nature genetics

Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease

Eleven susceptibility loci for late-onset Alzheimer's disease (LOAD) were identified by previous studies; however, a large portion of the genetic risk for this disease remains unexplained. We conducted a large, two-stage meta-analysis of genome-wide association studies (GWAS) in individuals of European ancestry. In stage 1, we used genotyped and imputed data (7,055,881 SNPs) to perform meta-analysis on 4 previously published GWAS data sets consisting of 17,008 Alzheimer's disease cases and 37,154 controls. In stage 2, 11,632 SNPs were genotyped and tested for association in an independent set of 8,572 Alzheimer's disease cases and 11,312 controls. In addition to the *APOE* locus (encoding apolipoprotein E), 19 loci reached genome-wide significance ($P < 5 \times 10^{-8}$) in the combined stage 1 and stage 2 analysis, of which 11 are newly associated with Alzheimer's disease.

Alzheimer's disease is a devastating neurological disorder primarily affecting the elderly. The disease manifests with progressive deterioration in cognitive functions, leading to loss of autonomy. The *APOE* gene (encoding apolipoprotein E) is a major genetic risk factor for Alzheimer's disease[1,2]. Previous GWAS in individuals of European ancestry identified nine other genomic regions associated with LOAD[3–7]. Recently, a rare susceptibility variant in *TREM2* was identified[8,9]. The search for additional genetic risk factors requires large-scale meta-analysis of GWAS to increase statistical power. Under the banner of I-GAP (International Genomics of Alzheimer's Project), we conducted a meta-analysis of 4 GWAS samples of European ancestry totaling 17,008 cases and 37,154 controls (stage 1) followed up by genotyping of 11,632 SNPs showing moderate evidence of association ($P < 1 \times 10^{-3}$ in stage 1) in an independent sample that included 8,572 cases and 11,312 controls (stage 2).

In the stage 1 meta-analysis, we used data from four consortia: the Alzheimer's Disease Genetic Consortium (ADGC), the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, the European Alzheimer's Disease Initiative (EADI) and the Genetic and Environmental Risk in Alzheimer's Disease (GERAD) Consortium (**Table 1**, Online Methods, **Supplementary Table 1** and **Supplementary Note**). We used European population reference (EUR) haplotype data from the 1000 Genomes Project (2010 interim release based on sequence data freeze from 4 August 2010 and phased haplotypes from December 2010) to impute genotypes for up to 11,863,202 SNPs per data set. We excluded SNPs that did not pass quality control in each study (**Supplementary Table 2** and **Supplementary Note**). Our meta-analysis included SNPs either genotyped or successfully imputed in at least 40% of the Alzheimer's disease cases and 40% of the control samples across all data sets (7,055,881 SNPs; Online Methods). In each data set, genotype dosages were analyzed as described in the **Supplementary Note** (**Supplementary Table 2**). We performed meta-analysis of the results after applying genomic control correction to each study. The genomic control inflation factor for the meta-analysis was 1.087 for the full set of SNPs and 1.082 after excluding SNPs within the *APOE* locus (chr. 19: 45,409,039–45,412,650) and within 500 kb of SNPs associated with Alzheimer's disease at a pre-specified level of genome-wide significance ($P < 5 \times 10^{-8}$) in stage 1 (see **Supplementary Fig. 1** for quantile-quantile plots).

In addition to the *APOE* locus, 14 genomic regions had associations that reached the genome-wide significance level (**Fig. 1**). Nine had been previously identified by GWAS as genetic susceptibility factors[3–7], and five (*HLA-DRB5–HLA-DRB1*, *PTK2B*, *SORL1*, *SLC24A4-RIN3* and *DSG2*) represent newly associated loci (**Table 2**). *SORL1* had previously been identified as an Alzheimer's disease gene through candidate gene approaches and in a GWAS combining ADGC and Asian samples[10]. Genes attributed to a signal were those closest to the most significantly associated SNP. However, we are aware that these are potentially not the causative genes. Detailed results for each region are given in **Supplementary Figures 2–7**.

In stage 2, we selected for genotyping all stage 1 SNPs with a *P* value less than 1×10^{-3}, excluding SNPs flanking *APOE* (chr. 19: 45,409,039–45,412,650) (n = 19,532; see URLs for database access). From the initial set of SNPs, 14,445 could be genotyped using Illumina iSelect technology. After quality control procedures (Online Methods), we considered 11,632 SNPs for association analysis. The stage 2 sample included 8,572 cases and 11,312 controls of European ancestry originating from Austria, Belgium, Finland, Germany, Greece, Hungary, Italy, Spain, Sweden, the UK and the United States (**Table 1** and **Supplementary Note**). We observed 116 SNPs showing the same risk allele and direction of association in stages 1 and 2 that were significantly associated with Alzheimer's disease risk in stage 2 after a strict Bonferroni correction for multiple testing ($P < 4.3 \times 10^{-6}$). Of these 116 SNPs, 80 had been associated at genome-wide significance with Alzheimer's disease risk in stage 1. Additionally, in analyses in stage 2, 2,562 SNPs were associated with Alzheimer's disease at a nominal level of significance ($P < 0.05$), having the same risk allele and direction of association as in stage 1.

The results from stages 1 and 2 and from the combined stage 1 and stage 2 data sets, which represent a secondary discovery effort,

A full list of authors and affiliations appears at the end of the paper.

Received 13 March; accepted 27 September; published online 27 October 2013; doi:10.1038/ng.2802

Table 1 Description of the consortium data sets used for stage 1 and stage 2

		Alzheimer's disease cases				Controls			
	Consortium	*N*	Percent women	Mean AAO (s.d.)		*N*	Percent women	Mean AAE (s.d.)	
Stage 1	ADGC	10,273	59.4	74.7	(7.7)	10,892	58.6	76.3	(8.1)
	CHARGE	1,315	63.6	82.7	(6.8)	12,968	57.8	72.8	(8.6)
	EADI	2,243	64.9	68.5	(8.9)	6,017	60.7	74.0	(5.4)
	GERAD	3,177	64.0	73.0	(8.5)	7,277	51.8	51.0	(11.8)
	N	17,008				37,154			
	Country	*N*	Percent women	Mean AAO (s.d.)		*N*	Percent women	Mean AAE (s.d.)	
Stage 2	Austria	210	61.0	72.5	(8.1)	829	43.3	65.5	(8.0)
	Belgium	878	66.1	75.4	(8.6)	661	59.5	65.7	(14.3)
	Finland	422	68.0	71.4	(6.9)	562	59.3	69.1	(6.2)
	Germany	972	63.9	73.0	(8.6)	2,378	53.1	69.5	(10.1)
	Greece	256	63.3	69.2	(8.0)	229	34.1	49.3	(16.4)
	Hungary	125	68.0	74.9	(6.8)	100	69.0	74.4	(6.5)
	Italy	1,729	66.5	71.5	(8.7)	720	55.7	70.0	(10.4)
	Spain	2,121	66.3	75.0	(8.3)	1,921	55.3	70.2	(10.8)
	Sweden	797	61.7	76.8	(8.1)	1,506	62.8	70.6	(8.7)
	UK	490	57.6	74.6	(8.7)	1,066	29.2	73.8	(6.5)
	United States	572	61.9	83.5	(7.6)	1,340	54.0	79.3	(6.8)
	N	8,572				11,312			

AAO, age at onset; AAE, age at examination.

are shown in **Table 2**. With the exception of *CD33* and *DSG2*, we nominally replicated all loci that surpassed the genome-wide significance level in stage 1. Inability to replicate *DSG2* is not surprising, as evidence of association for this locus was based on data for a single SNP and was not supported by data from surrounding SNPs in linkage disequilibrium (LD, $r^2 > 0.8$; **Supplementary Fig. 7b**). Moreover, seven new loci reached the genome-wide significance level in the combined analysis (**Table 2**). More detailed results for the seven newly identified LOAD loci are provided in **Supplementary Figures 8–11**. There was no significant heterogeneity across studies at any of the loci, except at *DSG2* (**Table 2** and **Supplementary Figs. 12–16**). To identify potential causative genes, we also examined all SNPs with association $P < 5\times 10^{-8}$ that were within 500 kb of the top SNP at each locus to identify *cis* expression quantitative trait locus (*cis*-eQTL) associations (Online Methods and **Supplementary Table 3**).

The results from the combined stage 1 and stage 2 data sets also identified 13 loci with suggestive evidence of association ($P < 1 \times 10^{-6}$) (**Supplementary Table 4**). Among these, we detected a signal for rs9381040 ($P = 6.3 \times 10^{-7}$), which is located approximately 5.5 kb away from the 3′ end of *TREML2* and 24 kb away from the 5′ end of *TREM2*. *TREM2* was recently reported to carry a rare variant (encoding p.Arg47His) associated with three- to fourfold increased risk of developing Alzheimer's disease[8,9]. This region also reached genome-wide significance in a study of cerebral spinal fluid levels of phosphorylated tau, a biomarker for Alzheimer's disease[11].

