DMC1:

```
                        170       180       190       200       210       220       230
                        |....|....|....|....|....|....|....|....|....|....|....|....|
Homo sapiens (M200V)    LRDIADRFNVDHDAVLDNVLYARAYTSEHQVELLDYVAAKFHEEAGIFKLLIIDSIMALFR
Homo sapiens (WT)       ..............................M..............................
Pan troglodytes         ---------------------------..................................
Canis familiaris        .............................................................
Rattus norvegicus       ....................................................V........
Mus musculus            ............E................................................
Bos taurus              ............N................................................
Gallus gallus           .............................................................
```

MSH5:

```
                                 10        20       29         40        50
                        ....|....|....|....|....|....|....|....|....|....|....
Homo sapiens (P29S)     MASLGANPRRTPQGPRPGAASSGFPSPASVPGPREAEEEEVEEEEELAEIHLCVLWNSG
Patient (WT)            ............................P..............................
Pan troglodytes         .......L.......G............................VSEGSRNLN.
Canis familiaris        ...VR.T.G........Q..LATCS....Q.HG......E...Q.P..A-...IYRN-
Rattus norvegicus       -MAFR.T.G...P..G.--GVPSASFSS.Q.AMAAPGGI.E.D...P.........S..
Mus musculus            -MAFR.T.G...P..G.RSGIPSASF.S.Q.PMAGPGGI.E.D...P.........S..
Bos taurus              -----------------------------------------------------------
Gallus gallus           ------------------M.ATSATSC.L.PLLAS.----Q..QGCS.T.MS...YA.
```

Figure 3 *DMC1* and *MSH5* multiple protein sequence alignments displaying the regions of exon 10 for *DMC1* and exon 2 for *MSH5* surrounding human mutations. Residues that differ from the wild-type human sequence are shaded in black. Dotted boxes show the positions of each mutation (amino acid numbers referred to human sequence).

sister of this patient, carriers of the same mutation with heterozygous status are fertile. A multiple sequence alignment showed that the amino acid M200V was highly conserved across vertebrates, from chicken to primates (Fig. 3), suggesting a functional or structural role. Accordingly, Polyphen program prediction suggested a potential damaging effect of this mutation. The encoded substitution of a methionine for a valine at this residue is non-conservative in nature and affects a Dmc1 protein region which has been shown to facilitate binding. In addition, preliminary results on yeast have shown that the DMC1 M200V mutation of the conserved residue when introduced into *Schizosaccharomyces pombe* causes a significant decrease in meiotic recombination frequencies (work in progress).

However, further investigations will be needed to confirm the pathological role of these mutations such as screening of increased numbers of patients, and controls and generation of *in vivo* models using knock-in alleles in which missense mutations are introduced that mimic the kinds of mutations found in POF patients.

Acknowledgements

The authors thank Jeremy Don, MD, François Morel MD, Gilbert Schaison MD, and Joelle Belaish MD for their contributions to the collection of patient samples. We also extend thanks to Jerome Dulon and Alexandra Cibot for their technical assistance. We are also grateful to Pr. Dieye, Pasteur Institute from Dakar-Senegal for providing control DNA samples. Supported by an INSERM grant (GIS Maladies Rares A02212KS).

References

1 Nelson LM, Covington SN & Rebar RW. An update: spontaneous premature ovarian failure is not an early menopause. *Fertility and Sterility* 2005 **83** 1327–1332.

2 Coulam CB. Premature gonadal failure. *Fertility and Sterility* 1982 **38** 645–655.

3 Vegetti W, Marozzi A, Manfredini E, Testa G, Alagna F, Nicolosi A, Caliari I, Taborelli M, Tibiletti MG, Dalpra L & Crosignani PG. Premature ovarian failure. *Molecular and Cellular Endocrinology* 2000 **161** 53–57.

4 Davison RM, Fox M & Conway GS. Mapping of the POF1 locus and identification of putative genes for premature ovarian failure. *Molecular Human Reproduction* 2000 **6** 314–318.

5 Marozzi A, Manfredini E, Tibiletti MG, Furlan D, Villa N, Vegetti W, Crosignani PG, Ginelli E, Meneveri R & Dalpra L. Molecular definition of Xq common-deleted region in patients affected by premature ovarian failure. *Human Genetics* 2000 **107** 304–311.

6 Rizzolio F, Bione S, Sala C, Goegan M, Gentile M, Gregato G, Rossi E, Pramparo T, Zuffardi O & Toniolo D. Chromosomal rearrangements in Xq and premature ovarian failure: mapping of 25 new cases and review of the literature. *Human Reproduction* 2006 **21** 1477–1483.

7 Fimiani G, Laperuta C, Falco G, Ventruto V, D'Urso M, Ursini MV & Miano MG. Heterozygosity mapping by quantitative fluorescent PCR reveals an interstitial deletion in Xq26.2-q28 associated with ovarian dysfunction. *Human Reproduction* 2006 **21** 529–535.

8 Portnoi MF, Aboura A, Tachdjian G, Bouchard P, Dewailly D, Bourcigaux N, Frydman R, Reyss AC, Brisset S & Christin-Maitre S. Molecular cytogenetic studies of Xq critical regions in premature ovarian failure patients. *Human Reproduction* 2006 **21** 2329–2334.

9 Lacombe A, Lee H, Zahed L, Choucair M, Muller JM, Nelson SF, Salameh W & Vilain E. Disruption of POF1B binding to nonmuscle actin filaments is associated with premature ovarian failure. *American Journal of Human Genetics* 2006 **79** 113–119.

10 Di Pasquale E, Beck-Peccoz P & Persani L. Hypergonadotropic ovarian failure associated with an inherited mutation of human bone morphogenetic protein-15 (BMP15) gene. *American Journal of Human Genetics* 2004 **75** 106–111.

11 Hui ES, Udofa EA, Soto J, Vanderhoof VH, Zachman K, Tong ZB & Nelson LM. Investigation of the human stem cell factor KIT ligand gene, KITLG, in women with 46,XX spontaneous premature ovarian failure. *Fertility and Sterility* 2006 **85** 1502–1507.

12 Shibanuma K, Tong ZB, Vanderhoof VH, Vanevski K & Nelson LM. Investigation of KIT gene mutations in women with 46,XX spontaneous premature ovarian failure. *BMC Women's Health* 2002 **2** 8.

13 Laissue P, Christin-Maitre S, Touraine P, Kuttenn F, Ritvos O, Aittomaki K, Bourcigaux N, Jacquesson L, Bouchard P, Frydman R, Dewailly D, Reyss AC, Jeffery L, Bachelot A, Massin N, Fellous M & Veitia RA. Mutations and sequence variants in GDF9 and BMP15 in patients with premature ovarian failure. *European Journal of Endocrinology* 2006 **154** 739–744.

14 Beau I, Touraine P, Meduri G, Gougeon A, Desroches A, Matuchansky C, Milgrom E, Kuttenn F & Misrahi M. A novel phenotype related to partial loss of function mutations of the FSH receptor. *Journal of Clinical Investigation* 1998 **102** 1352–1359.

15 Touraine P, Beau I, Gougeon A, Meduri G, Desroches A, Pichard C, Detoeuf M, Paniel B, Prieur M, Zorn JR, Milgrom E, Kuttenn F & Misrahi M. New natural inactivating mutations of the FSH receptor: correlations between receptor function and phenotype. *Molecular Endocrinology* 1999 **13** 1844–1854.

16 Aittomaki K, Lucena JL, Pakarinen P, Sistonen P, Tapanainen J, Gromoll J, Kaskikari R, Sankila EM, Lehvaslaiho H, Engel AR, Nieschlag E, Huhtaniemi I & de la Chapelle A. Mutation in the follicle-stimulating hormone receptor gene causes hereditary hypergonadotropic ovarian failure. *Cell* 1995 **82** 959–968.

17 Meduri G, Touraine P, Beau I, Meduri G, Touraine P, Beau I, Lahuna O, Desroches A, Vacher-Lavenu MC, Kuttenn F & Misrahi M. Delayed puberty and primary amenorrhea associated with a novel mutation of the human follicle-stimulating hormone receptor: clinical, histological, and molecular studies. *Journal of Clinical Endocrinology and Metabolism* 2003 **88** 3491–3498.

18 Doherty E, Pakarinen P, Tiitinen A, Kiilavuori A, Huhtaniemi I, Forrest S & Aittomaki K. A novel mutation in the FSH receptor inhibiting signal transduction and causing primary ovarian failure. *Journal of Clinical Endocrinology and Metabolism* 2002 **87** 1151–1155.

19 Layman LC, Amde S, Cohen DP, Jin M & Xie J. The finnish follicle-stimulating hormone receptor gene mutation is rare in North American women with 46,XX ovarian failure. *Fertility and Sterility* 1998 **69** 300–302.

20 Crisponi L, Deiana M, Loi A, Chiappe F, Uda M, Amati P, Bisceglia L, Zelante L, Nagaraja R, Porcu S, Ristaldi MS, Marzella R, Rocchi M, Nicolino M, Lienhardt-Roussie A, Nivelon A, Verloes A, Schlessinger D, Gasparini P, Bonneau D, Cao A & Pilia G. The putative forkhead transcription factor FOXL2 is mutated in blepharophimosis/ptosis/epicanthus inversus syndrome. *Nature Genetics* 2001 **27** 159–166.

21 De Baere E, Lemercier B, Christin-Maitre S, Durval D, Messiaen L, Fellous M & Veitia R. FOXL2 mutation screening in a large panel of POF patients and XX males. *Journal of Medical Genetics* 2002 **39** e43.

22 Gersak K, Harris SE, Smale WJ & Shelling AN. A novel 30 bp deletion in the FOXL2 gene in a phenotypically normal woman with primary amenorrhoea: case report. *Human Reproduction* 2004 **19** 2767–2770.

23 Dixit H, Rao KL, Padmalatha VV, Kanakavalli M, Deenadayal M, Gupta N, Chakrabarty BN & Singh L. Mutational analysis of the betaglycan gene-coding region in susceptibility for ovarian failure. *Human Reproduction* 2006 **21** 2041–2046.

24 Di Pasquale E, Rossetti R, Marozzi A, Bodega B, Borgato S, Cavallo L, Einaudi S, Radetti G, Russo G, Sacco M, Wasniewska M, Cole T, Beck-Peccoz P, Nelson LM & Persani L. Identification of new variants of human BMP15 gene in a large cohort of women with premature ovarian failure. *Journal of Clinical Endocrinology and Metabolism* 2006 **91** 1976–1979.

25 Kovanci E, Rohozinski J, Simpson JL, Heard MJ, Bishop CE & Carson SA. Growth differentiating factor-9 mutations may be associated with premature ovarian failure. *Fertility and Sterility* 2007 **87** 143–146.

26 Mlinar B, Gersak K, Karas N, Zitnik IP, Battelino T & Lukac-Bajalo J. Galactose-1-phosphate uridyl transferase gene mutations in women with premature ovarian failure. *Fertility and Sterility* 2005 **84** 253–255.

27 Harris SE, Chand AL, Winship IM, Gersak K, Nishi Y, Yanase T, Nawata H & Shelling AN. INHA promoter polymorphisms are associated with premature ovarian failure. *Molecular Human Reproduction* 2005 **11** 779–784.

