A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25

Rayjean J. Hung, James D. McKay, Valerie Gaborieau, Paolo Boffetta, Mia Hashibe, David Zaridze, Anush Mukeria, Neonilia Szeszenia-Dabrowska, Jolanta Lissowska, Peter Rudnai, Eleonora Fabianova, Dana Mates, Vladimir Bencko, Lenka Foretova, Vladimir Janout, Chu Chen, Gary Goodman, John K. Field, Triantafillos Liloglou, George Xinarianos, Adrian Cassidy, John McLaughlin, Geoffrey Liu, Steven Narod, Hans E. Krokan, Frank Skorpen, Maiken Bratt Elvestad, Kristian Hveem, Lars Vatten, Jakob Linseisen, Françoise Clavel-Chapelon, Paolo Vineis, H. Bas Bueno-de-Mesquita, Eiliv Lund, Carmen Martinez, Sheila Bingham, Torgny Rasmuson, Pierre Hainaut, Elio Riboli, Wolfgang Ahrens, Simone Benhamou, Pagona Lagiou, Dimitrios Trichopoulos, Ivana Holcátová, Franco Merletti, Kristina Kjaerheim, Antonio Agudo, Gary Macfarlane, Renato Talamini, Lorenzo Simonato, Ray Lowry, David I. Conway, Ariana Znaor, Claire Healy, Diana Zelenika, Anne Boland, Marc Delepine, Mario Foglio, Doris Lechner, Fumihiko Matsuda, Helene Blanche, Ivo Gut, Simon Heath, Mark Lathrop, Paul Brennan

Angaben zur Veröffentlichung / Publication details:

Hung, Rayjean J., James D. McKay, Valerie Gaborieau, Paolo Boffetta, Mia Hashibe, David Zaridze, Anush Mukeria, et al. 2008. "A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25." *Nature* 452 (7187): 633–37. https://doi.org/10.1038/nature06885.

A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25

Rayjean J. Hung[1,2]*, James D. McKay[1]*, Valerie Gaborieau[1], Paolo Boffetta[1], Mia Hashibe[1], David Zaridze[3], Anush Mukeria[3], Neonilia Szeszenia-Dabrowska[4], Jolanta Lissowska[5], Peter Rudnai[6], Eleonora Fabianova[7], Dana Mates[8], Vladimir Bencko[9], Lenka Foretova[10], Vladimir Janout[11], Chu Chen[12], Gary Goodman[12], John K. Field[13], Triantafillos Liloglou[13], George Xinarianos[13], Adrian Cassidy[13], John McLaughlin[14], Geoffrey Liu[15], Steven Narod[16], Hans E. Krokan[17], Frank Skorpen[17], Maiken Bratt Elvestad[17], Kristian Hveem[17], Lars Vatten[17], Jakob Linseisen[18], Françoise Clavel-Chapelon[19], Paolo Vineis[20,21], H. Bas Bueno-de-Mesquita[22], Eiliv Lund[23], Carmen Martinez[24], Sheila Bingham[25], Torgny Rasmuson[26], Pierre Hainaut[1], Elio Riboli[20], Wolfgang Ahrens[27], Simone Benhamou[28,29], Pagona Lagiou[30], Dimitrios Trichopoulos[30], Ivana Holcátová[31], Franco Merletti[32], Kristina Kjaerheim[33], Antonio Agudo[34], Gary Macfarlane[35], Renato Talamini[36], Lorenzo Simonato[37], Ray Lowry[38], David I. Conway[39], Ariana Znaor[40], Claire Healy[41], Diana Zelenika[42], Anne Boland[42], Marc Delepine[42], Mario Foglio[42], Doris Lechner[42], Fumihiko Matsuda[42], Helene Blanche[43], Ivo Gut[42], Simon Heath[43], Mark Lathrop[42,43] & Paul Brennan[1]

