Gene-Wide Analysis Detects Two New Susceptibility Genes for Alzheimer's Disease

Citation

Published Version
doi:10.1371/journal.pone.0094661

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:12406561

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story
The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story.

Accessibility
Gene-Wide Analysis Detects Two New Susceptibility Genes for Alzheimer’s Disease


1 Institute of Psychological Medicine and Clinical Neurosciences, MRC Centre for Neuropsychiatric Genetics & Genomics, Cardiff University, Cardiff, United Kingdom, 2 Inserm U744, Lille, France, 3 Université Lille 2, Lille, France, 4 Institut Pasteur de Lille, Lille, France, 5 Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, United States of America, 6 Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, United States of America, 7 Department of Epidemiology and Neurology, Erasmus MC University Medical Center, Rotterdam, the Netherlands, 8 Department of Biostatistics and Epidemiology and Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine, University of Pennsylvania,
Philadelphia, Pennsylvania, United States of America, Department of Medicine (Biomedical Genetics), Boston University School of Medicine, Boston, Massachusetts, United States of America, 9 Department of Ophthalmology, Boston University School of Medicine, Boston, Massachusetts, United States of America, 10 Department of Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, Washington, United States of America, 11 The John P. Hussman Institute for Human Genomics, University of Miami, Miami, Florida, United States of America, 12 Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, Florida, United States of America, 13 Functional Genomics Center Zurich, ETH/University of Zurich, Zurich, Switzerland, 14 Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, Tennessee, United States of America, 15 Department of Molecular Genetics, University of Iceland, Reykjavik, Iceland, 16 Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, Washington, United States of America, 17 Icelandic Heart Association, Reykjavik, Iceland, 18 Department of Neurology, Boston University School of Medicine, Boston, Massachusetts, United States of America, 19 Departments of Epidemiology, Neurology and Radiology, Erasmus MC University Medical Center, Rotterdam, the Netherlands, 20 Netherlands Consortium for Healthy Aging, Leiden, The Netherlands, 21 Centre National de Genotypage, Institut Genomique, Commissariat à l'énergie Atomique, Evry, France, 22 Fondation Jean Dausset-CEPH, Paris, France, 23 Institute for Molecular Biology and Biochemistry, Medical University of Graz, Graz, Austria, 24 Reta Lila Weston Research Laboratories, Department of Molecular Neuroscience, UCL Institute of Neurology, London, United Kingdom, 25 NHLBI Cardiovascular Epidemiology and Human Genomics Branch, The Framingham Heart Study, Framingham, Massachusetts, United States of America, 26 Memory Clinic of Fundació ACE, Institut Catala de Neurociències Aplicades, Barcelona, Spain, 27 Taub Institute on Alzheimer’s Disease and the Aging Brain, Department of Neurology, Columbia University New York, New York, United States of America, 28 Gertrude H. Sergievsky Center, Department of Neurology, Columbia University, New York, New York, United States of America, 29 CNR-MAJ, Centre Hospitalier Régional Universitaire de Lille, Lille, France, 30 University Paris Descartes, Sorbonne Paris V, Broca Hospital, Geriatrics department, Paris, France, 31 Departments of Epidemiology and Global Health, University of Washington, Seattle, Washington, United States of America, 32 Department of Neuroscience, Mount Sinai School of Medicine, New York, New York, United States of America, 33 Department of Psychiatry, Mount Sinai School of Medicine, New York, New York, United States of America, 34 Departments of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, United States of America, 35 CNR-MAJ, Inserm U1079, Rouen University Hospital, 76031 France, Rouen, France, 36 Department of Medicine, University of Washington, Seattle, Washington, United States of America, 37 German Center for Neurodegenerative Diseases (DZNE), Bonn, and Institute for Medical Biometry, Informatics and Epidemiology, University of Bonn, Bonn, Germany, 38 Department of Psychiatry and Hope Center Program on Protein Aggregation and Neurodegeneration, University of Washington School of Medicine, St. Louis, Missouri, United States of America, 39 Ageing Group, Centre for Public Health, School of Medicine, Dentistry and Biomedical Sciences, Queen’s University Belfast, Belfast, United Kingdom, 40 Department of Epidemiology, Erasmus MC University Medical Center, Rotterdam, the Netherlands, 41 INSERM U1061, Faculty of Medicine, Hôpital La Colombière, Montpellier, France, 42 Departments of Neurology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, United States of America, 43 Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Department of Neurology & Psychiatry, Brigham and Women’s Hospital and Harvard Medical School, Boston, Massachusetts, United States of America, 44 Program in Medical and Population Genetics, Broad Institute, Boston, Massachusetts, United States of America, 45 Rush Institute for Healthy Aging, Department of Internal Medicine, Rush University Medical Center, Chicago, Illinois, United States of America, 46 King’s College London, Institute of Psychiatry, Department of Neuroscience, De Crespigny Park, Denmark Hill, London, United Kingdom, 47 INSERM U897, Victor Segalen University, F-33076, Bordeaux, France, 48 Cambridge Institute for Medical Research, University of Cambridge, Cambridge, United Kingdom, 49 Neurodegenerative Brain Diseases Group, Department of Molecular Genetics, VIB, Antwerp, Belgium, 50 Laboratory of Neurogenetics, Institute Born-Bunge, University of Antwerp, Antwerp, Belgium, 