28 Watkins WJ, Harris SE, Craven MJ, Vincent AL, Winship IM, Gersak K & Shelling AN. An investigation into FOXE1 polyalanine tract length in premature ovarian failure. *Molecular Human Reproduction* 2006 **12** 145–149.

29 Watkins WJ, Umbers AJ, Woad KJ, Harris SE, Winship IM, Gersak K & Shelling AN. Mutational screening of FOXO3A and FOXO1A in women with premature ovarian failure. *Fertility and Sterility* 2006 **86** 1518–1521.

30 Chand AL, Robertson DM, Shelling AN & Harrison CA. Mutational analysis of betaglycan/TGF-betaRIII in premature ovarian failure. *Fertility and Sterility* 2007 **87** 210–212.

31 Telfer EE, Gosden RG, Byskov AG, Spears N, Albertini D, Andersen CY, Anderson R, Braw-Tal R, Clarke H, Gougeon A, McLaughlin E, McLaren A, McNatty K, Schatten G, Silber S & Tsafriri A. On regenerating the ovary and generating controversy. *Cell* 2005 **122** 821–822.

32 Skaznik-Wikiel M, Tilly JC, Lee HJ, Niikura Y, Kaneko-Tarui T, Johnson J & Tilly JL. Serious doubts over 'Eggs forever?' *Differentiation* 2007 **75** 93–99.

33 Pittman DL, Cobb J, Schimenti KJ, Wilson LA, Cooper DM, Brignull E, Handel MA & Schimenti JC. Meiotic prophase arrest with failure of chromosome synapsis in mice deficient for Dmc1, a germline-specific RecA homolog. *Molecular Cell* 1998 **1** 697–705.

34 Yoshida K, Kondoh G, Matsuda Y, Habu T, Nishimune Y & Morita T. The mouse RecA-like gene Dmc1 is required for homologous chromosome synapsis during meiosis. *Molecular Cell* 1998 **1** 707–718.

35 Edelmann W, Cohen PE, Kneitz B, Winand N, Lia M, Heyer J, Kolodner R, Pollard JW & Kucherlapati R. Mammalian MutS homologue 5 is required for chromosome pairing in meiosis. *Nature Genetics* 1999 **21** 123–127.

36 Rozen S & Skaletsky HJ. Primer3 on the WWW for general users and for biologist programmers. In *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, pp 365–386. EdsS Krawetz & S Misener. Totowa, NJ: Humana Press, 2000.

37 Takahashi M, Matsuda F, Margetic N & Lathrop M. Automated identification of single nucleotide polymorphisms from sequencing data. *Journal of Bioinformatics and Computational Biology* 2003 **1** 253–265.

38 Sato S, Seki N, Hotta Y & Tabata S. Expression profiles of a human gene identified as a structural homologue of meiosis-specific recA-like genes. *DNA Research* 1995 **2** 183–186.

39 Sato S, Kobayashi T, Hotta Y & Tabata S. Characterization of a mouse recA-like gene specifically expressed in testis. *DNA Research* 1995 **2** 147–150.

40 Habu T, Taki T, West A, Nishimune Y & Morita T. The mouse and human homologs of DMC1, the yeast meiosis-specific

homologous recombination gene, have a common unique form of exon-skipped transcript in meiosis. *Nucleic Acids Research* 1996 **3** 470–477.

41 Ogawa T, Yu X, Shinohara A & Egelman EH. Similarity of the yeast RAD51 filament to the bacterial RecA filament. *Science* 1993 **259** 1896–1899.

42 Her C & Doggett NA. Cloning, structural characterization, and chromosomal localization of the human orthologue of *Saccharomyces cerevisiae* MSH5 gene. *Genomics* 1998 **52** 50–61.

43 Di Giacomo M, Barchi M, Baudat F, Edelmann W, Keeney S & Jasin M. Distinct DNA-damage-dependent and -independent responses drive the loss of oocytes in recombination-defective mouse mutants. *PNAS* 2005 **102** 737–742.

44 Hollingsworth NM, Ponte L & Halsey C. MSH5, a novel MutS homolog, facilitates meiotic reciprocal recombination between homologs in *Saccharomyces cerevisiae* but not mismatch repair. *Genes and Development* 1995 **9** 1728–1739.

45 Kelly KO, Dernburg AF, Stanfield GM & Villeneuve AM. *Caenorhabditis elegans* msh-5 is required for both normal and radiation-induced meiotic crossing over but not for completion of meiosis. *Genetics* 2000 **156** 617–630.

46 Yi W, Lee TH, Tompkins JD, Zhu F, Wu X & Her C. Physical and functional interaction between hMSH5 and c-Abl. *Cancer Research* 2006 **66** 151–158.

47 Yi W, Wu X, Lee TH, Doggett NA & Her C. Two variants of MutS homolog hMSH5: prevalence in humans and effects on protein interaction. *Biochemical and Biophysical Research Communications* 2005 **332** 524–532 (Erratum in: *Biochemical and Biophysical Research Communications*) 2006 **340** 1018.

Received 3 September 2007
Accepted 7 October 2007

Vol 452 | 3 April 2008 | doi:10.1038/nature06885

LETTERS

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.

©2008 Nature Publishing Group

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.

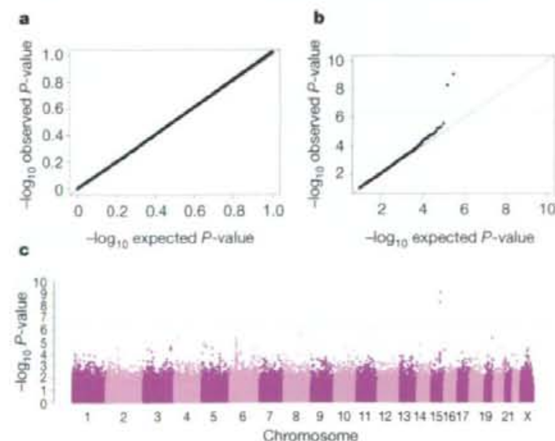

Figure 1 | Genome-wide association results in the central Europe study. **a–c**, Quantile–quantile plot for bottom 90% of *P*-values (**a**) and top 10% of *P*-values (**b**), as well as scatter plot (**c**) of *P*-values in −log scale from the trend test for 310,023 genotyped variants comparing 1,926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

©2008 Nature Publishing Group

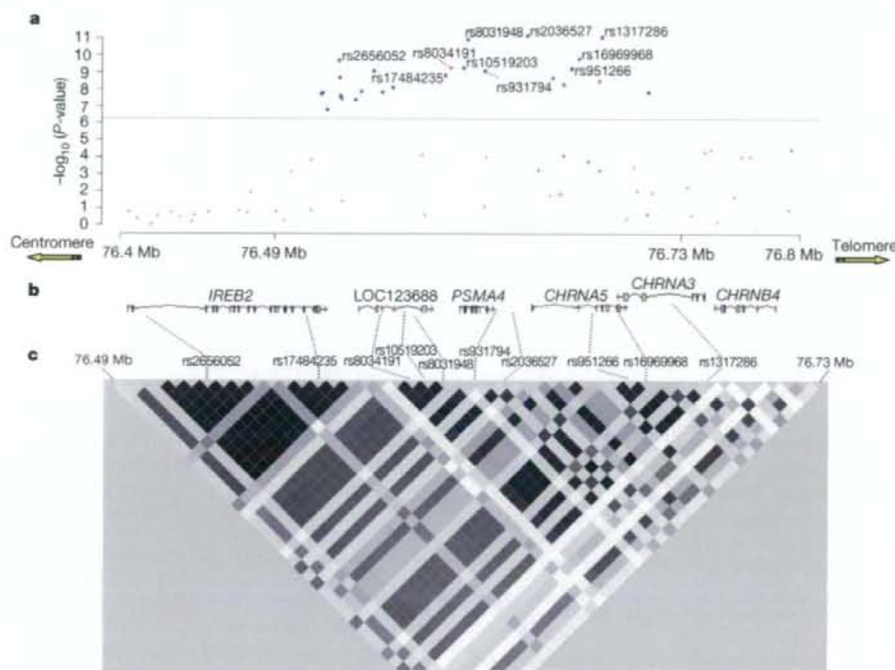

Figure 2 | Lung cancer area of interest across 15q25. **a**, *P*-values for SNPs genotyped in the 15q25 region (76.4–76.8 Mb). The dotted line indicates the genome-wide threshold of $P < 5 \times 10^{-7}$. Points labelled with rs numbers have a $P < 1 \times 10^{-9}$. Points in red are genotyped in the 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.

©2008 Nature Publishing Group

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).

©2008 Nature Publishing Group

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).

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

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. All authors contributed to the final paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.B. (brennan@iarc.fr).

©2008 **Nature Publishing Group**

Manganese Superoxide Dismutase Gene (*SOD2*) Polymorphism and Exudative Age-related Macular Degeneration in the Japanese Population

EDITOR:

THE ARTICLE BY KIMURA AND ASSOCIATES DEMONSTRATED that the manganese superoxide dismutase gene (*SOD2*) polymorphism, the C allele of rs4880, conferred the risk for exudative age-related macular degeneration (AMD) in the Japanese population.[1] Doubts regarding this variant as a risk factor for exudative AMD had been raised by a previous study of Esfandiary and associates with a study sample from Northern Ireland.[2] We attempted to replicate the study using a larger cohort that harbored the same genetic background as was seen in the original study that demonstrated positive results.

We recruited unrelated Japanese individuals: 215 with exudative AMD and 363 population-based controls. The rs4880 T/C was genotyped using the Taqman SNP assay (Applied Biosystems, Foster City, California, USA). For the exudative AMD, the genotype counts were: TT, 176; TC, 39; and CC, 0; whereas for the controls, the counts were: TT, 268; TC, 85; and CC, 10. Both the exudative AMD cases and the controls were within the Hardy-Weinberg equilibrium ($P > .22$, Hardy-Weinberg equilibrium exact test). In the exudative AMD group, there was a lower frequency of the C allele as compared with the control group (9.1% in exudative AMD vs 14.5% in the controls). Therefore, when compared by using the Chi-square test, the C allele proved to be protective in our cohort ($P = .0073$; odds ratio [OR], 0.59; 95% confidence interval [CI], 0.40 to 0.87).

The current results did not replicate the previously reported Japanese study. Using a multiplicative model ($\alpha = 0.05$), the reported result of the OR for homozygous CC as 10.14 was rejected with 100% power. One possible explanation for the discrepancy between the previous study and our own observations is that the number of samples in both studies still may be too low. To determine if any trends may be found within the data, we combined the results of the previous Japanese study and our currents results and performed a metaanalysis. The OR was found to be 0.92 (95% CI, 0.67 to 1.27) when using the Mantel-Haenszel method.

It has been postulated that *SOD2* may be the potential enzyme that is responsible for the pathogenesis of AMD, and some reports have accepted the risk effect of the C allele as the evidence. We would like to express our concerns that the evidence for the risk effect of C allele of *SOD2* rs4880 for exudative AMD seem to be very fragile.