Lung cancer is the most common cause of cancer death worldwide, with over one million cases annually[1]. To identify genetic factors that modify disease risk, we conducted a genome-wide association study by analysing 317,139 single-nucleotide polymorphisms in 1,989 lung cancer cases and 2,625 controls from six central European countries. We identified a locus in chromosome region 15q25 that was strongly associated with lung cancer ($P = 9 \times 10^{-10}$). This locus was replicated in five separate lung cancer studies comprising an additional 2,513 lung cancer cases and 4,752 controls ($P = 5 \times 10^{-20}$ overall), and it was found to account for 14% (attributable risk) of lung cancer cases. Statistically similar risks were observed irrespective of smoking status or propensity to smoke tobacco. The association region contains several genes, including three that encode nicotinic acetylcholine receptor subunits (*CHRNA5*, *CHRNA3* and *CHRNB4*). Such subunits are expressed in neurons and other tissues, in particular alveolar epithelial cells, pulmonary neuroendocrine cells and lung cancer cell lines[2,3], and they bind to *N*′-nitrosonornicotine and potential lung carcinogens[4]. A non-synonymous variant of *CHRNA5* that induces an amino acid substitution (D398N) at a highly conserved site in the second intracellular loop of the protein is among the markers with the strongest disease associations. Our results provide compelling evidence of a locus at 15q25 predisposing to lung cancer, and reinforce interest in nicotinic acetylcholine receptors as potential disease candidates and chemopreventative targets[5].

Lung cancer is caused predominantly by tobacco smoking, with cessation of tobacco consumption being the primary method for prevention. The risk among those who quit smoking remains elevated (although less than those who continue to smoke), and former smokers make up an increasing proportion of lung cancer patients in countries where tobacco consumption has declined[6,7]. Treatment strategies are of limited efficacy, with an overall 5-year survival rate of about 15%[8]. Lung cancer has an important heritable component[9], and identifying genes that are involved may suggest chemoprevention targets or allow for identification of groups at high risk. Despite a large number of studies including both sporadic and multi-case families, success in identifying genes that cause lung cancer has been extremely limited.

The availability of tagging single-nucleotide polymorphism (SNP) panels across the whole genome allows for efficient and comprehensive analysis of common genomic variation to be conducted without a priori hypotheses based on gene function or disease pathways. They

[1]International Agency for Research on Cancer (IARC), Lyon 69008, France. [2]School of Public Health, University of California at Berkeley, Berkeley, California 94720, USA. [3]Institute of Carcinogenesis, Cancer Research Centre, Moscow 115478, Russia. [4]Department of Epidemiology, Institute of Occupational Medicine, Lodz 90950, Poland. [5]M. Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw 02781, Poland. [6]National Institute of Environmental Health, Budapest 1097, Hungary. [7]Specialized Institute of Hygiene and Epidemiology, Banska Bystrica 97556, Slovakia. [8]Institute of Public Health, Bucharest 050463, Romania. [9]Charles University in Prague, First Faculty of Medicine, Institute of Hygiene and Epidemiology, Prague 2 12800, Czech Republic. [10]Department of Cancer Epidemiology and Genetics, Masaryk Memorial Cancer Institute, Brno 65653, Czech Republic. [11]Palacky University, Olomouc 77515, Czech Republic. [12]Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. [13]Roy Castle Lung Cancer Research Programme, University of Liverpool Cancer Research Centre, Liverpool L3 9TA, UK. [14]Cancer Care Ontario, and the Samuel Lunenfeld Research Institute, Toronto M5G 2L7, Canada. [15]Princess Margaret Hospital, Ontario Cancer Institute, Toronto M5G 2M9, Canada. [16]Women's College Research Institute, Toronto M5G 1N8, Canada. [17]Norwegian University of Science and Technology, Trondheim 7489, Norway. [18]Division of Cancer Epidemiology, German Cancer Research Centre (DKFZ), Heidelberg 69120, Germany. [19]INSERM ERI20, Institut Gustave Roussy, Villejuif 94805, France. [20]Department of Epidemiology and Public Health, Imperial College, London W2 1PG, UK. [21]Institute for Scientific Interchange (ISI), Torino 10133, Italy. [22]Centre for Nutrition and Health, National Institute of Public Health and the Environment, Bilthoven 3710 BA, The Netherlands. [23]Institute of Community Medicine, University of Tromsö, Tromsö 9037, Norway. [24]Andalusian school of Public Health and Ciber Epidemiology y Salud Publica, Granada 18011, Spain. [25]MRC Centre for Nutrition and Cancer, University of Cambridge, Department of Public Health and Primary Care and MRC Dunn Human Nutrition Unit, Cambridge CB2 0XY, UK. [26]Department of Radiation Sciences, Oncology, Umea University, Umea 90187, Sweden. [27]Epidemiological Methods and Etiologic Research, Bremen Institute for Prevention Research and Social Medicine, Bremen 28359, Germany. [28]INSERM U794, Fondation Jean Dausset-CEPH, Paris 75010, France. [29]CNRS FRE2939, Institute Gustave Roussy, Villejuif 94805, France. [30]Department of Hygiene and Epidemiology, University of Athens School of Medicine, Athens 11527, Greece, and Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts 02115, USA. [31]Institute of Hygiene and Epidemiology, Prague 2 12800, Czech Republic. [32]University of Turin, Turin 10126, Italy. [33]Cancer registry of Norway, Oslo 0310, Norway. [34]Institut Català d'Oncologia, Barcelona 08907, Spain. [35]University of Aberdeen School of Medicine, Aberdeen AB25 2ZD, UK. [36]Aviano cancer center, Aviano 33081, Italy. [37]Department of Environmental Medicine and Public Health, University of Padua, Padua 35131, Italy. [38]University of Newcastle Dental School, Newcastle NE2 4BW, UK. [39]University of Glasgow Dental School, Glasgow G2 3JZ, UK. [40]Croatian National Cancer Registry, National Institute of Public Health, Zagreb 10000, Croatia. [41]Trinity College School of Dental Science, Dublin 2, Ireland. [42]Centre National de Genotypage, Institut Genomique, Commissariat à l'énergie Atomique, Evry 91000, France. [43]Fondation Jean Dausset-CEPH, Paris 75010, France.
*These authors contributed equally to this work.

