◀ **Fig. 1** Location of single nucleotide polymorphisms (SNPs) in the *APOA1* (**a**), *CCL1* (**b**), *CCL2* (**c**), *ITGAM* (**d**), *ITGAX* (**e**), *ITGB7* (**f**), *LIPG* (**g**), *SCARB1* (**h**), *SELPLG* (**i**), *TGFB2* (**j**), *TGFB3* (**k**), *TGFBR2* (**l**), and *TGFBR3* (**m**) genes, indicated by *vertical lines*. Exons are indicated by a *solid rectangle*. The regions that have been sequenced are indicated by a *horizontal line*. The polymorphism numbers are accession numbers from the ThaiSNP database (correspondence to dbSNP rsIDs is given in Tables 2 and 3). The novel polymorphisms are indicated by an *asterisk*. The genomic sequences used for alignment are NT_033899 (*APOA1*), NT_010799 (*CCL1*), (*CCL2*), NT_010393 (*ITGAM and ITGAX*), NT_029419 (*ITGB7*), *NT*_010966 (*LIPG*), NT_009755 (*SCARB1*), NT_019546 (*SELPLG*), NT_021877 (*TGFB2*), *NT*_026437 (*TGFB3*), NT_032977 (*TGFBR2*), and NT_022517 (*TGFBR3*)

polymorphic in Thai, Caucasian, and African but not in Chinese and Japanese.

Linkage disequilibrium (LD) analysis and haplotype-block definition

LD statistics (D' or r^2) for the individual genotypes were calculated using the confidence intervals algorithm (Gabriel et al. 2002) implemented in the Haploview program for defining a haplotype block. To evaluate the effect the novel SNPs found in this study on the definition of haplotype blocks, we redefined the LD block in the Thai population using Thai SNPs excluding the novel SNPs (defined as known ThaiSNPs) and compared them with those from the combined Chinese–Japanese population. The populations were combined because the Japanese and Chinese populations were recently shown to be insignificantly different (The International HapMap 2005). Figure 3 shows haplotype-block definitions for the *ITGB7* gene using HapMap data, Thai population data with known SNPs, and Thai population data with novel SNPs. Both combined Chinese–Japanese and Thai population data had been defined with one block, with small differences in SNP members in the block. By introducing two novel SNPs from the Thai population to the SNP set and recalculation of the haplotype blocks, both novel SNPs appeared outside the original haplotype blocks. The haplotype block was then calculated for the rest of the 13 genes (data not shown); from this, 5.77% of novel SNPs were located within the LD block defined by the known Thai SNPs reported in dbSNP.

Tag SNP efficiency

Tag SNP efficiency was calculated by the number of tag SNPs from resequencing data and the number of tag SNPs from data verified by the NCBI. We identified that eight of 13 genes achieved 100% tag SNPs among discovered novel SNPs from the resequencing process, which means all SNPs newly discovered from the resequencing process were tagging SNPs (Table 4). In contrast, only one gene, *CCL2*, did not benefit from resequencing, because no novel SNPs were found. *ITGAM*, *TGFBR3*, *ITGAX*, and *LIPG* each had novel tag SNPs identified when the newly discovered SNPs were included. Our defined parameter, tag SNP efficiency (see "Materials and methods") of *ITGAM*, *TGFBR3*, *ITGAX*, and *LIPG* were 75.00%, 66.76%, 64.29%, and 50.00%, respectively. This suggested that a high percentage of novel SNPs define new tags. However, the overall tag SNP efficiency for all genes was 81.23%.

Functional polymorphism assessment

To assess the impact of amino acid substitutions on protein activity, we analyzed 16 nsSNPs, including eight novel nsSNPs, using three nsSNP functional prediction tools that utilize different algorithms: SIFT, PolyPhen, and SNPs3D (Table 5). Eleven nsSNPs were concordantly predicted to be intolerant and seven predicted to be neutral with these three functional prediction tools, whereas the remaining five were predicted to have a mixture of neutral and damaging activity. Allele frequencies were also used to classify the potential effects of nsSNPs. Interestingly, eight out of 16 nsSNPs were common SNPs (MAF > 5%). SNP rs2230429 located on *ITGAX* had the highest MAF observed (0.5) and was predicted to be damaging by all these tools.

Discussion

Cardiovascular disease is a complex disease that is influenced by many factors, including genetics. Candidate-gene-based association studies are the most common approach used in disease-causing gene identification research. Choosing markers for association approaches is based on extensive information on the distribution of SNPs across the genome. To obtain such information, 13 candidate genes, which had been associated to atherosclerosis, were resequenced in exon-flanking regions in the Thai population. We decreased all the possible known errors by using only high-quality chromatograms for the analysis. The sequencing data was obtained from both strands with difference primers. More than 80% of SNPs found were associated with both strands. We identified 59 novel polymorphisms (30%) by comparing them with dbSNP build 126. The percentage of novel polymorphisms found in this study was quite similar to the other SNP discovery studies (Michiels et al. 2007), but the allele frequencies were observed to have a high number of rare alleles; as many as 45 (76%) of those novel SNPs were rare alleles.

Table 2 Summary of 61 novel genetic variants identified in the 13 cardiovascular-related genes