NORIMOTO GOTOH
RYO YAMADA
FUMIHIKO MATSUDA
NAGAHISA YOSHIMURA
Kyoto, Japan
TOMOHIRO IIDA
Fukushima, Japan

REFERENCES

1. Kimura K, Isashiki Y, Sonoda S, Kakiuchi-Matsumoto T, Ohba N. Genetic association of manganese superoxide dismutase with exudative age-related macular degeneration. Am J Ophthalmol 2000;130:769–773.
2. Esfandiary H, Chakravarthy U, Patterson C, Young I, Hughes AE. Association study of detoxification genes in age-related macular degeneration. Br J Ophthalmol 2005;89:470–474.

REPLY

WE APPRECIATE GOTOH AND ASSOCIATES' CRITICAL READing of our article reporting an association of manganese superoxide dismutase (*SOD2*) gene polymorphism with exudative age-related macular degeneration (AMD).[1] Their failure to replicate our results remains to be elucidated. A polymerase chain reaction–restriction determination of the *SOD2* allele used in our study has been proved valid elsewhere,[2] and there is noticeably no substantial difference in the *SOD2* allelic distribution between their control subjects and ours. One possible explanation to be examined to determine the discrepancy between Gotoh and associates' and our observations is the differences in the genetic or anthropologic background of the sample population, that is, AMD patients in the southern region of Japan and those in the central region. Our analysis in terms of the mitochondrial D-loop haplotype suggests that approximately half of our sample population from Southern Japan is assigned to the phylogenetic cluster that is dominant for Okinawa or Thailand, but not for the central Japan.[3] The *SOD2* molecule may play a crucial role in the protection of the retinal pigment epithelium (RPE) against oxidative stress that has been thought to be one of the major factors involved in RPE cell death in AMD.[4] Therefore, it is justified to repeat studies to define further a possible molecular association with disease susceptibility in a larger number of samples with haplotype analysis of intragenic polymorphisms in a single gene.[5]

YASUSHI ISASHIKI
Aichi, Japan
KATSUAKI KIMURA
Kagoshima, Japan
NORIO OHBA
Nagoya, Japan

G/T Substitution in Intron 1 of the *UNC13B* Gene Is Associated With Increased Risk of Nephropathy in Patients With Type 1 Diabetes

David-Alexandre Trégouet,[1] Per-Henrik Groop,[2,3] Steven McGinn,[4] Carol Forsblom,[2,3] Samy Hadjadj,[5,6] Michel Marre,[7,8] Hans-Henrik Parving,[9] Lise Tarnow,[10] Ralph Telgmann,[11] Tiphaine Godefroy,[1] Viviane Nicaud,[1] Rachel Rousseau,[1] Maikki Parkkonen,[3] Anna Hoverfält,[3] Ivo Gut,[4] Simon Heath,[4] Fumihiko Matsuda,[4] Roger Cox,[12] Gbenga Kazeem,[13] Martin Farrall,[13] Dominique Gauguier,[13] Stefan-Martin Brand-Herrmann,[11] François Cambien,[1] Mark Lathrop,[4] and Nathalie Vionnet,[1] for the EURAGEDIC Consortium

OBJECTIVE—Genetic and environmental factors modulate the susceptibility to diabetic nephropathy, as initiating and/or progression factors. The objective of the European Rational Approach for the Genetics of Diabetic Complications (EURAGEDIC) study is to identify nephropathy susceptibility genes. We report molecular genetic studies for 127 candidate genes for nephropathy.

RESEARCH DESIGN AND METHODS—Polymorphisms were identified through sequencing of promoter, exon, and flanking intron gene regions and a database search. A total of 344 nonredundant SNPs and nonsynonymous variants were tested for association with diabetic nephropathy (persistent albuminuria ≥300 mg/24 h) in a large type 1 diabetes case/control (1,176/1,323) study from three European populations.

RESULTS—Only one SNP, rs2281999, located in the *UNC13B* gene, was significantly associated with nephropathy after correction for multiple testing. Analyses of 21 additional markers fully characterizing the haplotypic variability of the *UNC13B* gene showed consistent association of SNP rs13293564 (G/T) located in intron 1 of the gene with nephropathy in the three populations. The odds ratio (OR) for nephropathy associated with the TT genotype was 1.68 (95% CI 1.29–2.19) ($P = 1.0 \times 10^{-4}$). This association was replicated in an independent population of 412 case subjects and 614 control subjects (combined OR of 1.63 [95% CI 1.30–2.05], $P = 2.3 \times 10^{-5}$).

CONCLUSIONS—We identified a polymorphism in the *UNC13B* gene associated with nephropathy. *UNC13B* mediates apoptosis in glomerular cells in the presence of hyperglycemia, an event occurring early in the development of nephropathy. We propose that this polymorphism could be a marker for the initiation of nephropathy. However, further studies are needed to clarify the role of *UNC13B* in nephropathy. ***Diabetes* 57: 2843–2850, 2008**

From [1]INSERM, Paris, France, and Pierre and Marie Curie-Paris VI University, Paris, France; the [2]Helsinki University Central Hospital, Department of Medicine, Division of Nephrology, Helsinki, Finland; the [3]Folkhälsan Institute of Genetics, Folkhälsan Research Center, Biomedicum, Helsinki, Finland; the [4]CEA/Institute of Genomics-National Genotyping Center, Evry, France; [5]CHU Poitiers, Department of Diabetology, Poitiers, France; [6]INSERM U927, CHU Poitiers, Poitiers, France; the [7]Assistance Publique des Hôpitaux de Paris, Centre Hospitalier Universitaire Bichat-Claude Bernard, Paris, France; [8]Université Paris, INSERM U695, Paris, France; the [9]University Hospital of Copenhagen, Rigshospitalet, Department of Medical Endocrinology, Copenhagen, Denmark; the [10]Steno Diabetes Center, Copenhagen, Denmark; the [11]Leibniz Institute for Arteriosclerosis Research, Department of Molecular Genetics of Cardiovascular Disease, University of Muenster, Muenster, Germany; the [12]Mammalian Research Council, Mammalian Genetics Unit, Harwell, U.K.; and the [13]Wellcome Trust Center for Human Genetics, University of Oxford, Oxford, U.K.

Corresponding author: Nathalie Vionnet, vionnet@cng.fr.

Received 18 January 2008 and accepted 8 July 2008.

Published ahead of print at http://diabetes.diabetesjournals.org on 15 July 2008. DOI: 10.2337/db08-0073.

© 2008 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. See http://creativecommons.org/licenses/by-nc-nd/3.0/ for details.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Diabetic nephropathy, characterized by persistent albuminuria, a relentless decline in glomerular filtration rate and raised arterial blood pressure, affects approximately one-third of patients with diabetes (1). Nephropathy accounts for 40% of end-stage renal disease and is associated with high cardiovascular morbidity and mortality (2). Epidemiological and familial studies suggest that genetic factors influence the risk of diabetic nephropathy in both type 1 and type 2 diabetic patients (3–6). Despite rapid research progress, robust predictors of this complication are still lacking.

Phenotypic characterization of nephropathy is more accurate in patients with type 1 diabetes than in those with type 2 diabetes, where the kidney failure may often be caused by nondiabetic factors, mainly hypertension. Using a concerted effort including 2,499 patients with type 1 diabetes from the Danish, Finnish, and French populations, the European Rational Approach for the Genetics of Diabetic Complications (EURAGEDIC) consortium has established a large study for association with diabetic nephropathy that includes 1,176 case subjects and 1,323 control subjects (7). Single nucleotide polymorphisms (SNPs) located in 127 candidate genes selected through assessment of linkage studies, knowledge of metabolic pathways, and animal models were sought for association with nephropathy.

RESEARCH DESIGN AND METHODS

Patient populations. Three European centers, from Denmark, Finland, and France, contributed to the case/control study, with a total of 2,499 subjects with type 1 diabetes. Details for the recruitment of patients have previously been presented (7) and clinical characteristics of the patients are shown in online supplementary Table 1 (available in an online appendix at http://dx.doi.org/10.2337/dc08-0073). Type 1 diabetes was considered present if the age at onset of diabetes was ≤35 years and the time to definitive insulin therapy ≤1 year. Patients in the initial phase of type 1 diabetes, that is, duration of diabetes <5 years, were not included. Established diabetic nephropathy (case subjects) was defined by persistent albuminuria (≥300 mg/24 h or ≥200 μg/min or ≥200 mg/l) in two out of three consecutive

measurements on sterile urine. Patients with clinical or laboratory suspicion of nondiabetic renal or urinary tract disease were excluded. Absence of diabetic nephropathy (control subjects) was defined as persistent normoalbuminuria (urinary albumin excretion rate: <30 mg/24 h or <20 µg/min or <20 mg/l) after at least 15 years of diabetes duration in patients not treated with ACE inhibitors or angiotensin II receptor blockers.

Accordingly, for the initial study, Denmark contributed 952 patients with type 1 diabetes including 489 case subjects and 463 control subjects for diabetic nephropathy, Finland contributed 856 patients including 387 case subjects and 469 control subjects, and France contributed 691 patients including 300 case subjects and 391 control subjects, adding up to a total of 2,499 patients including 1,176 case subjects and 1,323 control subjects for nephropathy.

Two independent datasets were used for replication, the first one being an additional case/control group from the FinnDiane study (8), including 412 case subjects and 614 control subjects who matched the criteria used in the initial study. The second set consisted of 674 patients with type 1 diabetes and microalbuminuria (urinary albumin excretion rate 30–300 mg/24 h or 20–200 µg/min or 20–200 mg/l) from the Danish (n = 60), Finnish (n = 421), and French (n = 193) populations. Clinical characteristics for the replication datasets are presented in online supplementary Table 2.

Molecular screening and SNP selection. The study is a systematic investigation of 127 candidate genes selected through studies of susceptibility loci from linkage studies, metabolic pathways known to be affected in nephropathy, and data from animal models. A detailed list of genes and molecular analyses has been described elsewhere (7,9).

Single nucleotide polymorphisms (SNPs) in the genes selected for the study were identified through database searches and by direct SNP discovery. A total of 119 genes were screened by sequencing all exons, flanking intron sequences, 5′ and 3′ untranslated regions, and promoter regions in at least 32 DNA samples. The sample consisted of healthy French Caucasian subjects from the Epidemiological Study on the Genetics and Environment of Asthma (EGEA) (10). The sample size allowed us to detect SNPs with a minor allele frequency (MAF) of at least 5% with a probability of 96%. For 33 genes that were not initially included in the French National Genotyping Centre (CNG) resequencing effort of >15,000 human genes (http://www.cng.fr/en/teams/geneident/index.html), the screening was performed in 64 additional DNA samples from patients with type 1 diabetes. These included 24, 20, and 20 patients from Denmark, Finland, and France, respectively, half of them (n = 32) with nephropathy and the other half (n = 32) without. For sequencing, DNA samples of two individuals from the same population and with the same phenotype were pooled together. Accordingly, the screening was performed in 16 DNA pools for 86 genes and in 48 (16 + 32) DNA pools for 33 genes.