require very large series of cases and controls to ensure adequate statistical power, and multiple subsequent studies to confirm the initial findings. We conducted a genome-wide association study of lung cancer using the Illumina Sentrix HumanHap300 BeadChip containing 317,139 SNPs and estimated to tag approximately 80% of common genomic variation[10]. We initially genotyped 1,989 cases and 2,625 controls from the International Agency for Research on Cancer (IARC) central Europe lung cancer study. This was conducted in six countries between 1998 and 2002 and each centre followed an identical protocol to recruit newly diagnosed cases of primary lung cancer, as well as a comparable group of population or hospital controls (Supplementary Methods). We excluded samples that failed one of several quality control criteria (Supplementary Methods) or because they showed evidence of admixture with Asian ethnicity (Supplementary Fig. 1); we also excluded 7,116 problematic SNPs. This resulted in a comparison of 310,023 SNPs between 1,926 cases and 2,522 controls.

We analysed each SNP individually by calculating *P*-values for trend in a logistic regression model and incorporating additional parameters including country, age and sex (Supplementary Methods). The distribution of the bottom 90% of *P*-values was similar to the expected distribution, and the genomic control parameter was 1.03, implying that there was no systematic increase in false-positive findings owing to population stratification or any other form of bias (Fig. 1a). However, there was a marked deviation between the observed and expected *P*-values among the top 10% (Fig. 1b). In particular, two SNPs on chromosome 15q25, rs1051730 and rs8034191, were strongly associated with disease ($P = 5 \times 10^{-9}$ and $P = 9 \times 10^{-10}$, respectively), exceeding the genome-wide significance level of $P = 5 \times 10^{-7}$ (Fig. 1c). Further analysis incorporating adjustment by principal components indicated that population stratification was unlikely to account for this observation (Supplementary Methods).

The odds ratio (OR) and 95% confidence interval (CI) for carrying one copy of the most significant marker (rs8034191), adjusted by age, sex and country, was 1.27 (1.11–1.44) and for carrying two copies of the allele was 1.80 (1.49–2.18); the allelic OR was 1.32 (1.21–1.45). When the data were analysed separately by country of origin, we found a significant association in all countries except Romania, which had the smallest sample numbers, although the trend in Romania was similar and the association was significant under a dominant model (data not shown). There was no evidence of heterogeneity by country of origin ($P = 0.58$). Further adjustment was undertaken for various tobacco-related variables including duration of smoking, pack years (average number of cigarette packs per day multiplied by years of smoking) and age at onset of smoking. Adjustment by duration of smoking provided the best-fitting model to account for tobacco use based on the Akaike's information criteria (Supplementary Methods), although the adjusted estimates with duration of smoking (allelic OR = 1.28 (1.16–1.42)) were similar to the estimates adjusted by age, sex and country only.

