Published in final edited form as:

Nat Genet. 2010 May ; 42(5): 448–453. doi:10.1038/ng.573.

Sequence variants at *CHRNA3-CHRNA6* and *CYP2A6* affect smoking behavior

Thorgeir E. Thorgeirsson^{1,2}, Daniel F. Gudbjartsson¹, Ida Surakka^{3,4}, Jacqueline M. Vink⁵, Najaf Amin⁶, Frank Geller¹, Patrick Sulem¹, Thorunn Rafnar¹, Tõnu Esko^{7,8}, Stefan Walter⁶, Christian Gieger⁹, Rajesh Rawal⁹, Massimo Mangino¹⁰, Inga Prokopenko^{11,12}, Reedik Mägi^{11,12}, Kaisu Keskitalo¹³, Iris H. Gudjonsdottir¹, Solveig Gretarsdottir¹, Hreinn Stefansson¹, John R. Thompson¹⁴, Yuri S. Aulchenko⁶, Mari Nelis^{7,8}, Katja K. Aben^{15,16}, Martin den Heijer^{15,17}, Asger Dirksen¹⁸, Haseem Ashraf¹⁸, Nicole Soranzo^{10,19}, Ana M Valdes¹⁰, Claire Steves¹⁰, André G Uitterlinden^{6,20}, Albert Hofman⁶, Anke Tönjes^{21,22}, Peter Kovacs²³, Jouke Jan Hottenga⁵, Gonneke Willemsen⁵, Nicole Vogelzangs²⁴, Angela Döring⁹, Norbert Dahmen²⁵, Barbara Nitz⁹, Michele L. Pergadia²⁶, Berta Saez²⁷, Veronica De Diego²⁸, Victoria Lezcano²⁸, Maria D. Garcia-Prats²⁹, Samuli Ripatti^{3,4}, Markus Perola⁴, Johannes Kettunen¹⁹, Anna-Liisa Hartikainen³⁰, Anneli Pouta³¹, Jaana Laitinen³², Matti Isohanni³⁰, Shen Huei-Yi^{3,4}, Maxine Allen¹¹, Maria Krestyaninova³³, Alistair S Hall³⁴, Gregory T. Jones³⁵, Andre M. van Rij³⁵, Thomas Mueller³⁶, Benjamin Dieplinger³⁶, Meinhard Haltmayer³⁶, Steinn Jonsson³⁷, Stefan E. Matthiasson³⁸, Hogni Oskarsson³⁹, Thorarinn Tyrfinngsson⁴⁰, Lambertus A. Kiemeny^{15,16,41}, Jose I. Mayordomo⁴², Jes S Lindholt⁴³, Jesper Holst Pedersen⁴⁴, Wilbur A. Franklin⁴⁵, Holly Wolf⁴⁶, Grant W. Montgomery⁴⁷, Andrew C. Heath²⁶, Nicholas G. Martin⁴⁷, Pamela A.F. Madden²⁶, Ina Giegling⁴⁸, Dan Rujescu⁴⁸, Marjo-Riitta Järvelin^{31,49,50,51}, Veikko Salomaa⁴, Michael Stumvoll²¹, Tim D Spector¹⁰, H-Erich Wichmann^{9,52,53}, Andres Metspalu^{7,8}, Nilesh J. Samani⁵⁴, Brenda W. Penninx²⁴, Ben A. Oostra⁵⁵, Dorret I. Boomsma⁵, Henning Tiemeier⁶, Cornelia M. van Duijn⁶, Jaakko Kaprio^{3,13,56}, Jeffrey R. Gulcher¹, The ENGAGE Consortium⁵⁷, Mark I. McCarthy^{11,12}, Leena Peltonen^{3,19}, Unnur Thorsteinsdottir^{1,38,*}, and Kari Stefansson^{1,38,*}