For each gene, primers were defined for PCR amplification of the exon and promoter regions. PCR was performed in a 15-µl reaction mixture containing 25 ng pooled genomic DNA. Primer sequences are available from the authors on request. Sequencing reactions were performed according to the dye terminator method using an ABI PRISM 3700 DNA Analyzer (Applied Biosystems, Foster City, CA). Alignment of experimental results, SNP detection, and genotype calling were performed using the Genalys software (11) that allows for genotype calls obtained from pooled DNA.

For each gene, the haplotype structure and frequencies were determined from the genotypic data obtained from the control group and population groups using the expectation maximization (EM) algorithm (12). A total of 350 variants were selected to account for all estimated haplotypes with frequencies >5%. These tagSNPs represented a median genetic variation (haplotype diversity) by gene of 87% (range 64–100%). All were retained for further genotyping in the case/control study. In addition, all nonsynonymous variants that were detected in at least one diseased population were systematically investigated (n = 19).

For two genes (*RELA, TGFBR1*) for which no polymorphisms were identified, SNPs were selected using the SNPBrowser software v.2 (Applera Corporation). For eight additional genes (*CCR5, CNDP1, HNF4A, LTA, PON2, GCGR, INPPL1, PLA2G7*), polymorphisms were selected according to reported associations with phenotypes relevant for diabetic nephropathy (13–20). We also examined 94 SNP markers (genomic control markers) in nongenic regions spaced throughout the genome to control for possible stratification within each population (21,22).

A total of 21 additional SNPs in the *UNC13B* gene (GeneID#10497; full name unc-13 homolog B [*C. elegans*]) were genotyped after the initial positive association results from the first step. These additional SNPs were selected from the Hapmap project (http://www.hapmap.org) so that >95% of the *UNC13B* haplotypic variability was characterized.

Genotyping. Genomic DNA was isolated from human leukocytes using standard methods. SNP genotyping was performed at the French National Genotyping Center (CNG) using automated high-throughput methods including TaqMan, Amplifluor, MALDI-MS, and SNPlex methods. All liquid handling was performed robotically in 384-well plates with a BasePlate Robot (The Automation Partnership, Royston, U.K). For SNP genotyping by mass spectrometry, the GOOD assay was applied as previously described (23). TaqMan (assay-by-design) was carried out in a 5-µl volume according to the manufacturer's recommendations, with probes and mastermix from Applied Biosystems. For Amplifluor, primers were designed using "AssayArchitect" (http://www.assayarchitect.com). Primer sequences and conditions are available on request. End point fluorescence was detected for TaqMan and Amplifluor assays using an ABI7900HT reader (Applied Biosystems, Courtaboeuf, France), and genotypes were assigned with SDS 2.1 software. Genotyping with the SNPlex platform was performed according to the manufacturer's recommendations (Applied Biosystems, Courtaboeuf, France).

The genotyping success rate was >85% for all markers (<90% for 3% of the markers, between 90 and 95% for 17%, and >95% for 80% of the markers), and among 192 replicate samples genotyped blindly, no genotype differences were found. Hardy-Weinberg equilibrium was checked in case subjects and control subjects in all populations, and markers showing deviation from Hardy-Weinberg equilibrium at the 0.001 significance level were not considered in the case/control comparison.

Statistical analysis. Allele frequencies were estimated by gene counting, and deviation from Hardy-Weinberg equilibrium was tested by use of a χ^2 with 1 d.f. Difference in allele frequencies between case subjects and control subjects were tested by a χ^2 test with 1 d.f. separately in each population, and associated P values were combined across populations using Fisher's method (24) to produce an overall test of significance. Adjustment for multiple testing was carried out by correcting for the effective number of independent tests (25) to take into account the linkage disequilibrium (LD) between SNPs. Logistic regression analyses were performed to estimate genetic ORs, adjusted for age, sex, smoking, diabetes duration, and A1C. LD matrices were obtained using Haploview software (26), and haplotype association analyses were carried out using THESIAS software (27). Homogeneity of ORs across populations was investigated using the Mantel-Haenszel statistic (28).

Expression studies

Cell culture. Cell lines HepG2, MDCK I and II, MCF7, Cos7, HeLa, EAhy-926, SaOs-2, U2Os, SHSY5Y, and rat smooth muscle cells (rSMC) were maintained in Dulbecco's modified Eagle's medium (Sigma, Geissendorf, Germany) with 10% conditioned fetal calf serum (PAA, Cölbe, Germany), penicillin (100 units/ml), streptomycin (100 ng/ml), and L-glutamin (2 mmol $\cdot$ l^{-1} $\cdot$ ml^{-1}). HEK293T cells received iron-supplemented fetal calf serum (Cell Concepts, Umkirch, Germany). Suspension cell lines THP1, U937, K562, HL60, and RAW264.7 were maintained in RPMI-1640 medium (Sigma) with the same additions plus 1 × minimal essential medium (MEM) minimal amino acids (PAA). Differentiation of THP-1 monocytes into macrophages induced by stimulation with 10^{-8} mol/l phorbol 12-myristate 13-acetate (PMA) and differentiation of SaOs-2 osteosarcoma cells was induced by stimulation with 100 mmol/l glycerol-1-phosphate and 10 mmol/l ascorbic acid.

Isolation of total RNA and generation of cDNA. Total RNA from cells was isolated from 10^6 cells each with the RNeasy Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. RNA from human brain was extracted from the left frontal cortex of a 75-year-old male patient <24 h postmortem, and testis RNA was isolated 1 h after surgical operation from a 62-year-old patient who underwent orchidectomy for prostate cancer as described (29). RNA yield was controlled by TBE/agarose gel electrophoresis and adjusted nanophotometrically. For first-strand cDNA synthesis, 5 µg total RNA was used (Fermentas, St. Leon-Rot, Germany). Efficiency was routinely controlled by diagnostic PCR for ribosomal protein RP27. Podocyte cDNA was generated from an immortalized human podocyte cell line (30).

Diagnostic PCR. Exon-spanning primers for nested diagnostic PCR were designed from *UNC13B* sequence NM_006377 (sense primers S1: GTGCAC CACTCCTCATAACTT; S2: CAACCTACTGCTATGAGTGT; antisense primers A1: TGTGCAAGTCA GCAAAACTAAG, A2: AAGCCAAGGACAAAACAG GATC). PCR was conducted with Go*Taq*DNA-Polymerase (Promega) and 35 cycles of amplification. Integrity of the cDNA was controlled by diagnostic PCR for ribosomal protein 27 (rp27; sense primer: 5′-CCAGGATAAGGAAGG AATTCCTCCTG-3′, antisense primer: 5′-CCAGCACCACATTCATCAGAAGG-3′, not shown).

In silico analyses. For the prediction of putative transcription factor binding sites, a sequence of 25 bases to either side of the SNP was submitted for each SNP individually to a net-based search tool (Alibaba2.1, Transfac7.0; http://www.gene-regulation.com). Settings for core and pair similarities, matrix conservation, and factor class levels were adjusted according to factors predicted.

Five polymorphisms, rs10081672, rs10972356, rs13288912, rs12377498, and rs10972333, were in complete association with rs13293564 located in intron 1 of the *UNC13B* gene associated with nephropathy. They were detected on the NCBI B35 (http://www.ncbi.nlm.nih.gov) at the respective nucleotide positions 35145908, 35143435, 35143123, 35140841, and 35126146, with the beginning

of the 5′-UTR within exon 1 residing at nucleotide position 35151989. This start site was confirmed in all reference sequences in the University of California, Santa Cruz genome browser, with no indication of alternative upstream exons or presence of alternative promoters. Hence, the variants rs10081672, rs10972356, rs13288912, rs12377498, and rs10972333 are located, respectively, −6,081, −8,554, −8,866, −11,148, and −25,843 bp upstream of the transcription start site of the human *UNC13B* gene. Sequence homology scans and chromosomal neighborhood analyses were performed using the University of California, Santa Cruz genome browser (http://genome.ucsc.edu) covering chromosomal region 9:35,101,909–35,160,332. Special emphasis was put on placental mammal conserved elements in a 28-way multiz alignment. Results were cross-checked using rVista 2.0 software (http://rvista.dcode.org). There was no noticeable sequence conservation this far upstream of *UNC13B* exon 1 in either species.

RESULTS

SNP discovery, selection, and genotyping. A total of 119 genes were resequenced and 1,833 sequence variants were detected, including 1,673 SNPs and 160 insertion/deletion polymorphisms. A total of 773 (42.2%) of these variants were not present in the dbSNP (build 126) and therefore represent novel polymorphisms. All data have been cataloged in the dbSNP database and are available online at http://genecanvas.ecgene.net. They were located in the 5′-flanking region (n = 53), 5′UT (n = 31), intron (n = 1,166), nonsynonymous coding (n = 139), slice site (n =1), 3′UT (n = 221), and the 3′-flanking region (n = 40). The proportion of SNPs detected in exons was not different between the 773 newly discovered polymorphisms and the 1,060 variants in dbSNP build 126 (32.4 vs. 30.3%, χ^2 test: $P = 0.35$). As expected, newly discovered SNPs were mainly rare, 66.1% with MAF <5% compared with 12.3% of SNPs in dbSNP ($P < 10^{-4}$). The same held true for insertions/deletions (17.2% new vs. 2.5% in dbSNP, $P < 10^{-4}$).

A total of 532 haplotypes with a frequency >5% in at least one population were determined. For these 119 genes, a total of 369 polymorphisms, including haplotype-tagging SNPs and nonsynonymous variants, were selected for genotyping. For two genes with no variant identified through resequencing (*RELA, TGFBR1*), four SNPs were selected with SNPbrowser. In addition, 15 SNPs were selected in eight genes from previously reported associations with phenotypes relevant for diabetic nephropathy. We were not able to obtain data for 28 markers due to the impossibility of obtaining a genotyping assay, and 16 markers were excluded because they showed significant deviation from the Hardy-Weinberg equilibrium in case subjects and control subjects from the three populations.

Association studies. A total of 344 SNPs were investigated for association with nephropathy in the EURAGEDIC study. Allele frequencies in case and control groups from three populations are shown in supplementary Table 3 (in the online appendix). Nominally significant association across the three populations ($P < 0.05$) was observed for 33 SNPs of 344, with P values ranging from $P = 1.79 \times 10^{-5}$ to $P = 0.050$ (supplementary Table 3). Of the 15 polymorphisms in the eight genes selected from the literature, only one, rs1799987, located in the *CCR5* gene, showed nominal significant association across the three populations ($P = 0.025$). However, this association did not remain significant after correction for multiple testing. For the 119 remaining genes, the number of independent tests was estimated to be $N_{eff} = 317$, with a corresponding significance threshold of $P = 1.58 \times 10^{-4}$ ($P = 0.05/317$). Only one SNP, rs2281999, located in *UNC13B*, remained significantly associated with nephropathy ($P = 1.79 \times 10^{-5}$) after correction. This association was mainly observed in the Finnish sample, with a trend remaining in the Danish but not in the French samples (Table 1 and supplementary Table 3). Another *UNC13B* SNP (rs661712) showed nominal evidence for association with nephropathy ($P = 4 \times 10^{-4}$) but did not remain significant after correction for multiple testing. The association for the 94 genomic control markers was compatible with expectations under the null hypothesis of no association, indicating that stratification within one or more of the populations is an unlikely source of positive association results. Furthermore, correction of the association for the two significant eigenvectors identified by performing principal components analysis (using Eigenstrat) on the genomic control markers had no effect on the results.

