Citation for final published version:

Escott-Price, Valentina, Bellenguez, Céline, Wang, Li-San, Choi, Seung-Hoan, Harold, Denise, Jones, Lesley, Holmans, Peter, Gerrish, Amy, Vedernikov, Alexey, Richards, Alexander, DeStefano, Anita L., Lambert, Jean-Charles, Ibrahim-Verbaas, Carla A., Naj, Adam C., Sims, Rebecca, Jun, Gyungah, Bis, Joshua C., Beecham, Gary W., Grenier-Boley, Benjamin, Russo, Giancarlo, Thornton-Wells, Tricia A., Denning, Nicola, Smith, Albert V., Chouraki, Vincent, Thomas, Charlene, Ikram, M. Arfan, Zelenika, Diana, Vardarajan, Badri N., Kamatani, Yoichiro, Lin, Chiao-Feng, Schmidt, Helena, Kunkle, Brian, Dunstan, Melanie L., Vroonskaya, Maria, Johnson, Andrew D., Ruiz, Agustin, Bihoreau, Marie-Thérèse, Reitz, Christiane, Pasquier, Florence, Hollingworth, Paul, Hanon, Olivier, Fitzpatrick, Annette L., Buxbaum, Joseph D., Campion, Dominique, Crane, Paul K., Baldwin, Clinton, Becker, Tim, Gudnason, Vilmundur, Cruchaga, Carlos, Craig, David, Amin, Najaf, Berr, Claudine, Lopez, Oscar L., De Jager, Philip L., Deramecourt, Vincent, Johnston, Janet A., Evans, Denis, Lovestone, Simon, Letenneur, Luc, Hernández, Isabel, Rubinsztein, David C., Eiriksdottir, Gudny, Sleegers, Kristel, Goate, Alison M., Fiévet, Nathalie, Huentelman, Matthew J., Gill, Michael, Brown, Kristelle, Kamboh, M. Ilyas, Keller, Lina, Barberger-Gateau, Pascale, McGuinness, Bernadette, Larson, Eric B., Myers, Amanda J., Dufouil, Carole, Todd, Stephen, Wallon, David, Love, Seth, Rogaeva, Ekaterina, Gallacher, John, George-Hyslop, Peter St, Clarimon, Jordi, Lleo, Alberto, Bayer, Anthony, Tsuang, Debby W., Yu, Lei, Tsolaki, Magda, Bossù, Paola, Spalletta, Gianfranco, Proitsi, Petra, Collinge, John, Sorbi, Sandro, Garcia, Florentino Sanchez, Fox, Nick C., Hardy, John, Naranjo, Maria Candida Deniz, Bosco, Paolo, Clarke, Robert, Brayne, Carol, Galimberti, Daniela, Scarpini, Elio, Bonuccelli, Ubaldo, Mancuso, Michelangelo, Siciliano, Gabriele, Moebus, Susanne, Mecocci, Patrizia, Zompo, Maria Del, Maier, Wolfgang, Hampel, Harald, Pilotto, Alberto, Frank-García, Ana, Panza, Francesco, Solfrizzi, Vincenzo, Caffarra, Paolo, Nacmias, Benedetta, Perry, William, Mayhaus, Emanuel, Lannfelt, Lars, Hakonarson, Hakon, Pickler, Sabrina, Carrasquillo, Minerva M., Ingelsson, Martin, Beeley, Duane, Alvarez, Victoria, Zou, Fanggeng, Valladares, Otto, Younkin, Steven G., Coto, Eliecer, Hamilton-Nelson, Kara L., Gu, Wei, Razquin, Cristina, Pastor, Pau, Mateo, Ignacio, Owen, Michael J., Faber, Kelley M., Jonsson, Palmi V., Combarros, Onofre, O’Donovan, Michael C., Cantwell, Laura B., Soininen, Hilkka, Blacker, Deborah, Mead, Simon, Mosley Jr., Thomas H., Bennett, David A., Harris, Tamara B., Fratiiglioni, Laura, Holmes, Clive, de Bruijn, Renee F. A. G., Passmore, Peter, Montine, Thomas J., Bettens, Karolien, Rotter, Jerome L., Brice, Alexis, Morgan, Kevin, Foroud, Tatiana M., Kukull, Walter A., Hannequin, Didier, Powell, John F., Nalls, Michael A., Ritchie, Karen, Lunetta, Kathryn L., Kauwe, John S. K., Boerwinkle, Eric, Riemenschneider, Matthias, Boada, Mercè, Hiltunen, Mikko, Martin, Eden R., Schmidt, Reinhold, Rujescu, Dan, Dartigues, Jean-Francois, Mayeux, Richard, Tzourio, Christophe, Hofman, Albert, Nöthen, Markus M., Graff, Caroline, Psaty, Bruce M., Haines, Jonathan L., Lathrop, Mark, Pericak-Vance, Margaret A., Launer, Lenore J., Van Broeckhoven, Christine, Farrer, Lindsay A., van Duijn, Cornelia M.
Ramirez, Alfredo, Seshadri, Sudha, Schellenberg, Gerard D., Amouyel, Philippe and Williams, Julie 2014. Gene-wide analysis detects two new susceptibility genes for Alzheimer’s Disease. PLoS ONE 9 (6), e94661. 10.1371/journal.pone.0094661

