LETTERS

A variant associated with nicotine dependence, lung cancer and peripheral arterial disease

Thorgeir E. Thorgeirsson[1]*, Frank Geller[1]*, Patrick Sulem[1]*, Thorunn Rafnar[1]*, Anna Wiste[1,2], Kristinn P. Magnusson[1], Andrei Manolescu[1], Gudmar Thorleifsson[1], Hreinn Stefansson[1], Andres Ingason[1], Simon N. Stacey[1], Jon T. Bergthorsson[1], Steinunn Thorlacius[1], Julius Gudmundsson[1], Thorlakur Jonsson[1], Margret Jakobsdottir[1], Jona Saemundsdottir[1], Olof Olafsdottir[1], Larus J. Gudmundsson[1], Gyda Bjornsdottir[1], Kristleifur Kristjansson[1], Halla Skuladottir[3], Helgi J. Isaksson[4], Tomas Gudbjartsson[5], Gregory T. Jones[8], Thomas Mueller[9], Anders Gottsäter[10], Andrea Flex[11], Katja K. H. Aben[12,13], Femmie de Vegt[12], Peter F. A. Mulders[14], Dolores Isla[15], Maria J. Vidal[15], Laura Asin[16], Berta Saez[17], Laura Murillo[18], Thorsteinn Blondal[19], Halldor Kolbeinsson[6], Jon G. Stefansson[6], Ingunn Hansdottir[20], Valgerdur Runarsdottir[20], Roberto Pola[11,21], Bengt Lindblad[10], Andre M. van Rij[8], Benjamin Dieplinger[9], Meinhard Haltmayer[9], Jose I. Mayordomo[15,16,17], Lambertus A. Kiemeney[12,13,14], Stefan E. Matthiasson[22], Hogni Oskarsson[23], Thorarinn Tyrfingsson[20], Daniel F. Gudbjartsson[1], Jeffrey R. Gulcher[1], Steinn Jonsson[7], Unnur Thorsteinsdottir[1,22], Augustine Kong[1] & Kari Stefansson[1,22]

Smoking is a leading cause of preventable death, causing about 5 million premature deaths worldwide each year[1,2]. Evidence for genetic influence on smoking behaviour and nicotine dependence (ND)[3–8] has prompted a search for susceptibility genes. Furthermore, assessing the impact of sequence variants on smoking-related diseases is important to public health[9,10]. Smoking is the major risk factor for lung cancer (LC)[11–14] and is one of the main risk factors for peripheral arterial disease (PAD)[15–17]. Here we identify a common variant in the nicotinic acetylcholine receptor gene cluster on chromosome 15q24 with an effect on smoking quantity, ND and the risk of two smoking-related diseases in populations of European descent. The variant has an effect on the number of cigarettes smoked per day in our sample of smokers. The same variant was associated with ND in a previous genome-wide association study that used low-quantity smokers as controls[18,19], and with a similar approach we observe a highly significant association with ND. A comparison of cases of LC and PAD with population controls each showed that the variant confers risk of LC and PAD. The findings provide a case study of a gene–environment interaction[20], highlighting the role of nicotine addiction in the pathology of other serious diseases.

To perform a genome-wide association (GWA) study of smoking quantity (SQ), we used questionnaire data limited to basic questions on smoking behaviour that were available for a large number of lifetime smokers. The GWA scan comprises 10,995 Icelandic smokers who had been assayed with Infinium HumanHap300 SNP chips from Illumina. A set of 306,207 single-nucleotide polymorphisms (SNPs) fulfilling our quality criteria was tested. We focused on cigarette smoking, with SQ reported as cigarettes per day. All SQ data were clustered into categories (see Supplementary Information) and we refer to them as 'SQ levels'. The SQ levels were 0 (1–10 cigarettes per day), 1 (11–20), 2 (21–30) and 3 (31 or more). Each increment represents an increase in SQ of 10 cigarettes per day. Allele T of the SNP rs1051730 was most strongly associated with SQ, and the association was highly significant ($P = 5 \times 10^{-16}$). The SNP is within the *CHRNA3* gene in a linkage disequilibrium block also containing two other genes, *CHRNA5* and *CHRNB4*, that encode nicotinic acetylcholine receptors (ref. 18). Six other SNPs on chromosome 15q24 passed the threshold of genome-wide significance ($P < 2 \times 10^{-7}$), but they are all correlated with rs1051730 ($r^2 = 0.14$–0.93). After correction for rs1051730, none of these six SNPs showed a P value below 10^{-3} (see Supplementary Table 1). A quantile–quantile plot for the GWA scan (see Supplementary Fig. 1a) shows the observed excess of signals, whereas a quantile–quantile plot after removing 182 markers located within 1 megabase of rs1051730 is consistent with noise (see Supplementary Fig. 1b), illustrating that all of the strongest signals standing out in the first plot are located on chromosome 15q24. An additional 2,950 smokers from Iceland were genotyped for rs1051730, giving a total of 13,945 smokers (Table 1) with a mean variant frequency of 34.7%, which is not significantly different from the frequency of 34.4% observed in 4,203 individuals who were genotyped and who reported never having smoked (odds ratio (OR) 1.01, 95% confidence interval (CI) 0.96–1.07, $P = 0.60$). Indeed, the frequency of the variant in the 3,627 low-quantity smokers (10 or fewer cigarettes per day) is significantly less than that in those who do not smoke (OR 0.83, 95% CI 0.78–0.90, $P = 4.5 \times 10^{-7}$). The increase in

