Published in final edited form as:
Nat Genet. 2008 December ; 40(12): 1404–1406. doi:10.1038/ng.254.

Lung cancer susceptibility locus at 5p15.33

James D McKay[1,25], **Rayjean J Hung**[1,2,25], **Valerie Gaborieau**[1,25], **Paolo Boffetta**[1], **Amelie Chabrier**[1], **Graham Byrnes**[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], **John McLaughlin**[2,12], **Frances Shepherd**[13], **Alexandre Montpetit**[14], **Steven Narod**[15], **Hans E Krokan**[16], **Frank Skorpen**[16], **Maiken Bratt Elvestad**[16], **Lars Vatten**[16], **Inger Njølstad**[17], **Tomas Axelsson**[18], **Chu Chen**[19], **Gary Goodman**[19], **Matt Barnett**[19], **Melissa M Loomis**[19], **Jan Lubiński**[20], **Joanna Matyjasik**[20], **Marcin Lener**[20], **Dorota Oszutowska**[20], **John Field**[21], **Triantafillos Liloglou**[21], **George Xinarianos**[21], **Adrian Cassidy**[21], **EPIC Study**[24], **Diana Zelenika**[22], **Anne Boland**[22], **Marc Delepine**[22], **Mario Foglio**[22], **Doris Lechner**[22], **Fumihiko Matsuda**[22], **Helene Blanche**[23], **Ivo Gut**[22], **Simon Heath**[22], **Mark Lathrop**[22,23], **Paul Brennan**[1], **Paolo Vineis**[26,27], **Francoise Clavel-Chapelon**[28], **Domenico Palli**[29], **Rosario Tumino**[30], **Vittorio Krogh**[31], **Salvatore Panico**[32], **Carlos A González**[33], **José Ramón Quirós**[34], **Carmen Martínez**[35], **Carmen Navarro**[36,37], **Eva Ardanaz**[38], **Nerea Larrañaga**[39], **Kay Tee Kham**[40], **Timothy Key**[41], **H Bas Bueno-de-Mesquita**[42], **Petra H M Peeters**[43], **Antonia Trichopoulou**[44], **Jakob Linseisen**[45], **Heiner Boeing**[46], **Göran Hallmans**[47], **Kim Overvad**[48], **Anne Tjønneland**[49], **Merethe Kumle**[50], and **Elio Riboli**[27]

[1]International Agency for Research on Cancer (IARC), Lyon 69008, France [2]Samuel Lunenfeld Research Institute, Toronto M5T 3L9, Canada [3]Institute of Carcinogenesis, Cancer Research Centre, Moscow 115478, Russia [4]Department of Epidemiology, Institute of Occupational Medicine, Lodz 90950, Poland [5]The 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]Cancer Care Ontario, Toronto M5G 2L7, Canada [13]Princess Margaret Hospital, Ontario Cancer Institute, Toronto M5G 2M9, Canada [14]McGill University and Genome Quebec Innovation Centre, Montreal H3A 1A4, Canada [15]Women's College Research Institute, Toronto M5G 1N8, Canada [16]Norwegian University of Science and Technology, Trondheim 7489, Norway [17]Institute of Community Medicine, University of Tromsø, Tromsø 9037, Norway [18]Uppsala University, Department of Medical Sciences, SNP Technology Platform, Academic Hospital, Uppsala 751 85, Sweden [19]Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA [20]Pomeranian Medical University, Department of Genetics and Pathology, International Hereditary Cancer Center, Szczecin 70 115, Poland [21]Roy Castle Lung Cancer Research Programme, University of Liverpool Cancer Research Centre, Liverpool L3 9TA, UK [22]Commissariat à l'énergie Atomique, Institut Genomique, Centre National de Genotypage, Evry 91000, France [23]Fondation Jean Dausset-CEPH, Paris 75010, France [26]Servizio di Epidemiologia dei Tumori, Università di Torino and CPO-Piemonte, Turin 10126, Italy [27]Department of

Correspondence should be addressed to P.B. (brennan@iarc.fr)..
[24]A full list of authors appears at the end of this paper
[25]These authors contributed equally to this work.