Publishers page: http://dx.doi.org/10.1371/journal.pone.0094661

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher’s version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.
Gene-Wide Analysis Detects Two New Susceptibility Genes for Alzheimer’s Disease

Valentina Escott-Price1,9, Céline Bellenguez2,3,4,9, Li-San Wang5,9, Seung-Hoan Choi6,9, Denise Harold1, Lesley Jones1, Peter Holmans1, Amy Gerrish1, Alexey Vedernikov1, Alexander Richards1, Anita L. DeStefano5, Jean-Charles Lambert2,3,4, Carla A. Ibrahim-Verbaas7, Adam C. Naj8, Rebecca Sims1, Gyungah Jun6,9,10, Joshua C. Bis11, Gary W. Beecham12,13, Benjamin Grenier-Boley2,3,4, Giancarlo Russo14, Tricia A. Thornton-Wells15, Nicola Denning1, Albert V. Smith16,17, Vincent Chouraki2,3,4,18, Charlene Thomas1, M. Arfan Ikram19,20, Diana Zelenika21, Badri N. Vardarajan9,27,28, Yoichiro Kamatani22, Chiao-Feng Lin9, Helena Schmidt23, Brian Kunkle12, Melanie L. Dunstan1, Maria Vronskaya1, the United Kingdom Brain Expression Consortium24, Andrew D. Johnson25, Agustin Ruiz26, Marie-Thérèse Bihoue21, Christiane Reitz27,28, Florence Pasquier3,29, Paul Hollingworth1, Olivier Hanon30, Annette L. Fitzpatrick11,13, Joseph D. Buxbaum32,33,34, Dominique Campion35, Paul K. Crane36, Clinton Baldwin9, Tim Becker37, Vilmundur Gundnason16,17, Carlos Crucchaga38, David Craig39, Najaf Amin40, Claudine Berr41, Oscar L. Lopez42, Philip L. De Jager43,44, Vincent Deramecourt3,29, Janet A. Johnston39, Denis J. Evans45, Simon Lovestone46, Luc Letenneur17, Isabel Hernández26, David C. Rubinsztein48, Gudny Eiriksdottir19, Kristel Sleegers49,50, Alison M. Goate38, Nathalie Fiévet2,4, Matthew J. Huentelman51, Michael Gill52, Kristelle Brown53, M. Ilyas Kambho54,55, Lina Keller56, Pascale Barberger-Gateau46, Bernadette McGuinness39, Eric B. Larson36,57, Amanda J. Myers58, Carole Dufouil57, Stephen Todd39, David Wallon35, Seth Love59, Ekaterina Rogaeva60, John Gallacher61, Peter St George-Hyslop60,62, Jordi Clarimon63,64, Alberto Lleo63,64, Anthony Bayer61, Debby W. Tsuang65, Lei Yu66, Magda Tsolaki67, Paola Bossu68, Gianfranco Spalletta68, Petra