Gene (contig ID)	ThaiSNP ID	Position in contig	Allele	Frequency	AA change	Type	Gene (contig ID)	ThaiSNP ID	Position in contig	Allele	Frequency	AA change	Type
APOA1	th49	20270152	G:A	0.969:0.031	A/T	nsSNP	ITGB7	th1673	15730710	_:C	0.984:0.016	–	Intron
(NT_033899)	th47	20270625	CTC:_	0.984:0.016	–	UTR	(NT_029419)	th1670	15730869	T:C	0.969:0.031	N/N	synSNP
CCL1	th197	7425784	T:G	0.969:0.031	M/R	nsSNP		th1663	15738568	T:C	0.955:0.046	–	Locus
(NT_010799)													
ITGAM	th1612	22602325	G:A	0.984:0.016	–	Intron	SELPLG	th1720	32499354	T:C	0.922:0.078	Y/H	nsSNP
(NT_010393)	th1613	22646387	T:C	0.984:0.016	–	Intron	(NT_019546)	th1719	32499604	C:T	0.984:0.016	T/T	synSNP
	th1614	22648985	T:C	0.922:0.078	–	Intron		th1718	32499840	AACCAGTGCCCACGGA GGCACAGACCACTC:_	0.712:0.289	–	Intron
	th1621	22653328	A:T	0.659:0.341	–	Intron	SCARB1	th281	2694784	G:A	0.984:0.016	–	Intron
	th1622	22653447	A:G	0.563:0.438	–	Intron	(NT_009755)	th279	2702586	G:A	0.938:0.063	–	Intron
	th1623	22653599	C:T	0.969:0.031	–	Intron		th277	2704352	C:A	0.938:0.063	–	Intron
	th1624	22653688	G:T	0.984:0.016	M/I	nsSNP		th275	2704961	G:A	0.955:0.046	–	Intron
	th1626	22654062	A:G	0.977:0.023	–	Intron		th274	2706013	T:C	0.865:0.135	–	Intron
	th1629	22655695	C:G	0.984:0.016	–	Intron		th271	2709346	G:A	0.938:0.063	–	Intron
	th1633	22656933	T:C	0.984:0.016	–	UTR	LIPG	th243	28577675	C:G	0.969:0.031	–	UTR
	th1634	22657040	_:TTTAC	0.953:0.047	–	UTR	(NT_010966)	th245	28577928	C:T	0.641:0.359	–	Intron
	th1635	22657150	C:T	0.969:0.031	–	UTR		th246	28577932	C:G	0.641:0.359	–	Intron
	th1636	22657362	G:A	0.923:0.077	–	Locus		th247	28580851	C:T	0.981:0.019	R/C	nsSNP
ITGAX	th1638	22679349	A:T	0.984:0.016	–	Locus		th248	28580992	T:C	0.981:0.019	–	Intron
(NT_010393)	th1639	22681584	G:A	0.984:0.016	–	Intron		th252	28590899	G:A	0.967:0.033	P/P	synSNP
	th1640	22681658	C:T	0.984:0.016	C/C	synSNP		th259	28606442	C:T	0.984:0.016	–	UTR
	th1646	22695534	A:G	0.984:0.016	A/A	synSNP		th260	28606473	C:T	0.984:0.016	–	UTR
	th1647	22695594	G:A	0.984:0.016	R/R	synSNP		th266	28607810	A:_	0.983:0.017	–	UTR
	th1648	22695702	G:A	0.984:0.016	–	Intron	TGFB2	th1725	12077374	_:A	0.859:0.141	–	Intron
	th1649	22696175	G:C	0.983:0.017	L/L	synSNP	(NT_021877)	th1726	12077946	G:A	0.969:0.031	–	UTR
	th1650	22696182	C:G	0.983:0.017	P/A	nsSNP	TGFB3	th1736	57424151	T:C	0.984:0.016	–	Locus
	th1651	22696570	G:C	0.984:0.016	–	Intron	(NT_026437)	th1735	57424782	AGAC:_	0.984:0.016	–	UTR
	th1652	22696601	G:A	0.817:0.183	–	Intron		th1732	57431803	C:G	0.984:0.016	–	Intron
	th1654	22697835	C:T	0.933:0.067	–	Intron	TGFBR3	th1781	46001605	T:C	0.984:0.016	Y/Y	synSNP
	th1655	22705064	G:A	0.981:0.019	–	Intron	(NT_032977)	th1779	46004452	T:A	0.982:0.018	–	Intron
	th1659	22706509	G:A	0.95:0.05	–	UTR		th1776	46012430	C:T	0.969:0.031	–	Intron
	th1661	22707107	T:C	0.953:0.047	–	UTR		th1770	46019870	T:C	0.984:0.016	F/L	nsSNP
							TGFBR2 (NT_022517)	th1752	30653256	C:T	0.969:0.031	R/W	nsSNP

UTR untranslated region, *synSNP* synonymous single nucleotide polymorphism, *nsSNP* nonsynonymous single nucleotide polymorphism

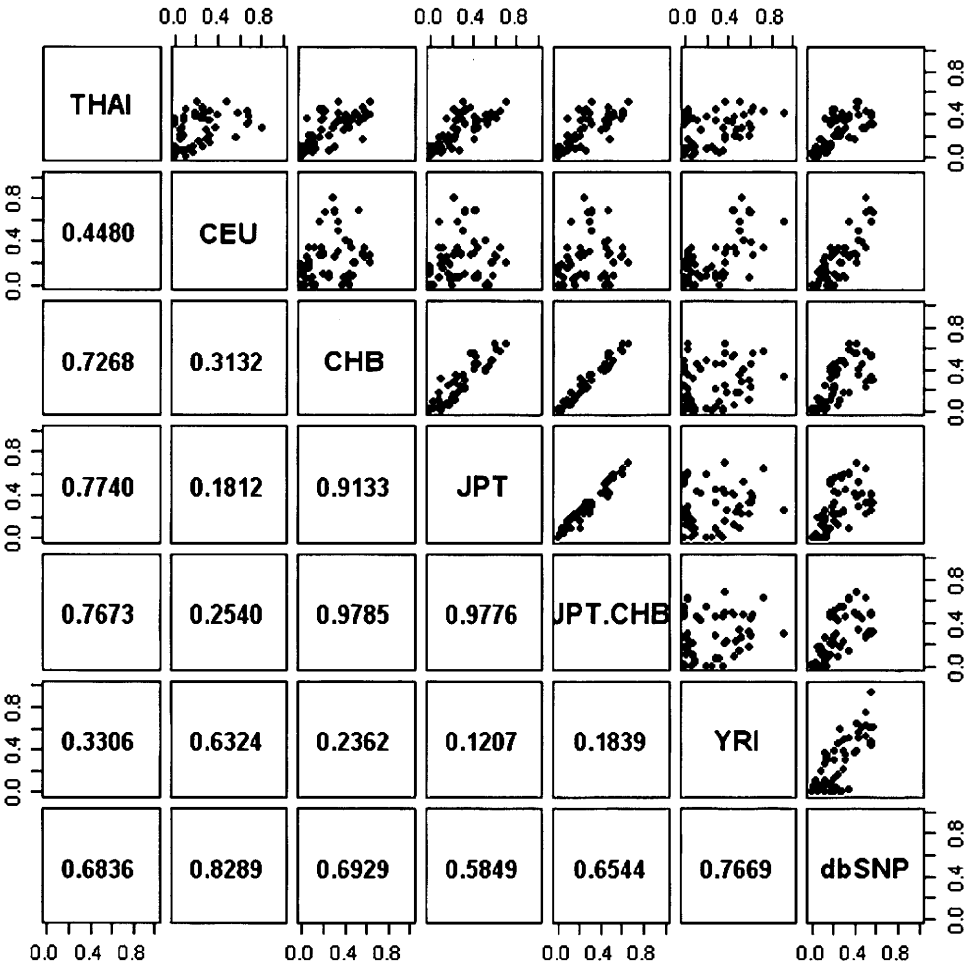

Fig. 2 Scatter plots of pairwise comparison of allele frequency distribution between Thai, Caucasian, Chinese, Japanese, Chinese + Japanese, African, and Single Nucleotide Polymorphism Database (dbSNP). Pearson's product moment correlation of allele frequency between the two populations is presented in the *lower diagonal matrix*

Table 3 Ethnic comparison of single nucleotide polymorphism (SNP) allele frequencies in the 13 cardiovascular-related genes