Beyond the already known, GWAS-defined genes (*ABCA7*, *BIN1*, *CD33*, *CLU*, *CR1*, *CD2AP*, *EPHA1*, *MS4A6A-MS4A4E* and *PICALM*), the most significant new association was in the *HLA-DRB5–DRB1* region (encoding major histocompatibility complex, class II, DRβ5 and DRβ1, respectively). This region is associated with immunocompetence and histocompatibility and, interestingly, with risk of both multiple sclerosis and Parkinson disease[12,13]. Owing to the complex genetic organization of the human leukocyte antigen (HLA) region on chromosome 6, we were unable to define which gene(s) are responsible for this signal (**Supplementary Fig. 6a**).

The second strongest signal was within the *SORL1* gene (encoding sortilin-related receptor, L(DLR class) 1). Our data clearly demonstrated that this gene was associated at genome-wide significance in European samples. *SORL1* is noteworthy, as it is associated with increased risk of both autosomal dominant and sporadic forms of Alzheimer's disease[14,15] and represents the first LOAD gene that directly connects aberrant trafficking and metabolism of the amyloid precursor protein (APP) to LOAD[14].

The third locus, *PTK2B* (encoding protein tyrosine kinase 2β), is only approximately 130 kb away from *CLU*, but we believe the two signals are independent because (i) the two most strongly associated SNPs within each of these two genes are not in LD ($D' = 0.06$ and $r^2 = 0.003$ as computed using 1000 Genomes Project data); (ii) a recombination peak exists between the two loci (**Fig. 2**); and (iii) conditional analysis in the stage 2 data confirmed the independence of the *PTK2B* association (**Supplementary Fig. 17** and **Supplementary Table 5**). The protein encoded by *PTK2B* may be an intermediate between neuropeptide-activated receptors or neurotransmitters that increase calcium flux and the downstream

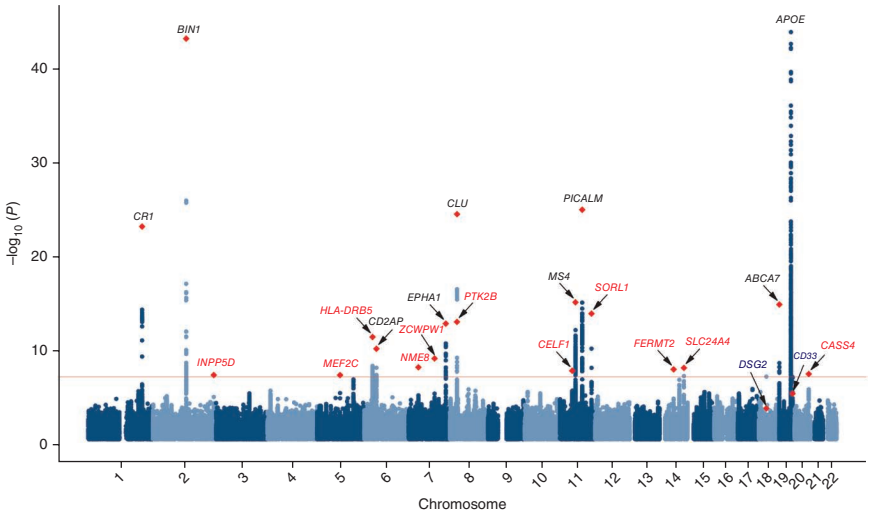

Figure 1 Manhattan plot of stage 1 for genome-wide association with Alzheimer's disease (17,008 cases and 37,154 controls). The threshold for genome-wide significance ($P < 5 \times 10^{-8}$) is indicated by the red line. Genes previously identified by GWAS are shown in black, and newly associated genes are shown in red. Red diamonds represent SNPs with the smallest *P* values in the overall analysis.

Table 2 Summary of stage 1, stage 2 and overall meta-analyses for SNPs reaching genome-wide significance after stages 1 and 2

						Stage 1		Stage 2		Overall		
SNP[a]	Chr.	Position[b]	Closest gene[c]	Major/minor alleles	MAF[d]	OR (95% CI)[e]	Meta *P* value	OR (95% CI)[e]	Meta *P* value	OR (95% CI)[e]	Meta *P* value	I^2 (%), *P* value[f]
Known GWAS-defined associated genes												
rs6656401	1	207692049	*CR1*	G/A	0.197	1.17 (1.12–1.22)	7.7×10^{-15}	1.21 (1.14–1.28)	7.9×10^{-11}	1.18 (1.14–1.22)	5.7×10^{-24}	0, 7.8×10^{-1}
rs6733839	2	127892810	*BIN1*	C/T	0.409	1.21 (1.17–1.25)	1.7×10^{-26}	1.24 (1.18–1.29)	3.4×10^{-19}	1.22 (1.18–1.25)	6.9×10^{-44}	28, 6.1×10^{-2}
rs10948363	6	47487762	*CD2AP*	A/G	0.266	1.10 (1.07–1.14)	3.1×10^{-8}	1.09 (1.04–1.15)	4.1×10^{-4}	1.10 (1.07–1.13)	5.2×10^{-11}	0, 9×10^{-1}
rs11771145	7	143110762	*EPHA1*	G/A	0.338	0.90 (0.87–0.93)	8.8×10^{-10}	0.90 (0.86–0.95)	2.8×10^{-5}	0.90 (0.88–0.93)	1.1×10^{-13}	14, 2.4×10^{-1}
rs9331896	8	27467686	*CLU*	T/C	0.379	0.86 (0.84–0.89)	9.6×10^{-17}	0.86 (0.82–0.90)	4.5×10^{-10}	0.86 (0.84–0.89)	2.8×10^{-25}	0, 4.9×10^{-1}
rs983392	11	59923508	*MS4A6A*	A/G	0.403	0.90 (0.87–0.93)	2.8×10^{-11}	0.90 (0.86–0.94)	4.5×10^{-6}	0.90 (0.87–0.92)	6.1×10^{-16}	1, 4.5×10^{-1}
rs10792832	11	85867875	*PICALM*	G/A	0.358	0.88 (0.85–0.91)	6.5×10^{-16}	0.85 (0.81–0.89)	1.1×10^{-11}	0.87 (0.85–0.89)	9.3×10^{-26}	0, 9.8×10^{-1}
rs4147929	19	1063443	ABCA7	G/A	0.190	1.14 (1.10–1.20)	1.7×10^{-9}	1.17 (1.10–1.24)	9.9×10^{-8}	1.15 (1.11–1.19)	1.1×10^{-15}	0, 9.4×10^{-1}
rs3865444[g]	19	51727962	CD33	C/A	0.307	0.91 (0.88–0.94)	5.1×10^{-8}	0.99 (0.94–1.04)	6.9×10^{-1}	0.94 (0.91–0.96)	3.0×10^{-6}	0, 6.9×10^{-1}
New loci reaching genome-wide significance in the discovery analysis												
rs9271192	6	32578530	*HLA-DRB5–HLA-DRB1*	A/C	0.276	1.11 (1.07–1.16)	1.6×10^{-8}	1.12 (1.06–1.18)	4.2×10^{-5}	1.11 (1.08–1.15)	2.9×10^{-12}	0, 5.4×10^{-1}
rs28834970	8	27195121	*PTK2B*	T/C	0.366	1.10 (1.07–1.14)	3.3×10^{-9}	1.11 (1.06–1.17)	4.3×10^{-6}	1.10 (1.08–1.13)	7.4×10^{-14}	10, 3.0×10^{-1}
rs11218343	11	121435587	*SORL1*	T/C	0.039	0.76 (0.70–0.83)	5.0×10^{-11}	0.78 (0.70–0.88)	4.0×10^{-5}	0.77 (0.72–0.82)	9.7×10^{-15}	0, 8.3×10^{-1}
rs10498633	14	92926952	*SLC24A4-RIN3*	G/T	0.217	0.90 (0.87–0.94)	1.5×10^{-7}	0.93 (0.88–0.98)	7.8×10^{-3}	0.91 (0.88–0.94)	5.5×10^{-9}	0, 6.3×10^{-1}
rs8093731[g]	*18*	*29088958*	*DSG2*	*C/T*	*0.017*	*0.54 (0.43–0.67)*	*4.6×10^{-8}*	*1.01 (0.80–1.28)*	*9.0×10^{-1}*	*0.73 (0.62–0.86)*	*1.0×10^{-4}*	38, 3.9×10^{-2}
New loci reaching genome-wide significance in the combined discovery and replication analysis												
rs35349669	2	234068476	*INPP5D*	C/T	0.488	1.07 (1.03–1.10)	9.6×10^{-5}	1.10 (1.05–1.15)	5.7×10^{-5}	1.08 (1.05–1.11)	3.2×10^{-8}	0, 8.0×10^{-1}
rs190982	5	88223420	*MEF2C*	A/G	0.408	0.92 (0.89–0.95)	2.5×10^{-6}	0.93 (0.89–0.98)	3.4×10^{-3}	0.93 (0.90–0.95)	3.2×10^{-8}	0, 6.4×10^{-1}
rs2718058	7	37841534	*NME8*	A/G	0.373	0.93 (0.90–0.96)	1.3×10^{-5}	0.91 (0.87–0.95)	6.3×10^{-5}	0.93 (0.90–0.95)	4.8×10^{-9}	0, 9.2×10^{-1}
rs1476679	7	100004446	*ZCWPW1*	T/C	0.287	0.92 (0.89–0.96)	7.4×10^{-6}	0.89 (0.85–0.94)	9.7×10^{-6}	0.91 (0.89–0.94)	5.6×10^{-10}	0, 7.0×10^{-1}
rs10838725	11	47557871	*CELF1*	T/C	0.316	1.08 (1.04–1.11)	6.7×10^{-6}	1.09 (1.04–1.14)	4.1×10^{-4}	1.08 (1.05–1.11)	1.1×10^{-8}	0, 7.6×10^{-1}
rs17125944	14	53400629	*FERMT2*	T/C	0.092	1.13 (1.07–1.19)	1.0×10^{-5}	1.17 (1.08–1.26)	1.6×10^{-4}	1.14 (1.09–1.19)	7.9×10^{-9}	10, 3.0×10^{-1}
rs7274581	20	55018260	*CASS4*	T/C	0.083	0.87 (0.82–0.92)	1.6×10^{-6}	0.89 (0.82–0.96)	4.1×10^{-3}	0.88 (0.84–0.92)	2.5×10^{-8}	0, 9.9×10^{-1}