¹Decode genetics, Sturlugata 8, Reykjavik, Iceland. ²Center for Biomolecular Science and Engineering, Jack Baskin School of Engineering, University of California, Santa Cruz, CA95064, USA. ³Institute for Molecular Genetics Finland, FIMM, University of Helsinki, Finland. ⁴National Institute for Health and Welfare, Helsinki, Finland. ⁵Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands. ⁶Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands. ⁷Estonian Genome Project of University of Tartu, Tiigi str 61b, Tartu 50410, Estonia. ⁸IMCB of University of Tartu and Estonian Biocentre, Riia str 23, Tartu 51010, Estonia. ⁹Institute of Epidemiology, Helmholtz Zentrum München, Ingolstaedter Landstr. 1, 85764 Munich/Neuherberg, Germany. ¹⁰Department of Twin Research and Genetic Epidemiology, King's College London, St. Thomas' Hospital Campus, London SE1 7EH, UK. ¹¹Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Old Road, OX3 7LJ, Oxford, UK. ¹²Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt drive, OX3 7BN, Oxford, UK. ¹³Department of Public Health, University of Helsinki, Helsinki, Finland. ¹⁴Department of Health Sciences and Genetics, University of Leicester, LE1 7RH Leicester, UK. ¹⁵Radboud University Nijmegen Medical Centre, Dept. of Epidemiology, Biostatistics and HTA, Nijmegen, The Netherlands. ¹⁶Comprehensive Cancer Centre East, Nijmegen, The Netherlands. ¹⁷Radboud University Nijmegen Medical Centre, Dept. of Endocrinology, Nijmegen, The Netherlands. ¹⁸Department of pulmonary Medicine Y, Gentofte University Hospital, DK-2900 Hellerup, Denmark. ¹⁹Wellcome Trust Sanger Institute, Hinxton, UK. ²⁰Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands. ²¹Department of Medicine, University of Leipzig, Liebigstr. 18, 04103, Leipzig,

Germany. ²²Coordination Centre for Clinical Trials, University of Leipzig, Härtelstr. 16-18, 04103, Leipzig, Germany. ²³Interdisciplinary Centre for Clinical Research, University of Leipzig, Inselstr. 22, 04103, Leipzig, Germany. ²⁴EMGO Institute/Department of Psychiatry, VU University Medical Center, Amsterdam, The Netherlands. ²⁵Department of Psychiatry, University of Mainz, Mainz, Germany. ²⁶Washington University School of Medicine, Department of Psychiatry, St Louis, Missouri, USA. ²⁷Instituto de Nanotecnología de Aragón, 50009 Zaragoza, Spain. ²⁸Division of Dermatology, Hospital San Pedro, 26006 Logroño, Spain. ²⁹Division of Pathology, San Jorge Hospital, 22004 Huesca, Spain. ³⁰Institute of Clinical Medicine, University of Oulu, Oulu, Finland. ³¹Lifecourse and service Department, National Institute of Health and Welfare, Oulu, Finland. ³²Finnish Institute of Occupational Health, Oulu, Finland. ³³European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK. ³⁴Multidisciplinary Cardiovascular Research Centre (MCRC), Leeds Institute of Genetics, Health and Therapeutics (LIGHT), University of Leeds, Leeds, LS2 9JT, UK. ³⁵Vascular Research Group, Otago Medical School, PO Box 913, Dunedin 9054, New Zealand. ³⁶Department of Laboratory Medicine, Konventhospital Barmherzige Brüder Linz, Seilerstaette 2-4, A-4020 Linz, Austria. ³⁷Landspítali University Hospital, Department of Medicine, Reykjavik, Iceland. ³⁸University of Iceland, School of Medicine, Reykjavik, Iceland. ³⁹Therapeia, 101 Reykjavik, Iceland. ⁴⁰Vogur SAA Addiction Treatment Center, Reykjavik, Iceland. ⁴¹Radboud University Nijmegen Medical Centre, Dept. of Urology, Nijmegen, The Netherlands. ⁴²University of Zaragoza, 50009 Zaragoza, Spain. ⁴³Vascular Research Unit, Viborg Hospital, DK-8800 Viborg, Denmark. ⁴⁴Department of Cardiothoracic Surgery RT, Rigshospitalet/ University of Copenhagen, DK-2100 Copenhagen, Denmark. ⁴⁵Department of Pathology, University of Colorado Denver, Aurora, CO 80045, USA. ⁴⁶Community & Behavioral Health, Colorado School of Public Health, University of Colorado Denver, Aurora, CO 80045, USA. ⁴⁷Queensland Institute of Medical Research, Queensland, Australia. ⁴⁸Department of Psychiatry, University of Munich (LMU), Munich, Germany. ⁴⁹Department of Epidemiology and Public Health, Imperial College, Faculty of Medicine, London, UK. ⁵⁰Institute of Health Sciences, University of Oulu, Oulu, Finland. ⁵¹Biocenter Oulu, University of Oulu, Oulu, Finland. ⁵²Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany. ⁵³Klinikum Grosshadern, Munich, Germany. ⁵⁴Department of Cardiovascular Sciences, University of Leicester, Clinical Sciences Wing, Glenfield Hospital, Groby Road, Leicester, LE3 9QP, UK. ⁵⁵Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands. ⁵⁶Department of Mental Health and Alcohol Abuse Services, National Institute for Health and Welfare, Helsinki, Finland. ⁵⁷A list of ENGAGE members appears in the Supplementary Note.