51 Neurogenomics Division, Translational Genomics Research Institute, Phoenix, Arizona, United States of America, 52 Discipline of Psychiatry, Trinity College, Dublin, Ireland, 53 Institute of Genetics, Queen’s Medical Centre, University of Nottingham, Nottingham, United Kingdom, 54 Department of Human Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America, 55 Alzheimer’s Disease Research Center, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America, 56 Aging Reasearch Center, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet and Stockholm University, Stockholm, Sweden, 57 Group Health Research Institute, Group Health, Seattle, Washington, United States of America, 58 Department of Psychiatry and Behavioral Sciences, Miller School of Medicine, University of Miami, Miami, Florida, United States of America, 59 University of Bristol Institute of Clinical Sciences, School of Clinical and Experimental Medicine, Bristol, United Kingdom, 60 Tanz Centre for Research in Neurodegenerative Disease, University of Toronto, Toronto, Ontario, Canada, 61 Institute of Primary Care and Public Health, Cardiff University, Neuadd Merionydd, University Hospital of Wales, Heath Park, Cardiff, United Kingdom, 62 Cambridge Institute for Medical Research and Department of Clinical Neurosciences, University of Cambridge, Cambridge, United Kingdom, 63 Neurology Department, IIB Sant Pau. Sant Pau Hospital. Universitat Autonoma de Barcelona, Barcelona, Spain, 64 Center for Networked Biomedical Research in Neurodegenerative Diseases (CIBERNED), Barcelona, Spain, 65 Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, Washington, United States of America, 66 Department of Neurological Sciences, Rush University Medical Center, Chicago, Illinois, United States of America, 67 3rd Department of Neurology, Aristotle University of Thessaloniki, Thessaloniki, Greece, 68 Clinical and Behavioral Neurology, Fondazione Santa Lucia, Roma, Italy, 69 MRC Prion Unit, Department of Neurodegenerative Disease, UCL Institute of Neurology, London, United Kingdom, 70 NEUROFARBA Department of Neuroscience, Psychology, Drug Research and Child Health, University of Florence, Florence, Italy, 71 Centro di Ricerca, Trasferimento e Alfa Formazione DENOTHE, University of Florence, Florence, Italy, 72 Department of Immunology, Hospital Universitario Dr. Negrin, Las Palmas de Gran Canaria, Spain, 73 Dementia Research Center, Department of Genetica molecular-Huca-Oviedo, Oviedo, Spain, 74 Department of Molecular Neuroscience, UCL Institute of Neurology, and Reta Lila Weston Laboratories, Institute of Neurology, London, United Kingdom, 75 IRCCS Associazione Oasi Maria SS, Troina, Italy, 76 Oxford Healthy Aging Project (OHAP), Clinical Trial Service Unit, University of Oxford, Oxford, United Kingdom, 77 Cognitive Function and Ageing Study (CFAS), Institute of Public Health, University of Cambridge, Cambridge, United Kingdom, 78 University of Milan, Fondazione Ca Granda, IRCCS Ospedale Policlinico, Milan, Italy, 79 Neurological Clinic, University of Pisa, Pisa, Italy, 80 Urban Epidemiology Institute, Medical for Informatics, Biometry and Epidemiology, University Hospital Essen, University Duisburg-Essen, Essen, Germany, 81 Section of Gerontology and Geriatrics, Department of Clinical and Experimental Medicine, University of Perugia, Perugia, Italy, 82 Section of Neuroscience and Clinical Pharmacology, Department of Biomedical Sciences, University of Cagliari, Cagliari, Italy, 83 Department of Psychiatry, University of Frankfurt am Main, Germany, 84 Department of Neurology, University of Frankfurt am Main, Frankfurt am Main, Germany (H.H.), 85 Department of Psychiatry, Ludwig-Maximilians University, Munich, Germany, 86 Gerontology and Geriatrics Research Laboratory, I.R.C.C.S. Casa Sollievo della Sofferenza, San Giovanni Rotondo (FG), Italy, 87 Centro de Biología Molecular Severo Ochoa (CSIC-UAM); Madrid, Spain, 88 Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED), Madrid, Spain, 89 Instituto de Investigación Sanitaria “Hospital la Paz” (IDIPaz), Madrid, Spain, 90 Department of Geriatrics,Center for Aging Brain,University of Bari, Bari, Italy, 91 Department of Neuroscience-University of Parma, Parma, Italy, 92 Center for Cognitive Disorders AUSL, Parma, Italy, 93 Department of Psychiatry, Ludwig-Maximilians University, Munich, Germany, 94 Department of Public Health/Geriatrics, Uppsala University, Uppsala, Sweden, 95 Center for Applied Genomics, Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania, United States of America, 96 Department of Neuroscience, Mayo Clinic, Jacksonville, Florida, United States of America, 97 National Alzheimer’s Coordinating Center, University of Washington, Seattle, Washington, United States of America, 98 Genetica molecular-Huca-Oviedo, Oviedo, Spain, 99 Department of Psychiatry, University Hospital, Salard, Germany, 100 Neurogenetics Laboratory, Division of Neurosciences, Center for Applied Medical Research, University of Navarra School of Medicine, Pamplona, Spain, 101 CIBERNED, Centro de Investigación Biomédica en Red de Enfermedades Neurodegenerativas, Instituto de Salud Carlos III, Madrid, Spain, 102 Neurology Service and CIBERNED, “Marqués de Valdecilla” University Hospital (University of Cantabria and IFIMAV), Santander, Spain, 103 Department of Medical and Molecular Genetics, Indiana University, Indianapolis, Indiana, United States of America, 104 Landskild University Hospital, Reykjavik, Iceland, 105 Institute of Clinical Medicine - Neurology, University of Eastern Finland, Kuopio, Finland, 106 Department of Neurology, Kuopio University Hospital, Kuopio, Finland, 107 Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, United States of America, 108 Department of Psychiatry, Massachusetts General Hospital/Harvard Medical School, Boston, Massachusetts, United States of America,
Abstract