Proitsi46, John Collinge69, Sandro Sorbi70,71, Florentino Sanchez Garcia72, Nick C. Fox73, John Hardy74, Maria Candida Deniz Naranjo72, Paolo Bosco75, Robert Clarke76, Carol Brayne77, Daniela Galimberti78, Elio Scarpini79, Ubaldino Bonuccelli79, Michelangelo Mancuso79, Gabriele Siciliano79, Susanne Moebus80, Patrizia Mecacci81, Maria Del Zompo82, Wolfgang Maier83, Harald Hampel84,85, Alberto Pilotto86, Ana Frank-Garcia87,88,89, Francesco Panza90, Vincenzo Solfrizzi90, Paolo Caffarra91,92, Benedetta Nacmias70,71, William Perry12,13, Manuel Mayhaus93, Lars Lannfelt94, Hakon Hakonarson95, Sabrina Pichler93, Minerva M. Carraquillo96, Martin Ingelsson94, Duane Beekly97, Victoria Alvarez98, Fanggeng Zou99, Otto Valladares5, Steven G. Younkin96, Eliette Coto98, Kara L. Hamilton-Nelson12, Wei Gu99, Cristina Razanou100, Pau Pastor100,101, Ignacio Mateo102, Michael J. Owen1, Kelvin M. Faber103, Palmi V. Jonsson16,104, Onofre Combarros102, Michael C. O’Donovan1, Laura B. Cantwell5, Hilkka Soininen105,106, Deborah Blacker107,108, Simon Mead69, Thomas H. Mosley, Jr.109, David A. Bennett56,110, Tamara B. Harris111, Laura Fratiglioni112,113, Clive Holmes114, Renee F. A. G. de Bruijn115, Peter Passmore39, Thomas J. Montine116, Karolien Bettens49,50, Jerome I. Rotter117, Alexis Brice118,119, Kevin Morgan53, Tatiana M. Foroud103, Walter A. Kuklinski20, Didier Hannequin35, John F. Powell16, Michael A. Nalls121, Karen Ritchie41,122, Kathryn L. Lunetta6, John S. K. Kauwe123, Eric Boerwinkle124,125, Matthias Riemenschneider99, Mercè Boada26,126, Mikko Hiltunen105,106, Eden R. Martin12,13, Reinhold Schmidt127, Dan Rujescu85, Jean-François Dartigues47,128, Richard Mayeux27,28, Christophe Tzourio129, Albert Hofman19,20, Markus M. Nöthen130, Caroline Graff113,131, Bruce M. Psaty11,132, Jonathan L. Haines133,134, Mark Lathrop10,22,135, Margaret A. Pericak-Vance12,13, Lenore J. Launer111, Christine Van Broeckhoven49,50, Lindsay A. Farrer5,9,10,136,137, Cornelia M. van Duijn20,40,138, Alfredo Ramirez139, Sudha Seshadri18,140,9, Gerard D. Schellenberg58,9, Philippe Amouyel23,4,29,141,9, Julie Williams1,9

