rs1051730[A]) represent independent association signals or that a variant(s) captured by a combination of these three variants remains to be identified. A SNP, rs578776, in LD with rs2036534 ($r^2 = 0.74$, $D' = 0.95$ in the

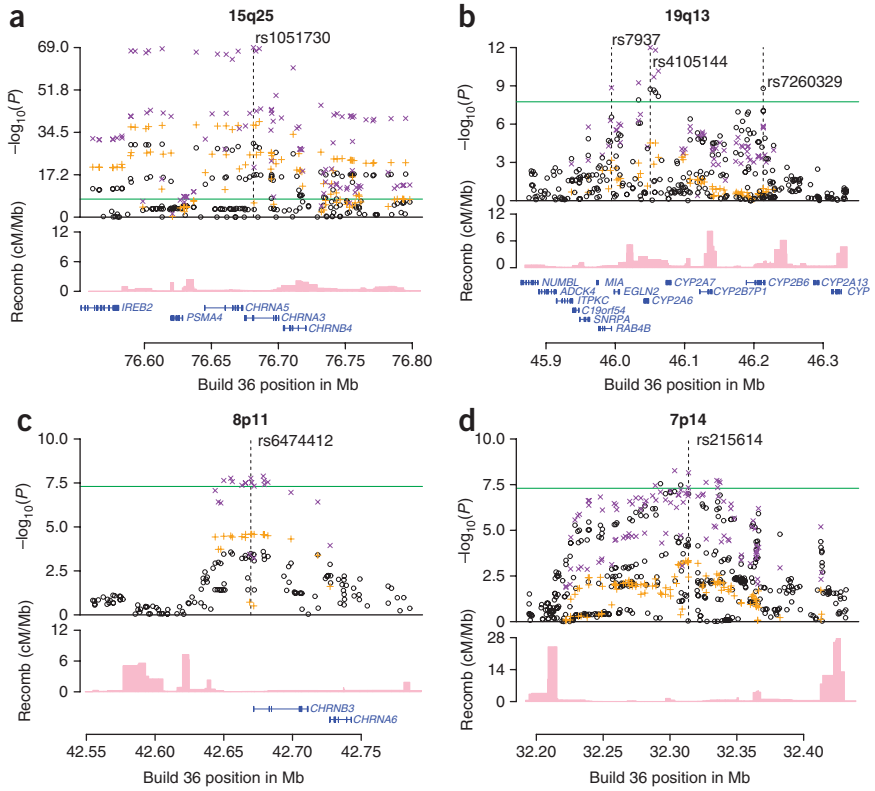

Figure 1 Regional plots of the CPD loci. The genomic regions of association with CPD on chromosomes 15q25 (**a**), 19q13 (**b**), and 8p11 (**c**) and 7p14 (**d**). Shown are the $-\log_{10}$ association *P* values of SNPs in the region with CPD from the ENGAGE meta-analysis (black circles), the *in silico* replication studies (orange crosses), the joint analysis of ENGAGE, TAG, and Ox-GSK GWA data (magenta crosses), the SNP build 36 coordinates, the genes in the region and their exons (in blue) and recombination rates in centimorgans (cM) per megabase (Mb) (pink histogram).

HapMap CEU samples) was previously reported to associate with nicotine dependence to an extent that could not be accounted for by rs1051730[15]. 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, $P = 1.4 \times 10^{-8}$) and rs13280604 (effect size = 0.31 CPD, $P = 1.3 \times 10^{-8}$), 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, $P = 6 \times 10^{-5}$)[2,3]. 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 *CHRNB3*, the gene encoding the β3 nAChR subunit, and belongs to a group of highly correlated SNPs that includes two SNPs in exons of *CHRNB3*: a synonymous SNP (rs4593) and a non-synonymous SNP (rs4952)[3]. Although *CHRNB3* 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[16]. Rodent studies have implicated CHRNA6 and CHRNB3 receptor subunits in nicotine-induced dopamine release[17]. Neither *CHRNA6* nor *CHRNB3* are expressed in lung tissue[18].

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[4]. For some of these variants, effects on smoking behavior have been suggested[4]. 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[4]. The *CYP2B6* levels in the human brain are higher than those of *CYP2A6* and are altered in smokers and alcoholics[4,19].

