Alzheimer's & Dementia 11 (2015) 658-671

Convergent genetic and expression data implicate immunity in Alzheimer's disease

International Genomics of Alzheimer's Disease Consortium (IGAP)[†]

Abstract

Background: Late-onset Alzheimer's disease (AD) is heritable with 20 genes showing genome-wide association in the International Genomics of Alzheimer's Project (IGAP). To identify the biology underlying the disease, we extended these genetic data in a pathway analysis.

Methods: The ALIGATOR and GSEA algorithms were used in the IGAP data to identify associated functional pathways and correlated gene expression networks in human brain.

Results: ALIGATOR identified an excess of curated biological pathways showing enrichment of association. Enriched areas of biology included the immune response ($P = 3.27 \times 10^{-12}$ after multiple testing correction for pathways), regulation of endocytosis ($P = 1.31 \times 10^{-11}$), cholesterol transport ($P = 2.96 \times 10^{-9}$), and proteasome-ubiquitin activity ($P = 1.34 \times 10^{-6}$). Correlated gene expression analysis identified four significant network modules, all related to the immune response (corrected $P = .002$–$.05$).

Conclusions: The immune response, regulation of endocytosis, cholesterol transport, and protein ubiquitination represent prime targets for AD therapeutics.

Keywords: Alzheimer's disease; Dementia; Neurodegeneration; Immune response; Endocytosis; Cholesterol metabolism; Ubiquitination; Pathway analysis; ALIGATOR; Weighted gene co-expression network analysis

1. Background

Alzheimer's disease (AD) affects more than five million Americans: one in eight at the age of >65 years and represents >60% of the six million dementia cases in Europe [1–3]. It is the commonest cause of dementia and imposes a large socioeconomic burden on individuals, their families, and society. Prevalence is estimated to treble by 2050; thus, understanding the mechanisms underlying this disease and developing treatments for it are essential. This study uses the largest genome-wide association study (GWAS) sample yet assembled for late-onset AD [4] and is the first to combine GWAS and expression data in a systematic search for the biological pathways underlying the genetic susceptibility to this disorder.

Much of our current understanding of the mechanisms that contribute to AD derives from the genetics of Mendelian forms of the disease: mutations in *APP*, *PSEN1*, and *PSEN2* cause early-onset forms of AD and underpin the amyloid cascade hypothesis [5]. Although amyloid deposition is diagnostic of AD, its etiologic contribution to the majority of common late-onset AD (LOAD) is unclear, and therapeutic strategies addressing the amyloid cascade hypothesis have been unsuccessful [6]. Therefore, other therapeutic avenues must be identified and targeted.

LOAD is genetically complex with 56% to 79% heritability [7]. In the Genetic and Environmental Risk in Alzheimer's Disease data set [8], approximately 20% of the total trait variance was accounted for by single-nucleotide polymorphisms (SNPs) on the GWAS chip outside the *APOE* region [9], with the ε4 allele of the *APOE* gene [10] accounting for a similar amount [9,11]. However, a substantial proportion of the genetic variance of late-onset AD is not accounted for by the 20 susceptibility genes currently identified [11]. The remaining genetic

[†]A list of members of the International Genomics of Alzheimer's Disease Consortium (IGAP) can be found in the Appendix at the end of this article.

*Corresponding authors: williamsj@cardiff.ac.uk and philippe.amouyel@pasteur-lille.fr

1552-5260/$ - see front matter © 2015 Published by Elsevier Inc. on behalf of The Alzheimer's Association.
http://dx.doi.org/10.1016/j.jalz.2014.05.1757

variance is likely to be due to both susceptibility genes of small effect that current sample sizes are insufficient to detect and rare variants, such as the coding variants in *TREM2* [12], that are poorly tagged by common variants in GWAS panels. In addition, individual genome-wide significant (GWS) genes identified in such studies may themselves not form good therapeutic targets, and the areas of biology that they highlight may only give a partial view of the potential therapeutic landscape. To gain the maximum useful information about causative pathways that may underpin LOAD and be prime targets for pharmaceutical intervention, we performed a robust pathway and integrated gene expression analysis using the largest available GWAS for AD [4].

2. Methods

2.1. Samples and genetic data

The sample comprised 17,008 AD cases and 37,646 control subjects in the primary GWAS analysis, with 8752 AD cases and 11,312 control subjects in the replication/extension sample and is described in detail elsewhere [4]. Only selected SNPs were genotyped in the replication/extension sample (see Online Methods).

2.2. Pathway analyses

We explored whether particular biological pathways were enriched for genetic associations [13,14] in the International Genomics of Alzheimer's Project (IGAP) data [4]. We used ALIGATOR [13,14] to test whether genes containing signals below the genome-wide significance threshold contribute to a pathway signal. ALIGATOR defines significant genes as having a best single-SNP P value less than a preset threshold. The resulting list of significant genes is compared with replicate gene sets generated by sampling SNPs randomly (thereby correcting for gene size). The method also controls for linkage disequilibrium (LD) between genes and multiple testing of nonindependent pathways (see Online Methods). Brown's method [15] was used to test pathway enrichment in the replication data. This method combines multiple SNPs together, explicitly correcting for both linkage disequilibrium (LD) between SNPs and the number of SNPs per gene (see Online Methods). Thus, correction for gene size was applied at both stages of the analysis. We interrogated the externally curated gene ontology and KEGG and MSigDB functional pathway collections (see Online Methods).

2.3. Expression correlation analyses

We used the expression data from Gibbs et al. [16] and performed weighted gene correlation network analysis (WGCNA) using the WGCNA package [17], separately on each tissue type to identify clusters of highly correlated genes called "modules." These modules were then tested for enrichment of genome-wide association signal in ALIGATOR.

3. Results

The sub-GWS variation in the IGAP data contains genetic signal, manifest by a significant excess of SNPs at all significance threshold up to $P = .05$ (Supplementary Table 1). This signal is unlikely to be due to uncorrected stratification because each of the individual Caucasian GWAS samples in the IGAP meta-analysis was corrected for ethnic variation using principal components [18].

We first identified a significant excess of biological pathways enriched for association signal in the IGAP data (Table 1 and Supplementary Table 2). Using the most significant 18,472 SNPs ($P < 8.32 \times 10^{-4}$) from IGAP [4], covering the top 5% of genes, 177 significantly enriched ($P < .01$) curated pathways were identified by ALIGATOR. To ensure that the excess of pathways was not an artifact of LD with genes of strong effect, we performed secondary enrichment analyses removing all genes that lay in the LD region of *APOE* or any of the GWS genes from the IGAP [4] study. A significant excess of enriched pathways remained (Table 1), showing that the pathways showed significant enrichment independent of the "known" AD genes. Likewise, a significant excess of enriched pathways was observed when the P-value criterion for defining significant SNPs and genes was varied (Supplementary Table 3).

Many of the 177 pathways with $P < .01$ in ALIGATOR are still significantly enriched after removing the *APOE*

Table 1
Significant excess of enriched pathways remain after removing *APOE* and the genome-wide significant genes

	Enrichment $P < .05$		Enrichment $P < .01$		Enrichment $P < .001$	
Genes removed (number of genes)	Number of pathways	P	Number of pathways	P	Number of pathways	P
None	542	<.0002	177	<.0002	40	<.0002
APOE + 1 Mb (77)	446	.0002	131	.0006	28	.0008
APOE + 1 Mb + GWS (98)	402	.0020	116	.0008	23	<.0002
APOE + 1 Mb + GWS+1 Mb (552)	336	.0094	93	.0066	22	.0018

Abbreviations: GWS, genome-wide significant; SNP, single-nucleotide polymorphism.

NOTE. Genes containing a SNP with $P < 8.32 \times 10^{-4}$ were counted as significant. This corresponds to the top 5% of genes (ranked by most significant SNP) when no genes are removed. The zero-kilobase window was used to assign SNPs to genes.

region and genes within 1 Mb of a GWS SNP (Table 2 and Supplementary Table 4). They remain significantly enriched under a range of *P*-value criteria for defining significant SNPs and are also significant under a GSEA analysis [19,20]. This robustness to analysis parameters and methods gives confidence that the enrichments observed by ALIGATOR are genuine.