Background: Alzheimer’s disease is a common debilitating dementia with known heritability, for which 20 late onset susceptibility loci have been identified, but more remain to be discovered. This study sought to identify new susceptibility genes, using an alternative gene-wide analytical approach which tests for patterns of association within genes, in the powerful genome-wide association dataset of the International Genomics of Alzheimer’s Project Consortium, comprising over 7 million genotypes from 25,580 Alzheimer’s cases and 48,466 controls.

Principal Findings: In addition to earlier reported genes, we detected genome-wide significant loci on chromosomes 8 (TP53INP1, p = 1.4 x 10^-6) and 14 (IGHV1-67 p = 7.9 x 10^-8) which indexed novel susceptibility loci.

Significance: The additional genes identified in this study, including aspects of energy metabolism, protein degradation and the immune system and add further weight to these pathways as potential therapeutic targets in Alzheimer’s disease.


Editor: Yong-Gang Yao, Kunming Institute of Zoology, Chinese Academy of Sciences, China

Received December 3, 2013; Accepted March 17, 2014; Published June 12, 2014

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Funding: The i-Select chips were funded by the French National Foundation on Alzheimer’s disease and related disorders. The French National Foundation on Alzheimer’s disease and related disorders supported several i-GAP meetings and communications. Data management involved the Centre National de Génomique, and was supported by the Institut Pasteur de Lille, Inserm, FRC (fonctionnement pour la recherche sur le cerveau) and Rotary. This work has been developed and supported by the LAXEB (laboratory of excellence program investment for the future) DISTALZ grant (Development of Innovative Strategies for a Transdisciplinary approach to Alzheimer’s disease) and by the LAXEB GENMED grant (Medical Genomics). The French National Foundation on Alzheimer’s disease and related disorders and the Alzheimer’s Association (Chicago, Illinois) grant supported IGAP in-person meetings, communication and the Alzheimer’s Association (Chicago, Illinois) grant provided some funds to each consortium for analyses. EADI The authors thank Dr. Anne Boland (CNG) for her technical help in preparing the DNA samples for analyses. This work was supported by the National Foundation for Alzheimer’s disease and related disorders, the Institut Pasteur de Lille and the Centre National de Génomique. The Three-City Study was performed as part of a collaboration between the Institut National de la Santé et de la Recherche Médicale (Inserm), the Victor Segalen Bordeaux II University and Sanofi-Synthélabo. The Fondation pour la Recherche Médicale funded the preparation and initiation of the study. The 3C Study was also funded by the Caisse Nationale Maladie des Travailleurs Salarisés, Direction Générale de la Santé, MGEN, Institut de la Longévité, Agence Française de Sécurité des Produits de Santé, l’Aquitaire and Bourgogne Regional Councils, Agence Nationale de la Recherche, ANR supported the COGNIT and COVADIS projects. Fondation de France and the joint French Ministry of Research/INSERM «Cohorts et collections de données biologiques» programme. Lille Génopôle received an unconditional grant from Eisai. The Three-city biological bank was developed and maintained by the laboratory for genomic analysis LAG-BRC - Institut Pasteur de Lille. Belgium sample collection: The patients were clinically and pathologically characterized by the neurologists Sebastiaan Engelborghs, Rik Vandenberghe and Peter P. De Deyn, and in part genetically by Caroline Van Cauwenberghen, Karolen Bettens and Kristel Sleeegers. Research at the Antwerp site is funded in part by the Belgian Science Policy Office Interuniversity Attraction Poles program, the Foundation Alzheimer Research (SAO-FRA), the Flemish Government initiated Methusalem Excellence Program, the Research Foundation Flanders (FWO) and the University of Antwerp Research Fund, Belgium. Karolien Bettens is a postdoctoral fellow of the FWO. The Antwerp site authors thank the personnel of the VIB Genetic Service Facility, the Biobank of the Institute Born-Bunge and the Departments of Neurology and Memory Clinics at the Hospital Network Antwerp and the University Hospitals Leuven. Finish sample collection: Financial support for this project was provided by the Health Research Council of the Academy of Finland, EVO grant 5772708 of Kuopio University Hospital, and the Nordic Centre of Excellence in Neurodegeneration. Italian sample collections: the Bologna site (FL) obtained funds from the Italian Ministry of research and University as well as Carimonte Foundation. The Florence site was supported by grant RF-2010-231972, grant from the Cassa di Risparmio di Pistoia e Pescia (Grant 2012) and the Cassa di Risparmio di Firenze (Grant 2012). The Milan site was supported by a grant from the...
Introduction

The prevalence of Alzheimer’s disease (AD) is increasing as more people live into old age. Hope for finding preventative and clinical therapies lies in the ability to gain a better understanding of the underlying biology of the disease, and genetics will provide a valuable starting point for advancement. Rare monogenic forms of AD, the majority of which are attributable to mutations in one of three genes, APP, PSEN1 and PSEN2, exist, but common, late-onset AD is genetically complex with heritability estimated to be between 56–79%[1,2]. Along with the APOE polymorphism[3], 20 common susceptibility loci have been identified associated with AD[4–9]. (This figure does not include CD33 as it did not show genome-wide significance in the original report[9].) Recently, a moderately rare variant in TREM2 has also shown evidence for association[10]. However, new variants remain to be found. This study sought to identify new susceptibility genes, using an alternative gene-wide analytical approach, which focuses on the pattern of association within gene regions.