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

Epidemiology and Public Health, Imperial College, London SW7, UK [28]INSERM, E3N-EPIC Group Institut Gustave Roussy, Villejuif 94805, France [29]Molecular and Nutrional Epidemiology Unit, Cancer Research and Prevention Institute (ISPO), Florence 50139, Italy [30]Cancer Registry and Histopathology Unit, Azienda Ospedaliera "Civile M.P. Arezzo," Ragusa 97100, Italy [31]Istituto Nazionale dei Tumori, Milan 20133, Italy [32]Dipartimento di Medicina Clinica e Sperimentale, Universita di Napoli, Federico II, Naples 80131, Italy [33]Servicio de Epidemiología y registro del Cáncer, Instituto Catalán de Oncología, Barcelona 08907, Spain [34]Jefe Sección Información Sanitaria, Consejería de Servicios Sociales, Principado de Asturias, Oviedo 33001, Spain [35]Escuela Andaluza de Salud Pública, Granada 18011, Spain [36]Epidemiology Department, Murcia Health Council, Murcia 18011, Spain [37]CIBER Epidemiologia y Salud Publica (CIBERESP), Barcelona 8003, Spain [38]Registro de Cáncer de Navarra, Instituto de Salud Pública, Gobierno de Navarra, Pamplona 31003, Spain [39]Subdirección de Salud Pública de Gipuzkoa, Gobierno Vasco, San Sebastian 20113, Spain [40]MRC Dunn Human Nutrition Unit, Cambridge CB2 0XY, UK [41]Cancer Research UK, University of Oxford, Oxford 0X3 7XP, UK [42]Centre for Food and Health, National Institute of Public Health and the Environment (RIVM), Bilthoven 3720 BA, The Netherlands [43]Julius Center for Health Sciences and Primary Care, Department of Epidemiology, University of Utrecht, Utrecht 3508 GA, The Netherlands [44]Department of Hygiene and Epidemiology, University of Athens, Athens 11527, Greece [45]Division of Clinical Epidemiology, German Cancer Research Centre, Heidelberg 69120, Germany [46]Department of Epidemiology, Deutsches Institut für Ernährungsforschung, Potsdam-Rehbrücke 14558, Germany [47]Department of Public Health and Clinical Medicine, University of Umeå, Umeå 90187, Sweden [48]Department of Epidemiology and Social Medicine, Aarhus University, Aarhus 8000, Denmark [49]The Danish Cancer Society, Institute of Cancer Epidemiology, Copenhagen 2100, Denmark [50]Institute of Community Medicine, University of Tromsø, Tromsø 9037, Norway

Abstract

We carried out a genome-wide association study of lung cancer (3,259 cases and 4,159 controls), followed by replication in 2,899 cases and 5,573 controls. Two uncorrelated disease markers at 5p15.33, rs402710 and rs2736100 were detected by the genome-wide data ($P = 2 \times 10^{-7}$ and $P = 4 \times 10^{-6}$) and replicated by the independent study series ($P = 7 \times 10^{-5}$ and $P = 0.016$). The susceptibility region contains two genes, *TERT* and *CLPTM1L*, suggesting that one or both may have a role in lung cancer etiology.

We and others have recently reported a susceptibility locus for lung cancer in gene region 15q25, an area that includes a cluster of nicotinic acetylcholine receptor genes[1-3]. In order to identify further susceptibility gene loci, we genotyped an additional 1,291 cases and 1,561 controls from three further studies (Toronto case-control study, HUNT2/Tromsø cohort study and CARET cohort study) for a total of 3,259 cases of lung cancer and 4,159 controls with genome-wide data (Table 1 and Supplementary Methods online). After exclusion of subjects because of genotyping quality or evidence of non-European ancestry (Supplementary Methods and Supplementary Fig. 1 online), we analyzed under a log-additive model 315,194 SNPs for 2,971 lung cancer cases and 3,746 controls, adjusting for age, sex and country (Supplementary Fig. 2 online). Using principal-component analysis (Supplementary Methods) to adjust for population stratification, we found only minor differences in the estimates of risk and significance (Supplementary Table 1 online).