Of the 177 pathways significant at $P < .01$ in the ALIGATOR analysis of the IGAP GWAS, 119 are significant ($P < .05$) in the replication sample. This is more than that expected by chance (see Online Methods), a further confirmation that the pathways highlighted by the ALIGATOR analysis contain genuine signals. Notably, pathway SNPs had significantly lower replication *P* values than nonpathway SNPs even after correcting for their *P* values in the original IGAP GWAS (two-sided $P = .0237$, see Online Methods). Thus, the pathway analyses highlighted which among a set of associated, but not GWS, SNPs are likely to replicate and therefore be enriched for true signals. To obtain the most strongly enriched pathways in the entire data set, the *P* values from the ALIGATOR analysis were combined with those from the replication study using the Fisher method and corrected for multiple testing of 9816 pathways using the Sidak formula. Forty-five pathways

Table 2
Clusters of significant pathways in combined IGAP GWAS and replication data (Sidak-corrected *P* value <.05)

Cluster	Pathway number	Number of genes	Number of sig	*P* value	*P* value no GWS	Description
1	GO: 2455	32	5	3.27×10^{-12}	5.72×10^{-1}	Humoral immune response mediated by circulating immunoglobulin
1	GO:50776	421	29	3.24×10^{-9}	1.57×10^{-4}	Regulation of immune response
1	GO: 2684	421	31	3.95×10^{-9}	2.11×10^{-4}	Positive regulation of immune system process
1	GO:50778	271	21	1.55×10^{-7}	6.65×10^{-4}	Positive regulation of immune response
1	KEGG 4664	78	13	5.76×10^{-4}	2.18×10^{-2}	Fc epsilon RI signaling pathway
2	GO:60627	140	20	1.31×10^{-11}	2.00×10^{-1}	Regulation of vesicle-mediated transport
2	GO:30100	88	14	6.76×10^{-10}	1.06×10^{-1}	Regulation of endocytosis
2	GO:45806	19	6	3.91×10^{-7}	1.77×10^{-2}	Negative regulation of endocytosis
2	GO:48261	6	3	3.89×10^{-6}	9.82×10^{-1}	Negative regulation of receptor-mediated endocytosis
2	GO:48259	30	6	6.19×10^{-5}	1.00	Regulation of receptor-mediated endocytosis
3	GO:30301	41	8	2.96×10^{-9}	2.51×10^{-1}	Cholesterol transport
3	GO:43691	16	5	3.90×10^{-9}	2.78×10^{-1}	Reverse cholesterol transport
3	GO:15918	42	8	3.91×10^{-9}	3.15×10^{-1}	Sterol transport
3	GO:34366	8	2	6.40×10^{-7}	N/A	Spherical high-density lipoprotein particle
4	KEGG 4640	81	11	1.05×10^{-8}	4.91×10^{-1}	Hematopoietic cell lineage
5	GO:32434	40	5	1.34×10^{-6}	1.00	Regulation of proteasomal ubiquitin-dependent protein catabolic process
5	GO:51437	70	9	2.60×10^{-3}	2.60×10^{-3}	Positive regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle
5	GO:51439	76	9	3.82×10^{-3}	3.82×10^{-3}	Regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle
5	REACT 440	108	11	3.89×10^{-3}	3.89×10^{-3}	REACTOME_CELL_CYCLE_CHECKPOINTS
5	GO:51443	77	9	9.62×10^{-3}	9.62×10^{-3}	Positive regulation of ubiquitin-protein ligase activity
6	REACT 539	261	25	2.95×10^{-5}	6.93×10^{-2}	REACTOME_HEMOSTASIS
7	GO:30131	31	7	1.20×10^{-3}	9.13×10^{-1}	Clathrin adaptor complex
7	GO:30119	32	7	1.53×10^{-3}	9.54×10^{-1}	AP-type membrane coat adaptor complex
7	GO:44433	301	31	1.01×10^{-2}	1.00	Cytoplasmic vesicle part
7	GO:30122	9	4	1.29×10^{-2}	1.00	AP-2 adaptor complex
7	GO:30118	39	7	1.35×10^{-2}	1.00	Clathrin coat
8	GO: 6457	200	12	1.60×10^{-3}	1.00	Protein folding

Abbreviations: IGAP, International Genomics of Alzheimer's Project; GWAS, genome-wide association study; GWS, genome-wide significant.

NOTE. To obtain the most strongly enriched pathways in the entire data set (IGAP GWAS and replication), the *P* values from the ALIGATOR analysis (counting the top 5% of genes as significant) were combined with those from the replication study using the Fisher method. The resulting *P* values from the combined samples were corrected for multiple testing of 9816 pathways using the Sidak formula. For each pair of gene sets, an overlap measure K was defined as the number of genes common to both sets divided by the number of genes in the smaller data set. A gene set was assigned to a cluster if the average K between it and the gene sets already in the cluster was greater than 0.4. If it was not possible to assign a gene set to an existing cluster, a new cluster was started. This procedure was carried out recursively, in descending order of enrichment significance. Clusters containing a significant pathway are listed here, and where >5 pathways are significant, only the five most significant pathways in each cluster are shown. A complete list of pathways significant at $P < .01$ in the ALIGATOR analysis of the IGAP GWAS data is given in Supplementary Table 4. "No GWS" refers to analyses in which genes containing an SNP genome-wide significant ($P < 5 \times 10^{-8}$) in the IGAP GWAS data set (and thus expected to be strongly significant in the replication data set) are removed from the analysis of the replication data.

were significant after multiple testing correction (Sidak $P < .05$) in the combined data set. These pathways are listed in Table 2, grouped into clusters by gene membership such that pathways with >40% of genes in common are gathered in a cluster.

This multiple testing correction may be considered conservative because it assumes that the pathways are independent, whereas in fact, there is considerable genetic overlap between them. Sidak-corrected *P* values for the combined IGAP GWAS and replication data sets are therefore given in Supplementary Table 4 for all 177 pathways significant at $P < .01$ in the ALIGATOR analysis of the IGAP GWAS. Redundant pathways (i.e., those with high genetic overlap with other pathways) were not pruned from our analysis because it is not clear a priori which pathways will give the most significant enrichment and should thus be retained. Pruning a posteriori (i.e., by choosing the most significant pathways) will bias the significance of the combined discovery and replication *P* values (making the correction for multiple testing of pathways anticonservative). The pathway clusters given in Table 2 and Supplementary Table 4 are intended to aid interpretation of our results in light of shared gene membership between pathways, by highlighting areas of biology rather than individual pathways.

The clusters of multiple pathways were related to the broad categories of immune response, regulation of endocytosis, cholesterol transport, protein ubiquitination, and clathrin, with the first three of these being particularly strongly enriched for signal. Because one would expect SNPs showing strong association to be significant on replication regardless of biology, the analysis was repeated removing genes containing a GWS SNP in the IGAP GWAS from the analysis of the replication data. From Table 2, it can be seen that the immune-related and ubiquitination pathways are still highly significant. Sidak-corrected *P* values for all 177 pathways significant at $P < .01$ in the ALIGATOR analysis are shown in Supplementary Table 4. The relationship between the enriched pathways is shown by their shared gene membership (Fig. 1). Table 3 lists genes in the clusters identified in Table 2 that are counted as significant (best single-SNP $P < 8.32 \times 10^{-4}$) in the ALIGATOR analysis of the IGAP GWAS and also gene-wide significant (gene-wide $P < .05$) in both the IGAP GWAS and the replication data. *P* values for all genes counted as significant in the ALIGATOR analysis from the 177 pathways enriched at $P < .01$ are given in Supplementary Table 5.

In contrast to ALIGATOR, GSEA uses all genes (rather than using a threshold) and weights these by their significance so may highlight different biological signals. We therefore performed a secondary analysis of all pathways using GSEA. Pathways significant under GSEA but not ALIGATOR are shown in Supplementary Table 6. Most of these pathways relate to areas of biology already highlighted in the ALIGATOR analysis, the exceptions being synapse, neuronal differentiation, and calcium signaling (Supplementary Table 6). Genes contributing to these pathway signals that are significant in both the IGAP GWAS and the replication studies are listed in Supplementary Table 7. Notably, these pathways contain large genes. In addition to the differences between ALIGATOR and GSEA described previously, the Simes correction for gene size used by GSEA is less stringent for large genes than that used by ALIGATOR, thereby explaining the discrepancy in the results between the methods.