Genome-wide association (GWA) studies to date have focused on single nucleotide polymorphisms (SNPs) as the unit of analysis. Single locus tests are the simplest to generate and to interpret, but have limitations. For example, if susceptibility is conferred by multiple variants within a locus[11,12], this gives rise to complex patterns of association that might not be reflected by association to the same SNPs in different samples, despite apparently reasonably powered tests[13,14]. In addition, rare risk-increasing variants may not be tagged by single SNPs, as is e.g. the case for CLU in which significant enrichment of rare variants in patients was observed independent of the single locus GWA signal[15]. It is therefore likely that the power to detect association might be enhanced by exploiting information from multiple signals within genes encompassed by gene-wide statistical approaches[12]. Disease risk may reflect the co-action of several loci but the number of loci involved at the individual or the population levels are unknown, as is the spectrum of allele frequencies and effect sizes[16]. The observations of multiple genome-wide significant or suggestive linkage signals for disorders, that do not readily replicate between studies but which are not randomly distributed across the genome[17,18] is compatible with the existence of multiple risk alleles of moderate effect that would implicate a locus in disease risk, when analysed together. Thus the first aim of this study is to test for gene-wide association with AD, using a powerful mega-meta-analysis of genome-wide datasets as part of the International Genomics of Alzheimer’s Project (IGAP) Consortium comprising four AD genetic consortia (see the full list of consortia members in Materials S1): Genetic and Environmental Risk in Alzheimer’s Disease (GERAD), European Alzheimer’s Disease Initiative (EADI), Cohorts for Heart and Aging in Genomic Epidemiology (CHARGE) and Alzheimer’s Disease Genetics Consortium (ADGC) (see full IGAP datasets description in Materials S2). A two stage study was undertaken. In Stage 1 the combined sample included 17,008 AD cases and 37,154 controls. In Stage 2 loci with p-values (combined over all SNPs at the locus) less than 10^{-8} were selected for replication for 8,572 AD cases and 11,312 controls of European ancestry. We observed evidence for genome-wide association at loci which implicate genes which already show genome-wide significant association from single SNP analysis (CRI, B3N1, HLA-DRB5/HLA-DRB1, CD2AP, EPHA1, PTK2B, CLU, MS4A6A, PICALM, SORL1, SLC24A4, ABCA7, APOE), three new genes in the vicinity of lately reported single SNP hits[9] (ZNF3, NDUF83, MTHF2) and two novel loci (TP53INP1, combined p = 1.4×10^{-6} and IGHV1-67 combined p = 7.9×10^{-6}).

Results

Initially, we tested for excess genetic signal revealed by the Stage 1 IGAP SNP GWAS study. We observed more SNPs at all significance intervals, and more genes at multiple significance thresholds, than expected by chance (Table S1). This is unlikely to be due to uncorrected stratification, since each of the individual GWAS samples in the IGAP Stage 1 analysis was corrected for...
Table 1. Overrepresentation of replication of significant genes/loci available at Stage 2, excluding all loci of 0.5 Mb around genes previously reported[4–8] and Stage 1 IGAP genes[9,19] containing genome-wide significant SNPs.

<table>
<thead>
<tr>
<th>GENES</th>
<th>LOCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1 significance level</td>
<td>Significant at Stage 1</td>
</tr>
<tr>
<td>p≤10^{-4}</td>
<td>27</td>
</tr>
<tr>
<td>p≤10^{-5}</td>
<td>74</td>
</tr>
<tr>
<td>p≤0.01</td>
<td>229</td>
</tr>
<tr>
<td>p≤0.05</td>
<td>390</td>
</tr>
<tr>
<td>Total (p≤1)</td>
<td>887</td>
</tr>
</tbody>
</table>

The observed p-values were calculated with chi-square/Fisher’s exact tests counting the genes within 0.5 Mb as one locus.
doi:10.1371/journal.pone.0094661.t001

Over-representation genetic signals, in addition to those detected by the primary analysis[9,19].

Next, we looked at overrepresentation of significant genes in the Stage 1 data. Table 1 gives the observed and expected numbers of significant genes at significance levels 10^{-4}, 10^{-5}, 10^{-6} when all genes are counted in the analyses and when the known genes (Table S1) and genes within 500kb of them are excluded, the observed numbers of genes are much larger than expected at all significance levels (all p≤0.001). Thus there are more loci associated with AD to find.

Furthermore, the number of independent nominally significant loci at Stage 2 (N=60, (13.5%)) was significantly greater than expected by chance (p = 4.6×10^{-12}). The percentage of replicated loci increased with the decrease of the gene-wise significance threshold at Stage 1 (see Table 2 for details).