Eight SNPs exceeded the genome-wide significance level of 5×10^{-7} (Supplementary Fig. 2b and Supplementary Table 1). Seven of these are located at 15q25.1, the locus previously reported as being associated with lung cancer[1-3], with the most prominent association with rs1051730 ($P = 1 \times 10^{-15}$). The eighth SNP, rs402710, is located at 5p15.33 ($P = 2 \times 10^{-7}$),

indicating a potentially new susceptibility locus for lung cancer. Three additional SNPs in the 5p15.33 region showed evidence of association $P < 5 \times 10^{-6}$ (Supplementary Table 1). Two of these, rs31489 and rs401681, were in strong linkage disequilibrium (LD) with rs402710 ($r^2 > 0.680$) in the 3,746 controls genotyped on the Illumina platform. In contrast, rs2736100 showed relatively little LD with rs402710 ($r^2 = 0.026$) (Supplementary Fig. 3 online).

We subsequently genotyped rs402710 and rs2736100 using Taqman in an additional 2,899 lung cancer cases and 5,573 controls from four separate studies (Table 1 and Supplementary Methods). These included the EPIC cohort study, the Liverpool case-control study, the Szczecin lung cancer study and, uniquely for rs402710 because of limited DNA availability, additional cases and controls from the CARET cohort study. This independent sample provided evidence for replication of the initial finding for both variants ($P = 7 \times 10^{-5}$ for rs402710 and $P = 0.016$ for rs2736100). A combined association using all 5,870 cases and 9,319 controls with correction for the 315,194 comparisons in the genome-wide analysis yielded P values of 4×10^{-6} for rs402710 and 0.02 for rs2736100. The estimated allelic odds ratio (OR) in the replication series was more modest than that of the initial GWA series, subject to the `winner's curse'. The more conservative OR in replication series is the preferred estimate.

More detailed information on the association between lung cancer and the SNPs rs402710 and rs2736100 is presented in Figure 1. The risk-associated allele was the more common allele of rs402710 and the less common allele of rs2736100. The association with rs402710 was prominent in never-smokers ($P = 0.01$), ex-smokers ($P = 0.0007$) and current smokers ($P = 0.0001$), and there was no evidence of any heterogeneity by study, histology, age or sex. There was no apparent geographical heterogeneity in the allele frequencies of rs402710. Adjustment for smoking exposure (pack years) had no effect on the observed association with a smoking-adjusted OR per allele of 1.19 (1.12-1.26). We also investigated rs402710 in the context of smoking intensity among controls and did not observe any association between number of cigarettes consumed per day and rs402710 ($p = 0.74$). The effects observed with rs2736100 were similar, with the associations for the less common (risk) allele being largely comparable to those for rs402710.

Several lines of evidence suggest that the associations observed with rs402710 and rs2736100 are independent. We found little LD between rs402710 and rs2736100 using all available controls. After incorporation of either one of these SNPs into the logistic regression, the association with the other remained significant, and there was no change in the risk estimate (OR per allele for rs402710 = 1.17 ($P = 2 \times 10^{-8}$) with adjustment for rs2736100 and OR per allele for rs2736100 = 1.11 ($P = 0.0004$) with adjustment for rs402710). Second, when cases and controls were compared for the number of risk alleles for rs402710 and rs2736100, there was an increasing trend with increasing number of risk alleles ($P = 2 \times 10^{-13}$) reaching an OR of 1.65 (1.34-2.02) for those who were homozygous for both risk variants (Supplementary Table 2 online). Finally, when we imputed genotypes (Supplementary Methods) at 5p15.33 in the 2,971 cases of lung cancer and 3,746 controls with genome-wide data, we did not identify any SNPs more strongly associated with risk than rs402710 (Supplementary Table 3 online). The top 11 imputed SNPs ($P < 0.0001$) were genotyped subsequently in the cases and controls of central European ancestry (Supplementary Methods) and comparison of haplotype frequencies from this direct genotyping indicated that the prevalence of two distinct haplotypes differed between cases and controls (Supplementary Table 4 online). One haplotype carried the minor allele of rs402710 and eight additional SNPs in high LD ($r^2 > 0.644$) with rs402710, and the second haplotype tagged the minor allele of both rs2736100 and a second SNP rs2736098. Nevertheless, the possibility remains that rs402710 and rs2736100, although only weakly associated with each other, are in LD with one or more causal variants in this region.