In the ALIGATOR analysis, 73.2% of the top 5% of genes mapped to a pathway, leaving a substantial minority of genes unannotated: in addition, many annotated genes may possess other functions not currently annotated. Genes with correlated expression patterns display functional similarities,

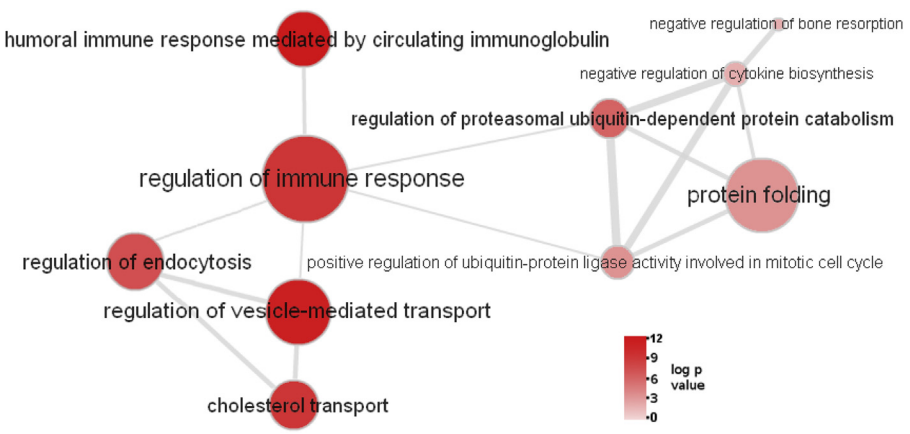

Fig. 1. The pathways highlighted by ALIGATOR ontology analyses are related. The network was generated in ReVIGO [32] using gene ontology processes identified in ALIGATOR only. Bubble size (and label font size) reflects the frequency of the gene ontology (GO) term in the GOA database and bubble color reflects pathway *P* value. Similar GO terms are linked by edges (lines) in the network in which line width reflects the degree of similarity between pathways but line length is arbitrary. Strong relationships are revealed between negative regulation of endocytosis and cholesterol transport, and many of the pathways are related to the immune response process.

Table 3
Genes in the significant ALIGATOR pathway clusters

Entrez ID	Gene symbol	Best *P* (IGAP)	Gene-wide *P* (IGAP)	Best *P* (REP)	Gene-wide *P* (REP)
Cluster 1: immune response					
1191	*CLU*	2.48×10^{-17}	5.14×10^{-15}	1.06×10^{-10}	2.60×10^{-8}
1378	*CR1*	3.65×10^{-15}	3.46×10^{-7}	3.82×10^{-11}	5.06×10^{-8}
2206	*MS4A2*	3.28×10^{-10}	3.68×10^{-9}	1.81×10^{-4}	6.54×10^{-6}
3117	*HLA-DQA1*	3.38×10^{-9}	1.20×10^{-5}	5.33×10^{-5}	8.89×10^{-3}
3123	*HLA-DRB1*	1.24×10^{-8}	6.54×10^{-6}	5.80×10^{-5}	1.13×10^{-2}
3127	*HLA-DRB5*	2.87×10^{-7}	4.78×10^{-5}	4.56×10^{-4}	5.23×10^{-3}
1380	*CR2*	9.35×10^{-7}	2.99×10^{-2}	5.76×10^{-5}	6.41×10^{-3}
3119	*HLA-DQB1*	2.97×10^{-6}	3.88×10^{-5}	3.58×10^{-4}	6.45×10^{-3}
3635	*INPP5D*	6.62×10^{-6}	3.33×10^{-3}	9.93×10^{-6}	1.02×10^{-4}
102	*ADAM10*	1.45×10^{-4}	2.90×10^{-2}	1.13×10^{-2}	2.71×10^{-2}
Cluster 2: endocytosis					
274	*BIN1*	3.72×10^{-16}	4.75×10^{-6}	3.15×10^{-11}	5.27×10^{-9}
8301	*PICALM*	1.91×10^{-12}	1.20×10^{-8}	2.57×10^{-7}	2.97×10^{-7}
2206	*MS4A2*	3.28×10^{-10}	3.68×10^{-9}	1.81×10^{-4}	6.54×10^{-6}
1265	*CNN2*	1.19×10^{-6}	3.07×10^{-3}	2.91×10^{-4}	2.11×10^{-3}
Cluster 3: cholesterol transport					
1191	*CLU*	2.48×10^{-17}	5.14×10^{-15}	1.06×10^{-10}	2.60×10^{-8}
10,347	*ABCA7*	1.70×10^{-9}	3.00×10^{-7}	1.43×10^{-6}	1.02×10^{-6}
Cluster 4: hematopoietic cell lineage					
1378	*CR1*	3.65×10^{-15}	3.46×10^{-7}	3.82×10^{-11}	5.06×10^{-8}
3123	*HLA-DRB1*	1.24×10^{-8}	6.54×10^{-6}	5.80×10^{-5}	1.13×10^{-2}
3127	*HLA-DRB5*	2.87×10^{-7}	4.78×10^{-5}	4.56×10^{-4}	5.23×10^{-3}
1380	*CR2*	9.35×10^{-7}	2.99×10^{-2}	5.76×10^{-5}	6.41×10^{-3}
Cluster 5: protein ubiquitination					
1191	*CLU*	2.48×10^{-17}	5.14×10^{-15}	1.06×10^{-10}	2.60×10^{-8}
5702	*PSMC3*	3.70×10^{-6}	3.04×10^{-5}	1.55×10^{-2}	1.15×10^{-2}
5434	*POLR2E*	1.94×10^{-5}	6.93×10^{-3}	1.08×10^{-3}	1.26×10^{-4}
6827	*SUPT4H1*	1.94×10^{-4}	2.26×10^{-2}	2.27×10^{-2}	2.27×10^{-2}
5706	*PSMC6*	2.98×10^{-4}	1.25×10^{-2}	3.99×10^{-2}	3.79×10^{-2}
6878	*TAF6*	4.22×10^{-4}	1.66×10^{-2}	6.41×10^{-4}	6.41×10^{-4}
Cluster 6: hemostasis					
1191	*CLU*	2.48×10^{-17}	5.14×10^{-15}	1.06×10^{-10}	2.60×10^{-8}
3635	*INPP5D*	6.62×10^{-6}	3.33×10^{-3}	9.93×10^{-6}	1.02×10^{-4}
Cluster 7: clathrin/AP2 adaptor complex					
1191	*CLU*	2.48×10^{-17}	5.14×10^{-15}	1.06×10^{-10}	2.60×10^{-8}
8301	*PICALM*	1.91×10^{-12}	1.20×10^{-8}	2.57×10^{-7}	2.97×10^{-7}
9179	*AP4M1*	2.16×10^{-4}	2.13×10^{-3}	3.74×10^{-4}	1.57×10^{-4}
Cluster 8: protein folding					
1191	*CLU*	2.48×10^{-17}	5.14×10^{-15}	1.06×10^{-10}	2.60×10^{-8}
664618	*HSP90AB4*	4.62×10^{-4}	2.30×10^{-3}	2.19×10^{-2}	4.48×10^{-2}

Abbreviation: IGAP, International Genomics of Alzheimer's Project; REP, replication.

NOTE. Gene-wide *P* values were calculated using the Brown method (see Online Methods). Genes shown are those counted as significant (best $P < 8.32 \times 10^{-4}$) in the ALIGATOR analysis of the IGAP GWAS data that are also significant (gene-wide $P < .05$) in the replication data. Genes in the vicinity of *APOE* are not included in this table because this region was not genotyped in the replication sample. Such genes were highly significant in the meta-analysis ($P < 1 \times 10^{-10}$) and comprise *APOC1/2* in cluster 2; *APOE* and *APOC1/2/4* in cluster 4; *APOE*, *PVRL*, *BCL3*, and *PVR* in cluster 7; *APOE* in cluster 8.

and Zhang et al. [21] highlighted modules of co-expressed genes as being important in the etiology of LOAD. Therefore, to overcome the annotation gap and access biologically related signal across the entire genome, we created modules of brain co-expressed genes and tested them for enrichment of association signal in the IGAP GWAS. The data set we used consisted of gene expression data from four brain regions in a sample of approximately 150 control brains [16] and was independent from that used by Zhang et al. [21]. We used control individuals rather than AD cases so that correlations between expression levels would not be confounded by neuron loss. A WGCNA [17] gave 117 modules of co-expressed genes in these data (see Online Methods and Supplementary Table 8): these 117 modules were tested for enrichment of association signal in the IGAP GWAS using ALIGATOR. Four modules were found to be significantly enriched after correcting for multiple testing, and these enrichments were robust to varying *P*-value criteria and analysis methods (Supplementary Table 9). The four significantly enriched modules, one from each brain region,