Combining the gene-wide p-values in both stages 1 and 2, using Fisher’s method revealed two new gene-based genome-wide significant (p<2.5×10^{-5}) loci TP53INP1 and IGHV1-67. The TP53INP1 gene is located on chromosome 8:95,938,200–95,961,615 and its combined gene-based p-value = 1.4×10^{-6} (Table 3). Table S3 provides details for each SNP contributing to the gene-based result. Out of 45 SNPs in the gene, three SNPs (rs4735333, rs1713669, rs896855) have p-value≤10^{-4}. Figure 1 shows the LD plot of this gene and suggests that there are at least two partially independent signals in the TP53INP1 gene (r^2 between the pairs of most significant SNPs rs4735333-rs1713669 and rs1713669- rs896855 are 0.65 and 0.6 respectively).

The IGHV1-67 gene on chromosome 14:107,136,620–107,137,059 has combined p-value = 7.9×10^{-6} (Tables 3). This gene is covered by two SNPs (rs2011167, rs1961901), both are significant at 10^{-4} level. LD plot in Figure 2 and Table S4 indicate that the two most significant SNPs in IGHV1-67 gene represent almost the same signal (r^2 = 0.92, calculated with SNAP software[20], 1000 genomes Pilot 1 dataset, CEU population panel, [http://www.broadinstitute.org/mpg/snap]).

To look at the gene expression patterns in these novel genes, we used the Webster-Myers expression dataset[21], available at http://labs.med.miami.edu/myers/LFuN/data%20ajhg.html. Comparing 137 AD vs 176 controls with temporal or frontal cortex expression values by t-test, t showed significantly higher TP53INP1 expression in cases compared to controls (p = 0.0128).

Further examination in the BRAINEAC database[22] (www.braineac.org) from the UK Brain Expression Consortium showed TP53INP1 to have a best cis-eQTL p-value of 6.8×10^{-6} (for rs4382532 SNP, which is about 7.6 kb upstream of the gene). The three SNPs with association p≤10^{-4} mentioned above (rs4735333, rs1713669, rs896855) had significant cis-eQTL p-values of 8.2×10^{-6}, 7.8×10^{-5} and 1.1×10^{-6} respectively in BRAINEAC brain expression data. The r^2 between the cis-eQTL and the three associated SNPs were 0.80, 0.65, and 0.81, respectively. Further analysis of additional independent brain expression and methylation datasets (see Methods S1) indicated significant cis-eQTLs and meQTLs for TP53INP1 (Tables S10 and S11). The probe for the meQTL is in a CpG island region that corresponds well with ENCODE DNAseq/Histone marks and is located upstream (~1.5 kb) of the TP53INP1 gene.

Table 2. Overrepresentation of significant loci, excluding regions of 0.5 Mb around previously reported[4–8] and Stage 1 IGAP genes[9,19] containing genome-wide significant SNPs.

<table>
<thead>
<tr>
<th>Numbers of loci (genes)</th>
<th>p≤10^{-4}</th>
<th>p≤10^{-5}</th>
<th>p≤10^{-6}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>2.5</td>
<td>0.25</td>
<td>0.025</td>
</tr>
<tr>
<td>Expected</td>
<td>0.001</td>
<td>0.0013</td>
<td>0.0003</td>
</tr>
</tbody>
</table>

The observed number of genes is calculated by combining significant loci within 0.5 Mb into one signal. The APOE region is excluded (CHR19; 44,411,940–46,411,945bp). The total number of genes after exclusions is 24,849.
doi:10.1371/journal.pone.0094661.t002
transcription start site. In combination these results suggest a possible epigenetic mechanism whereby the associated variants in the region influence TP53INP1 expression in several brain regions. These expression data provide further evidence supporting the functional relevance of TP53INP1 to AD susceptibility. The IGHV1-67 gene was not found in those databases.

In addition we detected two genome-wide significant loci 1) ZNF3 (chr7: 99,661,653–99,679,371; p = 8.6 × 10⁻⁷) and 2) two closely located genes on chromosome 11 MITCH2 (47,638,358–47,664,206, combined p = 2.5 × 10⁻⁶) and NDUFS3 (47,600,632–47,606,114, combined p = 4.3 × 10⁻⁷) (Table 4). None of these genes harbour genome-wide significant SNPs in the SNP GWAS analysis on its own (see Tables S5-S7). Figures S1-S3 show LD structure of these additional genes.

ZNF3 and NDUFS3, MITCH2 genes on chromosomes 7 and 11, respectively, lie close to rs1476679 (chr7: 100,004,446; ZCWPW1) and rs1083872 (chr11: 47,557,871; CELF1) SNPs, which are shown to be genome-wide significant in the IGAP study, when combining Stage 1 and Stage 2 data. Figures S1-S3 show LD structure of these genes in relation to the IGAP single genome-wide significant hits. (Note that the NDUFS3 gene on chromosome 11 was gene-based genome-wide significant already at Stage 1.) Although none of these SNPs actually lie within the genes mentioned above, it is possible that they may account for the gene-based signals through linkage disequilibrium. In order to test whether the gene-based signals are independent of these strongly-associated SNPs, we performed single-SNP association for each SNP annotated to these genes by regression, adjusting for the significant SNPs mentioned above, along with the other study covariates. The resulting p-values were combined into gene-based tests, as described previously. Under this conditional analysis ZNF3 gene does not show significant association, however NDUFS3 still shows a trend towards significance (p = 0.081) (see Table S8 for details).

Furthermore, five genes in chr11:47,593,749–47,615,961 (KBTBD4, NDUFS3, LOC100287127, FAM180B, C1QTNF4) all have p<0.05 with gene-based analysis ± 10 kb, when conditioning by the genome-wide significant hit rs10838725 in this region. This may partially be explained by the SNP rs10838731 (p = 1.2 × 10⁻³ after conditioning by rs10838725) which is shared by all latter five genes.