One of the ENGAGE studies (GenMetS) contained information on immune-reactive serum cotinine levels for a set of samples ($n = 485$)[20], which reflect the catalytic activity of the *CYP2A* and *CYP2B* gene products[4]. One of the SNPs associating 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

	n		Freq.			
Population	Case	Control	Case	Control	OR (95% c.i.)	*P*
rs6474412[T], chromosome 8p11						
Iceland	839	36,606	0.784	0.770	1.08 (0.96, 1.22)	0.19
Denver	192	856	0.805	0.790	1.09 (0.83, 1.44)	0.53
Spain	351	1,195	0.819	0.764	1.40 (1.13, 1.72)	0.0019
Netherlands	515	769	0.828	0.809	1.13 (0.92, 1.39)	0.23
IARC[a]	1,914	2,506	0.778	0.763	1.10 (0.99, 1.22)	0.072
Combined	4,403	41,340	–	–	1.12 (1.05, 1.20)	0.00060
rs215614[G], chromosome 7p14						
Iceland	839	36,606	0.366	0.355	1.05 (0.95, 1.16)	0.37
Denver	195	864	0.403	0.376	1.12 (0.89, 1.40)	0.33
Spain	450	1,281	0.370	0.335	1.17 (1.00, 1.37)	0.055
Netherlands	502	1,709	0.366	0.367	0.99 (0.86, 1.15)	0.92
IARC[a]	1,917	2,513	0.365	0.344	1.09 (1.00, 1.19)	0.057
Combined	4,499	42,377	–	–	1.07 (1.02, 1.13)	0.011
rs7937[T], chromosome 19q13						
Iceland	836	36,552	0.555	0.549	1.03 (0.93, 1.13)	0.60
Denver	193	864	0.567	0.595	0.89 (0.71, 1.12)	0.32
Spain	453	1,330	0.532	0.512	1.08 (0.93, 1.26)	0.31
Netherlands	528	1,629	0.552	0.548	1.02 (0.89, 1.17)	0.80
IARC	1,921	2,518	0.580	0.559	1.09 (1.00, 1.18)	0.048
Combined	4,528	42,296	–	–	1.05 (0.99, 1.10)	0.080
rs4105144[C], chromosome 19q13						
Iceland	839	36,606	0.713	0.705	1.04 (0.90, 1.20)	0.61
Denver	193	848	0.725	0.688	1.20 (0.94, 1.53)	0.14
Spain	437	1,288	0.669	0.620	1.24 (1.06, 1.46)	0.0085
Netherlands	513	1,665	0.638	0.640	0.99 (0.86, 1.15)	0.93
Combined	1,982	40,407	–	–	1.09 (1.00, 1.18)	0.040
rs7260329[G], chromosome 19q13						
Iceland	831	36,454	0.688	0.669	1.09 (0.98, 1.21)	0.11
Denver	189	808	0.728	0.694	1.18 (0.92, 1.51)	0.20
Spain	457	1,305	0.702	0.674	1.14 (0.97, 1.35)	0.11
Netherlands	519	1,660	0.701	0.678	1.12 (0.96, 1.30)	0.15
IARC	1,899	2,481	0.670	0.662	1.02 (0.95, 1.10)	0.61
Combined	4,477	42,126	–	–	1.06 (1.00, 1.12)	0.041

Shown are the number of cases and controls (*n*), the frequencies of the effect allele (see **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 4**.

[a]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 ($r^2 = 1$).

as a score of four or higher on the Fagerstrom Test for Nicotine Dependence (FTND) or the fulfillment of at least three of the seven Diagnostic and Statistical Manual of Mental Disorders 4th edition (DSM-IV) criteria. Allele frequencies for 1,979 Icelandic (deCODE) and 835 Dutch (NTR-NESDA) nicotine-dependent cases were compared to 36,202 Icelandic 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**).

As we had previously found association of the 15q25 region with lung cancer and PAD[8], 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 40,424 controls) in samples of European ancestry. The lung cancer data were also combined with summary-level data from the publicly available GWA dataset on lung cancer (2,518 cases and 1,921 controls) from the International Agency for Research on Cancer (IARC)[11] (**Table 2**). Nominally significant associations with lung cancer were observed for rs6474412[T] on 8p11 (OR = 1.12, 95% confidence interval (c.i.) 1.05–1.20, $P = 0.00060$), rs215614[G] on 7p14 (OR = 1.07, 95% c.i. 1.02–1.13, $P = 0.011$), and rs7260329[G] and rs4105144[C] on 19q13 (OR = 1.06, 95% c.i. 1.00–1.12, $P = 0.041$ and OR = 1.09, 95% c.i. 1.00–1.18, $P = 0.040$, respectively) (**Table 2**). Similar to the effect on CPD (**Table 1**), the effects of these variants on lung cancer is substantially weaker than that of the 15q25 variants (OR = 1.31, $P = 1.5 \times 10^{-8}$)[8,10,11] (**Table 2**), warranting further analysis in additional sample sets. No significant associations with PAD were observed for the markers tested (**Supplementary Table 4b**). The potential effect of rs7260329[G] and rs4105144[C] on lung cancer is notable in light of the fact that the *CYP2A6* gene product activates procarcinogenic nitrosamines[4].