The 5p15.33 locus contains two known genes: the *TERT* (human telomerase reverse transcriptase) gene and the *CLPTM1L* (alias *CRR9*; cleft lip and palate transmembrane 1 like) gene. There is no clear evidence to suggest that rs2736100 or rs402710 are themselves causative alleles. The rs2736100 variant is located in intron 1 of *TERT*, and rs402710 is located in a region of high LD that includes the proximal and putative promoter regions of *TERT*, as well as the entire coding region of the *CLPTM1L* gene (Supplementary Fig. 3). Current knowledge of the role of these genes would seem to implicate *TERT* as the more plausible candidate. *TERT* is the reverse transcriptase component of telomerase[4], making it essential for telomerase enzyme production and maintenance of telomeres[5]. The telomerase enzyme is responsible for telomere regeneration, and up to 90% of human tumor samples, including lung cancer[6], show telomerase activity, indicating that regeneration of telomeres is a vital step for most forms of carcinogenesis[7]. *TERT* expression is actively present in germ cells, although is found in very low levels for most types of normal cells[8]. Activation of the *TERT* promoter seems to be a key step in synthesis of the TERT protein and resulting telomerase activity[9]. Such activity may be measured with the telomeric repeat amplification protocol (TRAP) and has been associated with both lung cancer progression and prognosis[6,10,11]. Inhibitors of TERT are clearly of much interest for potential chemo-prevention and treatment of cancer, although their development has so far been unsuccessful[6]. DNA resequencing has shown that there is little common genetic variation in the *TERT* coding region which, along with its high conservation between species, implies that the gene itself is under strong evolutionary restraint[12]. Rare mutations in the *TERT* coding sequence have been implicated in dyskeratosis congenita[13], an autosomal dominant syndrome characterized by bone marrow abnormalities, but also pulmonary fibrosis and increased risk of some cancers[14].

The other gene in this region, *CLPTM1L*, named for its similarity to a gene implicated in susceptibility to cleft lip palate, was identified through screening for cisplatin (CDDP) resistance-related genes and was found to be upregulated in CDDP-resistant ovarian tumor cell lines and to induce apoptosis in CDDP-sensitive cells[15]. The *CLPTM1L* gene is well conserved and expressed in various tissues, including lung tissue. On the basis of these properties, it could be hypothesized that *CLPTM1L* induces apoptosis of lung cells under genotoxic exposures such as tobacco carcinogen-related stress.

In summary, we have identified a new susceptibility locus for lung cancer that comprises two potential candidate genes: *TERT*, an essential component of telomerase production and of carcinogenesis, and *CLPTM1L*, which may induce apoptosis. The nature of the causative alleles remains unclear. Further studies to identify the causal genetic variants and elucidate their function will aid our understanding of the etiology of lung cancer.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