1 Institute of Psychological Medicine and Clinical Neurosciences, MRC Centre for Neuropsychiatric Genetics & Genomics, Cardiff University, Cardiff, United Kingdom, 2 INSERM U744, Lille, France, Université Lille 2, Lille, France, Institut Pasteur de Lille, Lille, France, 3 Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, United States of America, 4 Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, United States of America, 5 Department of Epidemiology and Neurology, Erasmus MC University Medical Center, Rotterdam, the Netherlands, 6 Department of Biostatistics and Epidemiology and Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine, University of Pennsylvania,
109 Department of Medicine (Geriatrics), University of Mississippi Medical Center, Jackson, Mississippi, United States of America, 110 Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, Illinois, United States of America, 111 Laboratory of Epidemiology, Demography, and Biometry, National Institute of Health, Bethesda, Maryland, United States of America, 112 Aging Research Center, Department Neurobiology, Care Sciences and Society, Karolinska Institutet and Stockholm University, Stockholm, Sweden, 113 Department Geriatric Medicine, Genomics Unit, Karolinska University Hospital Huddinge, Stockholm, Sweden, 114 Division of Clinical Neurosciences, School of Medicine, University of Southampton, Southampton, United Kingdom, 115 Departments of Neurology and Epidemiology, Erasmus MC University Medical Center, Rotterdam, the Netherlands, 116 Department of Pathology, University of Washington, Seattle, Washington, United States of America, 117 Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, California, United States of America, 118 INSERM UMR_S975-CNRS UMR 7225, Université Pierre et Marie Curie, Centre de recherche de l’Institut du Cerveau et de la Moelle épi-nérale-CIRCIM, Hôpital de la Salpêtrière, Paris France, 119 AP-HP, Hôpital de la Pitié-Salpêtrière, Paris, France, 120 Department of Epidemiology, University of Washington, Seattle, Washington, United States of America, 121 Laboratory of Neurogenetics, Intramural Research Program, National Institute on Aging, Bethesda, Maryland, United States of America, 122 Imperial College, London, United Kingdom, 123 Department of Biology, Brigham Young University, Provo, Utah, United States of America, 124 Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, United States of America, 125 Human Genetics Center and Div. of Epidemiology, University of Texas Health Sciences Center at Houston, Houston, Texas, United States of America, 126 Hospital Universitari Vall d’Hebron - Institut de Recerca, Universitat Autònoma de Barcelona. (VHIR-UAB), Barcelona, Spain, 127 Department of Neurology, Medical University Graz, Graz, Austria, 128 Centre de Mémoire de Ressources et de Recherche de Bordeaux, CHU de Bordeaux, Bordeaux, France, 129 Inserm U708, Victor Segalen University, Bordeaux, France, 130 Institute of Human Genetics, Department of Genomics, Life and Brain Center, University of Bonn, and German Center for Neurodegenerative Diseases (DZNE, Bonn), Bonn, Germany, 131 Karolinska Institutet, Department of Neurobiology, Care Sciences and Society, KIADRC, Stockholm, Sweden, 132 Group Health Research Institute, Group Health Cooperative, Seattle, Washington, United States of America, 133 Vanderbilt Center for Human Genetics Research, Vanderbilt University, Nashville, Tennessee, United States of America, 134 Department of Epidemiology & Biostatistics, Case Western Reserve University, Cleveland, Ohio, United States of America, 135 McGill University and Génome Québec Innovation Centre, Montreal, Canada, 136 Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts, United States of America, 137 Department of Neurology, Boston University School of Medicine, Boston, Massachusetts, United States of America, 138 Center for Medical Systems Biology, Leiden, The Netherlands, 139 Department of Psychiatry and Psychotherapy and Institute of Human Genetics, University of Bonn, Bonn, Germany, 140 The Framingham Heart Study, Framingham, Massachusetts, United States of America, 141 Centre Hospitalier Regional Universitaire de Lille, Lille, France

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 m 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 (TP53INP1, p = 1.4 × 10^-6) and 14 (IGHV1-67 p = 7.9 × 10^-6) 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.