The 13 regions that were selected from the ENGAGE meta-analysis of CPD but did not reach GWS 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[21,22]. Some of these genes may play a role in smoking behavior.

In conclusion, we have discovered sequence variants associated with smoking behavior within regions harboring nAChR genes (*CHRNB3–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 nAChRs, and the key variants in each region associate with nicotine dependence and lung cancer, bringing up the question of whether the risk for lung cancer is through the effect on smoking behavior or whether it involves increased vulnerability to the harmful effects of smoking as well[8,23–27]. However, the dissection of the causal

pathway 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.

URLs. ENGAGE Consortium, http://www.euengage.org.

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-2004-005166). 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[28], which has greatly facilitated data exchange and annotation. Further acknowledgments are listed in the **Supplementary Note** online.

AUTHOR CONTRIBUTIONS

The study was designed by and the results interpreted by T.E.T., D.F.G., F.G., J.R.G., U.T., K.S., L.P. and M.I.M. The meta-analysis was performed by D.F.G. and F.G., and D.F.G., F.G., I.S., J.M.V., P.S., N.A., T.E., S.W., C.G., R.R., M.M., I.P., R.M., J. Kettunen, Y.S.A., N.S. and J.J.H. were responsible for data analysis in each of the ENGAGE samples. Stage 3 and smoking-related disease samples were coordinated by I.H.G., H.S., S.G. and T.R. Those responsible for case and control ascertainment, recruitment and phenotypic information and project management at the study sites are: J.R.T., W.A.F., H.W., G.W.M., A.C.H., N.G.M., P.A.F.M., K.K.A., M.d.H., L.A.K., G.T.J., A.M.v.R., T.M., B.D., M.H., S.J., T.R., S.E.M., S.G., A.M.V., C.S., A.G.U., A.H., A.T., P.K., G.W., N.V., A. Dirksen, N.D., B.N., M.L.P., B.S., S.R., M.P., J. Kettunen, A.-L.H., A.P., J.L., M.I., A.S.H., T.E.T., H.O., T.T., V.D.D., V.L., M.D.G.-P., J.I.M., A. Döring, H.A., J.S.L., J.H.P., I.G., D.R., M.-R.J., V.S., M.S., T.D.S., H.-E.W., A.M., M.N., N.J.S., B.W.P., B.A.O., D.I.B., H.T., C.M.v.D., J. Kaprio, J.R.G., M.I.M., L.P., U.T. and K.S. Data submission coordination was provided by S.H.-Y., M.A. and M.K. Authors T.E.T., D.F.G. and U.T. wrote the first draft of the paper. All authors contributed to the final version of the paper.

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://npg.nature.com/reprintsandpermissions/.