Gene	SNP(rs no.)	Type	Allele	Thai[a]	Hapmap[b]				Ethnic difference[c]
					Caucasian	Han Chinese	Japanese	African	
CCL1	rs2282691	Intron	A:T	0.375:0.625	0.404:0.596	0.405:0.595	0.523:0.477	0.551:0.449	0.176
CCL2	rs4586	synSNP	T:C	0.583:0.417	0.667:0.333	0.422:0.578	0.352:0.648	0.258:0.742	0.409
ITGAM	rs1143678	nsSNP	C:T	0.933:0.067	0.867:0.133	0.989:0.011	1.000:0.000	0.8:0.2	0.2
	rs3815801	Intron	A:G	0.672:0.328	0.667:0.333	0.822:0.178	0.739:0.261	0.608:0.392	0.214
	rs4077810	Intron	C:T	0.568:0.432	0.737:0.263	0.841:0.159	0.775:0.225	0.966:0.034	0.398
	rs4597342	UTR	C:T	0.607:0.393	0.724:0.276	0.8:0.2	0.705:0.295	0.967:0.033	0.36
	rs7184677	Intron	G:A	0.938:0.062	0.892:0.108	0.989:0.011	1.000:0.000	0.642:0.358	0.358
	rs7206295	Intron	C:T	0.65:0.35	0.717:0.283	0.8:0.2	0.705:0.295	0.967:0.033	0.317
	rs9936831	Intron	A:T	0.95:0.05	0.9:0.1	1.000:0.000	1.000:0.000	0.636:0.364	0.364
ITGAX	rs11150620	Intron	G:C	0.654:0.346	0.75:0.25	0.42:0.58	0.386:0.614	0.975:0.025	0.589
	rs1140195	UTR	A:G	0.609:0.391	0.737:0.263	0.367:0.633	0.398:0.602	0.975:0.025	0.608
	rs11574635	Intron	G:C	0.917:0.083	0.828:0.172	1.000:0.000	1.000:0.000	0.737:0.263	0.263
	rs2929	UTR	G:A	0.75:0.25	0.712:0.288	0.822:0.178	0.744:0.256	0.517:0.483	0.305
	rs4264407	Intron	G:C	1.000:0.000	0.892:0.108	1.000:0.000	1.000:0.000	0.933:0.067	0.108
ITGB7	rs11170465	Intron	G:A	0.839:0.161	0.949:0.051	0.92:0.08	0.802:0.198	0.975:0.025	0.173
	rs11170466	Intron	G:A	0.839:0.161	0.942:0.058	0.898:0.102	0.778:0.222	0.975:0.025	0.197
	rs11574541	synSNP	C:T	0.969:0.031	1.000:0.000	1.000:0.000	1.000:0.000	1.000:0.000	0.031

Table 3 continued

Gene	SNP(rs no.)	Type	Allele	Thai[a]	Hapmap[b]				Ethnic difference[c]
					Caucasian	Han Chinese	Japanese	African	
	rs2272299	Intron	G:A	0.85:0.15	N/A	0.9:0.1	0.761:0.239	0.884:0.116	0.139
	rs2272300	Intron	T:G	0.812:0.188	0.942:0.058	0.9:0.1	0.761:0.239	0.405:0.595	0.537
	rs2272301	Intron	C:G	0.95:0.05	0.851:0.149	0.942:0.058	0.872:0.128	0.982:0.018	0.131
	rs3817537	Intron	G:C	0.984:0.016	1.000:0.000	0.978:0.022	1.000:0.000	N/A	0.022
	rs3825084	Intron	A:C	0.906:0.094	0.821:0.179	0.94:0.06	0.805:0.195	0.949:0.051	0.144
LIPG	rs2000812	Intron	T:C	0.633:0.367	0.8:0.2	0.522:0.478	0.42:0.58	1.000:0.000	0.58
	rs2000813	nsSNP	C:T	0.667:0.333	0.692:0.308	0.656:0.344	0.761:0.239	0.966:0.034	0.31
	rs2276269	Intron	T:C	0.6:0.4	0.417:0.583	0.667:0.333	0.738:0.262	0.067:0.933	0.671
	rs3744843	UTR	A:G	0.922:0.078	N/A	0.744:0.256	0.726:0.274	0.667:0.333	0.255
	rs3786247	UTR	A:C	0.732:0.268	0.931:0.069	0.557:0.443	0.545:0.455	0.642:0.358	0.386
	rs3786248	UTR	A:G	0.938:0.062	0.933:0.067	0.756:0.244	0.727:0.273	1.000:0.000	0.273
	rs3819166	Intron	G:A	0.6:0.4	0.808:0.192	0.511:0.489	0.432:0.568	1.000:0.000	0.568
	rs3826577	UTR	A:T	0.938:0.062	0.933:0.067	0.756:0.244	0.727:0.273	1.000:0.000	0.273
	rs6507931	Intron	C:T	0.833:0.167	0.425:0.575	0.833:0.167	0.909:0.091	0.492:0.508	0.484
	rs9958734	UTR	T:C	0.75:0.25	0.946:0.054	0.605:0.395	0.583:0.417	0.839:0.161	0.363
SCARB1	rs11057824	Intron	C:T	0.667:0.333	1.000:0.000	0.622:0.378	0.486:0.514	1.000:0.000	0.514
	rs11057825	Intron	C:T	0.646:0.354	1.000:0.000	0.567:0.433	0.444:0.556	1.000:0.000	0.556
	rs3825140	UTR	C:T	0.783:0.217	1.000:0.000	0.622:0.378	0.524:0.476	N/A	0.476
	rs4765615	Intron	C:T	0.5:0.5	0.509:0.491	0.655:0.345	0.697:0.303	0.491:0.509	0.206
	rs5889	synSNP	C:T	0.708:0.292	0.992:0.008	0.58:0.42	0.464:0.536	1.000:0.000	0.536
	rs5892	synSNP	C:T	0.906:0.094	1.000:0.000	0.977:0.023	0.965:0.035	0.892:0.108	0.108
SELPLG	rs2228315	nsSNP	G:A	0.808:0.192	0.908:0.092	0.756:0.244	0.83:0.17	0.617:0.383	0.291
TGFB2	rs900	UTR	A:T	0.281:0.719	N/A	N/A	N/A	0.667:0.333	0.386
	rs10482823	Intron	T:C	0.984:0.016	1.000:0.000	0.989:0.011	0.989:0.011	1.000:0.000	0.016
	rs6684205	Intron	A:G	0.266:0.734	0.808:0.192	0.278:0.722	0.227:0.773	0.525:0.475	0.581
TGFB3	rs3917147	Intron	T:C	0.969:0.031	1.000:0.000	0.911:0.089	0.966:0.034	0.667:0.333	0.333
	rs3917187	Intron	G:A	0.55:0.45	0.731:0.269	0.44:0.56	0.616:0.384	0.357:0.643	0.374
	rs3917200	Intron	T:C	0.95:0.05	0.944:0.056	0.966:0.034	0.911:0.089	0.708:0.292	0.258
	rs3917201	Intron	A:G	0.839:0.161	0.708:0.292	0.444:0.556	0.58:0.42	0.625:0.375	0.395
TGFBR2	rs1155705	Intron	A:G	0.297:0.703	0.683:0.317	0.3:0.7	0.33:0.67	0.608:0.392	0.386
	rs1155708	Intron	G:A	0.297:0.703	0.675:0.325	0.3:0.7	0.33:0.67	0.608:0.392	0.378
	rs2276767	Intron	C:A	0.875:0.125	0.667:0.333	0.889:0.111	0.911:0.089	0.942:0.058	0.275
	rs2276768	Intron	C:T	0.567:0.433	0.898:0.102	0.8:0.2	0.67:0.33	0.707:0.293	0.331
	rs9843942	Intron	G:A	0.734:0.266	0.616:0.384	0.557:0.443	0.583:0.417	0.382:0.618	0.352
TGFBR3	rs10874913	Intron	C:T	0.969:0.031	1.000:0.000	1.000:0.000	1.000:0.000	N/A	0.031
	rs11165376	Intron	A:G	0.683:0.317	0.31:0.69	0.477:0.523	0.573:0.427	0.563:0.438	0.373
	rs11165377	Intron	C:T	0.906:0.094	0.742:0.258	0.689:0.311	0.9:0.1	0.92:0.08	0.231
	rs11165378	Intron	T:C	1.000:0.000	1.000:0.000	1.000:0.000	1.000:0.000	0.942:0.058	0.058
	rs12069176	Intron	A:G	0.367:0.633	0.686:0.314	0.533:0.467	0.411:0.589	0.467:0.533	0.319
	rs1805109	UTR	G:A	0.661:0.339	0.915:0.085	N/A	N/A	0.936:0.064	0.275
	rs1805110	nsSNP	C:T	0.661:0.339	0.907:0.093	0.551:0.449	0.557:0.443	0.877:0.123	0.356
	rs1805117	UTR	A:G	0.859:0.141	0.78:0.22	0.936:0.064	0.841:0.159	0.956:0.044	0.176
	rs2279455	Intron	T:C	0.589:0.411	0.333:0.667	0.778:0.222	0.67:0.33	0.381:0.619	0.445
	rs2296621	Intron	C:A	0.659:0.341	0.75:0.25	0.956:0.044	0.841:0.159	0.911:0.089	0.297
	rs2306886	Intron	G:A	0.733:0.267	N/A	0.644:0.356	0.789:0.211	N/A	0.145
	rs2306887	Intron	C:T	0.714:0.286	N/A	0.676:0.324	0.667:0.333	N/A	0.047
	rs2306888	synSNP	T:C	0.933:0.067	1.000:0.000	0.878:0.122	0.795:0.205	1.000:0.000	0.205