[1]deCODE Genetics, 101 Reykjavik, Iceland. [2]Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, GA 30322, USA. [3]Department of Medical Oncology, [4]Department of Pathology, [5]Department of Cardiothoracic Surgery, [6]Department of General Adult Psychiatry, and [7]Department of Medicine, Landspitali University Hospital, 101 Reykjavik, Iceland. [8]Department of Surgery, Dunedin School of Medicine, University of Otago, Dunedin 9054, New Zealand. [9]Department of Laboratory Medicine, Konventhospital Barmherzige Brueder, 4020 Linz, Austria. [10]Center for Vascular Diseases, University Hospital, MAS, 205 02 Malmö, Sweden. [11]Laboratory of Vascular Biology and Genetics, A. Gemelli University Hospital, Universita Cattolica del Sacro Cuore School of Medicine, 00168 Rome, Italy. [12]Department of Epidemiology and Biostatistics, Radboud University Nijmegen Medical Centre, 6500 HB Nijmegen, The Netherlands. [13]Comprehensive Cancer Center East, 6501 BG Nijmegen, The Netherlands. [14]Department of Urology, Radboud University Nijmegen Medical Centre, 6525 GA Nijmegen, The Netherlands. [15]Division of Medical Oncology, University Hospital, 5009 Zaragoza, Spain. [16]Nanotechnology Institute of Aragon, 5009 Zaragoza, Spain. [17]Health Science Institute, 5009 Zaragoza, Spain. [18]Division of Medical Oncology, Hospital Reina Sofia, 31500 Tudela, Spain. [19]Department of Lung Disease, Primary Health Care Centre, 101 Reykjavik, Iceland. [20]Vogur SAA Addiction Treatment Center, 112 Reykjavik, Iceland. [21]OASI Institute for Research and Care (IRCCS), 94018 Troina (EN), Italy. [22]School of Medicine, University of Iceland, 101 Reykjavik, Iceland. [23]Therapeia, 101 Reykjavik, Iceland.
*These authors contributed equally to this work.

Table 1 | Genotype status and SQ level of 13,945 Icelandic smokers

Parameter	Genotype of rs1051730			Total *n* (frequency)	Frequency of T allele
	GG	GT	TT		
Cigarettes per day (SQ level)					
1 to 10 (0)	1,743	1,558	326	3,627 (0.260)	0.305
11 to 20 (1)	2,727	2,865	810	6,402 (0.459)	0.350
21 to 30 (2)	1,145	1,416	427	2,988 (0.214)	0.380
31 and more (3)	341	448	139	928 (0.067)	0.391
All levels (frequency)	5,956 (0.427)	6,287 (0.451)	1,702 (0.122)	13,945 (1.000)	0.347
Mean SQ level (mean ± s.d.)	1.01 ± 0.85	1.12 ± 0.86	1.22 ± 0.85	1.09 ± 0.86	

frequency between levels varies, and the largest increase (4.5%) is observed between the lowest levels (0 and 1), whereas the increase between the highest levels (2 and 3) is only 1.1%. In the context of a case-control LC study, an additional 523 smokers from Spain and 1,375 smokers from The Netherlands were genotyped. We performed multiple regression analyses of SQ data from the three countries, with adjustment for sex and year of birth (Table 2). Results from Spain and The Netherlands combined gave an estimated increase of 0.074 SQ units ($P = 0.012$) for each copy of the variant, which is not significantly different ($P = 0.45$) from the estimate of 0.098 SQ units ($P = 10^{-18}$) based on the Icelandic data. Combining all results, each copy of the variant was estimated to increase SQ level by 0.095 units ($P = 6 \times 10^{-20}$), which corresponds to about one cigarette per day. A recent GWA study reported an association between SQ and SNP rs6495308 ($P = 6.9 \times 10^{-5}$, correlation between this SNP and rs1051730, $r^2 = 0.18$ in the HapMap project) for about 7,500 individuals from two study groups, as well as association between SQ and rs1317286 ($P = 2.6 \times 10^{-6}$, $r^2 = 0.90$ to rs1051730 in the HapMap project) in a candidate gene study based on 1,740 heavy smokers (more than 25 cigarettes per day) and 6,200 low-quantity smokers (fewer than 5 cigarettes per day)[21].