1. WHO Report on the Global Tobacco Epidemic, 2008. 8 <http://tobaccofreecenter.org/mpower-2008> (2008).
2. Bierut, L.J. *et al.* Novel genes identified in a high-density genome wide association study for nicotine dependence. *Hum. Mol. Genet.* **16**, 24–35 (2007).
3. Saccone, S.F. *et al.* Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum. Mol. Genet.* **16**, 36–49 (2007).
4. Ray, R., Tyndale, R.F. & Lerman, C. Nicotine dependence pharmacogenetics: role of genetic variation in nicotine-metabolizing enzymes. *J. Neurogenet.* **23**, 252–261 (2009).
5. Rose, R., Broms, U., Korhonen, T., Dick, D. & Kaprio, J. Genetics of Smoking Behavior. in *Handbook of Behavior Genetics* (ed. Kim, Y.) Part IV, 411–432 (Springer Science+Business Media, 2009).
6. Li, M.D., Cheng, R., Ma, J.Z. & Swan, G.E. A meta-analysis of estimated genetic and environmental effects on smoking behavior in male and female adult twins. *Addiction* **98**, 23–31 (2003).
7. Koopmans, J.R., Slutske, W.S., Heath, A.C., Neale, M.C. & Boomsma, D.I. The genetics of smoking initiation and quantity smoked in Dutch adolescent and young adult twins. *Behav. Genet.* **29**, 383–393 (1999).
8. Thorgeirsson, T.E. *et al.* A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* **452**, 638–642 (2008).
9. Berrettini, W. *et al.* Alpha-5/alpha-3 nicotinic receptor subunit alleles increase risk for heavy smoking. *Mol. Psychiatry* **13**, 368–373 (2008).
10. Amos, C.I. *et al.* Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat. Genet.* **40**, 616–622 (2008).
11. Hung, R.J. *et al.* A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* **452**, 633–637 (2008).
12. Pillai, S.G. *et al.* A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet.* **5**, e1000421 (2009).
13. Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.* **42**, 441–447 (2010).
14. Liu, J. *et al.* Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat. Genet.* **42**, 436–440 (2010).
15. Stevens, V.L. *et al.* Nicotinic receptor gene variants influence susceptibility to heavy smoking. *Cancer Epidemiol. Biomarkers Prev.* **17**, 3517–3525 (2008).
16. Collins, A.C., Salminen, O., Marks, M.J., Whiteaker, P. & Grady, S.R. The road to discovery of neuronal nicotinic cholinergic receptor subtypes. *Handb. Exp. Pharmacol.* **192**, 85–112 (2009).
17. Mineur, Y.S. & Picciotto, M.R. Genetics of nicotinic acetylcholine receptors: relevance to nicotine addiction. *Biochem. Pharmacol.* **75**, 323–333 (2008).
18. West, K.A. *et al.* Rapid Akt activation by nicotine and a tobacco carcinogen modulates the phenotype of normal human airway epithelial cells. *J. Clin. Invest.* **111**, 81–90 (2003).
19. Miksys, S., Lerman, C., Shields, P.G., Mash, D.C. & Tyndale, R.F. Smoking, alcoholism and genetic polymorphisms alter CYP2B6 levels in human brain. *Neuropharmacology* **45**, 122–132 (2003).
20. Keskitalo, K. *et al.* Association of serum cotinine level with a cluster of three nicotinic acetylcholine receptor genes (CHRNA3/CHRNA5/CHRNB4) on chromosome 15. *Hum. Mol. Genet.* **18**, 4007–4012 (2009).
21. Uhl, G.R. *et al.* Molecular genetics of successful smoking cessation: convergent genome-wide association study results. *Arch. Gen. Psychiatry* **65**, 683–693 (2008).
22. Uhl, G.R. *et al.* Molecular genetics of nicotine dependence and abstinence: whole genome association using 520,000 SNPs. *BMC Genet.* **8**, 10 (2007).
23. Chanock, S.J. & Hunter, D.J. Genomics: when the smoke clears. *Nature* **452**, 537–538 (2008).
24. Thorgeirsson, T.E. & Stefansson, K. Genetics of smoking behavior and its consequences: the role of nicotinic acetylcholine receptors. *Biol. Psychiatry* **64**, 919–921 (2008).
25. Spitz, M.R., Amos, C.I., Dong, Q., Lin, J. & Wu, X. The CHRNA5–A3 region on chromosome 15q24–25.1 is a risk factor both for nicotine dependence and for lung cancer. *J. Natl. Cancer Inst.* **100**, 1552–1556 (2008).
26. Lips, E.H. *et al.* Association between a 15q25 gene variant, smoking quantity and tobacco-related cancers among 17000 individuals. *Int. J. Epidemiol.* **39**, 563–577 (2010).
27. Thorgeirsson, T.E. & Stefansson, K. Commentary: gene-environment interactions and smoking-related cancers. *Int. J. Epidemiol.* **39**, 577–579 (2010).
28. Krestyaninova, M. *et al.* A System for Information Management in BioMedical Studies–SIMBioMS. *Bioinformatics* **25**, 2768–2769 (2009).