Table 3 continued

Gene	SNP(rs no.)	Type	Allele	Thai[a]	Hapmap[b]				Ethnic difference[c]
					Caucasian	Han Chinese	Japanese	African	
	rs284176	Intron	G:A	0.667:0.333	0.664:0.336	0.544:0.456	0.545:0.455	0.703:0.297	0.159
	rs3738441	Intron	C:T	0.5:0.5	0.792:0.208	0.367:0.633	0.3:0.7	0.619:0.381	0.492
	rs6696224	Intron	A:G	0.633:0.367	0.917:0.083	0.522:0.478	0.411:0.589	0.792:0.208	0.506
	rs6699304	Intron	C:T	0.969:0.031	0.825:0.175	1.000:0.000	0.9:0.1	0.9:0.1	0.175
	rs7524066	Intron	G:T	0.812:0.188	0.664:0.336	0.944:0.056	0.878:0.122	0.542:0.458	0.402

[a] Allele frequency of the Thai population was determined by direct sequencing of DNA-pooled two samples selected from 32 unrelated Thai

[b] Allele frequencies were determined from data obtained by searching the Hapmap database (http://www.hapmap.org). SNP of some genes are not included in this table because of lacked of information in the Hapmap database

[c] Ethnic differences in allele frequency were calculated by subtracting the lowest allele frequency of the minor allele from the highest allele frequency of the minor allele among the ethnic groups for each SNP site

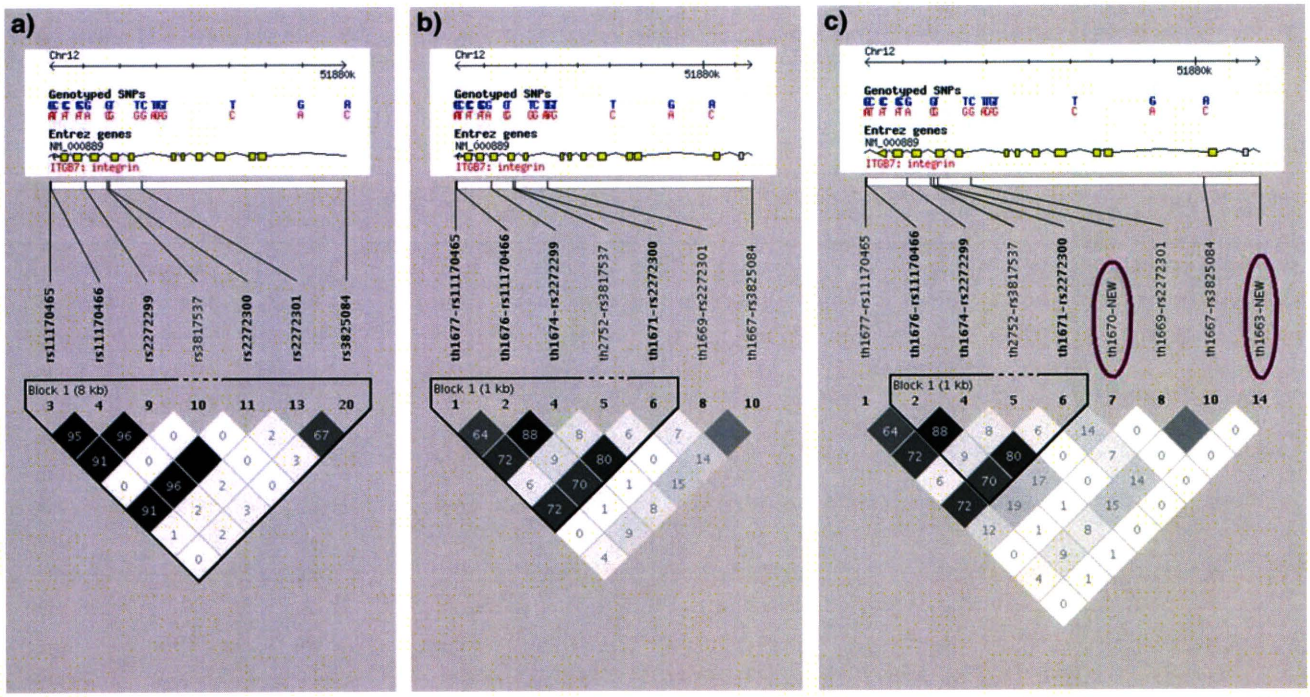

Fig. 3 The haplotype-block definition of the *ITGB7* gene comparing Japanese and Chinese HapMap data (**a**), Thai population without novel single nucleotide polymorphism (SNP) data (**b**), and Thai population with novel SNPs (**c**) using confidence intervals (Gabriel et al. 2002) in the Haploview software. SNP locations linked to the physical map on the chromosome are shown on the *white rectangle*. The novel SNPs are marked by *red ovals*

However, these novel SNPs are still important to consider for high-resolution association study design.

