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

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Published Version</td>
<td>doi:10.1371/journal.pone.0094661</td>
</tr>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:12406561">http://nrs.harvard.edu/urn-3:HUL.InstRepos:12406561</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>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 <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
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,
Discipline of Psychiatry, Trinity College, Dublin, Ireland, Neurodegenerative Brain Diseases Group, Department of Molecular Genetics, VIB, Antwerp, Departments of Epidemiology and Global Health, Department of Neuroscience-University of Parma, Parma, Italy, Centro di Ricerca, Trasferimento e Alta Formazione DENOTHE, University of
Department of Psychiatry, University of Frankfurt am Main, Frankfurt am Main, Germany (H.H.), Institute of Primary Care and Public Health, Cardiff University, Neuadd Meirionnydd, Departments of Neurology,
3rd Department of Neurology, Aristotle University of Thessaloniki, Thessaloniki, Greece, Department of Psychiatry and Behavioral Sciences, Miller School of Medicine, University of Miami, Miami, Florida, United States
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 Catalá de Neurociències Aplicades, Barcelona, Spain, 27 Taib 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-MAI, 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-MAI, Inserrm 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, Washington University 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 Neurosciences, De Crespigny Park, Denmark Hill, London, United Kingdom, 47 Inserrm U897, Victor Segal 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 Research 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 Neurosciences, Frimley Park Hospital, Frimley, 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, Neaund Meriöynnöll, 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 Network 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 Alta Formazione DENOTHE, University of Florence, Florence, Italy, 72 Department of Immunology, Hospital Universitari Dr. Negrit, Las Palmas de Gran Canaria, Spain, 73 Dementia Research Center, Department of Neurogenetics, University of California, Los Angeles, United States of America, 74 Department of Molecular Neuroscience 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 Cà Granda, IRCCS Ospedale Policlinico, Milan, Italy, 79 Neurological Clinic, University of Pisa, Pisa, Italy, 80 Urban Epidemiology Institute, Medical for Medical 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, Frankfurt am Main, Germany (H.H.), 84 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 Biológica en Red sobre Enfermedades Neuropenetrativas (CIBERNED), Madrid, Spain, 89 Instituto de Investigación Sanitaria “Hospital la Paz” (IDILPaz), 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 Hospital, Saarland, Germany, 94 Department of Public Health/Geriatrics, Universita Uppsala, 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 molecolare-Huca-Oviedo, Oviedo, Spain, 99 Department of Psychiatry, University Hospital, Sааrland, 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 Neuropenetrativas, 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 Landskapis 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 25,380 Alzheimer’s cases and 48,466 controls.

Principal Findings: In addition to earlier reported genes, we detected genome-wide significant loci on chromosomes 8 (TPS5N1P, p = 1.4 × 10^-6) and 14 (IGHV1-67 p = 7.9 × 10^-9) which indexed novel susceptibility loci.