Chr., chromosome.
[a]SNPs showing the best level of association after meta-analysis of stages 1 and 2. [b]Build 37, assembly hg19. [c]Genes located ± 100 kb of the top SNP. [d]Average in the discovery sample. [e]Calculated with respect to the minor allele. [f]Cochran's *Q* test. [g]Not replicated in stage 2.

signals regulating neuronal activity such as mitogen-activated protein kinase (MAPK) signaling[16]. *PTK2B* is involved in the induction of long-term potentiation in the hippocampal CA1 (cornu ammonis 1) region, a central process in the formation of memory[17]. We cannot, however, exclude the possibility that there are multiple signals in the *PTK2B-CLU* region that are functionally connected to a single gene. For instance, two SNPs associated with genome-wide significance in the *PTK2B-CLU* region are eQTLs for the gene *DPYSL2* that has been implicated in Alzheimer's disease[18] (**Supplementary Table 3**).

The fourth locus was *SLC24A4* (encoding solute carrier family 24 (sodium/potassium/calcium exchanger), member 4). The *SLC24A4* gene encodes a protein involved in iris development and hair and skin color variation in humans in addition to being associated with the risk of developing hypertension[19,20]. *SLC24A4* is also expressed in the brain and may be involved in neural development[21]. Of note, in the vicinity of the most strongly associated SNP is another gene called *RIN3* (encoding Ras and Rab interactor 3), and its gene product directly interacts with the *BIN1* gene product[22], a protein that may be connected to tau-mediated pathology[23].

In addition to these four loci reaching genome-wide significance in stage 1, seven new loci reached genome-wide significance in the combined analysis.

The strongest association at one of these new loci was intronic in the *ZCWPW1* gene (encoding zinc finger, CW type with PWWP

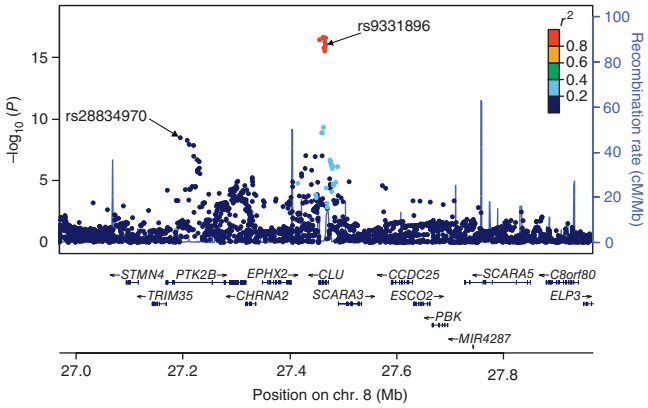

Figure 2 Regional plot for the *PTK2B-CLU* locus (17,008 cases and 37,154 controls).

domain 1), whose corresponding protein modulates epigenetic regulation[24]. However, the region defined by all the SNPs associated with Alzheimer's disease risk in our data is large and contains about ten genes (**Supplementary Fig. 9b**). Another interesting possible candidate gene in the *ZCWPW1* region is *NYAP1*, as disruption of the corresponding gene in mice affects brain size, neurite elongation and, more generally, neuronal morphogenesis[25]. Our data do not resolve which gene in this region may be causal.

A second locus was within the *CELF1* gene (encoding CUGBP, Elav-like family member 1), whose gene product is a member of the protein family that regulates pre-mRNA alternative splicing[26]. As with the *ZCWPW1* locus, the region of interest is large and contains about ten genes (**Supplementary Fig. 10a**). Among these genes is *MADD* (encoding MAP kinase–activating death domain), the reduced expression of which may affect long-term neuronal viability in Alzheimer's disease[27].

A discrete signal was observed adjacent to *NME8* (encoding NME/NM23 family member 8), which is responsible for primary ciliary dyskinesia type 6 (ref. 28).

The *FERMT2* gene (encoding fermitin family member 2) is expressed in the brain. Its corresponding protein localizes to cell matrix adhesion structures, activates integrins, is involved in the orchestration of actin assembly and cell shape modulation, and is an important mediator of angiogenesis[29]. An association between the *Drosophila melanogaster* ortholog of *FERMT2* (*fit1/fit2*) and tau-mediated toxicity was recently described[30].

We identified a fifth signal on chromosome 20 at *CASS4* (encoding Cas scaffolding protein family member 4). Little is known about the function of the encoded protein. However, the *Drosophila* CASS family ortholog (*p130CAS*) binds to CMS, the *Drosophila* ortholog of *CD2AP* (CMS), a known Alzheimer's disease susceptibility gene (**Table 2**) that is involved in actin dynamics[31].

Another locus was identified at *INPP5D* (encoding inositol polyphosphate-5-phosphatase, 145 kDa) on chromosome 2. *INPP5D* is expressed at low levels in the brain, but the encoded protein has been shown to interact with CD2AP, whose corresponding gene is one of the Alzheimer's disease genes previously identified by GWAS[32], and to modulate, along with GRB2, metabolism of APP[33].

We identified a seventh signal adjacent to *MEF2C* (encoding myocyte enhancer factor 2). Mutations at this locus are associated with severe mental retardation, stereotypic movements, epilepsy and cerebral malformation[34]. The MEF2C protein limits excessive synapse formation during activity-dependent refinement of synaptic connectivity and thus may facilitate hippocampal-dependent learning and memory[35].

In summary, our Alzheimer's disease GWAS meta-analysis has identified 11 new susceptibility loci in addition to the already known *ABCA7*, *APOE*, *BIN1*, *CLU*, *CR1*, *CD2AP*, *EPHA1*, *MS4A6A-MS4A4E* and *PICALM* genes. However, we were not able to replicate association of *CD33* in our stage 2 analysis ($P = 0.61$). We did not detect any biases in terms of imputation in our discovery data sets or genotyping in our replication data sets (data not shown), suggesting a potential statistical fluctuation across our populations as an explanation for the lack of replication. However, recent data suggest that genetically determined decreased *CD33* expression might reduce Alzheimer's disease risk and interfere with amyloid β peptide clearance[36], a dysfunction thought to be central in late-onset forms of Alzheimer's disease[37]. Further investigations in independent case-control studies will thus be required to confirm or refute the association of *CD33* with Alzheimer's disease.

The newly associated loci reinforce the importance of some previously suspected pathways such as APP (*SORL1* and *CASS4*) and tau (*CASS4* and *FERMT2*) in pathology. Several candidate genes at these loci are involved in pathways already shown to be enriched for association signal in Alzheimer's disease GWAS[38,39], such as immune response and inflammation (*HLA-DRB5–DRB1*, *INPP5D* and *MEF2C*), which is also supported by the described association of Alzheimer's disease with *CR1* (ref. 3) and *TREM2* (refs. 8,9), cell migration (*PTK2B*) and lipid transport and endocytosis (*SORL1*). Our results also suggest the existence of new pathways underlying Alzheimer's disease. These pathways could include hippocampal synaptic function (*MEF2C* and *PTK2B*), cytoskeletal function and axonal transport (*CELF1*, *NME8* and *CASS4*), regulation of gene expression and post-translational modification of proteins, and microglial and myeloid cell function (*INPP5D*).

Examining the genetic effect attributable to all the associated loci, we demonstrated that the most strongly associated SNPs at each locus other than *APOE* had population-attributable fractions (PAFs) or preventive fractions between 1.0–8.0% in the stage 2 sample (**Supplementary Table 6**). Strong efforts in sequencing and post-GWAS analyses will now be required to fully characterize the candidate genes and functional variants responsible for the association of these GWAS-identified loci with Alzheimer's disease risk and to understand their exact roles in the pathophysiology of Alzheimer's disease[40,41].