Abstract

Smoking is a risk factor for most of the diseases leading in mortality¹. We conducted genome-wide association (GWA) meta-analyses of smoking data within the ENGAGE consortium to search for common alleles associating with the number of cigarettes smoked per day (CPD) in smokers (N=31,266) and smoking initiation (N=46,481). We tested selected SNPs in a second stage (N=45,691 smokers), and assessed some in a third sample (N=9,040). Variants in three genomic regions associated with CPD ($P < 5 \cdot 10^{-8}$), including previously identified SNPs at 15q25 represented by rs1051730-A (0.80 CPD, $P = 2.4 \cdot 10^{-69}$), and SNPs at 19q13 and 8p11, represented by rs4105144-C (0.39 CPD, $P = 2.2 \cdot 10^{-12}$) and rs6474412-T (0.29 CPD, $P = 1.4 \cdot 10^{-8}$), respectively. Among the genes at the two novel loci, are genes encoding nicotine-metabolizing enzymes (*CYP2A6* and *CYP2B6*), and nicotinic acetylcholine receptor subunits (*CHRNA3* and *CHRNA6*) highlighted in previous studies of nicotine dependence²⁻³. Nominal associations with lung cancer were observed at both 8p11 (rs6474412-T, OR=1.09, $P=0.04$) and 19q13 (rs4105144-C, OR=1.12, $P=0.0006$).

Smoking behavior and nicotine dependence (ND) are under genetic influence⁴. While environmental influences play a strong role in the initiation of smoking⁵, the heritability of smoking persistence, smoking quantity and ND is high in most twin studies^{5,6}. Sequence variants within a cluster of nicotinic acetylcholine receptor (nAChR) genes on chromosome 15q25 have recently been shown to associate with the number of cigarettes smoked per day (CPD)⁷⁻⁸, ND^{3,7}, and smoking-related diseases such as lung cancer (LC)^{7,9-10}, peripheral arterial disease (PAD)⁷, and chronic obstructive pulmonary disease (COPD)¹¹.

To search for additional common variants associating with smoking behavior we performed meta-analyses of GWA studies, mainly using samples of European ancestry from the ENGAGE consortium (See URL section), 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, EGPOT, ERF, NFBC, KORA, NTR-NESDA, Rotterdam, SUSOD, TwinUK and WTCCC-CAD. For CPD we combined data from the same studies and the NL-BLC with 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 Methods online. 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. QQ-plots for CPD, excluding markers in the 15q25 region, displayed only modest inflation of the χ^2 -test statistic ($\lambda_{GC}=1.02$) (Supplementary Figure 1A). In addition to the 15q25 locus, SNPs at two loci, 19q13 and 7p14, were genome-wide significant (GWS) for CPD ($P<5\cdot10^{-8}$) in the meta-analysis data. The QQ-plot for smoking initiation displayed weak inflation of the χ^2 -test statistic ($\lambda_{GC}=1.03$) and no GWS associations (Supplementary Figure 1B).

We selected 15 regions for smoking initiation totaling 277 SNPs, and 14 regions totaling 443 SNPs for CPD, for *in silico* replication in samples from the Tobacco and Genetics (TAG) and the Oxford-Glaxo Smith Kline (Ox-GSK) consortia (see accompanying papers^{12,13}) (Supplementary Table 2). For CPD we included a region on chromosome 8p11 based on large number of SNPs exhibiting suggestive associations with CPD, strong candidacy of region genes (encoding nAChR subunits $\alpha 6$ and $\beta 3$ (*CHRNA6* and *CHRNA3*)), and prior suggestive evidence for association between SNPs within this region and ND^{3,4}.

In addition to the 15q25 locus, three novel loci, 7p14, 8p11 and 19q13, were GWS for CPD after combining the results from the ENGAGE meta-analysis set with those of TAG and OX-GSK (Table 1, Fig 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). Markers at 8p11 and the 19q13 loci had effects in the same directions, but not the marker on 7p14 (Table 1). After combining these data with ENGAGE results and the *in silico* replication the 8p11 and the 19q13 loci remained GWS but not the 7p14 locus (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 ($P=0.035$, $I^2=32\%$) and rs215614 at 7p14 ($P=0.018$, $I^2=34\%$). There

is little doubt that the 15q25 association with CPD is real. Therefore its heterogeneity must be due to differences between the study population 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” and the same sort of differences driving the heterogeneity at 15q25.