Our initial sequencing of the 39 exons of *UNC13B* has identified a total of 13 SNPs that could be tagged by four SNPs. These four SNPs, together with a nonsynonymous variant located in exon 28 (R1124Q), were genotyped in the whole EURAGEDIC sample. However, analysis of the available HapMap data revealed that these four SNPs were not sufficient to correctly characterize the haplotypic variability of the gene, which spans over ~240 kb on chromosome 9p12-p11. Therefore, 21 additional tagging SNPs spanning the whole gene were further genotyped to clarify the observed association of *UNC13B* SNPs with nephropathy (Fig. 1). The results of association analyses of all *UNC13B* SNPs (apart from two rare variants shown in supplementary Table 3) with nephropathy are summarized in Table 1. While most of the SNPs were associated with nephropathy in the whole study, none showed significant allelic association in the three populations, and only one, rs13293564, showed nominal allelic association in two populations: Denmark and Finland. In these two populations, homozygous carriers of the T allele were more frequent in case subjects than in control subjects (0.18 vs. 0.12 and 0.25 vs. 0.17, respectively) and were then at higher risk of nephropathy (OR 1.60 [1.10–2.32], $P = 0.013$; OR 1.57 [1.11–2.22], $P = 0.011$, respectively) (supplementary Table 4). Interestingly, French homozygous carriers of this allele also tended to be more frequent in case subjects than in control subjects (0.18 vs. 0.15), but the association failed to reach nominal significance (OR 1.26 [0.83–1.92], $P = 0.278$). These three ORs were not statistically different from each other ($P = 0.663$) and were therefore combined, leading to an increased risk of nephropathy associated with the TT genotype of OR 1.49 (95% CI 1.20–1.85) ($P = 0.0003$). Further adjustment for smoking and A1C did not modify these associations (supplementary Table 4).

One feature of the French patients with diabetes is that 76% of them had proliferative retinopathy, whereas this percentage was 49 and 58 in Denmark and Finland, respectively (7, supplementary Table 1). Further adjustment for retinopathy status strengthened the observed association, in particular in France, where the OR associated with the TT genotype was then similar to that observed in Denmark (Fig. 2), leading to a common OR for nephropathy associated with the TT genotype of 1.68 (95% CI 1.29–2.19) ($P = 0.0001$). This was explained by the slightly more pronounced difference in TT genotype frequencies between case subjects and control subjects observed in patients without proliferative retinopathy (0.25 vs. 0.15) than in patients with proliferative retinopathy (0.19 vs. 0.13) (supplementary Table 5). However, no heterogeneity was detected according to the retinopathy status ($P = 0.48$).

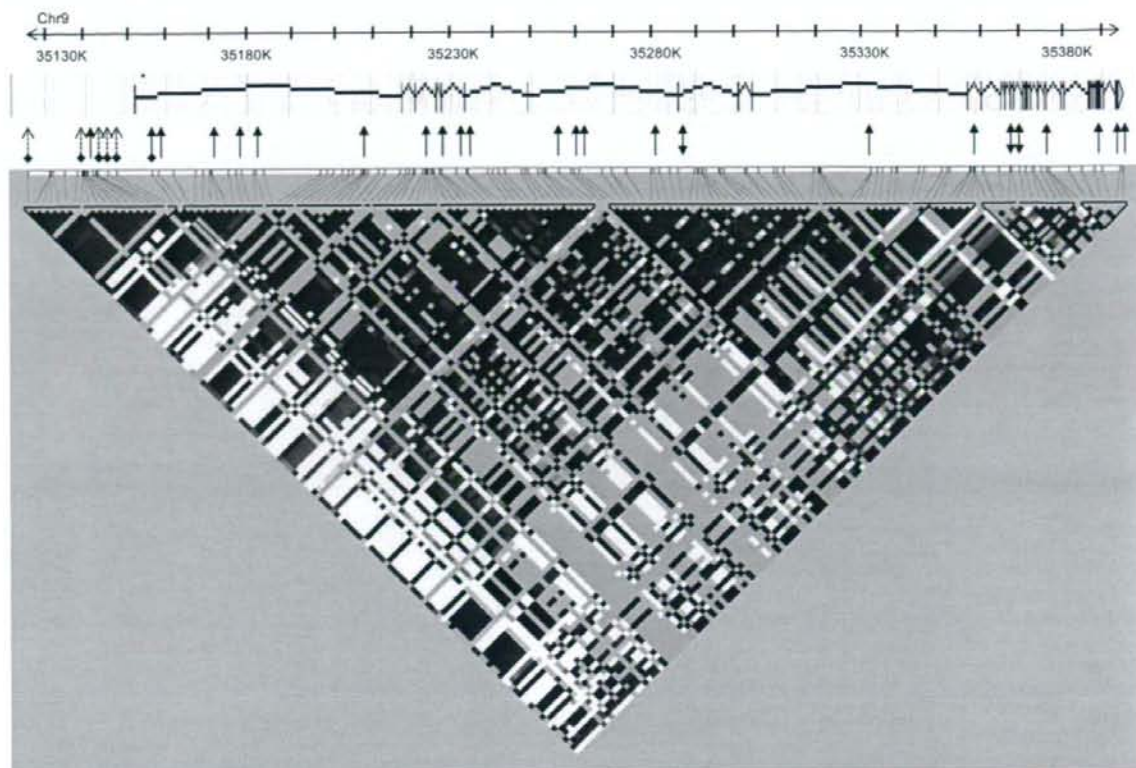

FIG. 1. Schematic representation of the *UNC13B* gene. The structure of UNC13B gene on chromosome 9 is presented with the respective positions of the 39 exons and of the 24 SNPs genotyped (rsID are given in Table 1), as well as the HapMap haplotype blocks (in D′). Arrows at both ends: SNPs selected through sequencing; single arrow: haplotype tagging SNPs selected from Hapmap; arrow with bullet: position of rs13293564, the SNP associated with nephropathy; dashed arrow with bullet: SNPs in complete association with rs13293564 (not typed).

A two-locus association analysis (Table 2) on the rs2281999 and rs13293564 SNPs showed that the difference in the genotype distribution between case subjects and control subjects mainly came from the rs13293564-TT genotype, suggesting that the initial association observed between the rs228199 and nephropathy was due to its LD with the rs13293564. All further LD, multilocus, and haplotype analyses converge to a unique recessive effect of the rs13293564 polymorphism (supplementary Tables 6–8, supplementary Fig. 1), an effect that occurs homogeneously in men and women (data not shown) and across the three EURAGEDIC populations.

The rs13293564 SNP was further investigated in an independent Finnish sample from the FinnDiane study (8) including 412 case subjects with nephropathy and 614 control subjects. In this population, the TT genotype was also associated with an increased risk of nephropathy (OR 1.45 [1.06–1.98]) ($P = 0.020$) that was hardly modified by further adjustment for smoking, A1C, and proliferative retinopathy (OR 1.51 [0.97–2.36]; $P = 0.070$). Finally, in the combined sample from the EURAGEDIC and FinnDiane studies, the adjusted OR for nephropathy associated with the TT genotype was 1.63 (95% CI 1.30–2.05) ($P = 2.3 \times 10^{-5}$) (Fig. 2).

The rs13293564 SNP was further investigated in 674 patients with type 1 diabetes and microalbuminuria from the three populations. The frequency of the TT genotype in patients with microalbuminuria was significantly higher than the frequency of this genotype in patients with normoalbuminuria (0.22 vs. 0.15, $P < 10^{-4}$) (supplementary Table 9). The frequency of the TT genotype was similar, whatever the stage of diabetic nephropathy, incipient nephropathy (microalbuminuria) (0.22), macroalbuminuria (0.21), or end-stage renal disease (ESRD) (0.23) (supplementary Table 9).

Assuming a minor allele frequency of 0.39 at the rs13293564 locus in patients with type 1 diabetes and an increased risk of 1.6 in homozygous carriers of the T allele, the population attributable risk for rs13293564 would be 8.3%.

Expression studies. A diagnostic PCR for *UNC13B* transcripts in various cell lines and tissues (supplementary Fig. 2) was performed. The strongest expression of *UNC13B* was detectable in human tissues from brain, testis, and podocytes, as well as the human immortalized podocyte cell line SHSy. Kidney cell lines COS7, and to a minor extent MDCK I and II, express *UNC13B*, but not embryonic kidney cell line HEK293T. Osteosarcoma cell lines (SaOs2, U2Os), liver (HepG2), and breast cancer (MCF7) show noticeable expression, whereas in monocytic cell lines, either differentiated or not, expression is strictly cell-line dependent. Choriocarcinoma cells HeLa do not express *UNC13B*.

In silico analyses. There was no feature to suggest that rs13293564 located in intron 1 of the *UNC13B* is the functional variant. Analyses of the five polymorphisms in

TABLE 1
Association analysis between *UNC13B* gene polymorphisms and diabetic nephropathy in the EURAGEDIC study