Citation: Escott-Price V, Bellenguez C, Wang L-S, Choi S-H, Harold D, et al. (2014) Gene-Wide Analysis Detects Two New Susceptibility Genes for Alzheimer’s Disease. PLoS ONE 9(6): e94661. doi:10.1371/journal.pone.0094661

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 was 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 (fondation pour la recherche sur le cerveau) and Rotary. This work has been developed and supported by the LABEX (laboratory of excellence program investment for the future) DISTALZ grant (Development of Innovative Strategies for a Transdisciplinary approach to ALZheimer’s disease) and by the LABEX GENEMED 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 Génotypage. 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-Synthelabo. 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 Sanitaire des Produits de Santé, the Aquitaine and Bourgogne Regional Councils, Agence Nationale de la Recherche, ANR supported the COGNUT and COVADIS projects. Fondation de France and the joint French Ministry of Research/INSERM « Cohortes et collections de données biologiques » programme. Lille Génopôle received an unconditional grant from Eisai. The Three-cities biologial 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 Caulenbergh, Karolien Bettens and Kristel Sleegers. 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 Carmicone Foundation. The Florence site was supported by grant RF-2010-2319722, grant from the Cassa di Risparmio di Pistoia e Pesca (Grant 2012) and the Cassa di Risparmio di Firenze (Grant 2012). The Milan site was supported by a grant from the...
New Susceptibility Genes for Alzheimer’s Disease
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^{-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-DRB5/HLA-DRB1, CD2AP, EPHA1, PTK2B, CLU, MSI4A6, PICALM, SORL1, SLC24A4, ABCA7, APOE), three new genes in the vicinity of lately reported single SNP hits[9] (ZNF3, NDUFS3, MTPC2) 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.

| GENES | Stage 1 significance level | Significant at Stage 1 | Replicated (ps≤0.05) at Stage 2 | Significant at Stage 1 | Replicated (ps≤0.05) at Stage 2 | Over-representation p-value |
|-------|-----------------------------|-------------------------|---------------------------------|-------------------------|---------------------------------|----------------------------|
|       | p≤10^{-4}                   | 27                      | 9 (33%)                        | 9                       | 3 (33%)                        | 0.109                      |
|       | p≤10^{-5}                   | 74                      | 17 (23%)                       | 36                      | 8 (22%)                        | 0.125                      |
|       | p≤0.01                      | 229                     | 49 (21%)                       | 102                     | 26 (25%)                       | 0.0001                     |
|       | p≤0.05                      | 390                     | 77 (20%)                       | 171                     | 33 (19%)                       | 0.007                      |
| Total | (p≤1)                       | 887                     | 124 (14%)                      | 444                     | 60 (13.5%)                     | 4.6×10^{-12}               |

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.

| Numbers of loci (genes) | ps≤10^{-4} | ps≤10^{-5} | ps≤10^{-6} |
|--------------------------|------------|------------|------------|
| Observed                 | 9 (27)     | 4 (8)      | 2 (2)      |
| Expected                 | 2.5        | 0.25       | 0.025      |
| p-value                  | 0.001      | 0.0013     | 0.003      |

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^{-3}, 10^{-2} 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-wise 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 (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} (Table 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,

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/FLuN/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 rs4582532 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^{-3} and 1.1×10^{-7} respectively in BRAINEAC brain expression data. The r² 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 DNase/ChIP-seq/Histone marks and is located upstream (~1.5 kb) of the TP53INP1
transcription start site. In combination these results suggest a possible epigenetic mechanism whereby the associated variants in the region influence \textit{TP53INP1} expression in several brain regions. These expression data provide further evidence supporting the functional relevance of \textit{TP53INP1} to AD susceptibility. The \textit{IGHV1-67} gene was not found in those databases.

In addition we detected two genome-wide significant loci 1) \textit{ZNF3} (chr7: 99,661,653–99,679,371; p = 8.6×10^{-7}) and 2) two closely located genes on chromosome 11 \textit{MTCH2} (47,638,358–47,674,206, combined p = 2.5×10^{-5}) and \textit{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 plots of these additional genes.