Significance: The additional genes identified in this study, have an array of functions previously implicated in Alzheimer’s disease, 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.
grants R01-HL087641, RC2-HL102419 (Boerwinkle, CHARGE-S), UC2 HL103010, U01-HL096917 (Mosley) and R01-HL093029; NHGRI contract U01-HG004402; and Heart, Lung, and Blood Institute contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, N01-HC-55022 and HJtaverned (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament). ASPS/PRODEM: The Austin Stroke Prevention Study and the Prospective Dementia Registry of the Austrian Alzheimer Society was supported by The Australian Science Foundation (IWF) grant number P20545-F05 (H. Schmidt) and P13180; The Austrian Alzheimer Society; The Medical University of Graz; Cardiovascular Health Study (CHS); This CHS research was supported by NHLBI contracts HHSN268201100036C, HHSN26820000007C, N01-HC-55022, N01HC85075, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, and HHSN268200600009C; and NHLBI grants HL080295, HL087562, HL105756 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through AG023629, AG19528, AG020758 and AG033193 (Seshadri) from the NIA. A full list of CH investigators and institutions can be found at http://www.nch-lhbi.org/p. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSA grant UL1TR000124, and the National Institute of Diabetes and Digestive and Kidney Disease Research Center (Contract) grant DK636491 to the Southern California Diabetes Endocrinology Research Center. Framingham Heart Study (FHS): This work was supported by the National Heart, Lung and Blood Institute's Framingham Heart Study (Contract No. N01-HC-25195) and its contract with A. yemetric, Inc for genotyping services (Contract No. N02-HL-6-4278). A portion of this research utilized the Linux Cluster for Genetic Analysis (LinGa-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. This study was also supported by grants from the National Institute on Aging; AG08112 and AG033193 (Seshadri). Drs. Seshadri and DeStefano were also supported by additional grants from the National Institute on Aging: (R01 AG16495; AG31287, 2004030), the National Institute of Neurological Disorders and Stroke (R01 NS17950), and the National Heart, Lung and Blood Institute (U01 HL09617; HL093029 and #2HL034844, RC2-HL102419 and UC2 HL103010. Fundacio ACE would like to thank patients and clinicians who participated in this project. The KORA (Kooperative Arbeitsgemeinschaft der Universitäten in der Region München, the Munich KORA) study has been funded by the German National Genome Research Network (NGFN); Alzheimer’s disease Integrated Genome Research Network; AD-IG: 01GS0465. Genotyping of the Bonn case-control sample was funded by the German centre for Neurodegenerative Diseases (DZNE), Germany. The GERAD Consortium also used samples and 9 million EUR grant of the Klinikum der Universität, Munich, Germany. The KORA F4 studies were financed by Helmholtz Zentrum München; German Research Center for Environmental Health; BMG: German National Genome Research Network and the Munich Center of Health Sciences. The Heinz Nixdorf Recall cohort was funded by the Heinz Nixdorf Foundation (Dr. Jur. G.Schmidt, Chairman) and BMF. Corell Repository is supported by NINDS and the Intramural Research Program of the National Institute on Aging. The authors would like to acknowledge the use of genotype data from the AIM study by the MRC and the Wellcome Trust which was supported by grants from the Wellcome Trust Case Control Consortium and the Type-1 Diabetes Genetics Consortium, sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Allergy and Infectious Diseases, National Human Genome Research Institute, National Institute of Child Health and Human Development and Juvenile Diabetes Research Foundation International. The Nottingham Group (RM) are supported by the Big Lottery. MRC CFAS is part of the consortium and data will be included in future analyses. ADGC. The National Institutes of Health, National Institute on Aging (NIA-NAA) supported this work through the following grants: ADGC, U01 AG032984, RC2 AG036528; NACC, U01 AG016976; NCRAD, U24 AG021886; NIA LOAD, U24 AG026395, R01 AG041779; MIRAGE R01 AG25259; Banner Sun Health Research Institute P30 AG19610; Boston University, P30 AG13046, U01 AG10483, R01 CA129769, R01 MH080295, R01 AG17173, R01 AG33193; Columbia University, P50 AG08702, R37 AG015473; Duke University, P30 AG028377, R01 AG05286; Group Health Research Institute, U01 AG06781, U13 AG004610; Indiana University, P30 