[1]deCODE Genetics, Reykjavik, Iceland. [2]Center for Biomolecular Science and Engineering, Jack Baskin School of Engineering, University of California, Santa Cruz, California, USA. [3]Institute for Molecular Medicine (FIMM), University of Helsinki, Finland. [4]National Institute for Health and Welfare, Helsinki, Finland. [5]Department of Biological Psychology, Vrije Universiteit (VU) Amsterdam, Amsterdam, The Netherlands. [6]Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands. [7]Estonian Genome Project of University of Tartu, Tartu, Estonia. [8]Estonian Biocentre, Tartu, Estonia. [9]Institute of Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany. [10]Department of Twin Research and Genetic Epidemiology, King's College London, St. Thomas' Hospital Campus, London, UK. [11]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK. [12]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [13]Department of Public Health, University of Helsinki, Helsinki, Finland. [14]Department of Health Sciences and Genetics, University of Leicester, Leicester, UK. [15]Radboud University Nijmegen Medical Centre, Department of Epidemiology, Biostatistics and Health Technology Assessment, Nijmegen, The Netherlands. [16]Comprehensive Cancer Centre East, Nijmegen, The Netherlands. [17]Radboud University Nijmegen Medical Centre, Department of Endocrinology, Nijmegen, The Netherlands. [18]Department of Pulmonary Medicine, Gentofte University Hospital, Hellerup, Denmark. [19]Wellcome Trust Sanger Institute, Hinxton, UK. [20]Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands. [21]Department of Medicine, University of Leipzig, Leipzig, Germany. [22]Coordination Centre for Clinical Trials, University of Leipzig, Leipzig, Germany. [23]Interdisciplinary Centre for Clinical Research, University of Leipzig, Leipzig, Germany. [24]Department of Psychiatry, VU University Medical Center, Amsterdam, The Netherlands. [25]Department of Psychiatry, University of Mainz, Mainz, Germany. [26]Washington University School of Medicine, Department of Psychiatry, St. Louis, Missouri, USA. [27]Instituto de Nanotecnología de Aragón, Zaragoza, Spain. [28]Division of Dermatology, Hospital San Pedro, Logroño, Spain. [29]Division of Pathology, San Jorge Hospital, Huesca, Spain. [30]Institute of Clinical Medicine, University of Oulu, Oulu, Finland. [31]Lifecourse and Service Department, National Institute of Health and Welfare, Oulu, Finland. [32]Finnish Institute of Occupational Health,

Oulu, Finland. [33]European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK. [34]Multidisciplinary Cardiovascular Research Centre (MCRC), Leeds Institute of Genetics, Health and Therapeutics (LIGHT), University of Leeds, Leeds, UK. [35]Vascular Research Group, Otago Medical School, Dunedin, New Zealand. [36]Department of Laboratory Medicine, Konventhospital Barmherzige Brueder Linz, Linz, Austria. [37]Landspitali University Hospital, Department of Medicine, Reykjavik, Iceland. [38]University of Iceland, Faculty of Medicine, Reykjavik, Iceland. [39]Therapeia, Reykjavik, Iceland. [40]National Center of Addiction Medicine, Vogur Hospital, Reykjavik, Iceland. [41]Radboud University Nijmegen Medical Centre, Department of Urology, Nijmegen, The Netherlands. [42]University of Zaragoza, Zaragoza, Spain. [43]Vascular Research Unit, Viborg Hospital, Viborg, Denmark. [44]Department of Cardiothoracic Surgery, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark. [45]Department of Pathology, University of Colorado Denver, Aurora, Colorado, USA. [46]Community and Behavioral Health, Colorado School of Public Health, University of Colorado Denver, Aurora, Colorado, USA. [47]Queensland Institute of Medical Research, Queensland, Australia. [48]Department of Psychiatry, University of Munich (LMU), Munich, Germany. [49]Department of Epidemiology and Public Health, Imperial College, Faculty of Medicine, London, UK. [50]Institute of Health Sciences, University of Oulu, Oulu, Finland. [51]Biocenter Oulu, University of Oulu, Oulu, Finland. [52]Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany. [53]Klinikum Grosshadern, Munich, Germany. [54]Department of Cardiovascular Sciences, University of Leicester, Clinical Sciences Wing, Glenfield Hospital, Leicester, UK. [55]Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands. [56]Department of Mental Health and Alcohol Abuse Services, National Institute for Health and Welfare, Helsinki, Finland. [57]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 samples[29] and was done mostly using IMPUTE[30]; however, some studies used MACH[31] (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) platform[32]. The quality of each Centaurus SNP assay was evaluated by genotyping each assay on the CEU samples and comparing the results with the HapMap data[29]. 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[8]. 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 SNPTEST[30], but MACH[31] (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 software[33] 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 multiplies[34]. 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 heterogeneity[35].

Correction for relatedness and stratification. We estimated an inflation factor for each GWA scan by calculating the average of the χ^2 statistics, which is a method of genomic control[36] 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 χ^2-test statistic and its expected value (0.675^2), 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.

29. International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
30. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).
31. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype imputation. *Annu. Rev. Genomics Hum. Genet.* **10**, 387–406 (2009).
32. Kutyavin, I.V. *et al.* A novel endonuclease IV post-PCR genotyping system. *Nucleic Acids Res.* **34**, e128 (2006).
33. Gretarsdottir, S. *et al.* The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. *Nat. Genet.* **35**, 131–138 (2003).
34. Rice, J.A. *Mathematical Statistics and Data Analysis.* 299–330 (Wadsworth, Belmont, California, USA, 1995).
35. Higgins, J.P. & Thompson, S. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* **21**, 1539–1558 (2002).
36. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).

doi:10.1038/ng.573