URLs. Database access, http://www.pasteur-lille.fr/en/recherche/u744/Igap_stage1.zip; IMPUTE2, http://mathgen.stats.ox.ac.uk/impute/impute_v2.html; MaCH, http://www.sph.umich.edu/csg/abecasis/MACH/; ProbABEL, http://www.genabel.org/packages/ProbABEL; SMARTPCA, http://www.hsph.harvard.edu/alkes-price/software/; GWAMA, http://www.well.ox.ac.uk/gwama/; LocusZoom, http://csg.sph.umich.edu/locuszoom/; PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/; SNPTEST, https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html; Aberrant, http://www.well.ox.ac.uk/software; Metal, http://www.sph.umich.edu/csg/abecasis/metal/; R, http://www.r-project.org/; R meta, http://cran.r-project.org/web/packages/rmeta/index.html; eQTL analyses (accessed 18 February 2013), http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

This work was made possible by the generous participation of the control subjects, the patients and their families. iSelect chips were funded by the French National Foundation on Alzheimer's Disease and Related Disorders. Data management involved the Centre National de Génotypage and was supported by the Institut Pasteur de Lille, INSERM, FRC (Fondation pour la Recherche sur le Cerveau) and Rotary. This work has been developed and supported by the LABEX (Laboratory of Excellence Program Investment for the Future) DISTALZ grant (Development of Innovative Strategies for a Transdisciplinary Approach to Alzheimer's Disease). The French National Foundation on Alzheimer's Disease and Related Disorders and the Alzheimer's Association (Chicago, Illinois) grant supported in-person meetings and communication for IGAP, and the Alzheimer's Association (Chicago, Illinois) grant provided some funds to each consortium for analyses.

GERAD was supported by the Wellcome Trust, the MRC, Alzheimer's Research UK (ARUK) and the Welsh government. ADGC and CHARGE were supported by the US National Institutes of Health, National Institute on Aging (NIH-NIA), including grants U01 AG032984 and R01 AG033193 (additional US National Institutes of Health grant numbers are listed in the **Supplementary Note**). CHARGE was also supported by Erasmus Medical Center and Erasmus University.

Complete acknowledgments are detailed in the **Supplementary Note**.

AUTHOR CONTRIBUTIONS

Study concept and design: J.-C.L., C.A.I.-V., D. Harold, A.C.N., A.L.D., J.C.B., A.V.S., M.A.I., H. Schmidt, A.L.F., V.G., O.L.L., D.W.T., D. Blacker, T.H.M., T.B.H., J.I.R., W.A.K., M. Boada, R. Schmidt, R.M., A.H., B.M.P., J.L.H., P.A.H., M.L., M.A.P.-V., L.J.L., L.A.F., C.M.v.D., V.M., S. Seshadri, J.W., G.D.S. and P.A. **Acquisition of data:** J.-C.L., C.A.I.-V., D. Harold, C. Bellenguez, R. Sims, G.J., B.G.-B., G.R., N.J., V.C., C. Thomas, D.Z., Y.K., A.G., H. Schmidt, M.L.D., M.-T.B., S.-H.C., P.H., V.G., C. Baldwin, C.C., C. Berr, O.L.L., P.L.D.J., D.E., L. Letenneur, G.E., K.S., A.M.G., N.F., M.J.H., M.I.K., E.B.L., A.J.M., C.D., S.T., S. Love, E.R., P.S.G.-H., L.Y., M.M.C., D. Beekly, F.Z., O.V., S.G.Y., W.G., M.J.O., K.M.F., P.V.J., M.C.O., L.B.C., D.A.B., T.B.H., R.F.A.G.d.B., T.J.M., J.I.R., K.M., T.M.F., W.A.K., J.F.P., M.A.N., K.R., J.S.K.K., E.B., M.R., M. Boada, L.-S.W., J.-F.D., C. Tzourio, M.M.N., B.M.P., L.J., J.L.H., M.L., L.J.L., L.A.F., A.H., C.M.v.D., S. Seshadri, J.W., G.D.S. and P.A. **Sample contribution:** A. Ruiz, F. Pasquier, A. Ramirez, O.H., J.D.B., D. Campion, P.K.C., C. Baldwin, T.B., C.C., D. Craig, V.D., J.A.J., S. Lovestone, F.J.M., D.C.R., K.S., A.M.G., N.F., M.G., K. Brown, M.I.K., L.K., P.B.-G., B.M., R.G., A.J.M., D.W., E.R., J.G., P.S.G.-H., J.C., A.L., A. Bayer, M.T., P. Bossù, G.S., P. Proitsi, J.C., S. Sorbi, F.S.-G., N.C.F., J.H., M.C.D.N., P. Bosco, R.C., C. Brayne, D.G., M. Mancuso, F.M., S. Moebus, P.M., M.D.Z., W.M., H. Hampel, A.P., M. Bullido, F. Panza, P.C., B.N., M. Mayhaus, L. Lannfelt, H. Hakonarson, S.P., M.M.C., M.I., V.A., S.G.Y., E.C., C. Razquin, P. Pastor, I.M., O.C., H. Soininen, S. Mead, D.A.B., L.F., C.H., P. Passmore, T.J.M., K. Bettens, A. Brice, D. Hannequin, K.R., M.R., M.H., D.R., C.G. and C.V.B. **Data analysis:** C.A.I.-V., D. Harold, A.C.N., R. Sims, C. Bellenguez, G.J., A.L.D., J.C.B., G.W.B., B.G.-B., G.R., T.A.T.-W., N.J., A.V.S., V.C., M.A.I., D.Z., Y.K., B.N.V., C.-F.L., A.G., B.K., C. Reitz, J.R.G., O.V., W.A.K., K.L.L., K.L.H.-N., E.R.M., L.-S.W., B.M.P., M.L., V.M. and J.W. **Statistical analysis and interpretation:** J.-C.L., C.A.I.-V., D. Harold, A.C.N., C. Bellenguez, G.J., A.L.D., J.C.B., G.W.B., T.A.T.-W., A.V.S., V.C., M.A.I., B.N.V., Y.K., C.-F.L., B.K., C. Reitz, A.L.F., N.A., J.R.G., R.F.A.G.d.B., W.A.K., K.L.L., E.R.M., L.-S.W., B.M.P., L.J., J.L.H., P.A.H., M.A.P.-V., L.J.L., L.A.F., C.M.v.D., V.M., S. Seshadri, J.W., G.D.S. and P.A. **Drafting of the manuscript:** J.-C.L., C.A.I.-V., D. Harold, A.C.N., C. Bellenguez, A.L.D., J.C.B., A.V.S., R.M., B.M.P., J.L.H., M.A.P.-V., L.J.L., L.A.F., C.M.v.D., C.V.B., S. Seshadri, J.W., G.D.S. and P.A.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Corder, E.H. *et al.* Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **261**, 921–923 (1993).
2. Genin, E. *et al. APOE* and Alzheimer disease: a major gene with semi-dominant inheritance. *Mol. Psychiatry* **16**, 903–907 (2011).
3. Lambert, J.-C. *et al.* Genome-wide association study identifies variants at *CLU* and *CR1* associated with Alzheimer's disease. *Nat. Genet.* **41**, 1094–1099 (2009).
4. Harold, D. *et al.* Genome-wide association study identifies variants at *CLU* and *PICALM* associated with Alzheimer's disease. *Nat. Genet.* **41**, 1088–1093 (2009).
5. Seshadri, S. *et al.* Genome-wide analysis of genetic loci associated with Alzheimer disease. *J. Am. Med. Assoc.* **303**, 1832–1840 (2010).
6. Hollingworth, P. *et al.* Common variants at *ABCA7*, *MS4A6A*/*MS4A4E*, *EPHA1*, *CD33* and *CD2AP* are associated with Alzheimer's disease. *Nat. Genet.* **43**, 429–435 (2011).
7. Naj, A.C. *et al.* Common variants at *MS4A4*/*MS4A6E*, *CD2AP*, *CD33* and *EPHA1* are associated with late-onset Alzheimer's disease. *Nat. Genet.* **43**, 436–441 (2011).
8. Guerreiro, R. *et al. TREM2* variants in Alzheimer's disease. *N. Engl. J. Med.* **368**, 117–127 (2013).
9. Jonsson, T. *et al.* Variant of *TREM2* associated with the risk of Alzheimer's disease. *N. Engl. J. Med.* **368**, 107–116 (2013).
10. Miyashita, A. *et al. SORL1* is genetically associated with late-onset Alzheimer's disease in Japanese, Koreans and Caucasians. *PLoS ONE* **8**, e58618 (2013).
11. Cruchaga, C. *et al.* GWAS of cerebrospinal fluid tau levels identifies risk variants for Alzheimer's disease. *Neuron* **78**, 256–268 (2013).
12. Sawcer, S. *et al.* Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**, 214–219 (2011).
13. Nalls, M.A. *et al.* Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet* **377**, 641–649 (2011).
14. Rogaeva, E. *et al.* The neuronal sortilin-related receptor *SORL1* is genetically associated with Alzheimer disease. *Nat. Genet.* **39**, 168–177 (2007).
15. Pottier, C. *et al.* High frequency of potentially pathogenic *SORL1* mutations in autosomal dominant early-onset Alzheimer disease. *Mol. Psychiatry* **17**, 875–879 (2012).
16. Pandey, P. *et al.* Activation of p38 mitogen-activated protein kinase by PYK2/related adhesion focal tyrosine kinase–dependent mechanism. *J. Biol. Chem.* **274**, 10140–10144 (1999).
17. Huang, Y. *et al.* CAKβ/Pyk2 kinase is a signaling link for induction of long-term potentiation in CA1 hippocampus. *Neuron* **29**, 485–496 (2001).
18. Williamson, R. *et al.* CRMP2 hyperphosphorylation is characteristic of Alzheimer's disease and not a feature common to other neurodegenerative diseases. *J. Alzheimers Dis.* **27**, 615–625 (2011).
19. Sulem, P. *et al.* Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat. Genet.* **39**, 1443–1452 (2007).
20. Adeyemo, A. *et al.* A genome-wide association study of hypertension and blood pressure in African Americans. *PLoS Genet.* **5**, e1000564 (2009).
21. Larsson, M. *et al.* GWAS findings for human iris patterns: associations with variants in genes that influence normal neuronal pattern development. *Am. J. Hum. Genet.* **89**, 334–343 (2011).
22. Kajiho, H. *et al.* RIN3: a novel Rab5 GEF interacting with amphiphysin II involved in the early endocytic pathway. *J. Cell Sci.* **116**, 4159–4168 (2003).
23. Chapuis, J. *et al.* Increased expression of BIN1 mediates Alzheimer genetic risk by modulating tau pathology. *Mol. Psychiatry* doi:10.1038/mp.2013.1 (12 February 2013).
24. He, F. *et al.* Structural insight into the zinc finger CW domain as a histone modification reader. *Structure* **18**, 1127–1139 (2010).
25. Yokoyama, K. *et al.* NYAP: a phosphoprotein family that links PI3K to WAVE1 signalling in neurons. *EMBO J.* **30**, 4739–4754 (2011).
26. Gallo, J.M. & Spickett, C. The role of CELF proteins in neurological disorders. *RNA Biol.* **7**, 474–479 (2010).
27. Del Villar, K. & Miller, C.A. Down-regulation of DENN/MADD, a TNF receptor binding protein, correlates with neuronal cell death in Alzheimer's disease brain and hippocampal neurons. *Proc. Natl. Acad. Sci. USA* **101**, 4210–4215 (2004).
28. Duriez, B. *et al.* A common variant in combination with a nonsense mutation in a member of the thioredoxin family causes primary ciliary dyskinesia. *Proc. Natl. Acad. Sci. USA* **104**, 3336–3341 (2007).
29. Pluskota, E. *et al.* The integrin coactivator kindlin-2 plays a critical role in angiogenesis in mice and zebrafish. *Blood* **117**, 4978–4987 (2011).
30. Shulman, J.M. *et al.* Functional screening in *Drosophila* identifies Alzheimer's disease susceptibility genes and implicates Tau-mediated mechanisms. *Hum. Mol. Genet.* doi:10.1093/hmg/ddt478 (9 October 2013).
31. Kirsch, K.H., Georgescu, M.M., Ishimaru, S. & Hanafusa, H. CMS: an adapter molecule involved in cytoskeletal rearrangements. *Proc. Natl. Acad. Sci. USA* **96**, 6211–6216 (1999).
32. Bao, M. *et al.* CD2AP/SHIP1 complex positively regulates plasmacytoid dendritic cell receptor signaling by inhibiting the E3 ubiquitin ligase Cbl. *J. Immunol.* **189**, 786–792 (2012).
33. Brauer, H. *et al.* Leukemia-associated mutations in SHIP1 inhibit its enzymatic activity, interaction with the GM-CSF receptor and Grb2, and its ability to inactivate PI3K/AKT signaling. *Cell. Signal.* **24**, 2095–2101 (2012).
34. Le Meur, N. *et al. MEF2C* haploinsufficiency caused by either microdeletion of the 5q14.3 region or mutation is responsible for severe mental retardation with stereotypic movements, epilepsy and/or cerebral malformations. *J. Med. Genet.* **47**, 22–29 (2010).
35. Akhtar, M.W. *et al. In vivo* analysis of MEF2 transcription factors in synapse regulation and neuronal survival. *PLoS ONE* **7**, e34863 (2012).
36. Griciuc, A. *et al.* Alzheimer's disease risk gene *CD33* inhibits microglial uptake of amyloid β. *Neuron* **78**, 631–643 (2013).
37. Lambert, J.C. & Amouyel, P. Genetics of Alzheimer's disease: new evidences for an old hypothesis? *Curr. Opin. Genet. Dev.* **21**, 295–301 (2011).
38. Lambert, J.-C. *et al.* Implication of the immune system in Alzheimer's disease: evidence from genome-wide pathway analysis. *J. Alzheimers Dis.* **20**, 1107–1118 (2010).
39. Jones, L. *et al.* Genetic evidence implicates the immune system and cholesterol metabolism in the aetiology of Alzheimer's disease. *PLoS ONE* **5**, e13950 (2010).
40. Brouwers, N. *et al.* Alzheimer risk associated with a copy number variation in the complement receptor 1 increasing C3b/C4b binding sites. *Mol. Psychiatry* **17**, 223–233 (2012).
41. Bettens, K., Sleegers, K. & Van Broeckhoven, C. Genetic insights in Alzheimer's disease. *Lancet Neurol.* **12**, 92–104 (2013).