Not surprisingly, the strongest associations observed with CPD in the combined analysis were with SNPs within the previously identified region on chromosome 15q25 (rs1051730-A, $P = 2.4 \cdot 10^{-69}$, effect size = 0.80 CPD) (Table 1, Figure 1). We searched for additional association signals in the 15q25 region, 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 LD, rs2869046-T ($P=4.8 \cdot 10^{-5}$, Effect=0.5 CPD) and rs2036534-T ($P=9.1 \cdot 10^{-5}$, Effect=0.5 CPD) ($r^2=0.080$ and $D'=0.65$ in the HapMap CEU samples) (Supplementary Figure 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 variants remains to be identified. A SNP, rs578776, in LD with rs2036534 ($r^2=0.74$, $D'=0.95$ in the HapMap CEU samples) was previously reported to associate with ND to an extent that could not be accounted for by rs1051730¹⁴. 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 ($r^2=0.0038$, $D'=0.41$ in the HapMap CEU samples) and thus does not explain the signal described here for rs2869046 at the 15q25 locus (Supplementary Figure 3).

The SNPs on chromosome 8p11 that reached genome-wide significance, rs6474412 (effect=0.29 CPD, $P=1.4 \cdot 10^{-8}$) and rs13280604 (effect=0.31 CPD, $P=1.3 \cdot 10^{-8}$), are in perfect LD in the HapMap CEU samples with a variant (rs13277254-T) that was previously highlighted as suggestively associated in a GWA study of ND using controls who had been exposed to smoking but had not developed ND (rs13277254-T, OR=1.19, $P=6 \cdot 10^{-5}$)²⁻³. The other SNPs in the region showing association with CPD (Figure 1) are all in strong LD with these SNPs. Rs6474412 is located about 2.1 kb from the 5' end of the $\beta 3$ nAChR subunit gene (*CHRNA3*), and belongs to a group of highly correlated SNPs, that includes two SNPs in exons of *CHRNA3*: a synonymous SNP (rs4593), and a non-synonymous SNP in *CHRNA3* (rs4952)³. Although the *CHRNA3* gene is implicated by the location of the associating SNPs, these markers could be tagging variation elsewhere within the LD block that also contains the $\alpha 6$ nAChR subunit gene (*CHRNA6*) (Figure 1).

Nine different nicotinic cholinergic receptor subunits ($\alpha 2$ - $\alpha 7$, $\beta 2$ - $\beta 4$) 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 nicotine¹⁵. Rodent studies have implicated *CHRNA6* and *CHRNA3* receptor subunits in nicotine-induced dopamine release¹⁶. Neither *CHRNA6* nor *CHRNA3* are expressed in lung tissue¹⁷.

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, as well as several other genes and pseudogenes belonging to the *CYP* gene family (Figure 1). A number of sequence variants in or near *CYP2A6* that reduce CYP2A6's enzymatic activity have been identified¹⁸. For some of

these variants, effects on smoking behavior have been suggested¹⁸. 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 (0.39 ± 0.06) is smaller than that of rs1801272 (0.68 ± 0.18) (Table 1), its association is more significant (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 the *RAB4B* gene, 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 that 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 its product converts nicotine to cotinine with about 10% of the catalytic activity of the *CYP2A6* enzyme, and also metabolizes several drugs of abuse, and bupropion, an atypical antidepressant also used as a smoking cessation aid¹⁸. The *CYP2B6* levels in the human brain are higher than those of *CYP2A6* and are altered in smokers and alcoholics^{18,19}.

One of the ENGAGE studies (GenMetS) had information on immune-reactive serum cotinine levels for a set of samples (N=485)²⁰, that reflects the catalytic activity of the *CYP2A* and *CYP2B* gene products¹⁸. One of the SNPs associating with CPD at 19q13 (rs7937-G) showed nominally significant association with cotinine levels (effect=1.16, $P=0.0031$), while the markers at 8p11 and 7p14 did not associate with cotinine levels in this sample ($P>0.26$). However, two SNPs at 19q13 showed a stronger association with cotinine levels than rs7937, rs2233152-A and rs2287692-A (effect=1.92, $P=0.00021$), but 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 novel regions associating with CPD for association with ND, defined as a score of four or higher on the Fagerstrom Test for Nicotine Dependence (FTND), or endorsement 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) ND cases were compared to 36,202 Icelandic and 611 Dutch population controls. SNPs on chromosome 8p11, and chromosome 7p14 associated nominally with ND, but none of the SNPs on chromosome 19q13 (Supplementary Table 4).