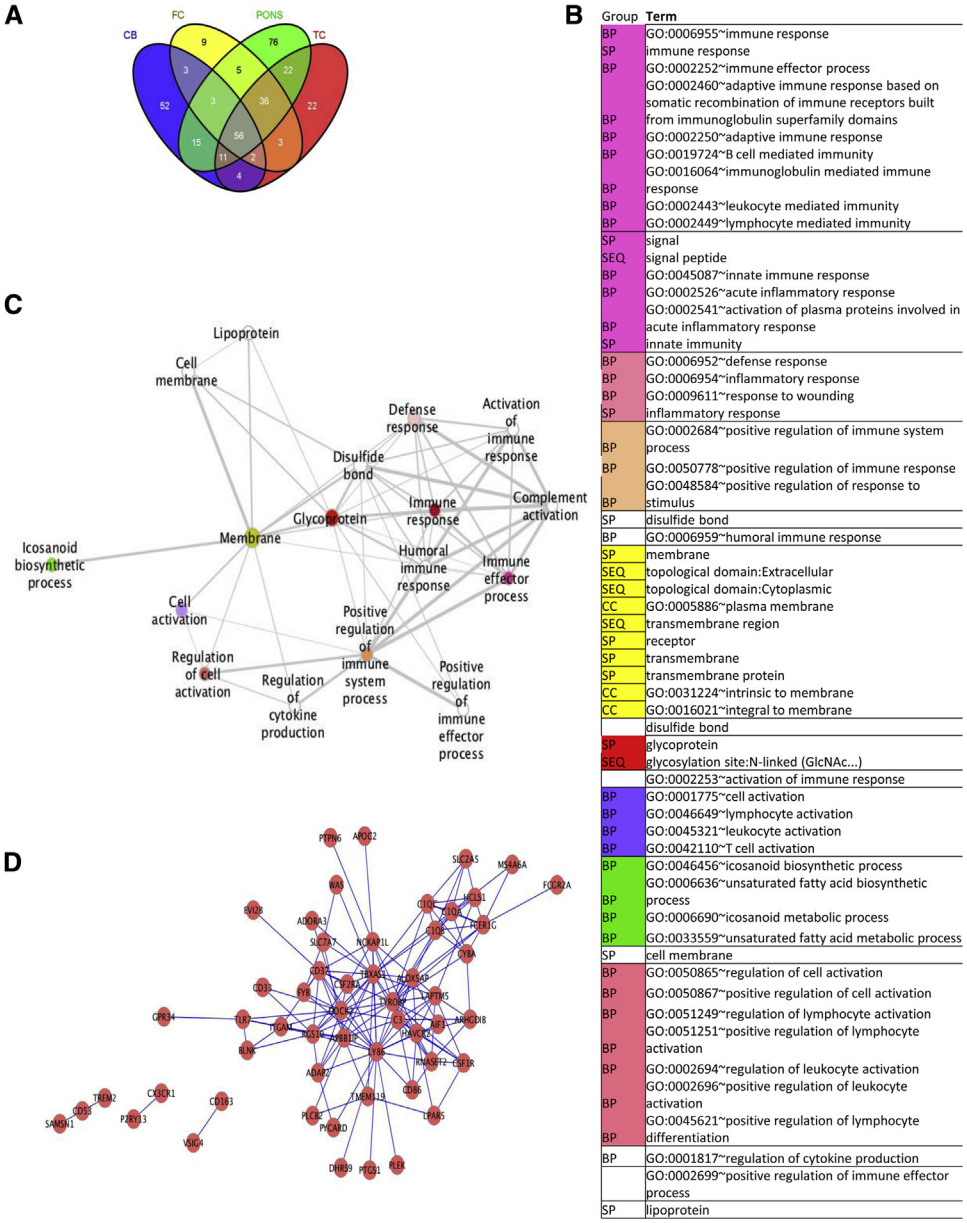

Fig. 2. The immune response is enriched in gene co-expression modules from human brain. (A) Venn diagram indicating the number of genes in common across the four modules that were found to be significantly enriched in the IGAP GWAS using ALIGATOR after correcting for multiple testing. Each significant module originates from a different brain region as indicated here. Colored nodes indicate a multiterm cluster: the related terms represented by each node are given in B, in increasing significance order. Sources of the functional terms are BP = GOTERM_BP_FAT: gene ontology biological processes in DAVID's GO Fat Database; CC = GOTERM_CC_FAT: cellular component terms in DAVID's GO Fat Database; SP = SP_PIR_KEYWORDS: keywords in the UniProt (Swiss-Prot/Protein Information Resource) database; SEQ = UP_SEQ_FEATURE: UniProt sequence annotation feature. The full data are available in Supplementary Table 8. (C) Network showing the pathways significantly enriched for gene membership among the 151 genes present in at least two of the four most significantly enriched expression modules: the principal biological themes were derived from DAVID [33,34] analysis. Terms from the analysis were filtered at 0.05% false discovery Rate (FDR), progressively clustered according to average gene similarity at a threshold of 90% and rendered on Cytoscape with the Enrichment Map plugin [35,36]. The diagram shows only the principal (lowest FDR) term for each of the clusters, and white nodes indicate a single term that does not cluster with other groups. (D) Network showing the strongest correlations in expression (>0.9 in at least one brain area) between genes present in at least two of the four most significantly enriched expression modules. Cb, cerebellum; FC, frontal cortex; TC, temporal cortex.

are related to the immune response and have overlapping gene membership (Fig. 2).

The extent to which the overlap in gene membership between these modules is related to the GWAS signal was investigated by examining genes that occurred in multiple modules and testing these for enrichment using ALIGATOR and GSEA (Supplementary Table 10). It can be seen that the set of 151 of 294 genes that are present in two or

more modules consistently showed the most significant enrichment of IGAP signal across a variety of test criteria. Conversely, the set of 143 genes present in only one module showed no significant enrichment for association signal, highlighting the utility of using multiple data sets to produce meaningful co-expression modules. Fig. 2 shows the strongest correlations (>0.9 in at least one brain area) between the 151 genes present in two or more modules. It can be seen that the *TYROBP* gene highlighted by Zhang et al. [21] as an important causal regulator is also a hub gene in this network. Pathways significantly enriched in the 151 genes present in two or more modules are shown in Fig. 2, clustered by gene membership. Many of the enriched pathways are immune related, but some are related to fatty acid metabolism and lipoprotein, further corroborating the results of our analysis of the IGAP GWAS data. A list of the 151 genes is shown in Supplementary Table 11.

We also directly tested the modules described by Zhang et al. [21] for enrichment of association signal in the IGAP GWAS data (Supplementary Table 12). No single module was significantly enriched after correction for multiple testing of modules ("corr P" $< .05$), but the most significantly enriched modules are immune related. Interestingly, the immune/microglia module highlighted by Zhang et al. [21] (#1, yellow) did not show significant enrichment for association signal in the IGAP GWAS under ALIGATOR analysis, although it did show moderate enrichment under GSEA. However, the 108 genes common to both this module and the set of 151 genes present in two or more of the four most significantly enriched modules in our analysis do show enrichment, which becomes progressively more significant as increasingly stringent criteria are used to select significant SNPs and genes (Supplementary Table 13). The genes that are in the Zhang module but not our set of 151 genes show no significant enrichment for association signal under either ALIGATOR or GSEA analysis.

4. Discussion

This analysis reveals pathways etiologically related to AD in addition to those identified previously [14,22]. The present sample [4] is larger than any used before and was imputed on a dense reference panel, giving improved gene coverage, and is therefore likely to be more powerful to detect real associations than any previous study. A larger set of pathways has been analyzed than previously and annotations have changed, so gene membership of pathways is not identical to previous studies, although a substantial proportion of genes still fall into the annotation gap and are not currently mapped to any pathway. In the present analysis, we also clustered genes that were within 1 Mb of each other together in ALIGATOR to prevent counting a single signal more than once. Secondary analyses were also performed removing genes in the *APOE* LD region and within 1 Mb of the GWS genes. This was done to prevent pathway enrichments being biased by LD between pathway genes and neighboring genes of strong effect and to test whether there were significant pathway enrichments independent of "known" AD genes. Such enrichments would increase the interest of novel pathways and genes highlighted by the main analysis. Despite these differences, many of the pathways previously identified [14] show enrichment in the IGAP data set (Supplementary Table 14). These include cholesterol transport, immunity, and the synaptic transmission, cholinergic pathway, the latter being the target of the cholinesterase class of drugs widely used in AD.