Jean-Charles Lambert[1–3,145], Carla A Ibrahim-Verbaas[4,5,145], Denise Harold[6,145], Adam C Naj[7,8,145], Rebecca Sims[6], Céline Bellenguez[1–3], Gyungah Jun[9–11], Anita L DeStefano[11], Joshua C Bis[12], Gary W Beecham[13,14], Benjamin Grenier-Boley[1–3], Giancarlo Russo[15], Tricia A Thornton-Wells[16], Nicola Jones[6], Albert V Smith[17,18], Vincent Chouraki[1–3], Charlene Thomas[6], M Arfan Ikram[4,5,19,20], Diana Zelenika[21], Badri N Vardarajan[9], Yoichiro Kamatani[22], Chiao-Feng Lin[23], Amy Gerrish[6], Helena Schmidt[24], Brian Kunkle[13], Melanie L Dunstan[6], Agustin Ruiz[25], Marie-Thérèse Bihoreau[21], Seung-Hoan Choi[11], Christiane Reitz[26,27], Florence Pasquier[2,28], Paul Hollingworth[6], Alfredo Ramirez[29,30], Olivier Hanon[31], Annette L Fitzpatrick[12,32,33], Joseph D Buxbaum[34–36], Dominique Campion[37], Paul K Crane[38], Clinton Baldwin[9], Tim Becker[39,40], Vilmundur Gudnason[17,18], Carlos Cruchaga[41,42], David Craig[43], Najaf Amin[5], Claudine Berr[44], Oscar L Lopez[45], Philip L De Jager[46,47], Vincent Deramecourt[2,28], Janet A Johnston[43], Denis Evans[48], Simon Lovestone[49], Luc Letenneur[50], Francisco J Morón[51], David C Rubinsztein[52], Gudny Eiriksdottir[18], Kristel Sleegers[53,54], Alison M Goate[41,42], Nathalie Fiévet[1,3], Matthew J Huentelman[55], Michael Gill[56], Kristelle Brown[57], M Ilyas Kamboh[58,59], Lina Keller[60], Pascale Barberger-Gateau[50], Bernadette McGuinness[43], Eric B Larson[38,61], Robert Green[62], Amanda J Myers[63], Carole Dufouil[50], Stephen Todd[43], David Wallon[37], Seth Love[64], Ekaterina Rogaeva[65], John Gallacher[66], Peter St George-Hyslop[52,65,67], Jordi Clarimon[68,69], Alberto Lleo[68,69], Anthony Bayer[66], Debby W Tsuang[70], Lei Yu[71], Magda Tsolaki[72], Paola Bossù[73], Gianfranco Spalletta[73], Petroula Proitsi[49], John Collinge[74], Sandro Sorbi[75,76], Florentino Sanchez-Garcia[77], Nick C Fox[78], John Hardy[79,80], Maria Candida Deniz Naranjo[77], Paolo Bosco[81], Robert Clarke[82], Carol Brayne[83], Daniela Galimberti[84], Michelangelo Mancuso[85], Fiona Matthews[86], European Alzheimer's Disease Initiative (EADI)[87], Genetic and Environmental Risk in Alzheimer's Disease (GERAD)[87], Alzheimer's Disease Genetic Consortium (ADGC)[87], Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE)[87], Susanne Moebus[88], Patrizia Mecocci[89], Maria Del Zompo[90], Wolfgang Maier[29,39], Harald Hampel[91,92], Alberto Pilotto[93], Maria Bullido[94–96], Francesco Panza[97], Paolo Caffarra[98,99], Benedetta Nacmias[75,76], John R Gilbert[13,14], Manuel Mayhaus[100], Lars Lannfelt[101], Hakon Hakonarson[102], Sabrina Pichler[100], Minerva M Carrasquillo[103], Martin Ingelsson[101], Duane Beekly[104], Victoria Alvarez[105], Fanggeng Zou[103], Otto Valladares[23], Steven G Younkin[103], Eliecer Coto[105], Kara L Hamilton-Nelson[13], Wei Gu[100], Cristina Razquin[106], Pau Pastor[106,107], Ignacio Mateo[108,109], Michael J Owen[6], Kelley M Faber[110], Palmi V Jonsson[17,111], Onofre Combarros[108,109], Michael C O'Donovan[6], Laura B Cantwell[23], Hilkka Soininen[112,113], Deborah Blacker[114,115], Simon Mead[74], Thomas H Mosley Jr[116], David A Bennett[71,117], Tamara B Harris[12], Laura Fratiglioni[60,118], Clive Holmes[119], Renee F A G de Bruijn[4,5,20], Peter Passmore[43], Thomas J Montine[120], Karolien Bettens[53,54], Jerome I Rotter[121], Alexis Brice[122,123], Kevin Morgan[57], Tatiana M Foroud[110], Walter A Kukull[32], Didier Hannequin[37], John F Powell[49], Michael A Nalls[124], Karen Ritchie[44,125], Kathryn L Lunetta[11], John S K Kauwe[126], Eric Boerwinkle[127–129], Matthias Riemenschneider[100], Mercè Boada[25,130], Mikko Hiltunen[112,113], Eden R Martin[13,14], Reinhold Schmidt[131], Dan Rujescu[132], Li-San Wang[23], Jean-François Dartigues[50,133], Richard Mayeux[26,27], Christophe Tzourio[134], Albert Hofman[5,20], Markus M Nöthen[135], Caroline Graff[117,136], Bruce M Psaty[12,32,61,137], Lesley Jones[6], Jonathan L Haines[16,138], Peter A Holmans[6], Mark Lathrop[21,22,139], Margaret A Pericak-Vance[13,14], Lenore J Launer[140], Lindsay A Farrer[9–11,141,142], Cornelia M van Duijn[5,20,143], Christine Van Broeckhoven[53,54], Valentina Moskvina[6], Sudha Seshadri[142,145,146], Julie Williams[6,145,146], Gerard D Schellenberg[23,145,146] & Philippe Amouyel[1–3,28,144–146]