Sex and year of birth are also strongly associated with SQ (Table 2). However, neither the interaction between variant and sex nor that between variant and year of birth is significant, indicating that the effect of the variant is similar for both sexes and is robust to population-wide changes in smoking habits over time. The phenotypic variance explained by the variant was highest in Iceland, amounting to 0.7%.

Association of the same variant with ND was previously reported in a candidate gene study involving 3,713 SNPs[18]. We assessed the association with ND, defined as a score of 4 or higher on the Fagerstrom Test for Nicotine Dependence (FTND)[22] or endorsement of at least three of the seven Diagnostic and Statistical Manual of Mental Disorders 4th edition (DSM-IV) criteria (see Supplementary Information). The variant is associated with ND in Iceland in a subset of 2,394 smokers from the SQ study tested against both 28,455 population controls (OR 1.17, 95% CI 1.10–1.25, $P = 3.3 \times 10^{-6}$) and 3,506 low-quantity smokers (OR 1.40, 95% CI 1.29–1.52, $P = 7 \times 10^{-15}$). The latter OR of 1.40 is comparable to the results of the candidate gene study[18]. They used non-ND smokers as controls (that is, individuals who had smoked but had an FTND score of 0), and reported association with rs1051730 (OR 1.3, $P = 10^{-3}$). rs1051730 is in strong linkage disequilibrium with rs16969968 ($r^2 = 0.90$ in the HapMap project), which was highlighted in the previous study with a similar result (OR 1.3, $P = 6 \times 10^{-4}$)[18]. Dependence on nicotine drives repeated self-administration of nicotine[23–25] and high SQ is a strong sign of ND and one of the criteria for a diagnosis of ND (SQ is included in the FTND scale). This, together with the fact that the ND subjects are part of the SQ study, means that the associations with ND and SQ cannot be considered independent results. Both the FTND and the DSM-IV scales include many items that are not based on SQ, and their total scores are measures of ND severity. In our ND group, positive scores on most items in both scales show a trend towards higher frequency of the variant, as does the total score on both the FTND and DSM-IV scales. Thus, the frequency of the variant increases with addiction severity, and is 46.8% and 43.8% for the highest deciles of FTND and DSM-IV, respectively (see Supplementary Table 2a, b). ND is believed to be the main reason for continued smoking. To explore the frequency of the variant in the context of the ability to quit smoking, we investigated differences between 6,388 current and 6,687 past smokers from the SQ analysis by a logistic regression model adjusting for sex and year of birth. The variant was associated with current smoking with an OR of 1.07 (95% CI 1.01–1.13, $P = 0.015$) (see Supplementary Table 3), and the effect is similar when corrected for SQ (OR 1.06, 95% CI 1.00–1.12, $P = 0.036$), indicating that carriers of the variant are less likely to quit smoking.

Smoking is a major risk factor for many diseases, and we decided to study the effect of the variant on LC and PAD risk directly. The LC study was based on 1,024 cases and 32,244 controls from Iceland, Spain and The Netherlands (Table 3); the PAD study was based on 2,738 cases and 29,964 controls from five caucasian populations (Iceland, New Zealand, Austria, Sweden and Italy) (Table 3). The results for LC and PAD (Table 4) represent the overall effect on LC and PAD including indirect effects through SQ and ND. Significant association was observed with LC for both the Icelandic data (OR 1.27, $P = 4.1 \times 10^{-5}$) and the data for Spain and The Netherlands combined (OR 1.39, $P = 6.6 \times 10^{-5}$). These two estimates are not significantly different from each other ($P = 0.34$), and combining results from all three groups gave an OR of 1.31 (95% CI 1.19–1.44, $P = 1.5 \times 10^{-8}$). There is no significant difference in frequency of the variant between histological types of LC, which is not surprising given the small number of cases per group (see Supplementary Table 4). Association with PAD was found both in the Icelandic data (OR 1.18, $P = 5.3 \times 10^{-5}$) and in the data for the foreign populations combined (OR 1.23, $P = 5.9 \times 10^{-4}$). These two estimates are not significantly different from each other ($P = 0.57$), and combining results from all five groups gave an OR of 1.19 (95% CI 1.12–1.27, $P = 1.4 \times 10^{-7}$).