As we had previously found association of the 15q25 region with LC and PAD⁷, we directly genotyped selected markers from the 7p14, 8p11 and 19q13 regions for association with LC (2,019 cases and 40,509 controls) and PAD (2,855 cases, 40,424 controls) in samples of European ancestry. The LC data were also combined with summary-level data from the publicly available GWA dataset on LC (2,518 case and 1,921 controls) from the International Agency for Research on Cancer (IARC)¹⁰ (Table 2). Nominally significant associations with LC were observed for rs6474412-T on 8p11 (OR =1.12, 95%CI: 1.05-1.20, $P=0.00060$), rs215614-G on 7p14 (OR =1.07, 95%CI: 1.02-1.13, $P=0.011$), and rs7260329-G and rs4105144-C on 19q13 (OR =1.06, 95%CI: 1.00-1.12, $P=0.041$ and (OR =1.09, 95%CI: 1.00-1.18, $P=0.040$) (Table 2). As for the effect on CPD (Table 1) the effects of these variants on LC is substantially weaker than that of the 15q25 variants (OR =1.31, $P=1.5 \times 10^{-8}$)^{7,9-10} (Table 2), warranting further analysis in additional sample sets. No significant associations with PAD were observed for the markers tested (Supplementary Table 5). The potential effect of rs7260329-G and rs4105144-C on LC, is interesting in light of the fact that *CYP2A6* gene product activates procarcinogenic nitrosamines¹⁸.

The 13 regions that were selected from the ENGAGE meta-analysis of CPD but were not GWS, nominate a number of interesting functional candidate genes, including genes encoding for γ -aminobutyric acid receptor subunits (*GABRA1* and *GABRG2*), and *PDE1C*, *CDH13*, and *A2BP1*, that were all highlighted in a GWA study of ND and smoking cessation²¹⁻²², and some of these genes may play a role in smoking behavior.

In conclusion, we have discovered sequence variants associating with smoking behavior within regions harboring nAChR genes (*CHRNA3-CHRNA6*, 8p11) and nicotine-metabolizing enzymes (*CYP2A6-CYP2B6*, 19q13). The 8p11 association is reminiscent of that with chromosome 15q25; both regions contain nAChR genes, and the key variants associate with ND and LC, bringing up the question of whether the risk for LC is through the effect on smoking behavior, or involves increased vulnerability to the harmful effects of smoking as well^{7,23-27}. 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 LC are smaller, and further studies are warranted.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Footnotes

***Corresponding authors:** Kari Stefansson, deCODE genetics, Sturlugata 8, 101 Reykjavík, Iceland. kstefans@decode.is, Phone: 354-5701900, fax 354-5701901. Unnur Thorsteinsdottir, deCODE genetics, Sturlugata 8, 101 Reykjavík, Iceland. unnur.thorsteinsdottir@decode.is, Phone: 354-5701992, fax 354-5701901

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., F.G., 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 was coordinated by I.H.G., H.S., S.G. and T.R.. Those responsible for patient 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.K., 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., G.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.

e-mails:

Thorgeir E. Thorgeirsson thorgeir.thorgeirsson@decode.is
 Daniel F. Gudbjartsson dfg@decode.is
 Ida Surakka ida.surakka@thl.fi
 Jacqueline M. Vink jm.vink@psy.vu.nl
 Najaf Amin n.amin@erasmusmc.nl
 Frank Geller frank.geller@decode.is
 Patrick Sulem patrick.sulem@decode.is
 Thorunn Rafnar thorunn.rafnar@decode.is
 Tõnu Esko tonu.esko@ut.ee
 Stefan Walter s.walter@erasmusmc.nl
 Christian Gieger christian.gieger@helmholtz-muenchen.de
 Rajesh Rawal rajesh.rawal@helmholtz-muenchen.de
 Massimo Mangino massimo.mangino@kcl.ac.uk
 Inga Prokopenko inga@well.ox.ac.uk
 Reedik Mägi reedik@well.ox.ac.uk
 Kaisu Keskitalo kaisu.keskitalo@helsinki.fi
 Iris H. Gudjonsdottir irish@decode.is
 Solveig Gretarsdottir solveig@decode.is
 Hreinn Stefansson hreinn@decode.is
 John R. Thompson trj@le.ac.uk
 Yurii S. Aulchenko i.aoultchenko@erasmusmc.nl
 Mari Nelis marinel@ut.ee