[1]INSERM, U744, Lille, France. [2]Université Lille 2, Lille, France. [3]Institut Pasteur de Lille, Lille, France. [4]Department of Neurology, Erasmus MC University Medical Center, Rotterdam, The Netherlands. [5]Department of Epidemiology, Erasmus MC University Medical Center, Rotterdam, The Netherlands. [6]Institute of Psychological Medicine and Clinical Neurosciences, Medical Research Council (MRC) Centre for Neuropsychiatric Genetics & Genomics, Cardiff University, Cardiff, UK. [7]Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [8]Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [9]Department of Medicine (Biomedical Genetics), Boston University School of Public Health, Boston, Massachusetts, USA. [10]Department of Ophthalmology, Boston University School of Medicine, Boston, Massachusetts, USA. [11]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA. [12]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, Washington, USA. [13]John P. Hussman Institute for Human Genomics, University of Miami, Miami, Florida, USA. [14]Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, Florida, USA. [15]Functional Genomics Center Zurich, Eidgenössische Technische Hochschule (ETH)/University of Zurich, Zurich, Switzerland. [16]Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, Tennessee, USA. [17]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [18]Icelandic Heart Association, Kopavogur, Iceland. [19]Department of Radiology, Erasmus MC University Medical Center, Rotterdam, The Netherlands. [20]Netherlands Consortium for Healthy Aging, Leiden, The Netherlands. [21]Centre National de Génotypage, Institut Génomique, Commissariat à l'Energie Atomique, Evry, France. [22]Fondation Jean Dausset, Centre d'Etude du Polymorphisme Humain (CEPH), Paris, France. [23]Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA. [24]Institute for Molecular Biology and Biochemistry, Medical University of Graz, Graz, Austria. [25]Memory Clinic of Fundació Alzheimer Centre Educacional (ACE). Institut Català de Neurociències Aplicades, Barcelona, Spain. [26]Taub Institute on Alzheimer's Disease and the Aging Brain,