The ethnic differences between these SNPs could be responsible for differences in gene regulation and differences in the prevalence of diseases among these ethnic groups. Because of this, allele frequencies in the Thai population were compared with Chinese, Japanese, Caucasian, African, and average allele frequencies in dbSNP. Not surprisingly, the results showed that the allele frequency distribution of the Thai population was more correlated to other Asian populations, Chinese and Japanese, than to Caucasian and African populations. Correlation coefficients were similar to other recent studies (Cha et al. 2004; Kim et al. 2005; Mahasirimongkol et al. 2006). When compared with a similar study performed on the Korean population (Kim et al. 2005), the allele frequency of Korean populations was very similar to that of the Japanese population (correlation coefficient $r = 0.907$), whereas it had very different patterns of allele frequency compared with Caucasian (correlation coefficient $r = 0.359$) or African (correlation coefficient $r = 0.156$) populations. When focusing only on the correlation

Table 4 Number of tag single nucleotide polymorphisms (SNPs) selected from resequencing data, number of tag SNPs selected from data verified by the National Center for Biotechnology Information (NCBI) and percentage of tag SNPs efficiency

Gene	No. of tag SNPs from resequencing data (N_{RT})	No. of total SNPs from resequencing process (N_R)	No. of tag SNPs from data verified by NCBI (N_{DT})	No. of total SNPs validated by NCBI (N_D)	Percentage of tag SNP efficiency (%)
APOA1	8	8	7	7	100.00
CCL1	4	4	3	3	100.00
CCL2	4	4	4	4	0.00
ITGAM	20	31	11	19	75.00
ITGAX	18	23	9	9	64.29
ITGB7	9	14	7	12	100.00
LIPG	18	24	15	18	50.00
SCARB1	17	17	11	11	100.00
SELPLG	3	3	1	1	100.00
TGFB2	4	5	3	4	100.00
TGFB3	6	6	4	4	100.00
TGFBR2	7	10	6	9	100.00
TGFBR3	19	26	17	23	66.67
TOTAL	137	175	98	124	81.23

Table 5 Assessment of the 16 nonsynonymous single nucleotide polymorphisms (SNPs) in the 13 cardiovascular-related genes using SIFT, PolyPhen, and SNP3D

Gene (protein ID)	ThaiSNP ID	NCBI SNP ID	Frequency	AA variant	SIFT score*	SIFT prediction	PolyPhen prediction	SNPs3D prediction
APOA1 (NP_000030)	th49	New	G[0.969]/A[0.031]	A61T	1.00	Tolerant	Benign	Neutral
CCL1 (NP_002972)	th197	New	T[0.969]/G[0.031]	I63R	0.00	Intol	PRB	NA
ITGAM (NP_000623)	th1604	rs1143679	G[0.969]/A[0.031]	R77H	0.47	Tolerant	Benign	Neutral
	th1617	rs7201448	C[0.933]/T[0.067]	A858V	0.19	Tolerant	Benign	Neutral
	th1624	New	G[0.984]/T[0.016]	M951I	0.38	Tolerant	Benign	Neutral
	th1630	rs1143678	C[0.933]/T[0.067]	P1146S	0.49	Tolerant	Benign	Neutral
ITGAX (NP_000878)	th1643	rs12928508	G[0.923]/A[0.077]	A251T	1.00	Tolerant	Benign	Neutral
	th1645	rs2230429	C[0.5]/G[0.5]	P517R	0.00	Intol	PRB	Damaging
	th1650	New	C[0.983]/G[0.017]	P720A	0.00	Intol	PRB	Damaging
SELPLG (NP_002997)	th1720	New	T[0.922]/C[0.078]	Y297H	0.25	Tolerant	Benign	Neutral
	th1717	rs2228315	G[0.808]/A[0.192]	M62I	0.00	Intol	Benign	Neutral
LIPG (NP_006024)	th247	New	C[0.981]/T[0.019]	R54C	0.00	Intol	PRB	Damaging
	th250	rs2000813	C[0.667]/T[0.333]	T111I	0.00	Intol	Benign	Neutral
TGFBR2 (NP_003233)	th1752	New	C[0.969]/T[0.031]	R193W	0.00	Intol	PRB	Neutral
TGFBR3 (NP_003234)	th1770	New	T[0.984]/C[0.016]	F142L	0.71	Tolerant	POS	NA
	th1758	rs1805110	C[0.661]/T[0.339]	S15F	0.03	Intol	Benign	NA

NCBI National Center for Biotechnology Information, *AA* amino acid, *Intol* Intolerant, *POS* possibly damaging, *PRB* probably damaging, *NA* no data or SNP analysis available

* TI scores ≤0.05 are predicted to be Intolerant, whereas TI scores >0.05 are tolerant variants

coefficient (r) among the Asian population, the Thai population and combined Chinese–Japanese frequencies had a higher correlation than between the Thai population and Chinese or Japanese populations. These results support the findings of the international HapMap consortium that the Chinese and Japanese populations are insignificantly different (The International HapMap 2005). Consistent with their population histories, the admixture event between Thai and Chinese is believed to have occurred quite some time ago. Although the allele frequency distribution of the Thai population was similar to the other Asian populations, there are some rare allele markers (rs11574541 and rs10874913) found only in the Thai population but not in other Asian populations.

For LD-block comparison, most novel SNPs were located out of the block defined by known SNPs. Some of them were formed a new LD block, but there were no LD-block formations of mostly novel SNPs. The number of

blocks might reflect population age, suggesting the greater LD block is the older population. We found a lower number of blocks than in other Asian populations, in which the referred age of the Thai population was younger than other Asian populations. Only 5.77% of the novel SNPs appear in the defined haplotype block. The LD blocks might be affected by the lower allele frequencies observed for the novel SNPs found in our study. These data indicate that using published SNP data alone would not have adequate coverage of the target region associated with disease. SNP discovery approaches can help identify causative SNPs, which were not reported in the public SNP databases.

When the number of tag SNPs of these two groups was compared, the average tag SNP efficiency (see "Materials and methods") was 81.23. Eight of 13 genes showed 100% tag SNP efficiency, that is, each of the newly discovered SNPs were also defined as tag SNPs. Therefore, additional SNP discovery is needed to assemble a map for use in the Thai population. The tag SNP results showed concordance to the LD-block results. Consequently, our data agree with the study of Carlson et al. (2003), who suggested that the study of populations other than European Americans required additional SNP discovery before conclusions can be drawn as to the adequacy of dbSNP for each population.

Additionally, SNP location is another factor that influences the selection of prominent SNPs for further association studies. Using three commonly known algorithms for protein function damaging SNP prediction, several novel SNPs located in coding regions were found to likely affect the alteration of protein function. This hypothesis should be investigated further using experimental functional assays to determine their corresponding effects.