Polymorphisms	Denmark			Finland			France			
	Allele frequency in control subjects	Allele frequency in case subjects	P*	Allele frequency in control subjects	Allele frequency in case subjects	P*	Allele frequency in control subjects	Allele frequency in case subjects	P*	Whole (P†)
rs13285401 (C/T)	0.584/0.416	0.633/0.367	0.0282	0.599/0.401	0.624/0.376	0.3046	0.619/0.381	0.626/0.374	0.7864	0.1248
rs13293564 (G/T)	0.643/0.357	0.587/0.413	0.0126	0.579/0.421	0.514/0.486	0.0072	0.605/0.395	0.591/0.409	0.5974	0.0032
rs10972365 (T/C)	0.721/0.279	0.748/0.252	0.1711	0.740/0.260	0.797/0.203	0.0060	0.736/0.264	0.767/0.233	0.2113	0.0097
rs4879877 (A/G)	0.872/0.128	0.821/0.179	0.0020	0.822/0.178	0.799/0.201	0.2298	0.869/0.131	0.866/0.134	0.8739	0.0158
rs4111859 (A/T)	0.872/0.128	0.822/0.178	0.0026	0.821/0.179	0.796/0.204	0.1945	0.870/0.130	0.868/0.132	0.9012	0.0174
rs3904435 (A/G)	0.850/0.150	0.801/0.199	0.0055	0.726/0.274	0.734/0.266	0.7131	0.840/0.160	0.849/0.151	0.6603	0.0639
rs12685290 (A/G)	0.594/0.406	0.578/0.422	0.4892	0.518/0.482	0.448/0.552	0.0044	0.590/0.410	0.605/0.395	0.5779	0.0373
rs17360668 (G/A)	0.744/0.256	0.781/0.219	0.0591	0.758/0.242	0.821/0.179	0.0016	0.754/0.246	0.772/0.228	0.4483	0.0026
rs10972396 (G/T)	0.873/0.127	0.824/0.176	0.0038	0.821/0.179	0.814/0.186	0.7247	0.861/0.139	0.873/0.127	0.5063	0.0407
rs10972397 (A/G)	0.873/0.127	0.819/0.181	0.0012	0.827/0.173	0.812/0.188	0.4190	0.861/0.139	0.867/0.133	0.7500	0.0150
rs7851161 (A/T)	0.556/0.444	0.560/0.440	0.8530	0.595/0.405	0.531/0.469	0.0085	0.511/0.489	0.532/0.468	0.4566	0.0761
rs10758301 (T/G)	0.586/0.414	0.576/0.424	0.6467	0.520/0.480	0.450/0.550	0.0041	0.598/0.402	0.614/0.386	0.5392	0.0414
rs10121009 (C/T)	0.810/0.190	0.814/0.186	0.8417	0.711/0.289	0.784/0.216	0.0007	0.826/0.174	0.804/0.196	0.3110	0.0085
rs10114937 (T/C)	0.672/0.328	0.708/0.292	0.0915	0.641/0.359	0.734/0.266	0.0000	0.715/0.285	0.696/0.304	0.4601	0.0000
rs10758303 (A/G)	0.541/0.459	0.523/0.477	0.4154	0.541/0.459	0.461/0.539	0.0010	0.568/0.432	0.561/0.439	0.8021	0.0136
rs661712 (C/T)	0.671/0.329	0.709/0.291	0.0763	0.648/0.352	0.739/0.261	0.0001	0.713/0.287	0.700/0.300	0.5904	0.0004
rs17296428 (C/G)	0.822/0.178	0.833/0.167	0.5156	0.857/0.143	0.819/0.181	0.0323	0.826/0.174	0.851/0.149	0.2332	0.0852
rs12684897 (T/C)	0.821/0.179	0.830/0.170	0.6083	0.861/0.139	0.828/0.172	0.0608	0.790/0.210	0.819/0.181	0.1841	0.1255
rs2282001 (G/C)	0.928/0.072	0.933/0.067	0.6768	0.883/0.117	0.915/0.085	0.0347	0.952/0.048	0.927/0.073	0.0569	0.0394
rs2281999 (C/T)	0.660/0.340	0.693/0.307	0.1396	0.627/0.373	0.725/0.275	0.0000	0.707/0.293	0.704/0.296	0.8933	0.00002
rs1927962 (T/C)	0.781/0.219	0.792/0.208	0.5489	0.826/0.174	0.768/0.232	0.0031	0.773/0.227	0.802/0.198	0.2083	0.0143
rs12339582 (G/T)	1.000/0.000	1.000/0.000	1.0000	1.000/0.000	1.000/0.000	1.0000	1.000/0.000	0.997/0.003	0.1065	0.6121
rs12726 (G/A)	0.780/0.220	0.774/0.226	0.7467	0.782/0.218	0.757/0.243	0.2181	0.751/0.249	0.739/0.261	0.6007	0.5895
rs10814234 (C/G)	0.864/0.136	0.882/0.118	0.2294	0.923/0.077	0.950/0.050	0.0220	0.877/0.123	0.882/0.118	0.7525	0.0839

*Difference in allele frequencies between case and control subjects was tested by a χ^2 test with 1 d.f., separately in each population. †For each tested SNP, the P values of the association tests obtained in the three populations were combined by Fisher's method.

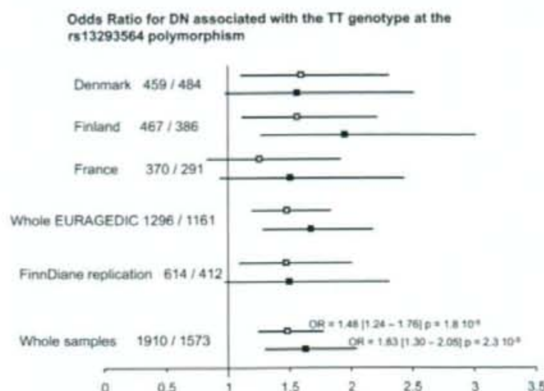

FIG. 2. ORs for diabetic nephropathy associated with the TT genotype at the rs13293564 polymorphism. Number of control subjects/number of case subjects is shown. □, OR adjusted for age and sex; ■, OR adjusted for age, sex, smoking, A1C, and proliferative retinopathy.

complete association with rs13293564 (rs10081672, rs10972356, rs13288912, rs12377498, and rs10972333) located in the putative promoter region showed that the rs10081672 and rs10972333 SNPs, located respectively at positions −6081 and −25843 bp upstream of the transcription start site of the human *UNC13B* gene, affect potential Sp1 and upstream stimulating factor binding sites, respectively.

DISCUSSION

We have shown for the first time that a common variation in the *UNC13B* gene was reproducibly associated with nephropathy. A recent study that assessed 115 candidate genes for nephropathy in 82 trios did not find significant results for any of the six *UNC13B* markers analyzed (31). This might be explained by the insufficient number of markers studied and a study of limited power.

The rs13293564 lies within the first intron of the *UNC13B* gene and is in strong LD with many other SNPs all along the gene (see supplementary data and HapMap data). However, it is in complete association with only five SNPs (rs10081672, rs10972356, rs13288912, rs12377498, and rs10972333) located in the putative promoter region. While there is no feature to suggest that rs13293564 is the functional variant, rs10081672 and rs10972333, located respectively at positions −6081 and −25843 bp upstream of the transcription start site of the human *UNC13B* gene, affect potential Sp1 and USF binding sites, respectively, suggesting that those SNPs could be the functional variants. It is possible that regulatory elements are located this far upstream of the proximal regulatory regions within the core promoter. Functional molecular analyses are needed to clarify their impact. It should also be stressed that an expressed repetitive element NM_001039797 of 3.032 kbp is located within *UNC13B* intron 1, 20 kb downstream of rs13293564, the function of which is unknown, and strong LD spans over this region. This opens up the possibility that genetic variation affects a regulatory element for *UNC13B* within the intronic region. Such mechanisms have recently been shown to be involved in susceptibility to different diseases (32–34).

The human *UNC13B* gene product (also called Hmunc13) has been cloned from the human kidney library (35) and is homologous to rat munc13 proteins, which are members of the protein kinase C (PKC) superfamily that lack a kinase domain. Human UNC13B has one C1 diacylglycerol (DAG) and three C2 (Ca^{2+}) binding domains. It is both upregulated and activated in the presence of hyperglycemia in renal cortical tubular cells and in glomerular mesangial cells (36). Using RT-PCR to further assess the expression of *UNC13B* in human tissues, we found that *UNC13B* was also highly expressed in podocytes (supplemental Fig. 2). It has been shown that DAG activation of *UNC13B*-expressing cells induces apoptosis. As hyperglycemia increases intracellular DAG levels and is associated with apoptosis in various tissues, including human kidney, it is plausible that *UNC13B* plays a role in mediating renal complications of diabetes (36). Apoptosis of glomerular cells occurs quite early in the natural history of diabetic nephropathy, and it is now recognized that it could be an inciting event rather than a late consequence caused by increasing proteinuria (37–41). Interestingly, TT carriers of UNC13B_rs13293564 tended to be at slightly higher risk of nephropathy if they had not yet developed proliferative retinopathy, a clinical control for longstanding diabetes, emphasizing the role of *UNC13B* at the early stage of the disease. This is further strengthened by our observation that the frequency of the UNC13B_rs13293564 TT genotype in patients with incipient diabetic nephropathy, an early stage of the renal complication characterized by the presence of microalbuminuria, was significantly higher than the frequency of this genotype in control subjects. The similar distribution of the UNC13B_rs13293564 polymorphism genotype in three groups of patients with different stages of diabetic nephropathy also suggests that this polymorphism is implicated in the initiation rather than in the progression of the disease to more severe stages. However, because of the cross-sectional design of this study, we cannot exclude a survival bias in the ESRD group that would artificially modify the allele frequency patterns of susceptibility genes in this group.

The molecular analyses of the other 118 candidate genes identified no further variant associated with nephropathy. However, the SNPs genotyped in this study do not tag the haplotype architecture of these loci; rather, they tag a subset of the haplotypes across these genes based on the SNPs identified around exons, flanking intronic sequence,

TABLE 2
Genotype distribution derived from rs2281999 and rs13293564 according to diabetic nephropathy status in the whole EURAGEDIC study

Control subjects	rs13293564			Case subjects	rs13293564		
rs2281999	GG (36%)	GT (50%)	TT (14%)	rs2281999	GG (32%)	GT (48%)	TT (20%)
CC (43%)	112 (9%)	261 (22%)	147 (12%)	CC (50%)	94 (9%)	241 (22%)	201 (19%)
CT (46%)	219 (18%)	323 (27%)	17 (2%)	CT (42%)	177 (17%)	263 (24%)	11 (1%)
TT (11%)	115 (9%)	15 (1%)	1	TT (8%)	81 (7%)	10 (1%)	0

untranslated regions, and promoters. To fully characterize the haplotypic variability of these genes, as would tagging SNPs selected from the Hapmap project, the genotyping of ~1,300 SNPs would be required, while 344 have been typed in this study. Therefore, we cannot exclude that these candidate genes contribute to the development of diabetic nephropathy.

Our strategy, based on the genetic assessment of many candidate genes in a large case/control study for diabetic nephropathy, including material from three different European populations, provided evidence of replicated association between *UNC13B* and nephropathy and allowed us to establish for the first time the involvement of *UNC13B* variants in nephropathy. To assess whether or not these variants are clinically relevant to predict the initiation of nephropathy or the progression to more advanced stages of nephropathy will require further investigation in cohorts of patients with follow-up data.

URLs. HapMap: http://www.hapmap.org; UCSC genome browser: http://genome.ucsc.edu; NCBI: http://www.ncbi.nlm.nih.gov; Alibaba2.1, Transfac7.0: http://www.gene-regulation.com; and http://genecanvas.ecgene.net: genotypic data for all markers, phenotypic covariates from Danish and French samples, and a table describing all variations identified in the EURAGEDIC study.

ACKNOWLEDGMENTS

The EURAGEDIC study was made possible through funding from the European Commission (contract number QLG2-CT-2001-01669). The FinnDiane study was supported by grants from the Folkhälsan Research Foundation, Wilhelm and Else Stockmann Foundation, Liv och Hälsa Foundation, Sigrid Juselius Foundation, and the Finnish Medical Society (Finska Läkaresällskapet). D.G. holds a Wellcome Trust Senior Fellowship in Basic Biomedical Science (057733). This study was also supported by a grant from the EU-Project Network of Excellence, FP6-2005-LIFESCIHEALTH-6, Integrating Genomics, In-Genious HyperCare proposal number 037093 to S.-M.B-H.

We thank the French Diabetes Association for its support (bourse Association Française des Diabétiques Recherche 2000). We acknowledge A. Boland, V. Dumaz, D. Lechner, A. Lemainque, L. Sobre, J.-A. Perrier, C. Plançon (CNG), and C. Mascarenhas (U525) for technical assistance as well as A. Thévard (U525) for data management. All patients with type 1 diabetes are sincerely acknowledged. We also thank the Danish Diabetes Association and the Sehested Hansen Foundation for their continued support of this study. We acknowledge all the physicians and nurses at each center participating in the collection of patients (online appendix supplemental material). Brain and testis tissue RNA was a gift of Dr. J. Kremerskothen, Münster, and podocyte cDNA was a gift from Dr. T. Weide, University Hospital, Münster.