Department of Neurology, Columbia University, New York, New York, USA. [27]Gertrude H. Sergievsky Center, Department of Neurology, Columbia University, New York, New York, USA. [28]Centre National de Reference pour les Malades Alzheimer Jeunes (CNR-MAJ), Centre Hospitalier Régional Universitaire de Lille, Lille, France. [29]Department of Psychiatry and Psychotherapy, University of Bonn, Bonn, Germany. [30]Institute of Human Genetics, University of Bonn, Bonn, Germany. [31]University Paris Descartes, Sorbonne Paris V, Broca Hospital, Geriatrics Department, Paris, France. [32]Department of Epidemiology, University of Washington, Seattle, Washington, USA. [33]Department of Global Health, University of Washington, Seattle, Washington, USA. [34]Department of Neuroscience, Mount Sinai School of Medicine, New York, New York, USA. [35]Department of Psychiatry, Mount Sinai School of Medicine, New York, New York, USA. [36]Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, New York, USA. [37]CNR-MAJ, INSERM, U1079, Rouen University Hospital, Rouen, France. [38]Department of Medicine, University of Washington, Seattle, Washington, USA. [39]German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany. [40]Institute for Medical Biometry, Informatics and Epidemiology, University of Bonn, Bonn, Germany. [41]Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri, USA. [42]Hope Center Program on Protein Aggregation and Neurodegeneration, Washington University School of Medicine, St. Louis, Missouri, USA. [43]Ageing Group, Centre for Public Health, School of Medicine, Dentistry and Biomedical Sciences, Queen's University, Belfast, UK. [44]INSERM, U1061, Faculty of Medicine, Hôpital La Colombière, Montpellier, France. [45]Departments of Neurology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA. [46]Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Department of Neurology & Psychiatry, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. [47]Program in Medical and Population Genetics, Broad Institute, Boston, Massachusetts, USA. [48]Rush Institute for Healthy Aging, Department of Internal Medicine, Rush University Medical Center, Chicago, Illinois, USA. [49]Institute of Psychiatry, King's College London, Denmark Hill, London, UK. [50]INSERM, U897, Victor Segalen University, Bordeaux, France. [51]Departamento de Genómica Estructural, Neocodex, Seville, Spain. [52]Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK. [53]Neurodegenerative Brain Diseases Group, Department of Molecular Genetics, VIB, Antwerp, Belgium. [54]Laboratory of Neurogenetics, Institute Born-Bunge, University of Antwerp, Antwerp, Belgium. [55]Neurogenomics Division, Translational Genomics Research Institute, Phoenix, Arizona, USA. [56]Mercer's Institute for Research on Aging, St. James Hospital and Trinity College, Dublin, Ireland. [57]Institute of Genetics, Queen's Medical Centre, University of Nottingham, Nottingham, UK. [58]Department of Human Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. [59]Alzheimer's Disease Research Center, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. [60]Aging Research Center, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet and Stockholm University, Stockholm, Sweden. [61]Group Health Research Institute, Group Health, Seattle, Washington, USA. [62]Department of Medicine and Partners Center for Personalized Genetic Medicine, Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. [63]Department of Psychiatry and Behavioral Sciences, Miller School of Medicine, University of Miami, Miami, Florida, USA. [64]University of Bristol Institute of Clinical Neurosciences, School of Clinical Sciences, Frenchay Hospital, Bristol, UK. [65]Tanz Centre for Research in Neurodegenerative Disease, University of Toronto, Toronto, Ontario, Canada. [66]Institute of Primary Care and Public Health, Cardiff University, Neuadd Meirionnydd, University Hospital of Wales, Heath Park, Cardiff, UK. [67]Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK. [68]Neurology Department. Biomedical Research Institute Sant Pau (IIB Sant Pau). Sant Pau Hospital. Universitat Autònoma de Barcelona, Barcelona, Spain. [69]Center for Networker Biomedical Research in Neurodegenerative Diseases (CIBERNED), Barcelona, Spain. [70]Geriatric Research, Education and Clinical Center (GRECC), Veteran Administration Puget Sound Health Care System (VAPSHCS), Seattle, Washington, USA. [71]Department of Neurological Sciences, Rush University Medical Center, Chicago, Illinois, USA. [72]Department of Neurology, Aristotle University of Thessaloniki, Thessaloniki, Greece. [73]Clinical and Behavioral Neurology, Fondazione Santa Lucia, Rome, Italy. [74]MRC Prion Unit, Department of Neurodegenerative Disease, University College London Institute of Neurology, London, UK. [75]NEUROFARBA (Department of Neuroscience, Psychology, Drug Research and Child Health), University of Florence, Florence, Italy. [76]Centro di Ricerca, Trasferimento e Alta Formazione DENOTHE, University of Florence, Florence, Italy. [77]Department of Immunology, Hospital Universitario Dr. Negrin, Las Palmas de Gran Canaria, Spain. [78]Dementia Research Centre, Department of Neurodegenerative Disease, University College London Institute of Neurology, London, UK. [79]Department of Molecular Neuroscience, Institute of Neurology, London, UK. [80]Reta Lilla Weston Laboratories, Institute of Neurology, London, UK. [81]Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Associazione Oasi Maria Santissima Srl, Troina, Italy. [82]Oxford Healthy Aging Project (OHAP), Clinical Trial Service Unit, University of Oxford, Oxford, UK. [83]Institute of Public Health, University of Cambridge, Cambridge, UK. [84]University of Milan, Fondazione Cà Granda, IRCCS Ospedale Policlinico, Milan, Italy. [85]Neurological Clinic, University of Pisa, Pisa, Italy. [86]MRC Biostatistics Unit, Cambridge, UK. [87]Full lists of members and affiliations appear in the **Supplementary Note**. [88]Urban Epidemiology, Institute for Medical Informatics, Biometry and Epidemiology, University Hospital Essen, University Duisburg-Essen, Essen, Germany. [89]Section of Gerontology and Geriatrics, Department of Clinical and Experimental Medicine, University of Perugia, Perugia, Italy. [90]Section of Neuroscience and Clinical Pharmacology, Department of Biomedical Sciences, University of Cagliari, Cagliari, Italy. [91]Department of Psychiatry, University of Frankfurt, Frankfurt am Main, Germany. [92]Department of Psychiatry, Ludwig Maximilians University, Munich, Germany. [93]Gerontology and Geriatrics Research Laboratory, IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy. [94]Centro de Biología Molecular Severo Ochoa, Consejo Superior de Investigaciones Científicas–Universidad Autónoma de Madrid, Madrid, Spain. [95]Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED), Madrid, Spain. [96]Instituto de Investigación Sanitaria Hospital la Paz (IdIPaz), Madrid, Spain. [97]Departement of Geriatrics, Center for Aging Brain, University of Bari, Bari, Italy. [98]Department of Neuroscience, University of Parma, Parma, Italy. [99]Center for Cognitive Disorders, Azienda Unita Sanitaria Local (AUSL), Parma, Italy. [100]Department of Psychiatry, Saarland University Hospital, Homburg, Germany. [101]Department of Public Health/Geriatrics, Uppsala University, Uppsala, Sweden. [102]Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. [103]Department of Neuroscience, Mayo Clinic, Jacksonville, Florida, USA. [104]National Alzheimer's Coordinating Center, University of Washington, Seattle, Washington, USA. [105]Genetica Molecular, Hospital Universitario Central Asturias, Oviedo, Spain. [106]Neurogenetics Laboratory, Division of Neurosciences, Center for Applied Medical Research, University of Navarra School of Medicine, Pamplona, Spain. [107]CIBERNED, Instituto de Salud Carlos III, Madrid, Spain. [108]Neurology Service, Marqués de Valdecilla University Hospital, University of Cantabria and Instituto de Formación e Investigación Marqués de Valdecilla (IFIMAV), Santander, Spain. [109]CIBERNED, Marqués de Valdecilla University Hospital, University of Cantabria and IFIMAV, Santander, Spain. [110]Department of Medical and Molecular Genetics, Indiana University, Indianapolis, Indiana, USA. [111]Geriatric Department, Landspitali University Hospital, Reykjavik, Iceland. [112]Institute of Clinical Medicine–Neurology, University of Eastern Finland, Kuopio, Finland. [113]Department of Neurology, Kuopio University Hospital, Kuopio, Finland. [114]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA. [115]Department of Psychiatry, Massachusetts General Hospital/Harvard Medical School, Boston, Massachusetts, USA. [116]Department of Medicine (Geriatrics), University of Mississippi Medical Center, Jackson, Mississippi, USA. [117]Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, Illinois, USA. [118]Department of Geriatric Medicine, Karolinska University Hospital Huddinge, Stockholm, Sweden. [119]Division of Clinical Neurosciences, School of Medicine, University of Southampton, Southampton, UK. [120]Department of Pathology, University of Washington, Seattle, Washington, USA. [121]Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA. [122]INSERM UMRS 975, CNRS UMR 7225, Université Pierre et Marie Curie, Centre de Recherche de l'Institut du Cerveau et de la Moëlle Epinière (CRICM), Hôpital de la Salpêtrière, Paris, France. [123]Assistance Publique des Hôpitaux de Paris (AP-HP), Hôpital de la Pitié–Salpêtrière, Paris, France. [124]Laboratory of Neurogenetics, Intramural Research Program, National Institute on Aging, Bethesda, Maryland, USA. [125]Faculty of Medicine, Imperial College, St. Mary's Hospital, London, UK. [126]Department of Biology, Brigham Young University, Provo, Utah, USA. [127]Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, USA. [128]Human Genetics Center, University of Texas Health Science Center at Houston, Houston, Texas, USA. [129]Division of Epidemiology, University of Texas Health Sciences Center at Houston, Houston, Texas, USA. [130]Hospital Universitari Vall d'Hebron–Institut de Recerca, Universitat Autònoma de Barcelona. (VHIR-UAB), Barcelona, Spain. [131]Department of Neurology, Medical University of Graz, Graz, Austria. [132]Department of Psychiatry, University of Halle, Halle, Germany. [133]Centre de Mémoire de Ressources et de Recherche de Bordeaux, Centre Hospitalier Universitaire (CHU) de Bordeaux, Bordeaux, France. [134]INSERM, U708, Victor Segalen University, Bordeaux, France. [135]Institute of Human Genetics, Department of Genomics, Life and Brain Center, University of Bonn, Bonn, Germany. [136]Department of Neurobiology, Karolinska Institutet, Care Sciences and Society, Karolinska Institute–Alzheimer's Disease Research Center (KIADRC), Stockholm, Sweden. [137]Department of Health Service, University of Washington, Seattle, Washington, USA. [138]Vanderbilt Center for Human Genetics Research, Vanderbilt University, Nashville, Tennessee, USA. [139]McGill University and Génome Québec Innovation Centre, Montreal, Quebec, Canada. [140]Laboratory of Epidemiology and Population Sciences, Intramural Research Program, National Institute on Aging, Bethesda, Maryland, USA. [141]Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts, USA. [142]Department of Neurology, Boston University School of Medicine, Boston, Massachusetts, USA. [143]Center for Medical Systems Biology, Leiden, The Netherlands. [144]Centre Hospitalier Régional Universitaire de Lille, Lille, France. [145]These authors contributed equally to this work. [146]These authors jointly directed this work. Correspondence should be addressed to P.A. (philippe.amouyel@pasteur-lille.fr) or J.W. (williamsj@cardiff.ac.uk).

ONLINE METHODS

All case-control studies are described in **Table 1**, in the **Supplementary Note** (see full description of the I-GAP data sets) and in **Supplementary Tables 1**, **7** and **8**. Written informed consent was obtained from study participants or, for those with substantial cognitive impairment, from a caregiver, legal guardian or other proxy, and the study protocols for all populations were reviewed and approved by the appropriate institutional review boards.

Imputation and SNP selection for stage 1 analysis. After quality control criteria were finalized for each individual and each sample collection (SNPs with call rates of <95% were excluded; **Supplementary Note**), IMPUTE2 (ref. 42) or MaCH/Minimac[43] software (**Supplementary Table 2**) was used to impute the genotypes of all participants with haplotypes derived from samples of European ancestry in the 1000 Genome Project (2010 interim release based on the sequence data freeze from 4 August 2010 and phased haplotypes from December 2010). In each data set, SNPs with R^2 or info score quality estimates of less than 0.3, as indicated by MaCH or IMPUTE2, respectively (with these two quality estimates described to be equivalent), were excluded from analyses. Similarly, SNPs with a MAF of <1% were also excluded. After these procedures, a maximum of 8,133,148 SNPs were retained that were present in at least 1 data set.

In each case-control data set, the association of LOAD with genotype dosage was analyzed by a logistic regression model including covariates for age, sex and principal components to adjust for possible population stratification (**Supplementary Table 2**). For the three CHARGE cohorts with incident Alzheimer's disease data, Cox proportional hazards models were used. The four consortia used different but analogous software for these analyses (PLINK[44], SNPTEST[45], ProbABEL[46] or R; **Supplementary Table 2**). Three of these tools were applied to the EADI data set for quality control, and very similar results were observed. After the exclusion of SNPs showing logistic regression coefficient $|\beta| > 5$ or P value equal to 0 or 1, the maximum number of SNPs in any data set was 8,131,643. Each consortium uploaded summarized results for each SNP to an internal I-GAP website for access by members of each consortium.