Table 2 | Multiple regression of SQ level as a function of rs1051730 genotype, sex and year of birth

Study group	*n*	Copies of T allele		Sex (male)		*P* for year of birth (categorical)	*P* for interactions: allele × sex, allele × age
		Estimate (95% CI)	*P*	Estimate (95% CI)	*P*		
Iceland	13,945	0.098 (0.076–0.120)	10^{-18}	0.411 (0.383–0.438)	$<10^{-16}$	$<10^{-16}$	0.53, 0.85
Spain	523	0.061 (−0.059–0.180)	0.32	0.504 (0.290–0.718)	$<10^{-5}$	0.006	0.80, 0.76
The Netherlands	1,375	0.078 (0.012–0.145)	0.021	0.326 (0.225–0.427)	$<10^{-9}$	$<10^{-4}$	0.68, 0.27
Foreign combined	1,898	0.074 (0.016–0.132)	0.012	NA	–	–	–
All combined	15,771	0.095 (0.075–0.115)	6×10^{-20}	NA	–	–	–

Multiple regression of SQ level on allele T, sex and year of birth, giving adjusted values for each explanatory variable adjusting for the others. For the tests of interaction, the interaction terms involving the variant were individually added to the initial model. NA, not available.

Table 3 | Demographics, age and phenotype breakdown

Study group	Cases			Controls		
	n	Male/female	Age (years)	*n*	Male/female	Age (years)
Smoking phenotypes						
Cigarettes per day available						
Iceland	13,945	6,134/7,811	58.7 ± 17.8			
Spain	523	354/169	54.0 ± 16.3			
The Netherlands	1,375	762/613	61.5 ± 10.2			
ND (Iceland)	2,394	800/1,594	48.1 ± 11.0	28,455	12,600/15,855	58.7 ± 21.8
Never smokers (Iceland)	4,203	1,273/2,930	55.4 ± 21.6			
LC						
Iceland	665	346/319	69.7 ± 11.1	28,752	12,174/16,578	56.8 ± 21.5
Spain	269	238/31	64.9 ± 11.6			
The Netherlands	90	71/19	68.5 ± 9.5			
PAD						
Iceland	1,503	926/577	74.2 ± 10.6	28,752	12,174/16,578	56.8 ± 21.5
New Zealand	441	251/189	70.6 ± 9.6	435	248/187	68.2 ± 6.4
Austria	457	322/135	68.4 ± 11.0	403	284/119	67.3 ± 10.7
Sweden	172	92/80	77.5 ± 9.9	140	64/76	67.9 ± 1.5
Italy	165	111/54	73.0 ± 9.3	234	162/72	72.6 ± 6.4

Ages are shown as means ± s.d.

Genotypic ORs for LC, PAD and ND did not deviate significantly from those obtained for the multiplicative model (see Supplementary Table 5), and no significant differences in the ORs between sexes were observed (see Supplementary Table 6).

According to our estimates for Icelandic LC patients, the correlation between SQ and LC is consistent with numbers reported in other studies[26,27]. Combining these estimates with our estimate of the association of the variant with SQ, the expected OR between the variant and LC is only about 1.05 in Iceland (see Supplementary Information), which is well below the direct OR estimate for LC of 1.27 (95% CI 1.13–1.43). A similar indirect estimate for PAD is 1.04, which again is substantially lower than the observed direct estimate of 1.18 (95% CI 1.09–1.27). It is not surprising that the ORs for LC and PAD cannot be explained by the effect of the variant on SQ alone, because the involvement of both SQ and the duration of smoking in LC and PAD was established in previous studies[15,28]. The SQ data for most individuals were derived from a single point in time and cannot be expected to cover all aspects of smoking behaviour affected by the variant and relevant to LC and PAD. An effect on other aspects of smoking behaviour, in particular smoking duration, is likely to account for the observed difference between the indirect and direct estimates of the LC and PAD risks. An alternative possibility is that the variant directly confers risk of LC and PAD, for example by increasing the vulnerability to tobacco smoke.

Thus, we have unequivocally demonstrated a correlation between, on the one hand, a sequence variant in the cluster of genes on chromosome 15 that encode nicotinic acetylcholine receptors and, on the other, SQ and ND. The variant does not influence smoking initiation; however, among smokers, carriers of the variant smoke more than non-carriers and have higher rates of ND. This variant was reported in a previous study of 1,050 ND cases and 879 controls who smoked and had an FTND score of 0 (refs 18, 19) and the authors concluded that the variant contributes to ND (ref. 18). This conclusion is put on firm ground by the highly significant OR of 1.40 ($P = 7 \times 10^{-15}$) for ND compared with low-quantity smokers (ten or fewer cigarettes per day). The direct measurement of the risk of LC and PAD revealed genome-wide significant associations with allelic ORs of 1.31 and 1.19, respectively. This demonstrates that a sequence variant associated with ND, a brain disorder, confers risk of lung and cardiovascular diseases through an effect on behaviour, which is an example of active gene–environment correlation[20] in the pathogenesis of disease. A calculation of the population attributable risk for the variant gives 18% for LC and 10% for PAD. Although these population attributable risks are at best approximate figures given the complex interplay between the variant, smoking, and smoking-related diseases, it is likely that the variant accounts for a substantial fraction of PAD and LC cases and the associated morbidity and mortality.