In summary, resequencing is a powerful method to discover novel SNPs and SNPs that are specific to certain ethnic groups. This approach could also reveal the distribution of SNPs along interesting genes, which will be useful for future association studies. Consequently, this SNP discovery project provided sufficient information for marker selection used in case-control association studies utilizing candidate gene approaches.

Acknowledgments We thank Justin Dantzer for designing figures used in this article and Lang Li and Jeesun Jung for helpful advice. CT received the postdoctoral fellowship from the National Center for Genetic Engineering and Biotechnology, Thailand, under the Thailand SNP Discovery Project. SP is funded by Thailand Center for Excellence in Life Sciences (TCELS) under the Postdoctoral Scholarships Program. SDM is funded by NIH K22LM009135.

References

Asztalos BF (2004) High-density lipoprotein metabolism and progression of atherosclerosis: new insights from the HDL Atherosclerosis Treatment Study. Curr Opin Cardiol 19:385–391

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265

Brown CM, Rea TJ, Hamon SC, Hixson JE, Boerwinkle E, Clark AG, Sing CF (2006) The contribution of individual and pairwise combinations of SNPs in the APOA1 and APOC3 genes to interindividual HDL-C variability. J Mol Med V84:561–572

Carlson CS, Eberle MA, Rieder MJ, Smith JD, Kruglyak L, Nickerson DA (2003) Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. Nat Genet 33:518–521

Cha PC, Yamada R, Sekine A, Nakamura Y, Koh CL (2004) Inference from the relationships between linkage disequilibrium and allele frequency distributions of 240 candidate SNPs in 109 drug-related genes in four Asian populations. J Hum Genet 49:558–572

de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D (2005) Efficiency and power in genetic association studies. Nat Genet 37:1217–1223

Department of Medical Service Thailand G (2006a) Statistical report. (http://www.dms.moph.go.th/statreport/index.html)

Department of Medical Service Thailand G (2006b) Statistical report of cardiovascular disease in Thai population. (http://www.dms.moph.go.th/statreport/2547/table0647.htm)

Disabella E, Grasso M, Marziliano N, Ansaldi S, Lucchelli C, Porcu E, Tagliani M, Pilotto A, Diegoli M, Lanzarini L, Malattia C, Pelliccia A, Ficcadenti A, Gabrielli O, Arbustini E (2006) Two novel and one known mutation of the TGFBR2 gene in Marfan syndrome not associated with FBN1 gene defects. Eur J Hum Genet 14:34–38

Do H, Vasilescu A, Diop G, Hirtzig T, Coulonges C, Labib T, Heath SC, Spadoni JL, Therwath A, Lathrop M, Matsuda F, Zagury JF (2006) Associations of the IL2Ralpha, IL4Ralpha, IL10Ralpha, and IFN (gamma) R1 cytokine receptor genes with AIDS progression in a French AIDS cohort. Immunogenetics 58:89–98

Fuki IV, Blanchard N, Jin W, Marchadier DH, Millar JS, Glick JM, Rader DJ (2003) Endogenously produced endothelial lipase enhances binding and cellular processing of plasma lipoproteins via heparan sulfate proteoglycan-mediated pathway. J Biol Chem 278:34331–34338

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. Science 296:2225–2229

Hoh J, Matsuda F, Peng X, Markovic D, Lathrop MG, Ott J (2003) SNP haplotype tagging from DNA pools of two individuals. BMC Bioinform 4:14

Ishida T, Choi SY, Kundu RK, Spin J, Yamashita T, Hirata K, Kojima Y, Yokoyama M, Cooper AD, Quertermous T (2004) Endothelial lipase modulates susceptibility to atherosclerosis in apolipoprotein-E-deficient mice. J Biol Chem 279:45085–45092

Jaye M, Lynch KJ, Krawiec J, Marchadier D, Maugeais C, Doan K, South V, Amin D, Perrone M, Rader DJ (1999) A novel endothelial-derived lipase that modulates HDL metabolism. Nat Genet 21:424–428

Kim JY, Moon SM, Ryu HJ, Kim JJ, Kim HT, Park C, Kimm K, Oh B, Lee JK (2005) Identification of regulatory polymorphisms in the TNF-TNF receptor superfamily. Immunogenetics 57:297–303

Laukkanen J, Yla-Herttuala S (2002) Genes involved in atherosclerosis. Exp Nephrol 10:150–163

Maglott D, Ostell J, Pruitt KD, Tatusova T (2005) Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res 33:D54–D58

Mahasirimongkol S, Chantratita W, Promso S, Pasomsab E, Jinawath N, Jongjaroenprasert W, Lulitanond V, Krittayapoositpot P,

Tongsima S, Sawanpanyalert P, Kamatani N, Nakamura Y, Sura T (2006) Similarity of the allele frequency and linkage disequilibrium pattern of single nucleotide polymorphisms in drug-related gene loci between Thai and northern East Asian populations: implications for tagging SNP selection in Thais. J Hum Genet 51:896–904

Matyas G, Arnold E, Carrel T, Baumgartner D, Boileau C, Berger W, Steinmann B (2006) Identification and in silico analyses of novel TGFBR1 and TGFBR2 mutations in Marfan syndrome-related disorders. Hum Mutat 27:760–769

McCarthy JJ, Lehner T, Reeves C, Moliterno DJ, Newby LK, Rogers WJ, Topol EJ (2003) Association of genetic variants in the HDL receptor, SR-B1, with abnormal lipids in women with coronary artery disease. J Med Genet 40:453–458

McDermott DH, Yang Q, Kathiresan S, Cupples LA, Massaro JM, Keaney JF Jr, Larson MG, Vasan RS, Hirschhorn JN, O'Donnell CJ, Murphy PM, Benjamin EJ (2005) CCL2 polymorphisms are associated with serum monocyte chemoattractant Protein-1 levels and myocardial infarction in the Framingham Heart Study. Circulation 112:1113–1120

Michiels S, Danoy P, Dessen P, Bera A, Boulet T, Bouchardy C, Lathrop M, Sarasin A, Benhamou S (2007) Polymorphism discovery in 62 DNA repair genes and haplotype-associations with risks for lung, and head and neck cancers. Carcinogenesis (in press)

Mizuguchi T, Collod-Beroud G, Akiyama T, Abifadel M, Harada N, Morisaki T, Allard D, Varret M, Claustres M, Morisaki H, Ihara M, Kinoshita A, Yoshiura K, Junien C, Kajii T, Jondeau G, Ohta T, Kishino T, Furukawa Y, Nakamura Y, Niikawa N, Boileau C, Matsumoto N (2004) Heterozygous TGFBR2 mutations in Marfan syndrome. Nat Genet 36:855–860

Ng PC, Henikoff S (2002) Accounting for human polymorphisms predicted to affect protein function. Genome Res 12:436–446

R Development Core Team G (2006) R: a language and environment for statistical computing

Ramensky V, Bork P, Sunyaev S (2002) Human non-synonymous SNPs: server and survey. Nucleic Acids Res 30:3894–900