SNPs genotyped or imputed in at least 40% of Alzheimer's disease cases and 40% of control samples were included in the meta-analysis. This threshold represented the best compromise between maximizing the total number of SNPs and maximizing the number of samples in which the given SNP was present. Indeed, analyzing all SNPs available in at least one study could have greatly increased the risk of false positives. On the other hand, studying SNPs only present in all studies could have led to the removal of SNPs of potential interest, even if those SNPs could have reached adequate statistical power in a more limited number of data sets (false negatives). This approach allowed us to increase homogeneity between studies for some SNPs by excluding poor quality data present only in a limited number of data sets of small size. This last selection step led to a final number of 7,055,881 SNPs in stage 1 analysis.

iSelect microarray design and stage 2 SNP quality control. SNPs associated with Alzheimer's disease risk and exhibiting P value $< 1 \times 10^{-3}$ in stage 1 analysis were selected for replication. A list of 19,532 SNPs was submitted to a devoted Illumina website to develop an iSelect microarray. A total of 16,732 SNPs exhibiting an Illumina score superior or equal to 0.4 were selected for microarray production. During the Illumina production process, 2,287 SNPs failed oligonucleotide synthesis, leading to a final number of 14,445 SNPs for which genotyping was attempted. Genotyping failure led to the further exclusion of an additional 1,999 SNPs as a result of the SNPs (i) having no intensity signal ($n = 559$), (ii) not being polymorphic ($n = 1{,}176$), (iii) only being found in a heterozygous state ($n = 248$) or (iv) having mismatched alleles compared to 1000 Genomes Project data ($n = 16$). Finally, several quality control measures were applied to the remaining 12,446 SNPs to detect potential biases in genotyping. We first tested for discrepancies in allelic frequency between the 1000 Genomes Project EUR reference panel and stage 2 data. Allele frequencies for stage 2 data were estimated on 10,750 controls (see "Stage 2 sample quality control") and after exclusion of Finnish individuals. The allelic test was performed with PLINK, and P values were computed by performing 4,500,000 permutations to avoid an assumption of Hardy-Weinberg equilibrium. In total, 798 SNPs showed a highly significant difference in terms of allele frequency between the 1000 Genomes Project EUR reference panel and stage 2 data ($P < 1 \times 10^{-5}$; **Supplementary Fig. 18**) and were excluded from the analysis.

Other SNP quality control steps were performed separately in data for each country. A SNP was considered of low genotyping quality in a country data set if it had missing genotype data for more than 10% of the individuals, if the P value for the Hardy-Weinberg test in controls was lower than 1×10^{-6} or if the P value for the test for differences in missingness between cases and controls was lower than 1×10^{-6} (see **Supplementary Table 9** for differences in missingness assessed for suggestive and significant hits across European populations). These quality control steps led to the removal of 16 SNPs with low genotyping quality in data from all countries.

After SNP quality control, 11,632 SNPs were considered to be of high genotyping quality in at least 1 country and were analyzed in stage 2. For imputed data sets, SNPs were considered to be of low imputation quality if their info score was <0.3.

Of note, of the 7,086 SNPs that we were unable to successfully genotype, only 471 were not tagged by another successfully genotyped variant (±100 kb) and associated with a P value at least 10 times higher than that of the missing SNP. Because the vast majority of the untagged SNPs exhibited stage 1 P values between 1×10^{-3} and 1×10^{-4} (92%), the likelihood of missing a true association was considered to be low.

Stage 2 sample quality control. The iSelect microarray contained 33,368 SNPs, of which 11,632 were devoted to stage 2. These supplementary SNPs included various genetic data that allowed us to further refine our quality control processes. On the basis of data for all of these SNPs, we excluded individuals who had more than 3% missing genotypes, showed a discrepancy between reported sex and sex estimated on the basis of genetic data (genetic sex) or showed evidence of non-European ancestry. Duplicated and related individuals were identified (**Supplementary Table 10**). Briefly, discrepancies in sex were examined using genetic sex as estimated by PLINK on 40 SNPs on chromosome X. We also removed 93 individuals from a single plate for whom an abnormal number of discrepancies in sex were observed, suggesting that sample mixing had occurred. Using a panel of 261 ancestry-informative markers (AIMs), we performed a principal-component analysis (PCA) on HapMap 2 data with the function SMARTPCA from EIGENSOFT 4.2 software[47]. For each country, individuals were projected onto the first two PCA axes to define their genetic ancestry. Individuals with evidence of non-European ancestry were then identified by applying a Bayesian clustering approach[48] to their coordinates on the first two axes. Identity by descent (IBD) was computed for all pairs of individuals using PLINK, and individuals in a pair with IBD greater than 0.98 were considered to be duplicates. If clinical data for duplicated individuals were discordant, both individuals were excluded. Otherwise, the individual with the greater proportion of missing genotype was excluded. Similarly, IBD was computed for all pairs of individuals in data from each country separately, using 6,764 autosomal SNPs with MAF of >1% and selected to minimize LD. Individuals in pairs with IBD greater than 0.2 were considered to be related and were iteratively removed so as to obtain a sample of unrelated individuals within each country data set.

Finally, individuals with missing clinical data and controls less than 25 years of age were excluded from the analysis. After sample quality control (**Supplementary Table 10**), 19,884 individuals (8,572 cases and 11,312 controls) were available for analysis in stage 2.

Statistical analysis. For the stage 1 meta-analysis, we undertook fixed-effects inverse variance-weighted meta-analysis with the standard errors of the β-coefficient scaled by the square roots of study-specific genomic inflation factors estimated before combining the summary statistics across data sets. Each consortium performed an independent stage 1 meta-analysis after downloading the data files available on the I-GAP website. Two software packages were used for meta-analysis: METAL[49] and GWAMA[50]. Very similar results were generated independently of the software used and as expected, perfect matching was observed between the analyses undertaken by each of the 4 consortia.

For stage 2, association tests were performed for each country for all high-quality genotyped SNPs under an additive model, using logistic regression as implemented in PLINK. Analysis was adjusted for age, sex and principal

components, when necessary. Using SMARTPCA, PCA was performed on individuals from each country separately. Difference in PCA coordinates between cases and controls were tested for the first four principal components, and analysis was further adjusted on principal components if the *P* value of this test was lower than 0.05. PCA for Bonn stage 2 samples was based on GWAS data. For imputed data sets, association tests were performed using likelihood score tests for missing data as implemented in SNPTEST. Genotyped and imputed German samples were analyzed separately, and results were then combined by fixed-effects meta-analysis using the inverse variance approach as implemented in METAL. Using this approach, a fixed-effects meta-analysis was then performed to combine stage 2 results from the different countries. We also performed the analysis separately for each center in stage 2 and combined the results by fixed-effects meta-analysis. Results were similar to those obtained when analysis was performed by country (data not shown).

We finally generated fixed-effects inverse variance–weighted meta-analyses by combining summary statistics across ADGC, CHARGE, EADI, GERAD and stage 2 data by country. At this point, we performed Cochran's *Q* test for heterogeneity and generated I^2 estimates with METAL to evaluate the possible effect of study heterogeneity on the results.

A graphic representation of the association signal in the stage 1 data was generated with LocusZoom software[51] for all the loci of interest reaching a genome-wide significant level after combined stage 1 and stage 2 analyses.

PAF was calculated using the Levin equation[52].

Annotation of I-GAP top SNPs for eQTLs. To gain further biological insights, we explored reported associations between SNPs in the top I-GAP loci and gene expression. We first selected all SNPs that reached genome-wide significance (P value $\leq 5 \times 10^{-8}$) in the combined stage 1 and stage 2 analysis and were located in a 500-kb window upstream or downstream of the top SNP at each locus (**Table 2**). We then searched for published data on gene expression associated with each of these SNPs in the eQTL database from the Pritchard laboratory (see URLs). For each reported eQTL gene and each type of eQTL association as defined in this database, we then counted the number of reported eQTL SNPs and selected the one with the lowest *P* value.

42. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
43. Li, Y., Willer, C.J., Ding, J., Scheet, P. & Abecasis, G.R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* **34**, 816–834 (2010).
44. Purcell, S. *et al.* PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
45. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies via imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).
46. Aulchenko, Y.S., Struchalin, M.V. & van Duijn, C.M. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics* **11**, 134 (2010).
47. Patterson, N., Price, A.L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
48. Bellenguez, C. *et al.* A robust clustering algorithm for identifying problematic samples in genome-wide association studies. *Bioinformatics* **28**, 134–135 (2012).
49. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
50. Mägi, R. & Morris, A.P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* **11**, 288 (2010).
51. Pruim, R.J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).
52. Levin, M.L. The occurrence of lung cancer in man. *Acta Unio Int. Contra Cancrum* **9**, 531–541 (1953).

doi:10.1038/ng.2802