The results of the study described here show that it is important to keep in mind, while attempting to shed light on the role of nature versus nurture in the pathogenesis of common or complex disease, that variants in the sequence of our genome influence not only how we respond to our environment but also our tendency to seek or avoid environment. The line between nature and nurture is therefore sometimes conspicuously absent.

METHODS SUMMARY

Subjects. Written informed consent was obtained from all subjects in the seven participating populations (Iceland, Spain, The Netherlands, Sweden, Italy, Austria and New Zealand; see Table 3). Inclusion in the study required the availability of genotypes from either GWA studies performed at deCODE

Table 4 | Association of rs1051730 allele T with LC and PAD

Study group	Controls		Cases		OR (95% CI)	*P*
	n	Frequency	*n*	Frequency		
LC						
Iceland	28,752	0.342	665	0.398	1.27 (1.13–1.43)	4.1×10^{-5}
Spain	1,474	0.390	269	0.483	1.46 (1.22–1.76)	5.4×10^{-5}
The Netherlands	2,018	0.314	90	0.350	1.18 (0.86–1.61)	0.31
Foreign combined	3,492	–	359	–	1.38 (1.18–1.62)	6.6×10^{-5}
All combined	32,244	–	1,024	–	1.31 (1.19–1.44)	1.5×10^{-8}
PAD						
Iceland	28,752	0.342	1,503	0.379	1.18 (1.09–1.27)	5.3×10^{-5}
New Zealand	435	0.274	441	0.337	1.35 (1.10–1.65)	0.0041
Austria	403	0.352	457	0.395	1.20 (0.99–1.46)	0.068
Sweden	140	0.304	172	0.331	1.14 (0.81–1.60)	0.46
Italy	234	0.378	165	0.412	1.15 (0.86–1.54)	0.33
Foreign combined	1,212	–	1,235	–	1.23 (1.09–1.39)	5.9×10^{-4}
All combined	29,964	–	2,738	–	1.19 (1.12–1.27)	1.4×10^{-7}

genetics or follow-up genotyping of rs1051730 in additional subjects. All subjects were of European descent. For details on recruitment periods and diagnostic criteria used for the ND, LC and PAD groups, see Methods and Supplementary Information.

Association analysis. For the genome-wide study of SQ the significance threshold was set at 2×10^{-7}, which is about 0.05 divided by 306,207, the number of SNPs passing quality control. Regressing SQ level as a quantitative variable on the number of copies of the allele carried (0/1/2), a likelihood ratio χ^2 statistic was used for testing. Evaluation of statistical significance took the relatedness of the individuals into account by dividing the χ^2 statistic by a correction factor based either on the method of genomic control[29] or on a simulation procedure using the known genealogy that we had previously employed[30] (see Supplementary Information). The variant did not correlate with sex or year of birth in the controls; the association analyses were therefore not adjusted for these factors.

ORs were assumed to have log-normal distributions, and the confidence intervals are test-based.

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

Received 17 December 2007; accepted 25 February 2008.

