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...
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 \( APOE \) genes remaining after removing \( APOE \) and the genome-wide significant genes.

<table>
<thead>
<tr>
<th>Genes removed (number of genes)</th>
<th>Enrichment ( P &lt; .05 )</th>
<th>Enrichment ( P &lt; .01 )</th>
<th>Enrichment ( P &lt; .001 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>542</td>
<td>.0002</td>
<td>177</td>
</tr>
<tr>
<td>( APOE + 1 ) Mb (77)</td>
<td>446</td>
<td>.0002</td>
<td>131</td>
</tr>
<tr>
<td>( APOE + 1 ) Mb + GWS (98)</td>
<td>402</td>
<td>.0020</td>
<td>116</td>
</tr>
<tr>
<td>( APOE + 1 ) Mb + GWS +1 Mb (552)</td>
<td>336</td>
<td>.0094</td>
<td>93</td>
</tr>
</tbody>
</table>

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](image)

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

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Pathway number</th>
<th>Number of genes</th>
<th>Number of sig</th>
<th>( P ) value</th>
<th>( P ) value no GWS</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GO: 2455</td>
<td>32</td>
<td>5</td>
<td>( 3.27 \times 10^{-12} )</td>
<td>( 5.72 \times 10^{-1} )</td>
<td>Humoral immune response mediated by circulating immunoglobulin</td>
</tr>
<tr>
<td>1</td>
<td>GO:50776</td>
<td>421</td>
<td>29</td>
<td>( 3.24 \times 10^{-9} )</td>
<td>( 1.57 \times 10^{-4} )</td>
<td>Regulation of immune response</td>
</tr>
<tr>
<td>1</td>
<td>GO: 2684</td>
<td>421</td>
<td>31</td>
<td>( 3.95 \times 10^{-9} )</td>
<td>( 2.14 \times 10^{-4} )</td>
<td>Positive regulation of immune system process</td>
</tr>
<tr>
<td>1</td>
<td>GO:50778</td>
<td>271</td>
<td>21</td>
<td>( 1.55 \times 10^{-7} )</td>
<td>( 6.65 \times 10^{-4} )</td>
<td>Positive regulation of immune response</td>
</tr>
<tr>
<td>1</td>
<td>KEGG 4664</td>
<td>78</td>
<td>13</td>
<td>( 5.76 \times 10^{-4} )</td>
<td>( 2.18 \times 10^{-2} )</td>
<td>( \text{Fc epsilon RI signaling pathway} )</td>
</tr>
<tr>
<td>2</td>
<td>GO:60627</td>
<td>140</td>
<td>20</td>
<td>( 1.31 \times 10^{-11} )</td>
<td>( 2.00 \times 10^{-1} )</td>
<td>Regulation of vesicle-mediated transport</td>
</tr>
<tr>
<td>2</td>
<td>GO:30100</td>
<td>88</td>
<td>14</td>
<td>( 6.76 \times 10^{-10} )</td>
<td>( 1.06 \times 10^{-1} )</td>
<td>Regulation of endocytosis</td>
</tr>
<tr>
<td>2</td>
<td>GO:45806</td>
<td>19</td>
<td>6</td>
<td>( 3.91 \times 10^{-7} )</td>
<td>( 1.77 \times 10^{-2} )</td>
<td>Negative regulation of endocytosis</td>
</tr>
<tr>
<td>2</td>
<td>GO:48261</td>
<td>6</td>
<td>3</td>
<td>( 3.89 \times 10^{-6} )</td>
<td>( 9.82 \times 10^{-1} )</td>
<td>Negative regulation of receptor-mediated endocytosis</td>
</tr>
<tr>
<td>2</td>
<td>GO:48259</td>
<td>30</td>
<td>6</td>
<td>( 6.19 \times 10^{-5} )</td>
<td>1.00</td>
<td>Regulation of receptor-mediated endocytosis</td>
</tr>
<tr>
<td>3</td>
<td>GO:30301</td>
<td>41</td>
<td>8</td>
<td>( 2.96 \times 10^{-9} )</td>
<td>( 2.51 \times 10^{-1} )</td>
<td>Cholesterol transport</td>
</tr>
<tr>
<td>3</td>
<td>GO:43691</td>
<td>16</td>
<td>5</td>
<td>( 3.90 \times 10^{-9} )</td>
<td>( 2.78 \times 10^{-1} )</td>
<td>Reverse cholesterol transport</td>
</tr>
<tr>
<td>3</td>
<td>GO:15918</td>
<td>42</td>
<td>8</td>
<td>( 3.91 \times 10^{-9} )</td>
<td>( 3.15 \times 10^{-1} )</td>
<td>Sterol transport</td>
</tr>
<tr>
<td>3</td>
<td>GO:34366</td>
<td>8</td>
<td>2</td>
<td>( 6.40 \times 10^{-7} )</td>
<td>N/A</td>
<td>Spherical high-density lipoprotein particle</td>
</tr>
<tr>
<td>4</td>
<td>KEGG 4640</td>
<td>81</td>
<td>11</td>
<td>( 1.05 \times 10^{-8} )</td>
<td>( 4.91 \times 10^{-1} )</td>
<td>Hematopoietic cell lineage</td>
</tr>
<tr>
<td>5</td>
<td>GO:32434</td>
<td>40</td>
<td>5</td>
<td>( 1.34 \times 10^{-6} )</td>
<td>1.00</td>
<td>Regulation of proteasomal ubiquitin-dependent protein catabolic process</td>
</tr>
<tr>
<td>5</td>
<td>GO:51437</td>
<td>70</td>
<td>9</td>
<td>( 2.60 \times 10^{-3} )</td>
<td>( 2.60 \times 10^{-3} )</td>
<td>Regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle</td>
</tr>
<tr>
<td>5</td>
<td>GO:51439</td>
<td>76</td>
<td>9</td>
<td>( 3.82 \times 10^{-3} )</td>
<td>( 3.82 \times 10^{-3} )</td>
<td>Regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle</td>
</tr>
<tr>
<td>5</td>
<td>REACT 440</td>
<td>108</td>
<td>11</td>
<td>( 3.89 \times 10^{-3} )</td>
<td>( 3.89 \times 10^{-3} )</td>
<td>REACTOME_CELL_CYCLE_CHECKPOINTS</td>
</tr>
<tr>
<td>5</td>
<td>GO:51443</td>
<td>77</td>
<td>9</td>
<td>( 9.62 \times 10^{-3} )</td>
<td>( 9.62 \times 10^{-3} )</td>
<td>Positive regulation of ubiquitin-protein ligase activity</td>
</tr>
<tr>
<td>6</td>
<td>REACT 539</td>
<td>261</td>
<td>25</td>
<td>( 2.95 \times 10^{-5} )</td>
<td>( 6.93 \times 10^{-2} )</td>
<td>REACTOME_HEMOSTASIS</td>
</tr>
<tr>
<td>7</td>
<td>GO:30131</td>
<td>31</td>
<td>7</td>
<td>( 1.20 \times 10^{-3} )</td>
<td>( 9.13 \times 10^{-1} )</td>
<td>Clathrin adaptor complex</td>
</tr>
<tr>
<td>7</td>
<td>GO:30119</td>
<td>32</td>
<td>7</td>
<td>( 1.53 \times 10^{-3} )</td>
<td>( 9.54 \times 10^{-1} )</td>
<td>AP-type membrane coat adapt complex</td>
</tr>
<tr>
<td>7</td>
<td>GO:44433</td>
<td>301</td>
<td>31</td>
<td>( 1.01 \times 10^{-2} )</td>
<td>1.00</td>
<td>Cytoplasmic vesicle part</td>
</tr>
<tr>
<td>7</td>
<td>GO:30122</td>
<td>9</td>
<td>4</td>
<td>( 1.29 \times 10^{-2} )</td>
<td>1.00</td>
<td>AP-2 adaptor complex</td>
</tr>
<tr>
<td>7</td>
<td>GO:30118</td>
<td>39</td>
<td>7</td>
<td>( 1.35 \times 10^{-2} )</td>
<td>1.00</td>
<td>Clathrin coat</td>
</tr>
<tr>
<td>8</td>
<td>GO: 6457</td>
<td>200</td>
<td>12</td>
<td>( 1.60 \times 10^{-3} )</td>
<td>1.00</td>
<td>Protein folding</td>
</tr>
</tbody>
</table>