\textit{ZNF3} and \textit{NDUFS3}, \textit{MTCH2} genes on chromosomes 7 and 11, respectively, lie close to rs1476679 (chr7:100,004,446; \textit{ZCWPW1}) and rs1083872 (chr11:47,557,871; \textit{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 \textit{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 \textit{ZNF3} gene does not show significant association, however \textit{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 (\textit{KBTD4, 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 (\textit{SPI1, SLC39A13, LOC100287086, PTPMT1, KBTD4, NDUFS3, LOC100287127, FAM180B}) and on chr7 with 6 genes (\textit{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.

| Gene Name | Chr | Position | Combined best SNP p-value | Stage 1 genotype-wide p-value | Stage 2 genotype-wide p-value | Combined gene-wide p-value | N of SNPs per gene | Combined p-value | Biological function |
|-----------|-----|----------|---------------------------|------------------------------|-------------------------------|---------------------------|------------------|----------------|-------------------|
| TP53INP1  | 8   | 95,938,200–95,961,615 | 1.7×10^{-4} | 6.5×10^{-4} | 4.5×10^{-6} | 1.7×10^{-7} | 45 | 1.4×10^{-6} | Regulation of apoptosis, cell cycle arrest |
| IGHV1-67  | 14  | 107,136,620–107,137,059 | 2.3×10^{-4} | 3.2×10^{-5} | 1.8×10^{-6} | 1.8×10^{-6} | 27 | 4.8×10^{-6} | Immunoglobulin heavy chain region: adaptive immunity |
| ZNF3      | 7   | 99,661,653–99,679,371 | 2.7×10^{-5} | 1.3×10^{-6} | 1.1×10^{-6} | 1.1×10^{-6} | 5 | 2.9×10^{-6} | Transcription factor, leucocyte activation |
| NDUFS3    | 11  | 47,600,632–47,606,114 | 2.2×10^{-6} | 2.2×10^{-7} | 2.2×10^{-6} | 2.2×10^{-6} | 34 | 7.2×10^{-6} | Mitochondrial inner membrane |
| MTCH2     | 11  | 42,653,808–42,664,206 | 2.5×10^{-10} | 8.7×10^{-10} | 8.7×10^{-10} | 8.7×10^{-10} | 7 | 2.5×10^{-10} | Mitochondrial electron transport, NADH to ubiquinone |

We did not observe genome-wide significance for CD33 gene. This gene was genome-wide significant in Stage 1 (p = 1.9×10^{-7}), 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 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. 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).  

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).

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).
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 0.5 Mb 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.

Figure 2. Linkage disequilibrium structure of IGHV1-67 gene ±5 kb. The SNPs which are significant at 10^{-4} level are circled in red. doi:10.1371/journal.pone.0094661.g002
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 adopted for set-based analysis of genetic data. 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 collects p-values of several tests and combines them by the method described in [43], taking LD into account, as were previously reported [9]. 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 $2.5 \times 10^{-8}$ to account for the number of tested genes.

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 $\sqrt{\alpha N}$ (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 (see [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 independent SNPs in intervals $z_1 < p \leq z_2$, $\alpha \approx 0$, $10^{-5}$, $10^{-6}$, ..., 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 $\alpha$ 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 $r^2 \geq 0.3$ or association p-value $\leq 10^{-3}$ 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 ±0.5Mb (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.

Acknowledgments

This work was made possible by the generous participation of the control subjects, the patients and the healthy subjects. 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 C. Bellengues LSW SHC D. Harold P. Holmans A. Richards AJ LAF S. Madi DZ BNV YK.