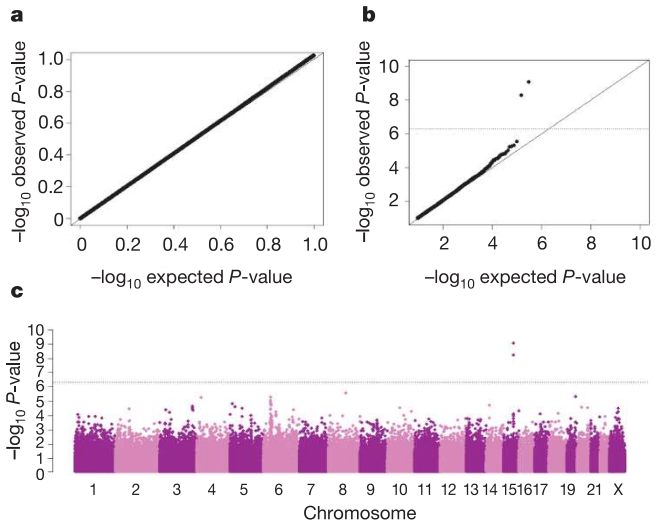

Figure 1 | Genome-wide association results in the central Europe study. **a–c**, Quantile–quantile plot for bottom 90% of *P*-values (**a**) and top 10% of *P*-values (**b**), as well as scatter plot (**c**) of *P*-values in −log scale from the trend test for 310,023 genotyped variants comparing 1,926 lung cancer cases and 2,522 controls.

We investigated further the association by genotyping 34 additional 15q25 markers that were selected as follows. First, we used an imputation method (see http://www.sph.umich.edu/csg/abecasis/MACH/index.html) to identify additional genetic variants from the Centre d'Etude du Polymorphism Humain Utah (CEU) HapMap data that are likely to have a strong disease association, but are not present in the HumanHap300 panel. We attempted genotyping of SNPs from the 15q25 region with an association *P*-value of the imputed data of $<10^{-6}$. Second, we included SNPs of *CHRNA5* and *CHRNA3* that had been included in a previous study of these genes in nicotine dependence[11]. Third, we attempted genotyping of all non-synonymous SNPs in dbSNP from the six genes within or near the association region. The results for all markers tested in the 15q25 region, including those in the HumanHap300 panel, are shown in Supplementary Table 1. Twenty-three of the additional genotyped markers showed evidence of association exceeding the genome-wide significance level of 5×10^{-7} (Fig. 2). These span more than 182 kilobases (kb) but are in strong linkage disequilibrium (pairwise $D' > 0.8$ and $r^2 > 0.6$) with two predominant haplotypes accounting for more than 85% of the haplotypes in patients and controls (Supplementary Table 2).

To confirm our findings we genotyped rs8034191 and rs16969968 (where rs16969968 is a second variant with a strong disease association) in five further independent studies of lung cancer: the European Prospective Investigation in Cancer and Nutrition (EPIC) cohort study (781 cases and 1,578 controls), the Beta-Carotene and Retinol Efficacy Trial (CARET) cohort study (764 cases and 1,515 controls), the Health Study of Nord-Trøndelag (HUNT) and Tromsø cohort studies (235 cases and 392 controls), the Liverpool lung cancer case-control study (403 cases and 814 controls), and the Toronto lung cancer case-control study (330 cases and 453 controls) (Supplementary Methods). We observed an increased risk for both heterozygous and homozygous variants of rs8034191 in all five replication samples (Table 1), with no evidence of any statistical heterogeneity between studies. After pooling across all six studies, the ORs (95% CI) were 1.21 (1.11–1.31) and 1.77 (1.58–2.00) for heterozygous and homozygous carriers, respectively, the allelic OR was 1.30 (1.23–1.37), and the *P*-value for trend was 5×10^{-20}. Further adjustment for duration of tobacco smoking did not alter the estimates: allelic OR = 1.30 (1.22–1.40). The genotype-specific model that estimated the OR for heterozygous and homozygous carriers separately was a significantly better fit than the model estimating the allelic OR ($P = 0.025$), suggesting a potential recessive effect.

The prevalence of the variant allele was 34%, resulting in 66% of the control participants carrying at least one copy, and the percentage of lung cancer explained by carrying at least one allele (that is, the population attributable risk) was 15% in the combined data set. We obtained a similar attributable risk in the central European study (16%) and in the replication studies (14%). The second variant with strong disease association (rs16969968) that was genotyped in the five replication studies gave very similar results, as expected from the strong linkage disequilibrium ($D' = 1.00$, $r^2 = 0.92$) among the disease-associated markers (allelic OR = 1.30 (1.23–1.38); $P = 1 \times 10^{-20}$).