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365–386

Ruano G, Seip RL, Windemuth A, Zollner S, Tsongalis GJ, Ordovas J, Otvos J, Bilbie C, Miles M, Zoeller R, Visich P, Gordon P, Angelopoulos TJ, Pescatello L, Moyna N, Thompson PD (2006) Apolipoprotein A1 genotype affects the change in high density lipoprotein cholesterol subfractions with exercise training. Atherosclerosis 185:65–69

Schneider M, Tognolli M, Bairoch A (2004) The Swiss-Prot protein knowledgebase and ExPASy: providing the plant community with high quality proteomic data and tools. Plant Physiol Biochem 42:1013–1021

Simon DI, Chen Z, Seifert P, Edelman ER, Ballantyne CM, Rogers C (2000) Decreased neointimal formation in Mac-1−/− mice reveals a role for inflammation in vascular repair after angioplasty. J Clin Invest 105:293–300

Singh NN, Ramji DP (2006) The role of transforming growth factor-beta in atherosclerosis. Cytokine Growth Factor Rev 17:487–499

Tabara Y, Kohara K, Yamamoto Y, Igase M, Nakura J, Kondo I, Miki T (2003) Polymorphism of the monocyte chemoattractant Protein (MCP-1) gene is associated with the plasma level of MCP-1 but not with carotid intima-media thickness. Hypertens Res 26:677–683

Tabibiazar R, Wagner RA, Ashley EA, King JY, Ferrara R, Spin JM, Sanan DA, Narasimhan B, Tibshirani R, Tsao PS, Efron B, Quertermous T (2005) Signature patterns of gene expression in mouse atherosclerosis and their correlation to human coronary disease. Physiol Genom 22:213–226

Takahashi M, Matsuda F, Margetic N, Lathrop M (2002) Automated identification of single nucleotide polymorphisms from sequencing data. Proc IEEE Comput Soc Bioinform Conf 1:87–93

The Innate Immunity PGA P (2000) IIPGA genetic data: ITGAM allele frequencies. (https://innateimmunity.net/IIPGA2/PGAs/InnateImmunity/ITGAM/allele_freqs?flavor=masked)

The International HapMap C (2005) A haplotype map of the human genome. Nature 437:1299–1320

Wang X, Le Roy I, Nicodeme E, Li R, Wagner R, Petros C, Churchill GA, Harris S, Darvasi A, Kirilovsky J, Roubertoux PL, Paigen B (2003) Using advanced intercross lines for high-resolution mapping of HDL cholesterol quantitative trait loci. Genome Res 13:1654–1664

Yue P, Melamud E, Moult J (2006) SNPs3D: candidate gene and SNP selection for association studies. BMC Bioinform 7:166

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.

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

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

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

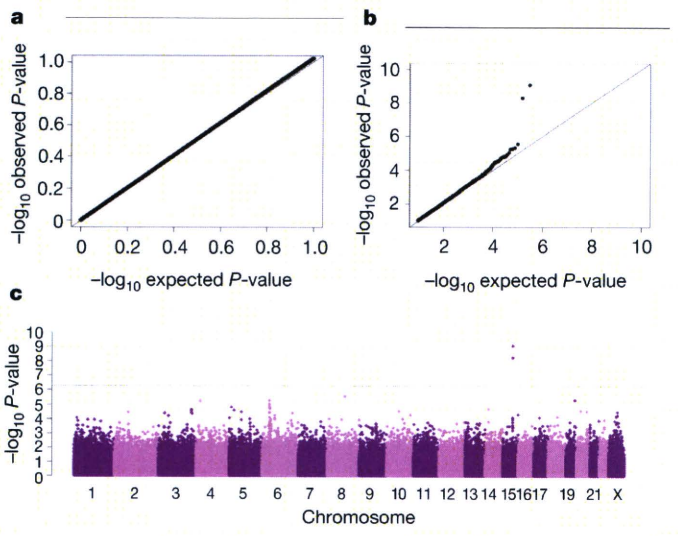

Figure 1 | Genome-wide association results in the central Europe study. **a–c**, Quantile–quantile plot for bottom 90% of *P*-values (**a**) and top 10% of *P*-values (**b**), as well as scatter plot (**c**) of *P*-values in −log scale from the trend test for 310,023 genotyped variants comparing 1,926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

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

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

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

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

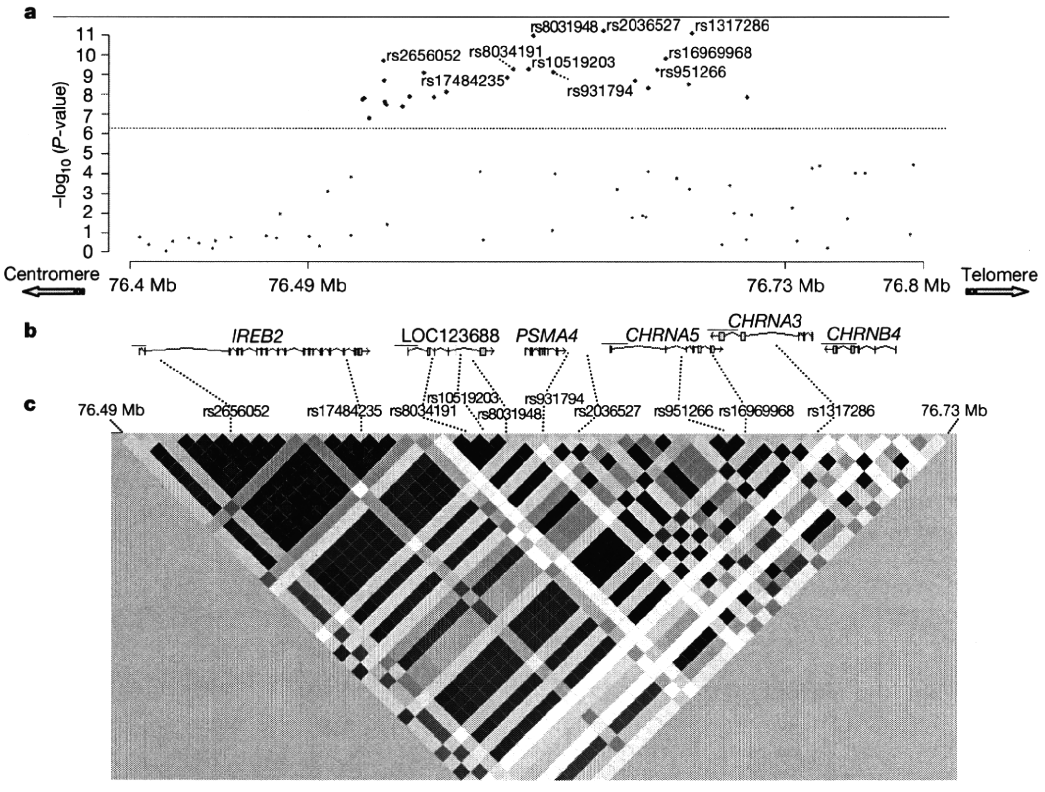

Figure 2 | Lung cancer area of interest across 15q25. a, *P*-values for SNPs genotyped in the 15q25 region (76.4–76.8 Mb). The dotted line indicates the genome-wide threshold of $P < 5 \times 10^{-7}$. Points labelled with rs numbers have a $P < 1 \times 10^{-9}$. Points in red are genotyped in the 317K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

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

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

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

Table 1 | Lung cancer risk and rs8034191 genotype

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

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

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

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

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

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

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

METHODS SUMMARY

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

Received 30 November 2007; accepted 7 March 2008.