AG10313; Johns Hopkins University, P50 AG005146, R01 AG206886; Massachusetts General Hospital, P50 AG005134; Mayo Clinic, P50 AG016574; Mount Sinai School of Medicine, P50 AG005138, P01 AG002219; New York University, P30 AG08051, MO1RR00095, and UL 1R028993; Northwestern University, P30 AG13854; Oregon Health & Science University, P30 AG08017, R01 AG026916; Rush University, P30 AG01061, R01 AG08905, R01 AG15819, R01 AG17971, R01 AG30146; Tgen, N01 NS059873; University of Alabama at Birmingham, P30 AG16582, UL1RR027776; University of Arizona, P30 AG10125; University of California, Davis, P50 AG10126; University of California, Los Angeles, P50 AG10127; University of California, Los Angeles, P50 AG10128; University of California, San Diego, P50 AG005113; University of California, San Francisco, P30 AG20351; University of Kentucky, P30 AG01724; University of Michigan, P50 AG08817; University of Pittsburgh, P30 AG010124; University of Tennessee, P50 AG05133, AG00563, AG041718; University of Southern California, P50 AG005142; University of Texas Southwestern, P30 AG010129; University of Washington, P30 AG010130; University of Wisconsin-Madison, P50 AG005124; University of Wisconsin, Madison, P50 AG005137; Vanderbilt University, R01 AG19985; and Washington University, P50 AG005681, P01 AG03991. The Kathleen Price Bryan Brain Bank at Duke University Medical Center is funded by NINDS grant 3R01 NS39764, NIH MH68041 and by Glaxo Smith Kline. Genotyping of the TGEN2 cohort was supported by Kronos Science. The Tgen series was also funded by NIA grant AG034504 to AJM, The Banner Alzheimer’s Foundation, The Johnnie B. Byrd Sr. Alzheimer’s Institute and Research Council, the University of Arizona, the University of Washington and the New Mexico Brain Tissue Bank (via the Medical Research Council, local NHS trusts and Newcastle University), MRC London Brain Bank for Neurodegenerative Diseases (funding via the Medical Research Council), South West Dementia Brain Bank (funding with the Newcastle University Trust and Research Innovation and Development Department), The Netherlands Brain Bank
(funding via numerous sources including Stichting MS Research, Brain Net Europe, Hersenstichting Nederland Breinbrekend Werk, International Parkinson Fonds, Internationale Stichting Alzheimer Onderzoek), Institut de Neurupatologia, Serveí Anatomia Patològica, Universitat de Barcelona. Marcelle Morrison-Bograd, PhD; Tony Phelps, PhD and Walter Kukulik PhD are thanked for helping to co-ordinate this collection. ADNI Funding for ADNI is through the Northern California Institute for Research and Education by grants from Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, Glaxo-SmithKline, Innogenetics, Johnson and Johnson, Eli Lilly and Co., Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer Inc, F. Hoffman-La Roche, Schering-Plough, Synarc, Inc., Alzheimer’s Association, Alzheimer’s Drug Discovery Foundation, the Dana Foundation, and by the National Institute of Biomedical Imaging and Bioengineering and NIA grants U01 AG024904, RC2 AG036533, K01 AG030514. Data collection and sharing for this project was funded by the ADNI (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; BioClinica, Inc; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc; Eli Lilly and Company; F. Hoffman-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, NV.; IBCO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC; Johnson & Johnson Pharmaceutical Research & Development LLC; Medpace, Inc.; Merck & Co., Inc.; Mes Scale Diagnostics, LLC; NeuroX Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer’s Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG10129 and K01 AG030514. The authors thank Drs. D. Stephen Snyder and Marilyn Miller from NIA who are ex-o, cia ADGC members. Support was also from the Alzheimer’s Association (LAF, IIRG-08-89720; MP-V, IIRG-05-14147) and the United States Department of Veterans Affairs Administration, Office of Research and Development, Biomedical Laboratory Research Program. Peter St George-Hyslop is supported by Wellcome Trust, Howard Hughes Medical Institute, and the Canadian Institute of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Bruce M. Patsy serves on the DSMB for a clinical trial of a device funded by the manufacturer (Zoll LifeCor) and on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. Data used in preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni. loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf. This does not alter the authors’ adherence to PLOS ONE policies on sharing data and materials.