The large number of patients in the combined data set allowed us to examine the association in different smoking categories and with respect to different histological subtypes (Supplementary Table 3 and

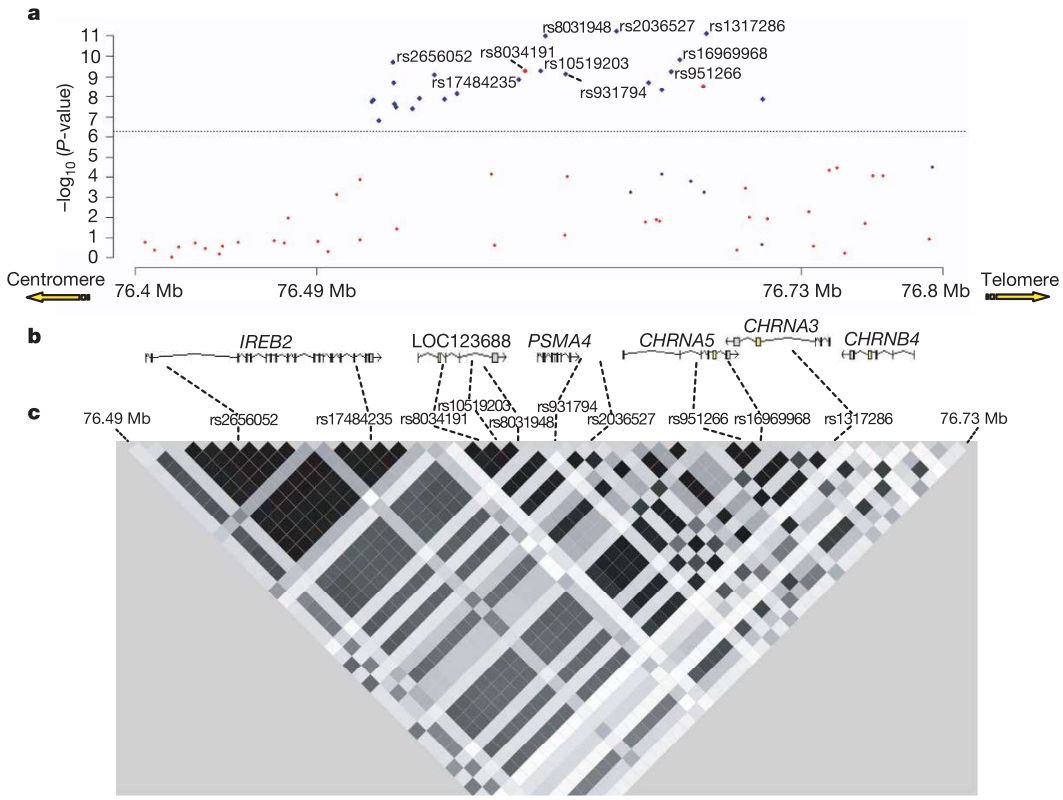

Figure 2 | Lung cancer area of interest across 15q25. **a**, *P*-values for SNPs genotyped in the 15q25 region (76.4–76.8 Mb). The dotted line indicates the genome-wide threshold of $P < 5 \times 10^{-7}$. Points labelled with rs numbers have a $P < 1 \times 10^{-9}$. Points in red are genotyped in the 317K Illumina panel; points in blue indicate additional genotyped SNPs (Taqman). **b**, Positions of the six known genes. **c**, Pairwise r^2 estimates for 46 common SNPs from 76.49 Mb to 76.73 Mb in controls from the central Europe IARC study, with increasing shades of grey indicating higher r^2 values. The majority of pairwise D' estimates for these SNPs exceed 0.8.

Supplementary Discussion). Increased risks were seen for former smokers ($P = 4 \times 10^{-7}$) and current smokers ($P = 3 \times 10^{-10}$), as well as a potential increased risk for people who had never smoked ($P = 0.013$). No appreciable variation of the risk was found across the main histological subtypes of lung cancer. We observed a similar risk after stratifying by age at diagnosis, and a slightly greater risk for women compared to men ($P = 0.06$) (Supplementary Table 3). Analysis of the susceptibility locus in additional lung cancer studies would be desirable to obtain further information on these patterns of risk, particularly with respect to smoking status, cumulative cigarette consumption, age and sex. Notably, the risk haplotype is rare in Asian (Japanese and Chinese) and not observed in African (Yoruba) data in the HapMap database[12] and many of the risk alleles have markedly varied allele frequencies in different populations (Supplementary Table 1). Thus, future examination of the association of these markers with lung cancer in different populations might contribute to refined mapping of the locus.

We further investigated whether the locus was associated with cancers of the head and neck including those of the oral cavity, larynx, pharynx and oesophagus. We analysed rs8034191 in two separate studies of head and neck cancer conducted in Europe, the first being conducted in five countries of central Europe and overlapping with the lung cancer controls from five of the six countries included in the present genome-wide association study (726 cases and 694 controls), and the second study being conducted in eight countries of Europe (the ARCAGE study) and including 1,536 cases and 1,443 controls. We observed no effect in either of the two studies separately or combined or in any of the cancer subgroups (Supplementary Fig. 2), implying that this association was specific for lung cancer. Similar results were also observed for rs16969968 (data not shown).