References

1. Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, et al. (2006) Role of genes and environments for explaining Alzheimer disease. Archives of General Psychiatry 63: 169–174.
2. Bettens K, Skegers K, Van Broeckhoven C (2013) Genetic insights in Alzheimer's disease. Lancet neurology 12: 92–104.
3. Corver EH, Saunders AM, Strottmaier WJ, Schmechel DE, Gaskell PC, et al. (1995) Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science 261: 921–925.
4. Harold D, Abraham R, Hollingsworth P, Sims R, Gerrish A, et al. (2009) Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nature genetics 41: 1089–1093.
5. Hollingsworth P, Harold D, Sims R, Gerrish A, Lambert JC, et al. (2011) Common variants at ARCA7, MS4A6A/MS4A6E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nature Genetics 43: 429–435.
6. Lambert JC, Heath S, Even G, Campion D, Skegers K, et al. (2009) Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. Nature Genetics 41: 1094–1106.
7. Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, et al. (2011) Common variants at MS4A4E, MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nature Genetics 43: 436–441.
8. Seshadri S, Fitzpatrick AL, Ichrak MA, DeStefano AL, Gudnason V, et al. (2010) Genome-wide analysis of genetic loci associated with Alzheimer disease. JAMA: the journal of the American Medical Association 303: 1832–1840.
9. Lambert JC, Ibrahim-Verhaas CH, Harold D, Naj AC, Sims R, et al. (2013) Meta-analysis of 74,496 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet 45: 1432–1438.
10. Guerreiro RJ, Hardy J (2011) The genetics of late-onset Alzheimer's disease. Nature review genetics 12: 76–88.
11. Ioannidis JP (2007) Non-replication and inconsistency in the genome-wide association studies. Nature review genetics 8: 933–937.
12. Guerreiro RJ, Hardy J (2011) The genetics of late-onset Alzheimer's disease. Nature review genetics 12: 76–88.
13. Moskvina V, O'Donovan MC (2007) Detailed analysis of the relative power of linkage and association mapping for complex traits. Human heredity 64: 1–12.
14. Terwilliger JD, Hiekkalinna T (2006) An utter refutation of the "fundamental problem of association". Hum Mol Genet 15: 2822–2825.
15. Bettens K, Brouwers N, Engelborghs S, Lambert JC, Rogaeva E, et al. (2012) Common variations at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Human genetics 131: 379–387.
16. Risch N (1990) Linkage strategies for genetically complex traits. I. Multilocus genetic models. American journal of human genetics 46: 222–228.
17. Lewis CM, Levinson DF, Wine LH, DeLisi LE, Straub RE, et al. (2005) Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. American journal of human genetics 73: 34–46.
18. Segalov R, Do Wing Leung SD, Levinson DF, Lewis CM, Gill M, et al. (2005) Genome scan meta-analysis of schizophrenia and bipolar disorder, part III: Bipolar disorder. American journal of human genetics 73: 49–62.
19. Lambert JC, et al. (2013) Extended meta-analysis of 74,538 individuals identifies 11 new susceptibility loci for Alzheimer's disease.
20. Johnson AD, Handaker RE, Polli SL, Nizzari MM, O'Donnell CJ, et al. (2008) SNAP: a web-based tool for identification and annotation of proxy SNPs using haplotype. Bioinformatics 24: 2932–2939.
21. Webster JA, Gibbs JR, Clarke J, Ray M, Zhang WX, et al. (2009) Genetic Control of Human Brain Transcript Expression in Alzheimer Disease. American Journal of Human Genetics 84: 443–458.
22. Trabzuni D, Ryten M, Walker R, Smith C, Irman S, et al. (2011) Quality control parameters on a large dataset of regionally dissected human control brains for whole genome expression studies. J Neurochem 119: 275–282.
23. Seux M, Peugeot S, Montero MP, Siret C, Rigot V, et al. (2011) TP53INP1 decreases pancreatic cancer cell migration by regulating SPARC expression. Oncogene 30: 3499–3506.