Gene-based analysis with ± 10 kb around genes did not reveal additional genome-wide significant loci in the Stage 1 data set. Moreover, the significance of the genes identified above did not improve in general, indicating that adding 10 kb flanking regions to genes introduces more noise to the gene-based signal. The combined Stage 1 and Stage 2 gene-based analysis provided further evidence for significant signals in the loci on chr 11 with 8 genes (SPH1, SLC39A13, LOC100287086, PTPMT1, KBTBD4, NDUFS3, LOC100287127, FAM180B) and on chr 7 with 6 genes (LOC100128334, MCM7, PILRB, PILRA, LOC100289298, C7orf51), all reaching genome-wide significance. This is likely to be due to the fact that including genes' flanking regions captures a greater number of the same SNPs or SNPs in high LD showing significant association.

The Manhattan plot of the gene-based p-values (Figure 3) gives a general overview of the gene-based results and shows the new loci in relation to previously reported genes (see also QQ-plots in Figure S4). The results of gene-wide analysis for the genes, which were previously reported as associated with AD[4-8] and those which are GWAS significant in the Stage 1 analysis are presented in Table S9. Out of 16 reported susceptibility genes, 15 are nominally significant with gene-wide analysis (almost all p-values are smaller than 10⁻⁴), however not all of them reach the gene-based genome-wide significance level (2.5 × 10⁻⁶) when the number of SNPs per gene and LD structure of the gene is taken into account.
Table 3. New genome-wide significant genes associated with AD.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Chr</th>
<th>Position</th>
<th>Combined best SNP p-value</th>
<th>Stage 1 gene-wide p-value</th>
<th>Stage 2 gene-wide p-value</th>
<th>N of SNPs per gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP53INP1</td>
<td>8</td>
<td>95,938,200–95,961,615</td>
<td>1.7×10^{-10}</td>
<td>1.7×10^{-10}</td>
<td>1.2×10^{-10}</td>
<td>45</td>
</tr>
<tr>
<td>IGHV1-67</td>
<td>14</td>
<td>107,136,620–107,137,059</td>
<td>2.2×10^{-4}</td>
<td>2.2×10^{-4}</td>
<td>2.2×10^{-4}</td>
<td>5</td>
</tr>
<tr>
<td>NDUFS3</td>
<td>7</td>
<td>47,638,88–47,664,206</td>
<td>2.5×10^{-10}</td>
<td>2.5×10^{-10}</td>
<td>2.5×10^{-10}</td>
<td>34</td>
</tr>
</tbody>
</table>

Table 3. New genome-wide significant genes associated with AD.

We did not observe genome-wide significance for CD33 gene. This gene was genome-wide significant in Stage 1 (p = 1.9×10^{-5}), but the association was attenuated when combining Stage 1 and Stage 2 data (p = 1.79×10^{-7}), similar to the single SNP association result in the SNP GWAS study[9,19].

**Discussion**

In this study we show that there are more signals in the GWAS imputed data at SNP- and gene-based levels than revealed by single SNP analysis. A gene-based analysis is a next logical step after the single SNP analyses in any attempt to combine possible several signals in genes and thus enhance the power of the association analyses.

The first new gene TP53INP1 (chromosome 8) encodes a protein that is involved in mediating autophagy-dependent cell death via apoptosis through altering the phosphorylation state of p53[23] and in modulating cell-extracellular matrix adhesion and cell migration[24]. TP53INP1 encodes a pro-apoptotic tumor suppressor and its antisense oligonucleotide has been used as potential treatment for castration-resistant prostate cancer[25]. This association is notable, given the potential inverse association between cancer and AD that has previously been reported[26,27].

The second new gene IGHV1-67 (chromosome 14) is a pseudogene in the immunoglobulin (IgG) variable heavy region of chromosome 14: its function is unknown but all genes in this region are most likely to be involved in IgG heavy chain VDJ recombinations that lead to the full repertoire of antigen-detecting immune cell clones[28].

The gene-based analysis in this study has shown its utility to enhance the information provided by single SNP analysis (i.e. NDUFS3 gene was genome-wide significant from Stage 1 using gene-based analysis whereas this gene was only genome-wide significant after combining the two stages of single SNP analysis). NF3 is a zinc-finger protein at the same locus on chromosome 7 as CWPW1 thus rendering it a candidate as the gene that contains the functional signal in this region. Although we can not identify which gene actually confers the risk to AD, it is interesting that NF3 function is unknown though it interacts with BAG3 which is involved in ubiquitin/proteasomal functions in protein degradation[29] and NF3 is regulated by upstream binding of BACH1 whose target genes have roles in the oxidative stress response and control of the cell cycle[30].

In the cluster of genes on chromosome 11, MITCH2 encodes one of the large family of inner mitochondrial membrane transporters[31] which is associated with mitochondrially-mediated cell death[32], adipocyte differentiation[33], insulin sensitivity[34] and has a genetic association with increased BMI[35]. NDUFS3 also has functions in the mitochondria as it encodes an iron-sulphur component of complex I (mitochondrial NADH:ubiquinone oxidoreductase) of the electron transport chain. A deficiency causes a form of Leigh syndrome[36] an early-onset progressive neurodegenerative disorder with a characteristic neuropathology consisting of focal lesions including areas of demyelination and gliosis[37].