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.

---
We used control individuals rather than AD cases so that cor-
neous regions in a sample of approximately 150 control brains [16] and was independent from that used by Zhang et al. [21].

To overcome the annotation gap and access biologically related 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 cor-
relations 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,
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 of the four most significantly enriched expression modules: the principal biological themes were derived from DAVID analysis [33,34] analysis. 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 ($r > 0.9$ 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.
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, SUP4H1, and TA6. 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, α2 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, α1D 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 phagocytic 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 Génopépide 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


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

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

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

Memory Clinic of Fundació ACE, Institut Català de Neurociències Aplicades, Barcelona, 8029, Spain

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

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

Centre Hospitalier Régional Universitaire de Lille, Lille, 59000, France

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

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

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

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

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

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

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

Department of Medicine, University of Washington, Seattle, WA, 98195, USA

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

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

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

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

INSERM U888, Hopital La Colombière, Montpellier, 34000, France

Departments of Neurology, University of Pittsburgh School of Medicine, Pittsburgh, PA, 15213, USA

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

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

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

King’s College London, Institute of Psychiatry, Department of Neuroscience, De Crespigny Park, Denmark Hill, London SE5

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

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

Cambridge Institute for Medical Research, University of Cambridge, Cambridge, CB2 OXY, UK

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

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

Neurogenomics Division, Translational Genomics Research Institute, Phoenix, AZ, 85004, USA

Mercer’s Institute for Research on Aging, St. James Hospital and Trinity College, Dublin, Dublin 8, Ireland

Institute of Genetics, Queen’s Medical Centre, University of Nottingham, NG7 2UH, UK

Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA, 15213, USA

Alzheimer’s Disease Research Center, University of Pittsburgh, Pittsburgh, PA, 15213, USA

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

Group Health Research Institute, Group Health, Seattle, WA, 98195, USA

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

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

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

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

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

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

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

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

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

Aristotle University of Thessaloniki, Despere 3, Thessaloniki, 54621, Greece