REFERENCES

1. Andersen AR, Christiansen JS, Andersen JK, Kreiner S, Deckert T: Diabetic nephropathy in type 1 (insulin-dependent) diabetes: an epidemiological study. *Diabetologia* 25:496–501, 1983
2. Borch-Johnsen K, Kreiner S: Proteinuria value as predictor of cardiovascular mortality in insulin dependent diabetes mellitus. *BMJ* 294:1651–1654, 1987
3. Seaquist ER, Goetz FC, Rich S, Barbosa J: Familial clustering of diabetic kidney disease: evidence for genetic susceptibility to diabetic nephropathy. *N Engl J Med* 320:1161–1165, 1989
4. Quinn M, Angelico MC, Warram JH, Krolewski AS: Familial factors determine the development of diabetic nephropathy in patients with IDDM. *Diabetologia* 39:940–945, 1996
5. Earle K, Walker J, Hill C, Viberti GC: Familial clustering of cardiovascular disease in patients with insulin-dependent diabetes and nephropathy. *N Engl J Med* 326:673–677, 1992
6. Diabetes Control and Complications Trial Research Group: Clustering of long-term complications in families with diabetes in the diabetes control and complications trial. *Diabetes* 46:1829–1839, 1997
7. Tarnow L, Groop PH, Hadjadj S, Kazeem G, Cambien F, Marre M, Forsblom C, Parving HH, Tregouet D, Thevard A, Farrall M, Gut I, Gauguier D, Cox R, Matsuda F, Lathrop M, Vionnet N, for the EURAGEDIC Consortium: European rational approach for the genetics of diabetic complications: EURAGEDIC: patient populations and strategy. *Nephrol Dial Transplant* 23:161–168, 2008
8. Thorn L, Forsblom C, Fagerudd J, Thomas MC, Pettersson-Fernholm K, Saraheimo M, Wadén J, Rönnback M, Rosengård-Bärlund M, af Björkesten C-G, Taskinen M-R, Groop P-H, on behalf of the FinnDiane Study Group: Metabolic syndrome in type 1 diabetes: association with diabetic nephropathy and glycaemic control (the FinnDiane Study). *Diabetes Care* 28:2019–2024, 2005
9. Vionnet N, Tregouët D, Kazeem G, Gut I, Groop PH, Tarnow L, Parving HH, Hadjadj S, Forsblom C, Farrall M, Gauguier D, Cox R, Matsuda F, Heath S, Thévard A, Rousseau R, Cambien F, Marre M, Lathrop M: Analysis of 14 candidate genes for diabetic nephropathy on chromosome 3q in European populations: strongest evidence for association with a variant in the promoter region of the adiponectin gene. *Diabetes* 55:3166–3174, 2006
10. Dizier MH, Bouzigon E, Guilloud-Bataille M, Bétard C, Bousquet J, Charpin D, Gormand F, Hochez J, Just J, Lemainque A, Le Moual N, Matran R, Neukirch F, Oryszczyn MP, Paty E, Pin I, Vervloet D, Kauffmann F, Lathrop M, Demenais F, Annesi-Maesano I: Genome screen in the French EGEA study: detection of linked regions shared or not shared by allergic rhinitis and asthma. *Genes Immun* 6:95–102, 2005
11. Takahashi M, Matsuda F, Margetic N, Lathrop M: Automated identification of single nucleotide polymorphisms from sequencing data. *J Bioinform Comput Biol* 1:253–265, 2003
12. Dempster AP, Laird NM, Rubin DB: Maximum likelihood from incomplete data via the EM algotrithm. *J R Stat Soc B* 39:1–38, 1977
13. Nakajima K, Tanaka Y, Nomiyama T, Ogihara T, Piao L, Sakai K, Onuma T, Kawamori R: Chemokine receptor genotype is associated with diabetic nephropathy in Japanese with type 2 diabetes. *Diabetes* 51:238–242, 2002
14. Janssen B, Hohenadel D, Brinkkoetter P, Peters V, Rind N, Fischer C, Rychlik I, Cerna M, Romzova M, de Heer E, Baelde H, Bakker SJ, Zirie M, Rondeau E, Mathieson P, Saleem MA, Meyer J, Köppel H, Sauerhoefer S, Bartram CR, Nawroth P, Hammes HP, Yard BA, Zschocke J, van der Woude FJ: Carnosine as a protective factor in diabetic nephropathy: association with a leucine repeat of the carnosinase gene CNDP1. *Diabetes* 54:2320–2327, 2005
15. Sale MM, Smith SG, Mychaleckyj JC, Keene KL, Langefeld CD, Leak TS, Hicks PJ, Bowden DW, Rich SS, Freedman BI: Variants of the transcription factor 7-like 2 (TCF7L2) gene are associated with type 2 diabetes in an African-American population enriched for nephropathy. *Diabetes* 56:2638–2642, 2007
16. Kanková K, Stejskalová A, Pácal L, Tschoplová S, Hertlová M, Krusová D, Izakovicová-Hollá L, Beránek M, Vaskù A, Barral S, Ott J: Genetic risk factors for diabetic nephropathy on chromosomes 6p and 7q identified by the set-association approach. *Diabetologia* 50:990–999, 2007
17. Pinizzotto M, Castillo E, Fiaux M, Temler E, Gaillard RC, Ruiz J: Paraoxonase2 polymorphisms are associated with nephropathy in type II diabetes. *Diabetologia* 44:104–107, 2001
18. Hager J, Hansen L, Vaisse C, Vionnet N, Philippi A, Poller W, Velho G, Carcassi C, Contu L, Julier C, Cambien F, Passa P, Lathrop M, Kindsvogel W, Demenais F, Nishimura E, Forguel P: A missense mutation in the glucagon receptor gene is associated with non-insulin-dependent diabetes mellitus. *Nat Genet* 9:299–304, 1995
19. Kaisaki PJ, Delépine M, Woon PY, Sebag-Montefiore L, Wilder SP, Menzel S, Vionnet N, Marion E, Riveline JP, Charpentier G, Schurmans S, Levy JC, Lathrop M, Farrall M, Gauguier D: Polymorphisms in type II SH2 domain-containing inositol 5-phosphatase (INPPL1, SHIP2) are associated with physiological abnormalities of the metabolic syndrome *Diabetes* 53:1900–1904, 2004
20. Ninio E, Tregouet D, Carrier JL, Stengel D, Bickel C, Perret C, Rupprecht HJ, Cambien F, Blankenberg S, Tiret L: Platelet-activating factor-acetylhydrolase and PAF-receptor gene haplotypes in relation to future cardiovascular event in patients with coronary artery disease. *Hum Mol Genet* 13:1341–1351, 2004
21. Devlin B, Roeder K: Genomic control for association studies. *Biometrics* 55:997–1004, 1999

22. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38:904–909, 2006
23. Sauer S, Lechner D, Berlin K, Plançon C, Heuermann A, Lehrach H, Gut IG: Full flexibility genotyping of single nucleotide polymorphisms by the GOOD assay. *Nucleic Acid Res* 28:E100, 2000
24. Fisher RA: *Statistical Methods for Research Workers*. 13th ed. London, Oliver & Loyd, 1925
25. Li J, Ji L: Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* 95:221–227, 2005
26. Barrett JC, Fry B, Maller J, Daly MJ: Haploview: analysis and visualization of LD and haplotypes maps. *Bioinformatics* 21:263–265, 2005
27. Tregouet DA, Garelle V: A new JAVA interface implementation of THESIAS: testing haplotype effects in association studies. *Bioinformatics* 23:1038–1039, 2007
28. Paul SR, Donner A: A comparison of tests of homogeneity of odds ratio in k 2x2 tables. *Stat Med* 8:1455–1468, 1989
29. Skryabin BV, Kremerskothen J, Vassilacopoulou D, Disotell TR, Kapitonov VV, Jurka J, Brosius J: The BC: 200 RNA gene and its neural expression are conserved in Anthropoidea (primates). *J Mol Evol* 47:677–685, 1998
30. Saleem MA, O'Hare MJ, Reiser J, Coward RJ, Inward CD, Farren T, Xing CY, Ni L, Mathieson PW, Mundel P: A conditionally immortalized human podocyte cell line demonstrating nephrin and podocin expression. *J Am Soc Nephrol* 13:630–638, 2002
31. Ewens KG, George RA, Sharma K, Ziyadeh FN, Spielman RS: Assessment of 115 candidate genes for diabetic nephropathy by transmission/disequilibrium test. *Diabetes* 54:3305–3318, 2005
32. Parle-McDermott A, Pangilinan F, Mills JL, Kirke PN, Gibney ER, Troendle J, O'Leary VB, Molloy AM, Conley M, Scott JM, Brody LC: The 19-bp deletion polymorphism in intron-1 of dihydrofolate reductase (DHFR) may decrease rather than increase risk for spina bifida in the Irish population. *Am J Med Genet A* 143:1174–1180, 2007
33. Greene E, Mahishi L, Entezam A, Kumari D, Usdin K: Repeat-induced epigenetic changes in intron 1 of the frataxin gene and its consequences in Friedreich ataxia. *Nucleic Acid Res* 35:3383–3390, 2007
34. Qu HQ, Montpetit A, Ge B, Hudson TJ, Polychronakos C: Toward further mapping of the association between the IL2RA locus and type 1 diabetes. *Diabetes* 56:1174–1176, 2007
35. Song Y, Ailenberg M, Silverman M: Cloning of a novel gene in the human kidney homologous to rat munc13s: its potential role in diabetic nephropathy. *Kidney Int* 53:1689–1695, 1998
36. Song Y, Ailenberg M, Silverman M: Human munc13 is a diacylglycerol receptor that induces apoptosis and may contribute to renal cell injury in hyperglycemia. *Mol Biol Cell* 10:1609–1619, 1999
37. Shankland SJ: The podocyte's response to injury: role in proteinuria and glomerulosclerosis. *Kidney Int* 69:2131–2147, 2006
38. Susztak K, Raff AC, Schiffer M, Bottinger EP: Glucose-induced reactive oxygen species cause apoptosis of podocytes and podocyte depletion at the onset of diabetic nephropathy. *Diabetes* 55:225–233, 2006
39. Meyer TW, Bennett PH, Nelson RG: Podocyte number predicts long-term urinary albumin excretion in Pima Indians with type II diabetes and microalbuminuria. *Diabetologia* 42:1341–1344, 1999
40. Pagtalunan ME, Miller PL, Jumping-Eagle S, Nelson RG, Myers BD, Rennke HG, Coplon NS, Sun L, Meyer TW: Podocyte loss and progressive glomerular injury in type II diabetes. *J Clin Invest* 99:342–348, 1997
41. Steffes MW, Schmidt D, McCrery R, Basgen JM, International Diabetic Nephropathy Study Group: Glomerular cell number in normal subjects and in type 1 diabetic patients. *Kidney Int* 59:2104–2113, 2001

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 AUGUST 21, 2008 VOL. 359 NO. 8

SLCO1B1 Variants and Statin-Induced Myopathy — A Genomewide Study

The SEARCH Collaborative Group*

ABSTRACT

BACKGROUND

Lowering low-density lipoprotein cholesterol with statin therapy results in substantial reductions in cardiovascular events, and larger reductions in cholesterol may produce larger benefits. In rare cases, myopathy occurs in association with statin therapy, especially when the statins are administered at higher doses and with certain other medications.