In summary, we report two novel genes TP53INP1 (chr8: 95,938,200–95,961,615; combined p = 1.4×10^{-5}) and IGHV1-67 (chr14: 107,136,620–107,137,059; combined p = 7.9×10^{-7}), which were not reported as genome-wide significant before. We also report NF3 gene on chromosome 7 and a cluster of genes on chromosome 11 (SPI1-MITCH2), showing gene-based genome-wide significant association with Alzheimer's disease. These genes are in proximity with, but not the same as, those detected by genome-wide significant SNPs, demonstrating support for the
signals identified by IGAP[9,19]. They have an array of functions previously implicated in AD including aspects of energy metabolism, protein degradation and the immune system and add further weight to these pathways as potential therapeutic targets in AD.

Materials and Methods

Stage 1 data

The main dataset was reported by the IGAP consortium[9,19] and consists in total of 17,008 cases and 37,154 controls. This sample of AD cases and controls comprises 4 data sets taken from genome-wide association studies performed by GERAD, EADI, CHARGE and ADGC (see primary IGAP manuscript[9,19] for more details). The full details of the samples and methods for conduct of the GWA studies are provided in the respective manuscripts[4-8].

Each of these datasets was imputed with Impute2[38] or MACH[39] software using the 1000 genomes data [release Dec2010] as a reference panel. In total 11,863,202 SNPs were included in the SNPs allelic association result file. To make our analysis as conservative as possible, we only included autosomal SNPs which passed stringent quality control criteria, i.e. we included only SNPs with minor allele frequencies (MAF) ≥0.01 and imputation quality score greater than or equal to 0.3 in each individual study, resulting in 7,055,881 SNPs which are present in at least 40% of the AD cases and 40% of the controls in the analysis. The summary statistics across datasets were combined using fixed-effects inverse variance-weighted meta-analysis. We corrected all individual SNPs p-values for genomic control (GC) \( \lambda = 1.087 \). These SNPs are well imputed on a large proportion of the sample, which increases confidence in the accuracy of the association analysis upon which gene-wide analysis is based.

Stage 2 data

11,632 SNPs with p-values <10^{-3} in the IGAP meta-analysis were successfully genotyped in a Stage 2 sample comprising 8,572 cases and 11,312 controls (see primary IGAP manuscript[9,19] for more details). An additional 771 SNPs were successfully genotyped to test all genes with gene-wide p-values <10^{-4} in the IGAP Stage 1 analysis, excluding genes reported prior to IGAP[4–8], the four loci reaching genome-wide significance in the Stage 1 IGAP meta-analysis[9,19] and the 0.5Mb regions around them (Table S2). These SNPs cover 887 genes and correspond to 444 independent loci where all genes within 0.5 Mb are counted as one locus.

Assignment of SNPs to genes

SNPs were assigned to genes if they were located within the genomic sequence lying between the start of the first and the end of the last exon of any transcript corresponding to that gene. The chromosome and location for all currently known human SNPs were taken from the dbSNP132 database, as was their assignment to genes (using build 37.1). In total, we retained 2,304,431 (39.7% of the total) SNPs which annotated 28,636 unique genes with 1–16,514 SNPs per gene. For the gene-wide analysis we have excluded genes which contain only one SNP in the IGAP Stage 1 analysis, leaving a total of 25,310 genes. If a SNP belongs to more than one gene, it was assigned to each of these genes. In order to account for possible signals which are correlated with those in a gene, gene-wide analysis was also performed using a 10 kb window around genes to assign SNPs to genes.
Table 4. New genome-wide significant genes associated with AD in the vicinity of recently reported single SNP genome-wide significant hits[9,19].

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Chr</th>
<th>Position</th>
<th>Combined best stage 1 p-value</th>
<th>Combined best stage 2 p-value</th>
<th>N of SNPs per gene</th>
<th>Combined best stage 1 and 2 p-value</th>
<th>Biological function</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZNF3</td>
<td>7</td>
<td>99,661,653-99,679,371</td>
<td>2.7×10^-7</td>
<td>2.7×10^-7</td>
<td>27</td>
<td>4.8×10^-7</td>
<td>Transcription factor, leucocyte activation</td>
</tr>
<tr>
<td>11</td>
<td>47,600,632-47,606,114</td>
<td>3.7×10^-6</td>
<td>3.7×10^-6</td>
<td>34</td>
<td>ND of SNPs</td>
<td>Mitochondrial electron transport, NADH to ubiquinone</td>
<td></td>
</tr>
<tr>
<td>NDUF3</td>
<td>11</td>
<td>47,638,858-47,664,206</td>
<td>1.7×10^-6</td>
<td>1.7×10^-6</td>
<td>1</td>
<td>10</td>
<td>Mitochondrial inner membrane</td>
</tr>
</tbody>
</table>

Gene-wide analysis

The gene-wide analysis was performed based on the summary p-values while controlling for LD and different number of markers per gene using an approximate statistical approach[40] adopted for set-based analysis of genetic data[41]. This is a method for calculating the significance of a set of SNPs in the absence of individual genotype data based on a theoretical approximation to Fisher’s statistic for combining p-values. Fisher’s statistic (\(\sum \ln(p_i)\)) combines probabilities and under the null hypothesis has a chi-square distribution with 2N degrees of freedom, where N is the number of markers, and the summation above is for \(i = 1,\ldots,N\). If Fisher’s statistic combines the results of several tests when the tests are independent, the approximate method combines non-independent tests and requires only the list of p-values for each SNP and knowledge of correlations between SNPs. Then the value of Fisher’s statistic and the number of degrees of freedom is corrected by the coefficient which depends upon the number of SNPs and correlations (LD) between them. This approximation was applied to the Stage 1 and Stage 2 samples separately, and the resulting gene-wide p-values combined using Fisher’s method (since these are independent). LD between markers was computed using 1000 genomes data. The gene-based genome-wide significant level was set to 2.5×10^-6 to account for the number of tested genes[42].