1. Anonymous. Annual smoking-attributable mortality, years of potential life lost, and productivity losses–United States, 1997–2001. *MMWR Morb. Mortal. Wkly. Rep.* **54,** 625–628 (2005).
2. World Health Organization. Economics of tobacco control. ⟨www.who.int/gb/fctc/PDF/wg1/e1t2.pdf⟩.
3. Carmelli, D., Swan, G. E., Robinette, D. & Fabsitz, R. Genetic influence on smoking—a study of male twins. *N. Engl. J. Med.* **327,** 829–833 (1992).
4. 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).
5. Kendler, K. S. *et al.* A population-based twin study in women of smoking initiation and nicotine dependence. *Psychol. Med.* **29,** 299–308 (1999).
6. Heath, A. C. & Martin, N. G. Genetic models for the natural history of smoking: evidence for a genetic influence on smoking persistence. *Addict. Behav.* **18,** 19–34 (1993).
7. Madden, P. A. *et al.* The genetics of smoking persistence in men and women: a multicultural study. *Behav. Genet.* **29,** 423–431 (1999).
8. Vink, J. M., Willemsen, G., Beem, A. L. & Boomsma, D. I. The Fagerstrom Test for Nicotine Dependence in a Dutch sample of daily smokers and ex-smokers. *Addict. Behav.* **30,** 575–579 (2005).
9. Carlsten, C. & Burke, W. Potential for genetics to promote public health: genetics research on smoking suggests caution about expectations. *J. Am. Med. Assoc.* **296,** 2480–2482 (2006).
10. Bierut, L. J. *et al.* Genetic research and smoking behavior. *J. Am. Med. Assoc.* **297,** 809 (2007).
11. Doll, R. & Hill, A. B. A study of the aetiology of carcinoma of the lung. *BMJ* **2,** 1271–1286 (1952).
12. Doll, R. & Peto, R. Cigarette smoking and bronchial carcinoma: dose and time relationships among regular smokers and lifelong non-smokers. *J. Epidemiol. Community Health* **32,** 303–313 (1978).
13. Wynder, E. L. & Graham, E. A. Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma; a study of 684 proved cases. *J. Am. Med. Assoc.* **143,** 329–336 (1950).
14. Parkin, D. M., Bray, F., Ferlay, J. & Pisani, P. Global cancer statistics, 2002. *CA Cancer J. Clin.* **55,** 74–108 (2005).
15. Hirsch, A. T. *et al.* ACC/AHA 2005 Practice Guidelines for the management of patients with peripheral arterial disease (lower extremity, renal, mesenteric, and abdominal aortic): a collaborative report from the American Association for Vascular Surgery/Society for Vascular Surgery, Society for Cardiovascular Angiography and Interventions, Society for Vascular Medicine and Biology, Society of Interventional Radiology, and the ACC/AHA Task Force on Practice Guidelines (Writing Committee to Develop Guidelines for the Management of Patients With Peripheral Arterial Disease): endorsed by the American Association of Cardiovascular and Pulmonary Rehabilitation; National Heart, Lung, and Blood Institute; Society for Vascular Nursing; TransAtlantic Inter-Society Consensus; and Vascular Disease Foundation. *Circulation* **113,** e463–e654 (2006).
16. Powell, J. T. *et al.* Risk factors associated with the development of peripheral arterial disease in smokers: a case-control study. *Atherosclerosis* **129,** 41–48 (1997).
17. Price, J. F. *et al.* Relationship between smoking and cardiovascular risk factors in the development of peripheral arterial disease and coronary artery disease: Edinburgh Artery Study. *Eur. Heart J.* **20,** 344–353 (1999).
18. 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).
19. 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).
20. Plomin, R., DeFries, J. C. & Loehlin, J. C. Genotype–environment interaction and correlation in the analysis of human behavior. *Psychol. Bull.* **84,** 309–322 (1977).
21. Berrettini, W. *et al.* α-5/α-3 nicotinic receptor subunit alleles increase risk for heavy smoking. *Mol. Psychiatry* (in the press).
22. Heatherton, T. F., Kozlowski, L. T., Frecker, R. C. & Fagerstrom, K. O. The Fagerstrom Test for Nicotine Dependence: a revision of the Fagerstrom Tolerance Questionnaire. *Br. J. Addict.* **86,** 1119–1127 (1991).
23. Rose, J. E. Nicotine addiction and treatment. *Annu. Rev. Med.* **47,** 493–507 (1996).
24. Russell, M. A. The nicotine addiction trap: a 40-year sentence for four cigarettes. *Br. J. Addict.* **85,** 293–300 (1990).
25. Stolerman, I. P. & Jarvis, M. J. The scientific case that nicotine is addictive. *Psychopharmacology (Berl.)* **117,** 2–10 (1995).
26. Haiman, C. A. *et al.* Ethnic and racial differences in the smoking-related risk of lung cancer. *N. Engl. J. Med.* **354,** 333–342 (2006).
27. Stellman, S. D. *et al.* Lung cancer risk in white and black Americans. *Ann. Epidemiol.* **13,** 294–302 (2003).
28. Peto, R. *et al.* Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies. *Br. Med. J.* **321,** 323–329 (2000).
29. Devlin, B., Bacanu, S. A. & Roeder, K. Genomic control to the extreme. *Nature Genet.* **36,** 1129–1130 (2004).
30. Grant, S. F. *et al.* Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nature Genet.* **38,** 320–323 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the participants in the genetic studies whose contributions made this work possible, the staff at Noatun (deCODE's recruitment centre), and personnel at deCODE's core facilities. This work was supported in part by funds from the National Institute of Drug Abuse and the European Commission.