Katja K. Aben k.aben@iko.nl
 Martin den Heijer m.denheijer@endo.umcn.nl
 Asger Dirksen adi@dadlnet.dk
 Haseem Ashraf haseem@dadlnet.dk
 Nicole Soranzo ns6@sanger.ac.uk
 Ana M Valdes ana.valdes@kcl.ac.uk
 Claire Steves claire.steves@kcl.ac.uk
 Andre G Uitterlinden a.g.utterlinden@erasmusmc.nl
 Albert Hofman a.hofman@erasmusmc.nl
 Anke Tönjes anke.toenjes@medizin.uni-leipzig.de
 Peter Kovacs peter.kovacs@medizin.uni-leipzig.de
 Jouke Jan Hottenga jj.hottenga@psy.vu.nl
 Gonneke Willemssen ahm.willemsen@psy.vu.nl
 Nicole Vogelzangs n.vogelzangs@ggzingeest.nl
 Angela Döring doering@helmholtz-muenchen.de
 Norbert Dahmen ndahmen@uni-mainz.de
 Barbara Nitz barbara.nitz@helmholtz-muenchen.de
 Michele L. Pergadia michelep@psychiatry.wustl.edu
 Berta Saez saez2001@yahoo.fr
 Veronica De Diego verodediego@hotmail.com
 Victoria Lezcano vlezcanobiosca@hotmail.com
 Maria de Garcia-Prats lolagarciaaprats@hotmail.com
 Samuli Ripatti samuli.ripatti@thl.fi
 Markus Perola markus.perola@thl.fi
 Johannes Kettunen jk5@sanger.ac.uk
 Anna-Liisa Hartikainen anna-liisa.hartikainen@oulu.fi
 Anneli Pouta anneli.pouta@ktl.fi
 Jaana Laitinen jaana.laitinen@ttl.fi
 Matti Isohanni matti.isohanni@oulu.fi
 Shen Huey-Yi huei-yi.shen@ktl.fi
 Maxine Allen maxine.allen@dr1.ox.ac.uk
 Maria Krestyaninova mariak@ebi.ac.uk
 Alistair S Hall cvsash@leeds.ac.uk
 Gregory T. Jones greg.jones@otago.ac.nz
 Andre M. van Rij andre.vanrij@stonebow.otago.ac.nz
 Thomas Mueller Thomas.Mueller@bs-lab.at
 Benjamin Dieplinger Benjamin.Dieplinger@bs-lab.at
 Meinhard Haltmayer Meinhard.Haltmayer@bs-lab.at
 Steinn Jonsson steinnj@landspitali.is
 Stefan E. Matthiasson sem@laekning.is
 Hogni Oskarson hogni@humus.is
 Thorarinn Tyrfingsson thorarinn@saa.is
 Lambertus A. Kiemeny B.Kiemeny@ebh.umcn.nl
 Jose I. Mayordomo josemayordomo@hotmail.com
 Jes S Lindholt Jes.S.Lindholt@Viborg.RM.dk
 Jesper Holst Pedersen jesper.holst.pedersen@rh.regionh.dk
 Wilbur A. Franklin wilbur.franklin@ucdenver.edu
 Holly Wolf Holly.Wolf@ucdenver.edu
 Grant W. Montgomery grantM@qimr.edu.au
 Andrew C. Heath aheath@wustl.edu
 Nicholas G. Martin nickM@qimr.edu.au
 Pamela A.F. Madden maddenp@psychiatry.wustl.edu
 Ina Giegling ina.giegling@med.uni-muenchen.de
 Dan Rujescu dan.rujesu@psy.med.uni-muenchen.de
 Marjo-Riitta Järvelin m.jarvelin@imperial.ac.uk
 Veikko Salomaa veikko.salomaa@thl.fi
 Michael Stumvoll michael.stumvoll@medizin.uni-leipzig.de
 Tim D Spector tim.spector@kcl.ac.uk
 H-Erich Wichmann wichmann@helmholtz-muenchen.de
 Andres Metspalu andres@ebc.ee
 Nilesh J Samani njs@le.ac.uk
 Brenda W. Penninx b.penninx@vumc.nl
 Ben A. Oostra b.oostra@erasmusmc.nl
 Dorret I. Boomsma di.boomsma@psy.vu.nl
 Henning Tiemeier h.tiemeier@erasmusmc.nl
 Cornelia M. van Duijn c.vanduijn@erasmusmc.nl
 Jaakko Kaprio jaakko.kaprio@helsinki.fi
 Jeffrey R. Gulcher jeffrey.gulcher@decode.is
 Mark I McCarthy mark.mccarthy@dr1.ox.ac.uk

Leena Peltonen leena@sanger.ac.uk
 Unnur Thorsteinsdottir unnur.thorsteinsdottir@decode.is
 Kari Stefansson kstefans@decode.is

COMPETING INTERESTS STATEMENT Authors whose affiliations are listed as deCODE genetics are employees of deCODE genetics, a biotechnology company.