Test for excess of associated SNPs/loci

The effective number N of independent SNPs in the whole genome (excluding genes with SNPs that are genome-wide significant in the Stage 1 IGAP dataset \(\pm 0.5\) Mb was estimated by the method described in [43] taking LD into account, as were the observed number of independent SNPs significant at each p-value criterion (adjusting individual SNP p-values for genomic control \(\lambda = 1.087\) before hand). LD was computed from the 1000 Genomes database (http://www.1000genomes.org/). In the absence of excess association, the expected number of independent SNPs significant at significance level \(\alpha\) is a normally distributed random variable whose mean and standard deviation (SD) can be calculated as \(\alpha N\) and \(\sqrt{\frac{\alpha}{2}}\) (mean and SD for a binomial distribution). The number of independent SNPs (and thus statistical tests) in the whole genome were estimated as \(\sim 3.7 \times 10^5\), \(\sim 3.6 \times 10^5\) and \(\sim 3.5 \times 10^5\) at significance levels below 0.1, between 0.05 and 0.1, and 0.2 and above respectively [43] for details on the dependence between the significance levels and the estimated number of independent tests). We then calculated mean of the expected number of significant SNPs in intervals \(z_1 < p \leq z_2\), \(z_1 = 0, 10^{-5}, 10^{-6}, \ldots, 0.5\) as difference between the expected numbers of independent SNPs at \(z_2\) and \(z_1\) significance levels and SD as the square root of sum of the corresponding variances.

We calculated the significance of the excess number of genes attaining the specified thresholds based upon the assumption that, under the null hypothesis of no association, the number of significant genes at a significance level of \(\alpha\) in a scan is distributed as a binomial \((N, \alpha)\), where \(N\) is the total number of genes, assuming that genes are independent. Genes within 0.5 Mb of each other are counted as one signal when calculating the observed number of significant genes. This prevents significance being inflated by LD between genes, where a single association signal gives rise to several significantly-associated genes. The total number of genes was not corrected for LD in this way, making the estimate of significance of the excess number of genes conservative.
### Supporting Information

**Table S1** Overrepresentation of significant SNPs excluding previously reported\([4–8]\) genes ±0.5Mb and the APOE region as above.

**Table S2** List of genes that are genome-wide significant in the IGAP stage 1 dataset and the flanking regions which included SNPs either in $r^2\geq0.3$ or association p-value $\leq10^{-3}$ whichever covers the largest region.

**Table S3** Detailed SNP information for TP53INP1 gene.

**Table S4** Detailed SNP information for IGHV1-67 gene.

**Table S5** Detailed SNP information for ZNF3 gene.

**Table S6** Detailed SNP information for NDUFS3 gene.

**Table S7** Detailed SNP information for MTCH2 gene.

**Table S8** Gene-based analysis results, when single SNPs p-values, contributing to the gene-based p-value were adjusted for the best genome-wide significant SNP in the nearby location.

**Table S9** Gene-wide analysis for genes which show GWAS significant association with AD in the stage 1 IGAP dataset.

**Table S10** Brain eQTL Tissues.

**Table S11** Brain Meth QTLs.

**Figure S1** ZNF3 gene with rs1476679 (ZCWPW1) reported by Lambert et al (2013) study. SNPs which are significant at 1e-3 level are circled in red, rs1476679 is highlighted in blue.

**Figure S2** NDUFS3 gene rs10838725 (CELF1) reported by Lambert et al (2013) study. SNPs which are significant at 1e-3 level are circled in red, rs10838725 is highlighted in blue.

**Figure S3** MTCH2 gene with rs10838725 (CELF1) reported by Lambert et al (2013) study. SNPs which are significant at 1e-3 level are circled in red, rs10838725 is highlighted in blue.

**Figure S4** QQ-plot of gene-wide p-values for all genes (A) and excluding previously reported\([4–8]\) GWAS significantly associated genes ±0.5Mb (B) in the discovery dataset. Genomic control $\lambda = 1.08$ and 1.07 respectively.

**Methods S1** Expression quantitative trait loci (eQTL) and Methylation quantitative trait loci (meQTL) analyses.

**Materials S1** Full IGAP datasets description.

**Materials S2** List of IGAP consortium members.
Materials S3 Acknowledgements.

(Doctoral Dissertation Title)

Acknowledgments

This work was made possible by the generous participation of the control subjects, the patients and their families. Complete acknowledgments are detailed in the Materials S3.

Author Contributions

Conceived and designed the experiments: VEP D. Harold P. Holmans S. Seshadri GDS PA JW. Analyzed the data: VEP JCL. Bellengues LSW SHC BGB GR TATW ND AVS VC. Contributed reagents/materials/analysis tools: VEP C. Bellengues LSW SHC D. Harold P. Holmans A. Richards AJ AV GR MV VC. Contributed Author Contributions detailed in the Materials S3.


References


8. The New Susceptibility Genes for Alzheimer’s Disease.