METHODS

We carried out a genomewide association study using approximately 300,000 markers (and additional fine-mapping) in 85 subjects with definite or incipient myopathy and 90 controls, all of whom were taking 80 mg of simvastatin daily as part of a trial involving 12,000 participants. Replication was tested in a trial of 40 mg of simvastatin daily involving 20,000 participants.

RESULTS

The genomewide scan yielded a single strong association of myopathy with the rs4363657 single-nucleotide polymorphism (SNP) located within *SLCO1B1* on chromosome 12 ($P=4\times10^{-9}$). *SLCO1B1* encodes the organic anion–transporting polypeptide OATP1B1, which has been shown to regulate the hepatic uptake of statins. The noncoding rs4363657 SNP was in nearly complete linkage disequilibrium with the nonsynonymous rs4149056 SNP ($r^2=0.97$), which has been linked to statin metabolism. The prevalence of the rs4149056 C allele in the population was 15%. The odds ratio for myopathy was 4.5 (95% confidence interval [CI], 2.6 to 7.7) per copy of the C allele, and 16.9 (95% CI, 4.7 to 61.1) in CC as compared with TT homozygotes. More than 60% of these myopathy cases could be attributed to the C variant. The association of rs4149056 with myopathy was replicated in the trial of 40 mg of simvastatin daily, which also showed an association between rs4149056 and the cholesterol-lowering effects of simvastatin. No SNPs in any other region were clearly associated with myopathy.

CONCLUSIONS

We have identified common variants in *SLCO1B1* that are strongly associated with an increased risk of statin-induced myopathy. Genotyping these variants may help to achieve the benefits of statin therapy more safely and effectively. (Current Controlled Trials number, ISRCTN74348595.)

Address reprint requests to the SEARCH Collaborative Group at the Clinical Trial Service Unit and Epidemiological Studies Unit, University of Oxford, Richard Doll Bldg., Old Road Campus, Roosevelt Dr., Oxford OX3 7LF, United Kingdom, or at search@ctsu.ox.ac.uk.

*The investigators and institutions participating in the Study of the Effectiveness of Additional Reductions in Cholesterol and Homocysteine (SEARCH) are listed in the Appendix and in the Supplementary Appendix, available with the full text of this article at www.nejm.org. The members of the writing committee (listed in the Appendix) assume responsibility for the overall content and integrity of the article.

This article (10.1056/NEJMoa0801936) was published at www.nejm.org on July 23, 2008.

N Engl J Med 2008;359:789-99.
Copyright © 2008 Massachusetts Medical Society.

Downloaded from www.nejm.org at KYOTO UNIV MED LIB on March 1, 2009 .
Copyright © 2008 Massachusetts Medical Society. All rights reserved.

EVIDENCE FROM LARGE-SCALE, RANDOMized studies shows that statin therapy reduces the incidence of heart attacks, strokes, and revascularization procedures by about one fifth for each reduction of 40 mg per deciliter (1 mmol per liter) in the low-density lipoprotein (LDL) cholesterol level.[1] In rare cases, statins can cause muscle pain or weakness in association with elevated creatine kinase levels (i.e., myopathy), and occasionally, this leads to muscle breakdown and myoglobin release (i.e., rhabdomyolysis), with a risk of renal failure and death.[2] The mechanisms by which statins cause myopathy remain unknown but appear to be related to statin concentrations in the blood. The incidence of myopathy is typically only about 1 case per 10,000 patients per year with standard doses of statins (e.g., 20 to 40 mg of simvastatin daily), but it increases with higher doses (e.g., 80 mg of simvastatin daily) and with concomitant use of certain drugs (e.g., cyclosporine, which can inhibit statin metabolism).[3,4] Although higher doses of statins may well result in larger reductions in the risk of vascular events,[1,5-9] large, long-term, randomized studies comparing different doses are needed to make a reliable assessment of the balance between efficacy and safety.

The ongoing Study of the Effectiveness of Additional Reductions in Cholesterol and Homocysteine (SEARCH), a randomized trial involving 12,064 participants with prior myocardial infarction, aims to determine whether a daily dose of 80 mg of simvastatin (Zocor, Merck) safely produces greater benefits than does a daily dose of 20 mg of simvastatin.[10] During an average follow-up of about 6 years among the 6031 participants who were assigned to receive 80 mg of simvastatin, there were 98 definite or incipient cases of myopathy; more than half occurred in the first year, and all of the patients had a full recovery. Interim analyses revealed a strong, previously unrecognized, association of myopathy with 80 mg of simvastatin daily and the concomitant use of amiodarone (relative risk of nearly 10). Consequently, participants who were taking amiodarone were given 20 mg of simvastatin daily (irrespective of their original assignment), and treatment with amiodarone is now contraindicated with higher doses of simvastatin.[11] We hypothesized that similarly strong associations might exist between myopathy with high-dose statin regimens and genetic variants, especially those affecting blood statin levels.

Previous studies have considered the relevance to myopathy of various candidate genes, such as *CYP3A4*, which is involved in the metabolism of certain statins,[3] genes encoding organic anion–transporting polypeptides,[12] some of which are associated with statin elimination, and genes involved in ubiquinone (coenzyme Q_{10}) deficiency.[13] Genetic associations with statin-induced myopathy, myalgia, or intolerance have been reported,[14,15] but none were statistically convincing, owing to the large numbers of candidate genes and single-nucleotide polymorphisms (SNPs) assessed.[13-17] Moreover, the apparent differences in the risk of myopathy in those studies may have been confounded by differences in statin regimens and concomitant use of other drugs.[4] The comparatively large number of cases of myopathy among patients who were taking a high dose of simvastatin in SEARCH and the inclusion of well-matched controls from the same population allowed us to conduct a genomewide association study with good power to detect genetic variants that have plausibly large effects.

METHODS

PARTICIPANTS AND SAMPLES IN SEARCH

Between September 1998 and October 2001, investigators in the SEARCH trial randomly assigned 12,064 participants from the United Kingdom who had had a myocardial infarction to receive either 80 mg or 20 mg of simvastatin daily.[10] Approval was obtained from the ethics committees of the participating institutions, and all participants gave written informed consent. At each follow-up assessment (at 2, 4, 8, and 12 months and then every 6 months), participants were questioned about new, unexplained muscle pain or weakness, and blood was drawn for measurements of creatine kinase and alanine aminotransferase levels at a central laboratory. By September 2006, "definite" myopathy (i.e., muscle symptoms, with creatine kinase levels that were more than 10 times the upper limit of the normal range) had developed in 49 of the 6031 participants who had been assigned to receive 80 mg of simvastatin. An additional 49 participants were considered to have "incipient" myopathy on the basis of their safety blood profile (a creatine kinase level that

Downloaded from www.nejm.org at KYOTO UNIV MED LIB on March 1, 2009 .
Copyright © 2008 Massachusetts Medical Society. All rights reserved.

was more than both 3 times the upper limit of the normal range and 5 times the baseline level, plus an alanine aminotransferase level that was more than 1.7 times the baseline value without an elevated alanine aminotransferase level alone at any other visit), irrespective of whether there were muscle symptoms. In contrast, only 2 definite and 6 incipient cases were identified among the 6033 participants who had been assigned to receive 20 mg of simvastatin.

The myopathy study was designed, analyzed, and interpreted jointly by members of the writing committee. The data were gathered as part of randomized trials designed and conducted by the Clinical Trial Service Unit, Oxford, United Kingdom, and genotyping was performed by the Centre National de Génotypage, Paris. The members of the writing committee drafted this report on behalf of the SEARCH Collaborative Group (see the Supplementary Appendix, available with the full text of this article at www.nejm.org) and take responsibility for the completeness, accuracy, and integrity of the data.

GENOTYPING AND SEQUENCING IN SEARCH

The genomewide association study was restricted to the 96 participants in whom myopathy (definite in 48 and incipient in 48) developed while they were taking 80 mg of simvastatin daily as part of SEARCH and for whom buffy-coat samples were available. Among the remaining participants who were assigned to receive 80 mg of simvastatin, 96 controls were selected who were matched with the case subjects with respect to sex, age, estimated glomerular filtration rate, and use or nonuse of amiodarone at baseline. Case subjects and controls were not known to be related; one case subject who had classified himself as having non-European ancestry was excluded. Multidimensional scaling of the matrix of genomewide identity-by-state distances was used to identify persons with potentially different ancestry or other outliers[18]; four participants appeared to cluster away from the remaining case subjects and controls (and a sensitivity analysis was performed with these participants excluded). The DNA concentration was measured by means of fluorescence staining (PicoGreen method, Invitrogen), and its quality examined by means of gel electrophoresis and polymerase-chain-reaction (PCR) amplification of two microsatellite markers; we obtained adequate DNA from 85 case subjects and 90 controls (see Table 1 in the Supplementary Appendix). The Sentrix HumanHap300-Duo BeadChip (Illumina), which contains 318,237 SNPs, was used,[19] and 316,184 SNPs (99.4%) passed data-quality checks (after exclusion of 1098 SNPs that were not successfully genotyped in any participant, 813 that were missing in more than 10% of the participants, 139 that were monomorphic in this population, and 3 that deviated from Hardy–Weinberg equilibrium among controls [$P<1.6\times10^{-7}$; $P<0.05$ with the Bonferroni correction]).

After the genomewide analysis, we resequenced exons within *SLCO1B1* in 83 case patients and 89 controls who had adequate DNA available and included an additional 38 genotyped and 141 imputed[20] variants (using HapMap CEU[21] as a reference population) with a nonzero minor-allele frequency in the case–control analysis. The Illumina panel does not cover variation in *CYP3A4*, which is a plausible candidate for statin-induced myopathy,[3] so we also resequenced this gene in 54 case subjects and 62 controls with adequate DNA remaining and included 20 genotyped and 11 imputed variants in the analysis. (Further details are available in Table 2 and the Methods section in the Supplementary Appendix.)

REPLICATION IN THE HEART PROTECTION STUDY

Between July 1994 and May 1997, a total of 20,536 patients in the United Kingdom with preexisting occlusive vascular disease or diabetes were randomly assigned to receive either 40 mg of simvastatin daily or placebo as part of the Heart Protection Study.[22] At each follow-up assessment (at 4, 8, and 12 months and then every 6 months), participants were questioned about any new, unexplained muscle pain or weakness, and blood was drawn for measurements of creatine kinase and alanine aminotransferase levels at a central laboratory. During an average follow-up of 5 years, 24 cases of myopathy (10 definite plus 14 incipient) were identified among the 10,269 participants who were assigned to receive 40 mg of simvastatin (with 23 of the cases identified while the participants were taking the study statin) versus 12 cases (4 definite plus 8 incipient) among the 10,267 who were assigned to receive placebo (with 3 of the cases identified while the participants were taking nonstudy statin). DNA was extracted from 19,856 participants (97%), and the rs4149056 and

Downloaded from www.nejm.org at KYOTO UNIV MED LIB on March 1, 2009 .
Copyright © 2008 Massachusetts Medical Society. All rights reserved.