Author Contributions T.E.T., F.G., P.S., and T.R. contributed equally to this work. T.E.T., F.G., P.S, T.R., A.W., D.F.G., A.K. and K.S. wrote the first draft of the paper. Ha.S., H.J.I., T.G. and S.J. recruited and diagnosed the Icelandic lung cancer patients. S.E.M. recruited and diagnosed the Icelandic peripheral arterial disease patients. T.B., H.K., J.G.S., I.H, V.R., H.O., T.T. and S.J. recruited and diagnosed nicotine addiction subjects. K.K.H.A., F.d.V., P.F.A.M. and L.A.K. recruited and diagnosed the subjects from The Netherlands. D.I., M.J.V., L.A., B.S., L.M. and J.I.M. recruited and diagnosed the Spanish subjects. G.T.J. and A.M.v.R. recruited and diagnosed the subjects from New Zealand. T.M., B.P. and M.H. recruited and diagnosed subjects from Austria. A.G. and B.L. recruited and diagnosed subjects from Sweden. A.F. and R.P. recruited and diagnosed subjects from Italy. A.W., A.I., S.N.S., J.T.B., S.T., J.G., M.J., J.S., O.O. and S.N.S. performed genotyping and experimental work. L.J.G., G.B. and K.K. incorporated phenotypic data into a database and analysed it. T.E.T., F.G., P.S., T.R., A.W., K.P.M., A.M., G.T., D.F.G. and A.K. analysed the data. T.E.T., F.G., P.S., T.R., K.P.M., Hr.S., T.J., J.I.M., L.K., H.O., T.T., J.R.G., S.J., D.G., U.T., A.K. and K.S. planned, supervised and coordinated the work. All authors contributed to the final version of the paper.

Author Information The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to T.E.T. (thorgeir@decode.is) or K.S. (kari.stefansson@decode.is).

METHODS

Icelandic subjects. For all studies involving Icelandic subjects, the study protocols were approved by the National Bioethics Committee (NBC) and the Data Protection Authority (DPA) of Iceland. The DPA encrypted all personal identifiers associated with information or blood samples with the use of a third-party encryption system[31]. In all, the Icelandic study involves 10,995 subjects with information on SQ available in the GWA, an additional 2,950 subjects with information on SQ, and 4,203 never-smokers. In the studies of LC and PAD, 665 and 1,503 patients, respectively, and 28,752 population controls were used (for details see Table 3).

Smoking. All Icelandic subjects in the study of smoking-related phenotypes, including Icelandic population controls, were originally recruited for different genetic studies conducted over 11 years (1996–2007) at deCODE Genetics, and information on the number of cigarettes smoked per day was available from questionnaires. The information on cigarettes smoked per day was categorized into SQ levels and used as a quantitative variable. Detailed information on SQ was also available for the foreign LC populations (Supplementary Information), but not for the foreign PAD populations.

Nicotine dependence. For a subset of the Icelandic smokers, information on the criteria used to diagnose ND was available from ongoing studies of ND and anxiety/depression[32]. We excluded individuals with diagnoses of other substance dependence or abuse, giving a total of 2,394 ND subjects. A score of 4 or higher on the FTND[22], or endorsement of three or more DSM criteria, were used to assign affected status for ND. Additional information on the Icelandic smoking and ND study group is available in the Supplementary Information.

Lung cancer. Iceland: recruitment was initiated in the year 1998 with a nationwide list from the Icelandic Cancer Registry (ICR). About 1,265 LC patients were alive during the period of recruitment, and 665 participated in the project. Information in the ICR includes year and age at diagnosis, year of death, SNOMED (Systematized Nomenclature of Medicine) code and ICD-10 (International Statistical Classification of Diseases and Related Health Problems, 10th revision) classification. Histological and cytological verification was available for 647 cases; the remaining 18 cases were diagnosed clinically.

The Netherlands: the 90 patients and 2,018 controls were identified retrospectively through three different ongoing studies on genetic risk factors of disease. All three study protocols were approved by the Institutional Review Board of the Radboud University Nijmegen Medical Centre.

Spain: patients were recruited from the Oncology Department of Zaragoza Hospital, from June 2006 to June 2007, and of 330 patients who were invited to participate, 292 enrolled (88%). Clinical information including age at onset and histology were collected from medical records. The 1,474 control individuals were approached at Zaragoza University Hospital. Study protocols were approved by the Institutional Review Board of Zaragoza University Hospital.

PAD. Iceland: patients have been recruited over the past nine years, as part of a genetic study at deCODE, from a registry of individuals diagnosed with PAD at the major hospital in Reykjavik, the Landspitali University Hospital, during the years 1983–2006. Diagnosis was confirmed by vascular imaging or segmental pressure measurements.

Austria: patients and controls were recruited through the Linz Peripheral Arterial Disease (LIPAD) study during 2000 to 2002, at the Department of Surgery, St John of God Hospital. Of the patients admitted for evaluation of suspected or definite PAD, all patients with chronic atherosclerotic occlusive disease of the lower extremities associated with typical symptoms—such as claudication or leg pain on exertion, rest pain, or minor or major tissue loss—were included on the basis of the final clinical diagnosis established by attending vascular surgeons. The diagnosis was verified by interview, physical examination, noninvasive techniques, and angiography[33]. All control subjects were patients at the same hospital and fulfilled the following criteria: no clinical indication of PAD by history and physical examination, and systolic brachial blood pressure equal to or less than the blood pressure in each of the right and left anterior tibial and posterior tibial arteries (that is, ankle brachial index ≥ 1.0)[33]. Smoking status was assessed as described in ref. 34.