24. Seullier M, Peugeot S, Gayet O, Gauthier C, NGuessen P, et al. (2012) TP53INP1, a tumor suppressor, interacts with LC3 and ATG8-family proteins through the LC3-interacting region (LIR) and promotes autophagy-dependent cell death. Cell death and differentiation 19: 1295–1305.
25. Giussano S, Baylot V, Andreiu C, Fazi L, Gleave M, et al. (2012) TP53INP1 as a new therapeutic target in castration-resistant prostate cancer. Prostate 72: 1286–1294.
26. Diouf YA, Beiser A, Au R, Kroger BE, Splandy GL, et al. (2012) Inverse association between cancer and Alzheimer's disease: results from the Framingham Heart Study. British Medical Journal 344.
27. Roe CM, Fitzpatrick AL, Xiong C, Sirih W, Kuller L, et al. (2010) Cancer linked to Alzheimer disease but not vascular dementia. Neurology 74: 106–112.
28. Watson CT, Reden F (2012) The immunoglobulin heavy chain locus: genetic variation, missing data, and implications for human disease. Genes and immunity 13: 363–373.
29. Chen Y, Yang LN, Cheng L, Ts S, Guo SJ, et al. (2013) BAG3 Interactome Analysis Reveals a New Role in Modulating Prostate Cancer Activity. Molecular & cellular proteomics. MCP.
30. Wannat HJ, Schmidt D, Manke T, Piccini I, Sultan M, et al. (2011) The BTB and CNC homology 1 (BACH1) target genes are involved in the oxidative stress response and in control of the cell cycle. The Journal of biological chemistry 286: 23521–23532.
31. Palmieri F (2013) The mitochondrial transporter family SLC25: identification, properties and physiopathology. Molecular aspects of medicine 34: 465–493.
32. Katz C, Zulman-Amir Y, Mostizky Y, Kollet N, Gross A, et al. (2012) Molecular basis of the interaction between proapoptotic truncated BID (dBid) protein and mitochondrial carrier homologue 2 (MTC2H) protein; key players in mitochondrial death pathway. The Journal of biological chemistry 287: 15016–15023.
33. Bernhardt F, Landgraf K, Kloting X, Berthold A, Buttner P, et al. (2013) Functional relevance of genes implicated by obesity genome-wide association studies for health as adipocyte biology. Diabetesologia 56: 311–322.
34. Fall T, Arulogun J, Berne C, Ingelsson E (2012) The role of obesity-related genetic loci in insulin sensitivity. Diabetic medicine: a journal of the British Diabetic Association 29: e62–66.
35. Haupt A, Thamer C, Henri M, Machicao F, Machmann J, et al. (2010) Novel obesity risk loci do not determine distribution of body fat depots: a whole-body MRI/MRS study. Obesity 18: 1212–1217.
36. Benit P, Slama A, Cartault F, Giurega I, Chretien D, et al. (2004) Mutant NDUFS3 subunit of mitochondrial complex I causes Leigh syndrome. Journal of medical genetics 41: 14–17.
37. Dahl HH (1998) Getting to the nucleus of mitochondrial disorders: identification of respiratory chain-enzyme genes causing Leigh syndrome. American journal of human genetics 63: 1594–1597.
38. Hossie BN, Donnelly P, Marinihi J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS genetics 5: e1000529.
39. Li Y, Wilner CJ, Ding J, Scheret P, Abecasis GR (2010) MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genetic epidemiology 34: 816–834.
40. Brown MB (1975) A method for combining non-independent, one-sided tests of significance. Biometrics 31: 978–982.
41. Moskvina V, O’Dushlaine C, Purcell S, Craddock N, Holmans P, et al. (2011) Evaluation of an approximation method for assessment of overall significance of multiple-dependent tests in a genomewide association study. Genetic epidemiology 35: 861–866.

42. Kiezun A, Garimella K, Do R, Stitziel NO, Neale BM, et al. (2012) Exome sequencing and the genetic basis of complex traits. Nature genetics 44: 623–630.

43. Moskvina V, Schmidt KM (2008) On multiple-testing correction in genomewide association studies. Genetic epidemiology 32: 567–573.