P. Brennan and M. Lathrop designed the study. J.D.M., R.J.H., V.G., M.B.E., A.B. and H. Blanche coordinated the preparation and inclusion of all biological samples. J.D.M., S.H. and V.G. undertook the statistical analysis. Bioinformatics analysis was undertaken by F.M., M.F. and S.H. D.Z., D.L. and I.G. coordinated the genotyping of the central Europe samples. A.M. and R.J.H. coordinated the genotyping of the Toronto samples. J.D.M., D.Z., M.D., A.C., T.A. and H.E.K. coordinated the genotyping of the other studies. All other coauthors coordinated the initial recruitment and management of the studies. M. Lathrop obtained financial support for genotyping of the central Europe study; P. Brennan, R.J.H. and H.E.K. obtained financial support for genotyping of the other studies. P. Brennan and J.D.M. drafted the manuscript with substantial contributions from R.J.H. and M. Lathrop. All authors contributed to the final paper. The authors thank all of the participants who took part in this research and the funders and support and technical staff who made this study possible. Support for the central Europe, HUNT2/Tromsø and CARET genome-wide studies and follow-up genotyping was provided by Institut National du Cancer, France. Support for the HUNT2/Tromsø genome-wide study was also provided by the European Community (Integrated Project DNA repair,

grant no. LSHG-CT-2005-512113), the Norwegian Cancer Association and the Functional Genomics Programme of Research Council of Norway. Funding for the Toronto genome-wide study was provided by the Ontario Institute of Cancer Research. Funding for the Szczecin/Poland replication study was provided by European Community program "Marie-Curie Host Fellowships for the Transfer of Knowledge," grant no. MTKD-CT-2004-510114. Additional funding for study coordination, genotyping of replication studies and statistical analysis was provided by the US National Cancer Institute (R01 CA092039).

References

1. Hung RJ, et al. Nature 2008;452:633–637. [PubMed: 18385738]
2. Amos CI, et al. Nat. Genet 2008;40:616–622. [PubMed: 18385676]
3. Thorgeirsson T, et al. Nature 2008;452:638–642. [PubMed: 18385739]
4. Weinrich SL, et al. Nat. Genet 1997;17:498–502. [PubMed: 9398860]
5. Greider CW, et al. Cell 1985;43:405–413. [PubMed: 3907856]
6. Lantuéjoul S, et al. Int. J. Cancer 2007;120:1835–1841. [PubMed: 17311257]
7. Hanahan D, Weinberg RA. Cell 2000;100:57–70. [PubMed: 10647931]
8. Greider CW, et al. Nature 1989;337:331–337. [PubMed: 2463488]
9. Janknecht R. FEBS Lett 2004;564:9–13. [PubMed: 15094035]
10. Mavrogiannou E, et al. Clin. Chem 2007;53:53–61. [PubMed: 17130181]
11. Wu TC, et al. Lung Cancer 2003;41:163–169. [PubMed: 12871779]
12. Savage SA, et al. Hum. Mutat 2005;26:343–350. [PubMed: 16110488]
13. Armanios M, et al. Proc. Natl. Acad. Sci. USA 2005;102:15960–15964. [PubMed: 16247010]
14. Garcia CK, et al. Nucleic Acids Res 2007;35:7406–7416. [PubMed: 17913752]
15. Yamamoto K, et al. Biochem. Biophys. Res. Commun 2001;280:1148–1154. [PubMed: 11162647]