Sweden: patients and controls were recruited at the Department of Vascular Diseases at Malmö University Hospital, a single referral centre for all patients with critical limb ischaemia in the three southernmost health-care districts in Sweden (723,750 inhabitants in 2001). The diagnosis of critical limb ischaemia was made in accordance with Trans-Atlantic Inter-Society Consensus scientific criteria[35] of ulceration, gangrene, or rest pain caused by PAD proved by ankle pressure (less than 50 to 70 mmHg), reduced toe pressure (less than 30 to 50 mmHg) or reduced transcutaneous oxygen tension. Diagnosis was confirmed by an experienced vascular surgery consultant. The control group consisted of healthy individuals without symptomatic PAD included in a health screening programme for a preventive medicine project[36].

Italy: patients and controls were recruited from subjects admitted to the Department of Medicine of the A. Gemelli University hospital of Rome, from 2000 to 2001. Inclusion criteria for the PAD group were Caucasian origin and presence of PAD, diagnosed in accordance with established criteria[37]. All patients had an ankle/arm pressure index lower than 0.8 and were at Fontaine's stage II, with intermittent claudication and no rest pain or trophic lesions. Inclusion criteria for the control group were caucasian origin, absence of PAD and CAD and no relationship to cases. Additional exclusion criteria from the study were tumours, chronic inflammatory diseases, and autoimmune diseases[38].

New Zealand: patients were recruited from the Otago–Southland region, and PAD was confirmed by an ankle brachial index of less than 0.7, pulse volume recordings and angiography/ultrasound imaging. The control group consisted of elderly individuals with no history of vascular disease from the same geographical region. Controls were asymptomatic for PAD and had ankle brachial indexes of more than 1. An abdominal ultrasound scan excluded concurrent abdominal aortic aneurysm from both the PAD and control groups, and Anglo-European ancestry was required for inclusion.

Genotyping. All 10,995 samples in the GWA study of SQ were genotyped with genotyping systems and specialized software (Human Hap300 and Human Hap300-duo+ Bead Arrays; Illumina)[39]. rs1051730 was genotyped with a Centaurus assay (Nanogen) for 8,566 Icelandic samples and all samples in the foreign study groups. Information on the genotyping and quality control is given in the Supplementary Information.

31. Gulcher, J. R., Kristjansson, K., Gudbjartsson, H. & Stefansson, K. Protection of privacy by third-party encryption in genetic research in Iceland. *Eur. J. Hum. Genet.* **8**, 739–742 (2000).
32. Thorgeirsson, T. E. *et al.* Anxiety with panic disorder linked to chromosome 9q in Iceland. *Am. J. Hum. Genet.* **72**, 1221–1230 (2003).
33. Mueller, T. *et al.* Factor V Leiden, prothrombin G20210A, and methylenetetrahydrofolate reductase C677T mutations are not associated with chronic limb ischemia: the Linz Peripheral Arterial Disease (LIPAD) study. *J. Vasc. Surg.* **41**, 808–815 (2005).
34. Rutherford, R. B. *et al.* Recommended standards for reports dealing with lower extremity ischemia: revised version. *J. Vasc. Surg.* **26**, 517–538 (1997).
35. Dormandy, J. A. & Rutherford, R. B. Management of peripheral arterial disease (PAD). TASC Working Group. TransAtlantic Inter-Society Consensus (TASC). *J. Vasc. Surg.* **31**, S1–S296 (2000).
36. Barani, J., Nilsson, J. A., Mattiasson, I., Lindblad, B. & Gottsater, A. Inflammatory mediators are associated with 1-year mortality in critical limb ischemia. *J. Vasc. Surg.* **42**, 75–80 (2005).
37. Anonymous.. Suggested standards for reports dealing with lower extremity ischemia. Prepared by the Ad Hoc Committee on Reporting Standards, Society for Vascular Surgery/North American Chapter, International Society for Cardiovascular Surgery. *J. Vasc. Surg.* **4**, 80–94 (1986).
38. Flex, A. *et al.* The −174 G/C polymorphism of the interleukin-6 gene promoter is associated with peripheral artery occlusive disease. *Eur. J. Vasc. Endovasc. Surg.* **24**, 264–268 (2002).
39. Barrett, J. C. & Cardon, L. R. Evaluating coverage of genome-wide association studies. *Nature Genet.* **38**, 659–662 (2006).