*E-mail: WilliamsJ@cf.ac.uk (JW); philippe.amouyel@pasteur-lille.fr (PA); gerardsc@mail.med.upenn.edu (GDS)

These authors contributed equally to this work.

Membership of the UK Brain Expression consortium is provided in Materials S1.

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 forCLU 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^{-4} were selected for replication for 8,572 AD cases and 11,312 controls of European ancestry. We observed evidence for gene-wide association at loci which implicate genes which already show genome-wide significant association from single SNP analysis (CR1, BIN1, HLA-DRB1/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, NDUFS3, MITCH2) 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>

Over-representation 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

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>9(27)</td>
<td>4(8)</td>
<td>2(2)</td>
</tr>
<tr>
<td>Expected</td>
<td>2.5</td>
<td>0.25</td>
<td>0.025</td>
</tr>
<tr>
<td>p-value</td>
<td>0.001</td>
<td>0.00001</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

ethnic variation. Thus it is likely that the sample contains novel 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 significance levels (all p < 0.05). LD plot in Figure 2 and Table S4 indicate that the two most significant SNPs in IGHV1-67 gene represent almost the same signal (r² = 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 expression and methylation datasets (see Methods S1) indicated that there are at least two partially independent signals in the TP53INP1 gene (r² 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^{-8} (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 (r² = 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 expression and methylation datasets (see Methods S1) indicated that there are at least two partially independent signals in the TP53INP1 gene (r² between the pairs of most significant SNPs rs4735333-rs1713669 and rs1713669- rs896855 are 0.65 and 0.6 respectively).

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 MTCH2 (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^{-5}) (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, MTCH2 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 singe 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 chr11 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.
We did not observe genome-wide significance for CD33 gene. This gene was genome-wide significant in Stage 1 ($p = 1.9 \times 10^{-7}$), but the association was attenuated when combining Stage 1 and Stage 2 data ($p = 1.79 \times 10^{-7}$), 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 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 chain 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., NDUF3 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).

The second new gene IG2M (chromosome 14) is a pseudogene in the immunoglobulin (IgG) variable heavy chain 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].

In the cluster of genes on chromosome 11, MTCH2 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]. NDUF3 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 $p = 1.4 \times 10^{-7}$) and IGHHV1-67 (chr14: 107,136,620–107,137,059; combined $p = 7.9 \times 10^{-5}$), which were not reported as genome-wide significant before. We also report NDUF3 gene on chromosome 7 and a cluster of genes on chromosome 11 (SPI1-MTCH2), 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) $\geq 0.01$ 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 2,804,431 (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.
New genome-wide significant genes associated with AD in the vicinity of recently reported single SNP genome-wide significant hits[9,19].

Table 4.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Chr</th>
<th>Position</th>
<th>Combined best SNP p-value</th>
<th>Combined best SNP p-value per gene</th>
<th>Biological function</th>
</tr>
</thead>
</table>
| ZNF3      | 2   | 1.8      | 2.2 × 10^{-7}             | 1.3 × 10^{-7}                     | Transcription factor,
|           |     |          |                           |                                   | leucocyte activation |
| MTO1      | 5   | 1.1      | 2.5 × 10^{-7}             | 1.1 × 10^{-7}                     | Mitochondrial electron transport,
|           |     |          |                           |                                   | NADH to ubiquinone    |
| MTCH2     | 11  | 1.2      | 4.8 × 10^{-7}             | 4.8 × 10^{-7}                     | Mitochondrial inner membrane |
| NDUFS3    | 7   | 2.5      | 2.5 × 10^{-6}             | 2.5 × 10^{-6}                     | Mitochondrial inner membrane |

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 (\(\sum l_{i} \ln(p_{i})\)) 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 \(i = 1, ..., N\). 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 ± 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 λ = 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 𝜋 is a normally distributed random variable whose mean and standard deviation (SD) can be calculated as \(2N\) and \(\sqrt{2N(1-\pi)}\) (mean and SD for a binomial distribution). The number of independent SNPs (and thus statistical tests) in the whole genome were estimated as \(3.7 \times 10^{6} - 3.6 \times 10^{6}\) and \(3.5 \times 10^{6}\) at significance levels below 0.1, between 0.05 and 0.1, and 0.2 and above respectively (see [43] for details on the dependence between the significance levels and the estimated number of independent tests). We then calculated the mean of the expected number of significant SNPs in intervals \(z_{21} < p \leq z_{2}, (z_{1}, z_{2} = 0, 10^{-5}, 10^{-6}, ..., 0.5)\) 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 \(z\) in a scan is distributed as a binomial \((N, z)\), 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 ±0.5 Mb 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 r² ≥ 0.3 or association p-value ≤ 10⁻³ 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 ±0.5 Mb (B) in the discovery dataset. Genomic control λ = 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.

(DOCX)

Acknowledgments

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

Author Contributions


References


New Susceptibility Genes for Alzheimer’s Disease

PLOS ONE | www.plosone.org 12 June 2014 | Volume 9 | Issue 6 | e94661