The disease-associated markers span six known genes, including the nicotinic acetylcholine receptor subunits *CHRNA5*, *CHRNA3* and *CHRNB4*, the *IREB2* iron-sensing response element, *PSMA4*, which is implicated in DNA repair, and LOC123688, a gene of

Table 1 | Lung cancer risk and rs8034191 genotype

			T/C versus T/T genotype		C/C versus T/T genotype		Co-dominant model			
	Cases*	Controls*	OR	95% CI	OR	95% CI	OR	95% CI	*P*-values	*P*-heterogeneity
Overall	4,435	7,272	1.21	1.11–1.31	1.77	1.58–2.00	1.30	1.23–1.37	5×10^{-20}	
By study										0.951
Central Europe	1,922	2,520	1.27	1.11–1.44	1.80	1.49–2.18	1.32	1.21–1.45	9×10^{-10}	
Toronto	330	453	1.20	0.85–1.68	1.84	1.14–2.97	1.32	1.05–1.65	0.017	
EPIC	781	1,578	1.18	0.97–1.43	1.68	1.29–2.19	1.27	1.12–1.44	2×10^{-4}	
CARET	764	1,515	1.31	1.08–1.58	1.77	1.34–2.34	1.33	1.16–1.51	2×10^{-5}	
Liverpool	403	814	1.04	0.80–1.34	1.65	1.11–2.44	1.20	1.00–1.44	0.047	
HUNT/ Tromsö	235	392	1.09	0.77–1.54	2.02	1.21–3.37	1.32	1.04–1.68	0.022	

Odds ratio (OR) and 95% confidence interval (CI) for lung cancer comparing heterozygous (T/C) and homozygous (C/C) genotypes of rs8034191 to homozygous (T/T) genotype, overall and separately for each of the six studies. ORs are standardized by age, sex and country. *P*-values are derived from the co-dominant model.
* Subjects with valid call for rs8034191.

unknown function (Fig. 2). It is not possible to identify likely causal alleles or genes based on the differences in the strength of the statistical association because of the strong linkage disequilibrium. However, the nicotinic acetylcholine receptor subunits are strong candidate genes. *CHRNA5* was the only gene found to contain a non-synonymous variant (rs16969968 in exon 5) with strong disease association ($P = 3 \times 10^{-9}$). *CHRNA3* contained a synonymous variant in exon 5 (rs1051730) that was also strongly associated with disease ($P = 5 \times 10^{-9}$); the r^2 between these two variants being 0.99. Although the other markers with a strong disease association either resided in introns or were inter-genic, we cannot exclude the possibility that they could have a biological effect on one or more of the genes from the region. However, other lines of evidence support a possible role for the nicotinic acetylcholine receptor subunit genes.

Nicotinic acetylcholine receptor subunit genes code for proteins that form receptors present in neuronal and other tissues, in particular alveolar epithelial cells, pulmonary neuroendocrine cells, and lung cancer cell lines[2,3], and they bind to nicotine and nicotine derivatives including *N′*-nitrosonornicotine. An association of *CHRNA3* and *CHRNA5* variants with nicotine dependence has been reported[11,13]. The associated markers include the non-synonymous *CHRNA5* SNP, rs16969968, which is one of our markers of lung cancer risk. This SNP introduces a substitution of aspartic acid (D) to asparagine (N) at amino acid position 398 (D398N) of the CHRNA5 protein, located in the central part of the second intracellular loop. Although the function of the second intracellular loop and the possible biological consequences of the D398N alteration remain to be elucidated, this amino acid is highly conserved across species, suggesting that it could have functional importance (Supplementary Fig. 3). A T529A substitution in the second intracellular loop of α4 nAchR, another nicotinic acetylcholine receptor subunit, is known to lead to altered responses to nicotine exposure in the mouse[14].