We used both GWAS and expression data to detect functional pathways associated with AD. ALIGATOR analysis of combined IGAP-GWAS and replication samples highlights four main areas of biology: the immune response, regulation of endocytosis, cholesterol transport, and protein ubiquitination. The immune response is particularly significant in the replication sample, even when GWS genes from the IGAP GWAS are excluded. The replication SNPs were not chosen for pathway membership and do not survey the genome randomly, so the lack of significance in some pathway clusters once the GWS genes are removed does not mean that there is no excess signal in these pathways: this may simply not have been measured. However, these data indicate that further genes that are involved in the immune response are likely to be implicated in LOAD. Both regulation of endocytosis and cholesterol transport are functions also implicated by the GWS genes, whereas the immune response and protein ubiquitination contain fewer GWS signals [4]. The most significant signals in the GSEA analysis relate to the same biology but add some additional categories related to neurologic biology including the synapse and neuronal projection development along with calcium-related signaling, not revealed by ALIGATOR. It is notable that these areas of biology are linked by common gene membership (Fig. 1), and their interdependence may also be important in susceptibility to AD.

The additional immune response genes implicated in cluster 1 (Table 3) are plausible AD risk genes: *CR2* encodes complement receptor 2, which is present on subsets of B-cells as is the GWS *CR1*. *HLA-DQB1* is in the chromosome 6 HLA locus in common with several GWS loci. *INPP5D* is GWS once replication analyses are taken into account [4]. As well as being annotated as having immune system activity, *ADAM10* has been proposed as a candidate α-secretase that cleaves amyloid precursor protein to prevent the production of β-amyloid [23]. The protein ubiquitination cluster 5 (Table 3) includes two ATPase subunits of the 19S proteasome, *PSMC3* and *PSMC6*, and three proteins involved in transcriptional control, *POLR2E, SUPT4H1*, and *TAF6*. *CNN2*, encoding calponin 2, believed to regulate the actin cytoskeleton [24] appears in the endocytosis cluster, although it can also

regulate phagocytosis in macrophages [25]. The additional genes from GSEA include *CHRNA2* encoding the neuronal cholinergic receptor, nicotinic, α 2 and *RAPSN*, the receptor-associated protein of the synapse, both of which appear in the synaptic transmission, cholinergic pathway (Supplementary Table 13). *CAV1* encodes caveolin 1, which can interact with apoE [26] and is found in caveolae, areas of cholesterol-rich lipid raft involved in endocytosis. *CACNA1D* encodes the calcium channel, voltage-dependent, L-type, α 1D subunit, one of a series of α subunits that confer channel-specific properties; influences insulin secretion; and is a risk gene for type 2 diabetes [27]. Finally, *APP* itself is highlighted in this analysis: it is annotated in both the synapse and neuronal clusters. Recent findings show that there is at least one rare protective coding variant in *APP* seen in late-onset AD [28], and this signal may reflect this or other relatively rare variants.

Convergent evidence for the importance of the immune response in AD susceptibility was obtained by performing WGCNA on expression data from four brain regions. The four modules that were significantly enriched for association in the IGAP GWAS after multiple testing correction were related to the immune response and shared multiple genes in common: *INPP5D* is GWS [4] and was the only GWS gene found in these modules. The enrichment for association was driven by genes that occurred in two or more of these modules. None of the modules from Zhang et al. [21] was significantly enriched for genetic association after multiple testing correction, although the immune-related modules in their study gave the strongest signal. However, although the microglia module highlighted by Zhang et al. [21] did not show significant enrichment for association, the genes shared in common with our significant expression modules did, highlighting the utility of using multiple expression data sets in generating biologically meaningful modules. The *TYROBP* gene highlighted by Zhang et al. as an important causal regulator is also a hub gene in this network [29].

Regulation of endocytosis, cholesterol transport, and ubiquitination were not strongly represented in our WGCNA modules, which may relate to the large size of the modules and the use of only brain gene expression. In addition, coordinated gene expression in brain may well reflect differences in distribution of specific cell types or subtypes [30]. The brain expression signatures we used came from nonneurologically compromised brains, but it is likely that changes in microglial composition or fate in response to inflammation or infection in these subjects could propagate such coordinate changes in gene expression. *TREM2* is one of the 151 genes that occur in two or more expression modules (Fig. 2), and rare variants in *TREM2* are associated with a significant increase in AD susceptibility [12]. *TREM2* regulates the phenotype of microglia controlling their downstream activation to an inflammatory or phagocytotic fate, believed to promote or inhibit AD pathogenesis, respectively [31]. Thus, the expression signature we detect through genome-wide association may also be a marker for changes in microglial phenotypes that act to enhance or inhibit the susceptibility of individuals to AD.

As the main motivation for genetic analysis of complex traits is to understand the biology of disease and inform the search for treatments, interpretation of genetic signals in a biologically meaningful way is essential. Pathway analyses that integrate multiple dense sources of data provide a means of starting to do this. Identifying strong susceptibility targets also highlights potential drug targets. Although expression analyses alone can provide important clues about etiology of disease, integrating them with genetic data that identify causative factors underlying susceptibility to disease ensures that the gene expression signatures revealed are related to disease etiology rather than secondary effects, making the pathways highlighted by the analysis primary targets for therapy. This study implicates regulation of endocytosis and protein ubiquitination, in addition to cholesterol metabolism, as potential therapeutic targets in AD. It strongly reinforces the critical role of the immune system in conferring AD susceptibility: gaining a detailed mechanistic understanding of the events within the immune system that predispose to AD and investigating how to address these mechanisms should now be a priority for AD research.

Acknowledgments

This work was made possible by the generous participation of the control subjects, the patients, and their families. GERAD was supported by the Wellcome Trust, Medical Research Council, Alzheimer's Research UK, and the Welsh Assembly Government. ADGC and CHARGE were supported by the National Institutes of Health, National Institute on Aging (NIH-NIA). CHARGE was also supported by Erasmus Medical Center and Erasmus University. IGAP was funded by the French National Foundation on Alzheimer's Disease and Related Disorders, the Centre National de Génotypage and 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 provided support for I-GAP meetings, communication and some funds to each consortium to support analyses.

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jalz.2014.05.1757.

RESEARCH IN CONTEXT

1. Systematic review: As the main motivation for genetic analysis of complex traits is to understand the biology of disease and inform the search for treatments, interpretation of genetic signals in a biologically meaningful way is essential. Pathway analyses that integrate multiple dense sources of data provide a means of starting to do this. Identifying strong susceptibility targets also highlights potential drug targets.
2. Interpretation: Although expression analyses alone can provide important clues about etiology of disease, integrating them with genetic data that identify causative factors underlying susceptibility to disease ensures that the gene expression signatures revealed are related to disease etiology rather than secondary effects, making the pathways highlighted by the analysis primary targets for therapy. This study implicates regulation of endocytosis and protein ubiquitination, in addition to cholesterol metabolism, as potential therapeutic targets in Alzheimer's disease (AD). It strongly reinforces the critical role of the immune system in conferring AD susceptibility.
3. Future directions: A detailed mechanistic understanding of the events within the immune system that predispose to AD and investigating how to address these mechanisms should now be a priority for AD research.

References

[1] Hebert LE, Scherr PA, Bienias JL, Bennett DA, Evans DA. Alzheimer disease in the US population: prevalence estimates using the 2000 census. Arch Neurol 2003;60:1119–22.

[2] Plassman BL, Langa KM, Fisher GG, Heeringa SG, Weir DR, Ofstedal MB, et al. Prevalence of dementia in the United States: the aging, demographics, and memory study. Neuroepidemiology 2007; 29:125–32.

[3] Wilson RS, Weir DR, Leurgans SE, Evans DA, Hebert LE, Langa KM, et al. Sources of variability in estimates of the prevalence of Alzheimer's disease in the United States. Alzheimer's & dementia. The Journal of the Alzheimer's Association 2011;7:74–9.

[4] Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet 2013; 45:1452–8.

[5] Hardy J. The amyloid hypothesis for Alzheimer's disease: a critical reappraisal. J Neurochem 2009;110:1129–34.

[6] Giacobini E, Gold G. Alzheimer disease therapy–moving from amyloid-beta to tau. Nature Reviews Neurology 2013;9:677–86.

[7] Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S, et al. Role of genes and environments for explaining Alzheimer disease. Arch Gen Psychiatry 2006;63:168–74.

[8] Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet 2009;41:1088–93.

[9] Lee SH, Harold D, Nyholt DR, , Consortium ANInternational Endogene C, Genetic, et al. Estimation and partitioning of polygenic variation captured by common SNPs for Alzheimer's disease, multiple sclerosis and endometriosis. Hum Mol Genet 2013;22:832–41.

[10] Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science 1993;261:921–3.

[11] So HC, Gui AH, Cherny SS, Sham PC. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. Genet Epidemiol 2011;35:310–7.

[12] Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogaeva E, Majounie E, et al. TREM2 variants in Alzheimer's disease. The New England Journal of Medicine 2013;368:117–27.