URLs. ENGAGE Consortium (www.euengage.org).

Acknowledgments

We thank all the participants in the genetic studies whose contributions made this work possible. This work was supported in part by NIH (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 platform²⁸ that has greatly facilitated data exchange and annotation. Additional acknowledgements are listed in the Supplementary Note online.

References

1. WHO Report on the Global Tobacco Epidemic, 2008. 2008.
2. Bierut LJ, et al. Novel genes identified in a high-density genome wide association study for nicotine dependence. *Hum Mol Genet.* 2007; 16:24–35. [PubMed: 17158188]
3. Saccone SF, et al. Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum Mol Genet.* 2007; 16:36–49. [PubMed: 17135278]
4. Rose, R.; Broms, U.; Korhonen, T.; Dick, D.; Kaprio, J. Genetics of Smoking Behavior. In: Kim, Y., editor. *Handbook of Behavior Genetics*. Springer Science+Business Media; 2009.
5. Li MD, Cheng R, Ma JZ, Swan GE. A meta-analysis of estimated genetic and environmental effects on smoking behavior in male and female adult twins. *Addiction.* 2003; 98:23–31. [PubMed: 12492752]
6. Koopmans JR, Slutske WS, Heath AC, Neale MC, Boomsma DI. The genetics of smoking initiation and quantity smoked in Dutch adolescent and young adult twins. *Behav Genet.* 1999; 29:383–93. [PubMed: 10857244]
7. Thorgeirsson TE, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature.* 2008; 452:638–42. [PubMed: 18385739]
8. Berrettini W, et al. Alpha-5/alpha-3 nicotinic receptor subunit alleles increase risk for heavy smoking. *Mol Psychiatry.* 2008; 13:368–73. [PubMed: 18227835]
9. Amos CI, et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet.* 2008; 40:616–22. [PubMed: 18385676]
10. Hung RJ, et al. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature.* 2008; 452:633–7. [PubMed: 18385738]
11. Pillai SG, et al. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet.* 2009; 5:e1000421. [PubMed: 19300482]
12. Furberg H. TAG study. *Nat Genet.* 2010
13. Marchini J. OX-GSK Study. *Nat Genet.* 2010
14. Stevens VL, et al. Nicotinic receptor gene variants influence susceptibility to heavy smoking. *Cancer Epidemiol Biomarkers Prev.* 2008; 17:3517–25. [PubMed: 19029397]
15. Collins AC, Salminen O, Marks MJ, Whiteaker P, Grady SR. The road to discovery of neuronal nicotinic cholinergic receptor subtypes. *Handb Exp Pharmacol.* 2009; 85–112. [PubMed: 19184647]
16. Mineur YS, Picciotto MR. Genetics of nicotinic acetylcholine receptors: Relevance to nicotine addiction. *Biochem Pharmacol.* 2008; 75:323–33. [PubMed: 17632086]