Within the ARCAGE study (see above), all participants were asked a series of questions relating to tobacco addiction based on the Fagerstrom tolerance questionnaire[15], and we used these to examine whether the chromosome 15q25 locus might be implicated in lung cancer through involvement in tobacco dependence. Two of these questions ('time to first cigarette' and 'numbers of cigarettes per day') have been shown to be particularly strongly associated with nicotine dependence, and responses to both questions result in a 'heaviness of smoking index (HSI)' with a score of between 0 and 6 (ref. 16). We did not observe an association in the ARCAGE controls between rs16969968 and any of the individual Fagerstrom indices of nicotine addiction, or when comparing controls with a HSI of 0 to those with a HSI of 3 or more (Supplementary Table 4). Almost identical patterns were observed for rs8034191 (data not shown). Thus, our data do not support an important role for the locus in nicotine addiction. However, a previous study of a large number of candidate gene markers (4,309 SNPs) identified a possible association between rs16969968 and addiction (uncorrected *P*-value $= 6.4 \times 10^{-4}$) using contrasting extreme phenotypes as measured by the Fagerstrom test for nicotine dependence (FTND)[11]. A second study also identified an association between variants in the region of chromosome 15q25 and numbers of cigarettes smoked per day, although it did not assess directly rs16969968[13]. The FTND and HSI measures of nicotine dependence are highly correlated together, and with cigarettes per day[17], and additional studies to clarify the relationship between chromosome 15q25 variants and tobacco dependence are warranted in light of these results.

Our observation of an increased risk with the chromosome 15q25 locus and lung cancer in non-smokers, as well as the lack of an association with smoking-related head and neck cancers, would indicate that the disease mechanism with lung cancer is unlikely to be explained by an association with tobacco addiction. Independent biological data also suggest that nicotinic acetylcholine receptors could be involved in lung cancer through other mechanisms. It has been suggested that *N′*-nitrosonornicotine and nitrosamines may facilitate neoplastic transformation by stimulating angiogenesis and tumour growth mediated through their interaction with nicotinic acetylcholine receptors[18–20]. The expression of these receptors can also be inhibited by nicotine receptor antagonists, which, if confirmed to be involved in disease aetiology through such a mechanism, implies possible chemoprevention opportunities for lung cancer[5].

No markers outside of those on chromosome 15q25 exceeded the genome-wide significance level for association with lung cancer, although a further 29 had a significance level of $P < 5 \times 10^{-5}$ (Supplementary Table 5). Although most were isolated markers, ten were found to be clustered in a segment of approximately 1 megabase (Mb) on chromosome 6p (28.5–29.5 Mb) within an extended region of high linkage disequilibrium around the major histocompatibility complex. Genotyping of the most significant SNP from the 6p region (rs4324798) in the other five studies provided independent evidence of association ($P = 4 \times 10^{-3}$). In the combined data set, the trend test reached genome-wide significance ($P = 4 \times 10^{-7}$; see Supplementary Fig. 4). The region contains up to 20 documented genes and identification of causal variants is complicated by strong linkage disequilibrium between variants within neighbouring human leukocyte antigen (HLA) and non-HLA genes[21]. Further analyses in multiple diverse populations will be required to confirm this locus and to identify additional lung cancer susceptibility variants. To aid in this, we have made our genome-wide association results available through a publicly accessible website (http://www.ceph.fr/cancer).

METHODS SUMMARY

A detailed description of the component studies can be found in the Supplementary Methods. The genotyping of the IARC central Europe study was conducted using Illumina Sentrix HumanHap300 BeadChip. We excluded variants with a call rate of less than 95% or whose allele distributions deviated strongly from Hardy–Weinberg equilibrium among controls. We also excluded subjects with a completion rate less than 90% or whose reported sex did not match with the inferred sex based on the heterozygosity rate from the X chromosomes. Unexpected duplicates and unexpected first-degree relatives were also excluded from the analysis. Additional quality control measures were applied as described in the Supplementary Methods. Population outliers were detected using STRUCTURE[22] with HapMap subjects as internal controls, and were subsequently excluded from the analysis. Additional analyses for population stratification were undertaken with EIGENSTRAT[23]. Odds ratios (OR) and 95% confidence intervals (CI) were calculated using multivariate unconditional logistic regression models. CEU HapMap SNPs were imputed using MACH (http://www.sph.umich.edu/csg/abecasis/MACH/index.html). Genotyping of additional markers was undertaken with Taqman or Amplifluor assays. Genotyping for all five replication studies was conducted for rs8034191 and rs16969968, and effect estimates from all six lung cancer studies were combined using a fixed-effect model. All *P*-values are two-sided.

Received 30 November 2007; accepted 7 March 2008.