[13] Holmans P, Green EK, Pahwa JS, Ferreira MA, Purcell SM, Sklar P, et al. Gene ontology analysis of GWA study data sets provides insights into the biology of bipolar disorder. Am J Hum Genet 2009;85:13–24.

[14] Jones L, Holmans PA, Hamshere ML, Harold D, Moskvina V, Ivanov D, et al. Genetic evidence implicates the immune system and cholesterol metabolism in the aetiology of Alzheimer's disease. PloS One 2010;5:e13950.

[15] Brown MB. Method for Combining Non-Independent, One-Sided Tests of Significance. Biometrics 1975;31:987–92.

[16] Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, Lai SL, et al. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. PLoS Genetics 2010; 6:e1000952.

[17] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008;9:559.

[18] Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904–9.

[19] Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet 2003;34:267–73.

[20] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America 2005;102:15545–50.

[21] Zhang B, Gaiteri C, Bodea LG, Wang Z, McElwee J, Podtelezhnikov AA, et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. Cell 2013;153:707–20.

[22] Lambert JC, Grenier-Boley B, Chouraki V, Heath S, Zelenika D, Fievet N, et al. Implication of the immune system in Alzheimer's disease: evidence from genome-wide pathway analysis. Journal of Alzheimer's disease. J Atten Disord 2010;20:1107–18.

[23] O'Brien RJ, Wong PC. Amyloid precursor protein processing and Alzheimer's disease. Annu Rev Neurosci 2011;34:185–204.

[24] Wu KC, Jin JP. Calponin in non-muscle cells. Cell Biochem Biophys 2008;52:139–48.

[25] Huang QQ, Hossain MM, Wu K, Parai K, Pope RM, Jin JP. Role of H2-calponin in regulating macrophage motility and phagocytosis. The Journal of Biological Chemistry 2008;283:25887–99.

[26] Yue L, Bian JT, Grizelj I, Cavka A, Phillips SA, Makino A, et al. Apolipoprotein E enhances endothelial-NO production by modulating caveolin 1 interaction with endothelial NO synthase. Hypertension 2012;60:1040–6.

[27] Reinbothe TM, Alkayyali S, Ahlqvist E, Tuomi T, Isomaa B, Lyssenko V, et al. The human L-type calcium channel Cav1.3 regulates insulin release and polymorphisms in CACNA1D associate with type 2 diabetes. Diabetologia 2013;56:340–9.

[28] Jonsson T, Atwal JK, Steinberg S, Snaedal J, Jonsson PV, Bjornsson S, et al. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. Nature 2012;488:96–9.
[29] Langfelder P, Mischel PS, Horvath S. When is hub gene selection better than standard meta-analysis? PloS one 2013;8:e61505.
[30] Kuhn A, Thu D, Waldvogel HJ, Faull RL, Luthi-Carter R. Population-specific expression analysis (PSEA) reveals molecular changes in diseased brain. Nature Methods 2011;8:945–7.
[31] Forabosco P, Ramasamy A, Trabzuni D, Walker R, Smith C, Bras J, et al. Insights into TREM2 biology by network analysis of human brain gene expression data. Neurobiol Aging 2013;34:2699–714.
[32] Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. PloS One 2011;6:e21800.
[33] Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols 2009;4:44–57.
[34] Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 2009;37:1–13.
[35] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498–504.
[36] Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PloS One 2010;5:e13984.

Appendix: International Genomics of Alzheimer's Disease Consortium (IGAP)

Lesley Jones,[a,1] Jean-Charles Lambert,[b,c,d,1] Li-San Wang,[t,1] Seung-Hoan Choi,[I] Denise Harold,[fffff] Alexey Vedernikov,[a] Valentina Escott-Price,[a] Timothy Stone,[a] Alexander Richards,[a] Céline Bellenguez,[b,c,d] Carla A. Ibrahim-Verbaas,[e] Adam C. Naj,[f] Rebecca Sims,[a] Amy Gerrish,[a] Gyungah Jun,[g,h,I] Anita L. DeStefano,[I] Joshua C. Bis,[j] GaryW. Beecham,[f,k] Benjamin Grenier-Boley,[b,c,d] Giancarlo Russo,[l] Tricia A. Thornton-Wells,[m] Nicola Jones,[a] Albert V. Smith,[n,o] Vincent Chouraki,[b,c,d] Charlene Thomas,[a] M Arfan Ikram,[p,q] Diana Zelenika,[r] Badri N. Vardarajan,[g] Yoichiro Kamatani,[s] Chiao-Feng Lin,[t] Helena Schmidt,[u] Brian W. Kunkle,[f] Melanie L. Dunstan,[a] Agustin Ruiz,[v] Marie-Thérèse Bihoreau,[r] Christiane Reitz,[w,x] Florence Pasquier,[c,y] Paul Hollingworth,[a] Olivier Hanon,[aa] Annette L. Fitzpatrick,[j,bb] Joseph D. Buxbaum,[cc,dd,ee] Dominique Campion,[ff] Paul K. Crane,[gg] Tim Becker,[hh] Vilmundur Gudnason,[n,o] Carlos Cruchaga,[ii] David Craig,[jj] Najaf Amin,[kk] Claudine Berr,[ll] Oscar L. Lopez,[mm] Philip L. De Jager,[nn,oo] Vincent Deramecourt,[b,y] Janet A. Johnston,[jj] Denis Evans,[pp] Simon Lovestone,[qq] Luc Letteneur,[rr] Johanes Kornhuber,[ss] Lluís Tárraga,[v] David C. Rubinsztein,[tt] Gudny Eiriksdottir,[o] Kristel Sleegers,[uu,vv] Alison M. Goate,[ii] Nathalie Fiévet,[b,d] Matthew J. Huentelman,[ww] Michael Gill,[xx] Valur Emilsson,[n,o] Kristelle Brown,[yy] M Ilyas Kamboh,[zz,aaa] Lina Keller,[bbb] Pascale Barberger-Gateau,[rr] Bernadette McGuinness,[jj] Eric B. Larson,[gg,ccc] Amanda J. Myers,[ddd] Carole Dufouil,[rr] Stephen Todd,[jj] David Wallon,[ff] Seth Love,[eee] Pat Kehoe,[eee] Ekaterina Rogaeva,[fff] John Gallacher,[ggg] Peter St George-Hyslop,[fff,hhh] Jordi Clarimon,[iii,jjj] Alberti Lleò,[iii,jjj] Anthony Bayer,[ggg] Debby W. Tsuang,[kkk] Lei Yu,[lll] Magda Tsolaki,[mmm] Paola Bossù,[nnn] Gianfranco Spalletta,[nnn] Petra Proitsi,[qq] John Collinge,[ooo] Sandro Sorbi,[ppp,qqq] Florentino Sanchez Garcia,[rrr] Nick Fox,[sss] John Hardy,[ttt] Maria Candida Deniz Naranjo,[rrr] Cristina Razquin,[aaaaaa] Paola Bosco,[uuu] Robert Clarke,[vvv] Carol Brayne,[www] Daniela Galimberti,[xxx] Michelangelo Mancuso,[yyy] MRC CFAS,[www] Susanne Moebus,[zzz] Patrizia Mecocci,[aaaa] Maria del Zompo,[bbbb] Wolfgang Maier,[cccc] Harald Hampel,[dddd] Alberto Pilotto,[eeee] Maria Bullido,[ffff,gggg,hhhh] Francesco Panza,[iiii] Paolo Caffarra,[iiij,kkkk] Benedetta Nacmias,[ppp,qqq] John R. Gilbert,[f,llll] Manuel Mayhaus,[mmmm] Frank Jessen,[cccc] Martin Dichgans,[ddddd] Lars Lannfelt,[nnnn] Hakon Hakonarson,[oooo] Sabrina Pichler,[mmmm] Minerva M. Carrasquillo,[pppp] Martin Ingelsson,[nnnn] Duane Beekly,[qqqq] Victoria Alavarez,[rrrr] Fanggeng Zou,[pppp] Otto Valladares,[t] Steven G. Younkin,[pppp] Eliecer Coto,[rrrr] Kara L. Hamilton-Nelson,[f] Ignacio Mateo,[ssss] Michael J. Owen,[a] Kelley M. Faber,[tttt] Palmi V. Jonsson,[uuuu] Onofre Combarros,[ssss] Michael C. O'Donovan,[a] Laura B. Cantwell,[t] Hilkka Soininen,[vvvv,wwww] Deborah Blacker,[xxxx,zzzz], Simon Mead,[ooo] Thomas H. Mosley, Jr.,[bbbbb] David A. Bennett,[lll,bbbbb] Tamara B. Harris,[ccccc] Laura Fratiglioni,[ddddd,eeeee] Clive Holmes,[fffff] Renee FAG. de Bruijn,[ggggg] Peter Passmore,[jj] Thomas J. Montine,[hhhhh] Karolien Bettens,[uu,vv] Jerome I. Rotter,[iiiii] Alexis Brice,[iiiij,kkkkk] Kevin Morgan,[yy] Tatiana M. Foroud,[tttt] Walter A. Kukull,[llllll] Didier Hannequin,[ff] John F. Powell,[qq] Michael A. Nalls,[mmmmm] Karen Ritchie,[ll] Kathryn L. Lunetta,[I] John SK. Kauwe,[tt] Eric Boerwinkle,[nnnnn,ooooo] Matthias Riemenschneider,[ppppp] Mercè Boada,[qqqqq] Mikko Hiltunen,[vvvv,wwww] Eden R. Martin,[f,k] Pau Pastor,[bbbbbb,cccccc] Reinhold Schmidt,[rrrrr] Dan Rujescu,[ssssss] Jean-François Dartigues,[rr,ttttt] Richard Mayeux,[w,x] Christophe Tzourio,[uuuuu] Albert Hofman,[p,q] Markus M. Nöthen,[vvvvv] Caroline Graff,[wwwww,eeeee] Bruce M. Psaty,[j,eeeeee] Jonathan L. Haines,[m,xxxxx] Mark Lathrop,[r,s,yyyyy] Margaret A. Pericak-Vance,[f,k] Lenore J. Launer,[ccccc] Lindsay A. Farrer,[g,h,I,yyyy,zzzzz] Cornelia M. van Duijn,[kk] Christine Van Broeckhoven,[uu,vv] Alfredo Ramirez,[z] Gerard D. Schellenberg,[t] Sudha Seshadri,[zzzzz] Philippe Amouyel,[b,c,d,y,1,*] Julie Williams,[a,1,*] Peter A. Holmans,[a]