17. West KA, et al. Rapid Akt activation by nicotine and a tobacco carcinogen modulates the phenotype of normal human airway epithelial cells. *J Clin Invest.* 2003; 111:81–90. [PubMed: 12511591]
18. Ray R, Tyndale RF, Lerman C. Nicotine dependence pharmacogenetics: role of genetic variation in nicotine-metabolizing enzymes. *J Neurogenet.* 2009; 23:252–61. [PubMed: 19169923]
19. Miksys S, Lerman C, Shields PG, Mash DC, Tyndale RF. Smoking, alcoholism and genetic polymorphisms alter CYP2B6 levels in human brain. *Neuropharmacology.* 2003; 45:122–32. [PubMed: 12814665]
20. Keskitalo K, et al. Association of serum cotinine level with a cluster of three nicotinic acetylcholine receptor genes (CHRNA3/CHRNA5/CHRNA4) on chromosome 15. *Hum Mol Genet.* 2009; 18:4007–12. [PubMed: 19628476]
21. Uhl GR, et al. Molecular genetics of successful smoking cessation: convergent genome-wide association study results. *Arch Gen Psychiatry.* 2008; 65:683–93. [PubMed: 18519826]
22. Uhl GR, et al. Molecular genetics of nicotine dependence and abstinence: whole genome association using 520,000 SNPs. *BMC Genet.* 2007; 8:10. [PubMed: 17407593]
23. Chanock SJ, Hunter DJ. Genomics: when the smoke clears. *Nature.* 2008; 452:537–8. [PubMed: 18385720]
24. Thorgeirsson TE, Stefansson K. Genetics of smoking behavior and its consequences: the role of nicotinic acetylcholine receptors. *Biol Psychiatry.* 2008; 64:919–21. [PubMed: 19000575]
25. Spitz MR, Amos CI, 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.* 2008; 100:1552–6. [PubMed: 18957677]
26. Lips EH, et al. Association between a 15q25 gene variant, smoking quantity and tobacco-related cancers among 17 000 individuals. *Int J Epidemiol.* 2009
27. Thorgeirsson TE, Stefansson K. Commentary: Gene-environment interactions and smoking-related cancers. *Int J Epidemiol.* 2010
28. Krestyaninova M, et al. A System for Information Management in BioMedical Studies--SIMBioMS. *Bioinformatics.* 2009; 25:2768–9. [PubMed: 19633095]
29. A haplotype map of the human genome. *Nature.* 2005; 437:1299–320. [PubMed: 16255080]
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.* 2007; 39:906–13. [PubMed: 17572673]
31. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. *Annu Rev Genomics Hum Genet.* 2009; 10:387–406. [PubMed: 19715440]
32. Kutayin IV, et al. A novel endonuclease IV post-PCR genotyping system. *Nucleic Acids Res.* 2006; 34:e128. [PubMed: 17012270]
33. Rice, JA. *Mathematical Statistics and Data Analysis.* Wadsworth Inc.; Belmont, CA: 1995.
34. Higgins J, Thompson S. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine.* 2002; 21:1539–58.
35. Devlin B, Roeder K. Genomic control for association studies. *Biometrics.* 1999; 55:997–1004. [PubMed: 11315092]

Figure 1.

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), and 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).

Table 1

Association of markers in 4 chromosomal regions with CPD. Results are given for the ENGAGE analysis (ENGAGE), the *in-silico* replication obtained by combining results from TAG and OX-GSK (*in-silico*), 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 and position, the estimated allelic effects on CPD and their standard errors in CPD (Effect and SE), the P value for the test of association (P), the P value for the test for heterogeneity in effect size, and an estimate of the proportion of total variation in study estimates that is due to heterogeneity (I^2)

SNP	Allele		Freq	Chr	Position	ENGAGE (meta-analysis) (N=31,266)		TAG and OX-GSK (In silico replication) (N=45,691)		ISL-AUS-DEN-GER- SPA (direct genotyping) (N=9,040)		Combined (N=85,997)		I^2
	Effect	Other				Effect±SE	P	Effect±SE	P	Effect±SE	P	Effect±SE	P	
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			0.80±0.05	2.4·10 ⁻⁶⁹	0.035
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	0.29±0.05	1.4·10 ⁻⁸	0.24
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			0.31±0.05	1.3·10 ⁻⁸	0.24
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	0.22±0.04	2.1·10 ⁻⁷	0.018
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			0.26±0.04	5.4·10 ⁻⁹	0.12
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	0.24±0.04	2.4·10 ⁻⁹	0.45
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			0.68±0.18	1.1·10 ⁻⁴	0.50
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	0.39±0.06	2.2·10 ⁻¹²	0.51
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	0.20±0.04	5.5·10 ⁻⁶	0.12

Table 2

Association of SNPs in 4 chromosomal regions with lung cancer in four populations. 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 (OR and 95%CI), the P value for the test of association (P). The results for Peripheral Arterial Disease are shown in Supplementary Table 4.

Population	N		Freq		OR (95% CI)	P
	case	control	case	control		
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	2,506	1,914	0.763	0.778	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	2,513	1,917	0.344	0.365	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	2,518	1,921	0.559	0.580	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

Population	N		Freq		OR (95% CI)	P
	case	control	case	control		
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
<i>rs7260329-G, chromosome 19q13</i>						
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	2,481	1,899	0.662	0.670	1.02 (0.95, 1.10)	0.61
Combined	4,477	42,126	-	-	1.06 (1.00, 1.12)	0.041

⁴For 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 ($r^2=1$).