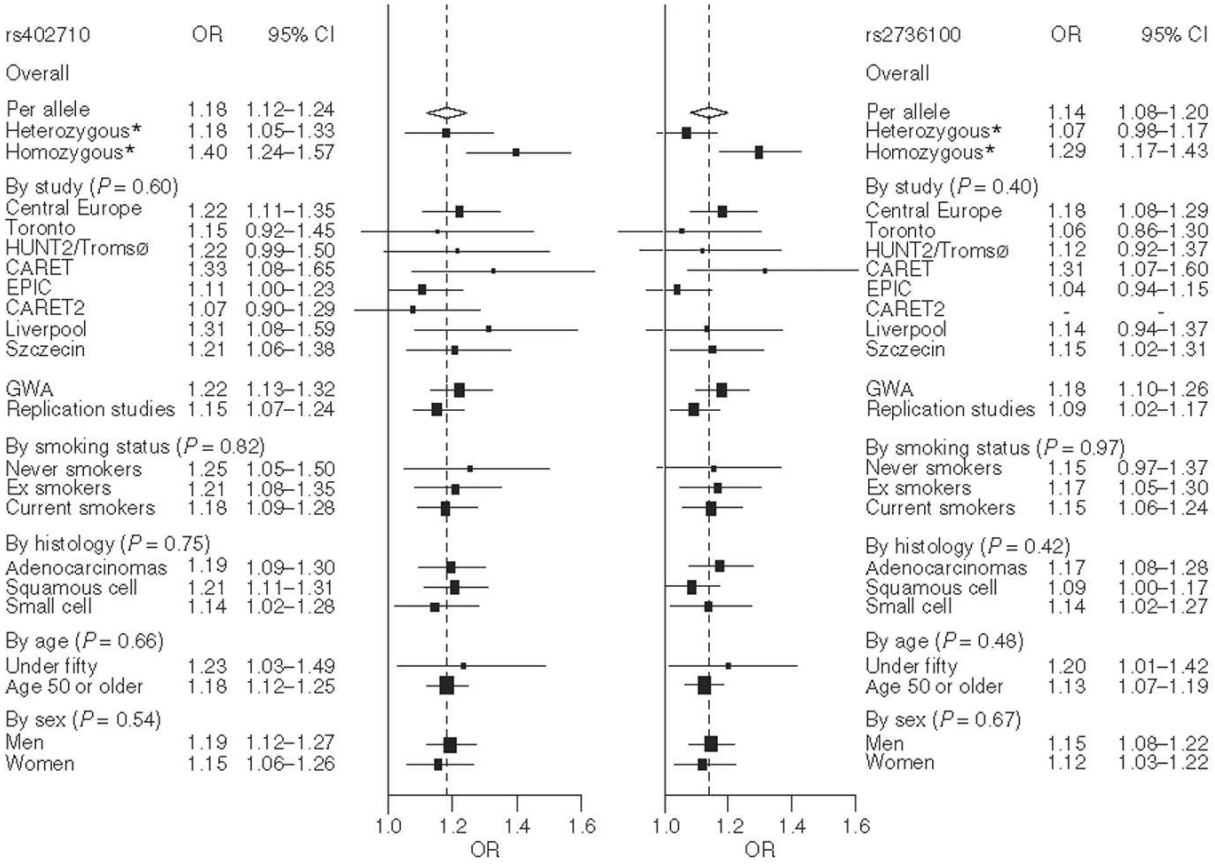

Figure 1.

Forest plot representing lung cancer risk and the two variants in the 5p region (rs402710 and rs2736100). Apart from the odds ratios for heterozygous and homozygous effect (*), odds ratios and 95% confidence intervals are derived from the per-allele model. All models are adjusted for age, sex and country. The overall OR is shown by the broken vertical line. *P* values are from heterogeneity tests.

Table 1

Description of the seven studies contributing to the genome-wide and replication analysis

Study	No. of subjects on ILLUMINA		No. of subjects passing QC		Location	Study design
	Cases	Controls	Cases	Controls		
Genome-wide association studies						
Central Europe	1,968	2,598	1,841	2,441	Romania, Hungary, Poland, Russia, Slovakia, Czech Republic	Case control
Toronto	438	709	330	500	Greater Toronto area (Canada)	Case control
HUNT2/Tromsø	433	433	403	412	North Trondelag County (Norway) and Tromsø city in Tromsø County (Norway)	Cohort
CARET	420	419	397	393	United States	Cohort
Total	3,259	4,159	2,971	3,746		
Replication studies						
EPIC	-	-	1,213	2,591	Sweden, Netherlands,UK, France, Germany, Spain, Italy, Norway, Denmark, Greece	Cohort
Szczecin	-	-	908	1,037	Poland	Case control
CARET2	-	-	363	1,128	United States	Cohort
Liverpool	-	-	415	817	UK	Case control
Total			2,899	5,573		
Total overall			5,870	9,319		

For quality control (Supplementary Methods), we excluded samples with call rate <95%, sex discrepancy or non-European ancestry. We also excluded non-expected duplicates and first-degree relatives from the final analysis.