[1]These authors contributed equally

[a]Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University, Hadyn Ellis Building, Maindy Road, Cardiff, CF24 4HQ

[b]Inserm U744, Lille, 59000, France

[c]Université Lille 2, Lille, 59000, France

[d]Institut Pasteur de Lille, Lille, 59000, France

[e]Department of Epidemiology, Clinical Genetics and Neurology, Erasmus MC University Medical Center, Rotterdam, 3015 CE, The Netherlands

[f]The John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL, 33124, USA

[g]Department of Medicine (Biomedical Genetics), Boston University School of Medicine, Boston, MA, 02218, USA

[h]Department of Ophthalmology, Boston University School of Medicine, Boston, MA, 02215, USA

[I]Department of Biostatistics, Boston University School of Public Health, Boston, MA, 02118, USA

[j]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, 98101, USA

[k]Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL, 33124, USA

[l]Functional Genomics Center, Zurich, 8038, Switzerland

[m]Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN, 37212, USA

[n]University of Iceland, Faculty of Pharmaceutical Sciences, Reykjavik, 525-4000, Iceland

[o]Icelandic Heart Association, Kopavogur, 525-4000, Iceland

[p]Departments of Epidemiology, Neurology and Radiology, Erasmus MC University Medical Center, Rotterdam, 3015 CE, The Netherlands

[q]Netherlands Consortium for Healthy Aging, Leiden, The Netherlands

[r]Centre National de Genotypage, Institut Genomique, Commissariat à l'énergie Atomique, Evry, 91300, France

[s]Fondation Jean Dausset- CEPH, Paris, 75010, France

[t]Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, 19104, USA

[u]Institute for Molecular Biology and Biochemistry, Medical University of Graz, Graz, 8010, Austria

[v]Memory Clinic of Fundació ACE. Institut Català de Neurociències Aplicades, Barcelona, 8029, Spain

[w]Taub Institute on Alzheimer's Disease and the Aging Brain, Department of Neurology, Columbia University New York, NY, 10027, USA

[x]Gertrude H. Sergievsky Center, Columbia University, Department of Neurology, Columbia University, New York, NY, 10027, USA

[y]Centre Hospitalier Régional Universitaire de Lille, Lille, 59000, France

[z]Department of Psychiatry and Psychotherapy and Institute of Human Genetics, University of Bonn, Bonn, 53113, Germany

[aa]UMR 894 Inserm, Faculté de Médecine, Université Paris Descartes, Paris, 75014, France

[bb]Departments of Epidemiology and Global Health, University of Washington, Seattle, WA, 98101, USA

[cc]Department of Neuroscience, Mount Sinai School of Medicine, New York, NY, 10027, USA

[dd]Department of Psychiatry, Mount Sinai School of Medicine, New York, NY, 10027, USA

[ee]Departments of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, NY, 10027, USA

[ff]INSERM U614, Faculté de Médecine, Centre Hospitalier du Rouvray, Rouen, 76000, France

[gg]Department of Medicine, University of Washington, Seattle, WA, 98195, USA

[hh]German Center for Neurodegenerative Diseases (DZNE, Bonn) and Institute for Medical Biometry, Informatics and Epidemiology, University of Bonn, Bonn, Germany, 53113

[ii]Department of Psychiatry and Hope Center Program on Protein Aggregation and Neurodegeneration, Washington University School of Medicine, St. Louis, MO, 98122, USA

[jj]Ageing Group, Centre for Public Health, School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, BT7 1NN, UK

[kk]Department of Epidemiology, Erasmus MC University Medical Center, Rotterdam, 3015 CE, The Netherlands

[ll]INSERM U888, Hopital La Colombière, Montpellier, 34000, France

[mm]Departments of Neurology, University of Pittsburgh School of Medicine, Pittsburgh, PA, 15213, USA

[nn]Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Department of Neurology & Psychiatry, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, 02215, USA

[oo]Program in Medical and Population Genetics, Broad Institute, Boston, MA, 02215, USA

[pp]Rush Institute for Healthy Aging, Department of Internal Medicine, Rush University Medical Center, Chicago, IL, 60612, USA

[qq]King's College London, Institute of Psychiatry, Department of Neuroscience, De Crespigny Park, Denmark Hill, London SE5

[rr]Inserm U897, Victor Segalen University, F-33076, Bordeaux, 33000, France

[ss]Department of Psychiatry and Psychotherapy, Friedrich-Alexander-University Erlangen-Nuremberg, Germany

[tt]Cambridge Institute for Medical Research, University of Cambridge, Cambridge, CB2 OXY, UK

[uu]Neurodegenerative Brain Diseases Group, Department of Molecular Genetics, VIB, Antwerp, 2000, Belgium

[vv]Laboratory of Neurogenetics, Institute Born-Bunge, University of Antwerp, Antwerp, 2000, Belgium

[ww]Neurogenomics Division, Translational Genomics Research Institute, Phoenix, AZ, 85004, USA

[xx]Mercer's Institute for Research on Aging, St. James Hospital and Trinity College, Dublin, Dublin 8, Ireland

[yy]Institute of Genetics, Queen's Medical Centre, University of Nottingham, NG7 2UH, UK

[zz]Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA, 15213, USA

[aaa]Alzheimer's Disease Research Center, University of Pittsburgh, Pittsburgh, PA, 15213, USA

[bbb]Aging Reasearch Center, Dept Neurobiology, Care Sciences and Society, Karolinska Institutet and Stockholm University, Stockholm, 17177, Sweden

[ccc]Group Health Research Institute, Group Health, Seattle, WA, 98195, USA

[ddd]Department of Psychiatry and Behavioral Sciences, Miller School of Medicine, University of Miami, Miami, FL, 33124, USA

[eee]University of Bristol, Clinical Neurosciences, Level 1 Learning and Research Building BS10 5NB

[fff]Tanz Centre for Research in Neurodegenerative Disease, University of Toronto, Toronto, ON M5S 1A1, Canada

[ggg]Institute of Primary Care and Public Health, Cardiff University, Neuadd Meirionnydd, University Hospital of Wales, Heath Park, Cardiff, CF14 4XN, UK

[hhh]Cambridge Institute for Medical Research and Department of Clinical Neurosciences, University of Cambridge, Cambridge, CB2 1TN, UK

[iii]Neurology Department, IIB Sant Pau, Sant Pau Hospital, Universitat Autònoma de Barcelona, Barcelona, 08026, Spain

[jjj]Center for Networker Biomedical Research in Neurodegenerative Diseases (CIBERNED), Barcelona, Spain

[kkk]Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, 98195, USA

[lll]Department of Neurological Sciences, Rush University Medical Center, Chicago, IL, 60612, USA