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

11. Saccone, S. F. *et al.* Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum. Mol. Genet.* **16,** 36–49 (2007).
12. International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449,** 851–861 (2007).
13. Berretini, W. *et al.* α-5/α-3 nicotinic receptor subunit alleles increase risk for heavy smoking. *Mol. Psychiatry* advance online publication doi:10.1038/sj.mp.4002154 (29 January 2008).
14. Tritto, T., Stitzel, J. A., Marks, M. J., Romm, E. & Collins, A. C. Variability in response to nicotine in the LSxSS RI strains: potential role of polymorphisms in α4 and α6 nicotinic receptor genes. *Pharmacogenetics* **12,** 197–208 (2002).
15. Fagerstrom, K. O. & Schneider, N. G. Measuring nicotine dependence: a review of the Fagerstrom Tolerance Questionnaire. *J. Behav. Med.* **12,** 159–182 (1989).
16. Heatherton, T. F., Kozlowski, L. T., Frecker, R. C. & Fagerstrom, K. O. A Fagerstrom test for nicotine dependence: a revision of the Fagerstrom tolerance questionnaire. *Br. J. Addict.* **86,** 1119–1127 (1991).
17. Chabrol, H. *et al.* Comparison of the Heavy Smoking Index and of the Fagerstrom test for nicotine dependence in a sample of 749 cigarette smokers. *Addict. Behav.* **30,** 1474–1477 (2005).
18. Lam, D. C. *et al.* Expression of nicotinic acetylcholine receptor subunit genes in non-small-cell lung cancer reveals differences between smokers and nonsmokers. *Cancer Res.* **67,** 4638–4647 (2007).
19. West, K. A. *et al.* Rapid Akt activation by nicotine and a tobacco carcinogen modulates the phenotype of normal human airway epithelial cells. *J. Clin. Invest.* **111,** 81–90 (2003).
20. Dasgupta, P. & Chellappan, S. P. Nicotine-mediated cell proliferation and angiogenesis: new twists to an old story. *Cell Cycle* **5,** 2324–2328 (2006).
21. De Bakker, P. I. *et al.* A high resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nature Genet.* **38,** 1166–1172 (2006).
22. Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164,** 1567–1587 (2003).
23. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genet.* **38,** 904–909 (2006).

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).

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

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

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.

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 apopotosis 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.

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 · l^{-1} · 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).

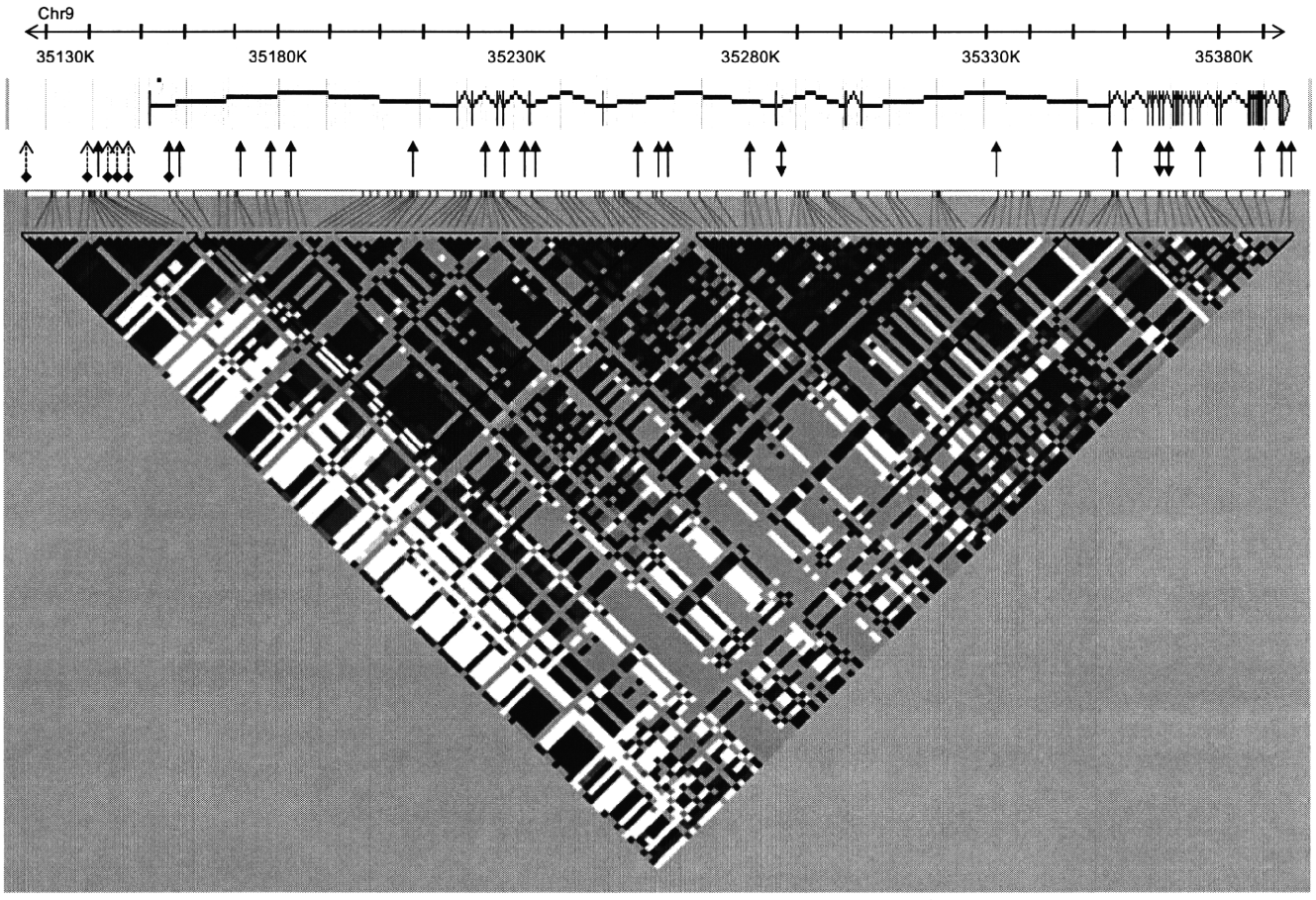

FIG. 1. Schematic representation of the *UNC13B* gene. The structure of UNC13B gene on chromosome 9 is presented with the respective positions of the 39 exons and of the 24 SNPs genotyped (rsID are given in Table 1), as well as the HapMap haplotype blocks (in D′). Arrows at both ends: SNPs selected through 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

	Denmark			Finland			France			
Polymorphisms	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.