1. Ferlay, J., Bray, F., Pisani, P. & Parkin, M. *GLOBOCAN 2002. IARC CancerBase No 5, version 2.0* (IARC, Lyon, 2004).
2. Minna, J. D. Nicotine exposure and bronchial epithelial cell nicotinic acetylcholine receptor expression in the pathogenesis of lung cancer. *J. Clin. Invest.* 111, 31–33 (2003).
3. Wang, Y. *et al.* Human bronchial epithelial and endothelial cells express α7 nicotinic acetylcholine receptors. *Mol. Pharmacol.* 60, 1201–1209 (2001).
4. Schuller, H. M. Nitrosamines as nicotinic receptor ligands. *Life Sci.* 80, 2274–2280 (2007).
5. Russo, P., Catassi, A., Cesario, A. & Servent, D. Development of novel therapeutic strategies for lung cancer: targeting the cholinergic system. *Curr. Med. Chem.* 13, 3493–3512 (2006).
6. International Agency for Research on Cancer. Reversal of risk after quitting smoking. *IARC Handbooks of Cancer Prevention* Vol. 11, 15–27 (IARC, Lyon, 2007).
7. International Agency for Research on Cancer. Tobacco smoke and involuntary smoking. *IARC Monographs* Vol. 83, 33–47 (IARC, Lyon, 2004).
8. Coleman, M. P. *et al.* EUROCARE Working Group. EUROCARE-3 summary: cancer survival in Europe at the end of the 20th century. *Ann. Oncol.* 14 (suppl. 5), v128–v149 (2003).
9. Matakidou, A., Eisen, T. & Houlston, R. S. Systematic review of the relationship between family history and lung cancer risk. *Br. J. Cancer* 93, 825–833 (2005).
10. Barrett, J. C. & Cardon, L. R. Evaluating coverage of genome-wide association studies. *Nature Genet.* 38, 659–662 (2006).

11. 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).
12. International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
13. Berretini, W. *et al.* α-5/α-3 nicotinic receptor subunit alleles increase risk for heavy smoking. *Mol. Psychiatry* advance online publication doi:10.1038/sj.mp.4002154 (29 January 2008).
14. Tritto, T., Stitzel, J. A., Marks, M. J., Romm, E. & Collins, A. C. Variability in response to nicotine in the LSxSS RI strains: potential role of polymorphisms in α4 and α6 nicotinic receptor genes. *Pharmacogenetics* **12**, 197–208 (2002).
15. Fagerstrom, K. O. & Schneider, N. G. Measuring nicotine dependence: a review of the Fagerstrom Tolerance Questionnaire. *J. Behav. Med.* **12**, 159–182 (1989).
16. Heatherton, T. F., Kozlowski, L. T., Frecker, R. C. & Fagerstrom, K. O. A Fagerstrom test for nicotine dependence: a revision of the Fagerstrom tolerance questionnaire. *Br. J. Addict.* **86**, 1119–1127 (1991).
17. Chabrol, H. *et al.* Comparison of the Heavy Smoking Index and of the Fagerstrom test for nicotine dependence in a sample of 749 cigarette smokers. *Addict. Behav.* **30**, 1474–1477 (2005).
18. Lam, D. C. *et al.* Expression of nicotinic acetylcholine receptor subunit genes in non-small-cell lung cancer reveals differences between smokers and nonsmokers. *Cancer Res.* **67**, 4638–4647 (2007).
19. 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).
20. Dasgupta, P. & Chellappan, S. P. Nicotine-mediated cell proliferation and angiogenesis: new twists to an old story. *Cell Cycle* **5**, 2324–2328 (2006).
21. De Bakker, P. I. *et al.* A high resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nature Genet.* **38**, 1166–1172 (2006).
22. Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567–1587 (2003).
23. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genet.* **38**, 904–909 (2006).

Acknowledgements The authors thank all of the participants who took part in this research and the funders and support staff who made this study possible. We also thank R. Peto for his comments on the manuscript. Funding for the initial genome-wide study was provided by INCa, France. Additional funding for replication studies was provided by the US NCI (R01 CA092039) and the Ontario Institute for Cancer Research (OICR).

Author Contributions P.B. and M.L. designed the study. R.J.H., J.D.M., A.B. and H.B. coordinated the preparation and inclusion of all biological samples. R.J.H., J.D.M., V.G. and S.H. undertook the statistical analysis. Bioinformatics analysis was undertaken by F.M., M.F. and S.H., D.Z. and M.D. coordinated the genotyping of the central Europe samples, and J.D.M, R.J.H. and V.G. coordinated the genotyping of the other studies. All other co-authors coordinated the initial recruitment and management of the studies. M.L. obtained financial support for genotyping of the central Europe study, and P.B. and R.J.H. obtained financial support for genotyping of the other studies. P.B. and M.L. drafted the manuscript with substantial contributions from R.J.H. and J.D.M.