[mmm]Aristotle University of Thessaloniki, Despere 3, Thessaloniki, 54621, Greece

[nnn]Clinical and Behavioral Neurology, Fondazione Santa Lucia, Roma, 00142, Italy
[ooo]MRC Prion Unit, Department of Neurodegenerative Disease, UCL Institute of Neurology, London, WC1N 3BG, UK
[ppp]Department of Neurological and Psychiatric Sciences, University of Florence, Florence, 50121, Italy
[qqq]Centro di Ricerca, Trasferimento e Alta Formazione DENOTHE, University of Florence, Florence, Italy
[rrr]Department of Immunology, Hospital Universitario Dr. Negrin, Las Palmas de Gran Canaria, 35012, Spain
[sss]Dementia Research Centre, Department of Neurodegenerative Disease, UCL Institute of Neurology, London, WC1N 3BG, UK
[ttt]Department of Molecular Neuroscience and Reta Lilla Weston Laboratories, Institute of Neurology, London, WC1N 3BG, UK
[uuu]IRCCS Associazione Oasi Maria SS, Troina, 94018, Italy
[vvv]Oxford Healthy Aging Project (OHAP), Clinical Trial Service Unit, University of Oxford, Oxford, OX1 2JD, UK
[www]Cognitive Function and Ageing Study (CFAS), Institute of Public Health, University of Cambridge, Cambridge, CB2 1TN, UK
[xxx]University of Milan, Fondazione Cà Granda, IRCCS Ospedale Policlinico, Milan, 20122, Italy
[yyy]Neurological Clinic, University of Pisa, 56126, Italy
[zzz]Urban Epidemiology, Institute for Medical Informatics, Biometry and Epidemiology, University Hospital Essen, University Duisburg-Essen, Essen, 45141, Germany
[aaaa]Section of Gerontology and Geriatrics, Department of Clinical and Experimental Medicine, University of Perugia, Perugia, 06100, Italy
[bbbb]Section of Neuroscience and Clinical Pharmacology, Department of Biomedical Sciences, University of Cagliari, Cagliari, 09123, Italy
[cccc]Department of Psychiatry and Psychotherapy, University of Bonn, Germany and German Center for Neurodegenerative Diseases (DZNE, Bonn), Bonn, 53113, Germany
[dddd]Université Pierre et Marie Curie, Département de Neurologie, Institut de la Mémoire et de la Maladie d'Alzheimer, Paris, France
[eeee]Gerontology and Geriatrics Research Laboratory, I.R.C.C.S. Casa Sollievo della Sofferenza, San Giovanni Rotondo (FG), 71013, Italy
[ffff]Centro de Biología Molecular Severo Ochoa (CSIC-UAM), Madrid, 28046, Spain
[gggg]Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED), Madrid, Spain
[hhhh]Instituto de Investigación Sanitaria "Hospital la Paz" (IdIPaz), Madrid, Spain
[iiii]Departement of Geriatrics, Center for Aging Brain, University of Bari, Bari, 70121, Italy
[jjjj]Department of Neuroscience-University of Parma, Parma, 43100, Italy
[kkkk]Center for Cognitive Disorders AUSL, Parma, Italy
[llll]Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL, 33124, USA
[mmmm]Dept. Of Psychiatry, University Hospital, 66424, Saarland
[nnnn]Dept. of Public Health/Geriatrics, Uppsala University, Uppsala, 75312, Sweden
[oooo]Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA
[pppp]Department of Neuroscience, Mayo Clinic, Jacksonville, FL, 32224, USA
[qqqq]National Alzheimer's Coordinating Center, University of Washington, Seattle, WA, 98195, USA
[rrrr]Genetica molecular-Huca-Oviedo, Oviedo, 33006, Spain
[ssss]Neurology Service and CIBERNED, "Marqués de Valdecilla" University Hospital (University of Cantabria and IFIMAV), Santander, 39005, Spain
[tttt]Department of Medical and Molecular Genetics, Indiana University, Indianapolis, IN, 46202, USA
[uuuu]Landspitali University Hospital, Reykjavik, 543-1000, Iceland
[vvvv]Institute of Clinical Medicine - Neurology, University of Eastern Finland, Kuopio, 70211, Finland
[wwww]Department of Neurology, Kuopio University Hospital, FIN-70211, Kuopio, Finland
[xxxx]Department of Epidemiology, Harvard School of Public Health, Boston, MA, 02115, USA
[yyyy]Department of Epidemiology, Boston University School of Public Health, Boston, MA, 02118, USA
[zzzz]Department of Psychiatry, Massachusetts General Hospital/Harvard Medical School, Boston, MA, 02115, USA
[aaaaa]Department of Medicine (Geriatrics), University of Mississippi Medical Center, Jackson, MS, 39216, USA
[bbbbb]Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, 60612, USA
[ccccc]Laboratory of Epidemiology, Demography, and Biometry, National Institute of Health, Bethesda, MD, 20892, USA
[ddddd]Aging Reasearch Center, Dept Neurobiology, Care Sciences and Society, Karolinska Institutet and Stockholm University, Stockholm, 11418, Sweden
[eeeee]Dept Geriatric Medicine, Genetics Unit, Karolinska University Hospital Huddinge, S-14186, Stockholm
[fffff]Division of Clinical Neurosciences, School of Medicine, University of Southampton, Southampton, SO17 1BJ, UK
[ggggg]Departments of Neurology and Epidemiology, Erasmus MC University Medical Center, Rotterdam, 3015 CE, The Netherlands
[hhhhh]Department of Pathology, University of Washington, Seattle, WA, 98195, USA
[iiiii]Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, CA, 90048, USA

[jjjjj]CRicm-UMRS975, Paris, France
[kkkkk]AP-HP, Hopital de la Pitiê-Salpêtrière, Paris, 75013, France
[lllll]Department of Epidemiology, University of Washington, Seattle, WA, 98195, USA
[mmmmm]Laboratory of Neurogenetics, Intramural Research Program, National Institute on Aging, Bethesda, MD, 20892, USA
[nnnnn]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, 77030, USA
[ooooo]Human Genetics Center and Div. of Epidemiology, University of Texas Health Sciences Center at Houston, Houston, TX, 77030, USA
[ppppp]Department of Psychiatry, University Hospital, 66424, Saarland
[qqqqq]Memory Clinic of Fundació ACE, Institut Català de Neurociències Aplicades, Barcelona, 8029, Spain
[rrrrr]Department of Neurology, Medical University of Graz, Graz, 8010, Austria
[sssss]Department of Psychiatry, University of Munich, Munich and University of Halle, Halle, 80336, Germany
[ttttt]Centre de Mémoire de Ressources et de Recherche de Bordeaux, CHU de Bordeaux, Bordeaux, France
[uuuuu]Inserm U708, Victor Segalen University, F-33076, Bordeaux, 33000, France
[vvvvv]Institute of Human Genetics, Department of Genomics, Life and Brain Center, University of Bonn, and German Center for Neurodegenerative Diseases (DZNE, Bonn), Bonn, 53113, Germany
[wwwww]Karolinska Institutet, Dept Neurobiology, Care Sciences and Society, KIADRC, Novum floor 5, S14186 Stockholm, 17177, Sweden
[xxxxx]Vanderbilt Center for Human Genetics Research, Vanderbilt University, Nashville, TN, 37212, USA
[yyyyy]McGill University and Génome Québec Innovation Centre, Montreal, Canada
[zzzzz]Department of Neurology, Boston University School of Medicine, Boston, MA, 02215, USA
[aaaaaa]Neurogenetics Laboratory, Division of Neurosciences, Center for Applied Medical Research, University of Navarra School of Medicine, Pamplona, 28023, Spain
[bbbbbb]Neurogenetics Laboratory, Division of Neurosciences, Center for Applied Medical Research and Department of Neurology, Clínica Universidad de Navarra, University of Navarra School of Medicine, Pamplona, Spain
[cccccc]CIBERNED, Centro de Investigación Biomédica en Red de Enfermedades Neurodegenerativas, Instituto de Salud Carlos III, Spain
[dddddd]Institute for Stroke and Dementia Research, Klinikum der Universität München, Munich, Germany and German Center for Neurodegenerative Diseases (DZNE, Munich), Munich, 80336, Germany
[eeeeee]Departments of Epidemiology and Health Services, University of Washington, Seattle, WA, USA, Group Health Research Institute, Group Health, Seattle, WA, USA
[ffffff]Neuropsychiatric Genetics Group, Department of Psychiatry, Trinity Centre for Health Sciences, St James's Hospital, Dublin 8