Family-based Association Analyses of Imputed Genotypes Reveal Genome-Wide Significant Association of Alzheimer's disease with *OSBPL6, PTPRG* and *PDCL3*

Christine Herold, PhD[1,3,§], **Basavaraj V. Hooli, PhD**[2,§], **Kristina Mullin, MS**[2], **Tian Liu, PhD**[4], **Johannes T Roehr, MS**[4], **Manuel Mattheisen, PhD**[5], **Antonio R. Parrado, PhD**[6], **Lars Bertram, MD**[4,7,8], **Christoph Lange, PhD**[3,*], and **Rudolph E. Tanzi, PhD**[2,*]

[1] German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany

[2] Genetics and Aging Research Unit, Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts, 02114 USA

[*]**To whom correspondence should be addressed:** Christoph Lange, Ph.D., Professor of Biostatistics, Department of Biostatistics, Harvard T.H. Chan School of Public Health, 665 Huntington Avenue, Boston, Massachusetts 02115, Tel: +1 (617) 432-4925, clange@hsph.harvard.edu, and, Rudolph E. Tanzi, Ph.D., Director, Genetics and Aging Research Unit, Department of Neurology, Massachusetts General Hospital, Charlestown, Massachusetts 02129, USA, Joesph P. and Rose F. Kennedy Professor of Neurology, Harvard Medical School, Boston, Massachusetts 02114, USA, Tel: +1 (617) 726-6845, tanzi@helix.mgh.harvard.edu.
[§]Contributed equally to the work
C. Herold, PhD, Christine.Herold@dzne.de
B. Hooli, PhD, BHooli@helix.MGH.Harvard.edu
K. Mullin1, BS, KMullin1@Partners.org
T. Liu, PhD, tianliu@mpib-berlin.mpg.de
J. T. Roehr, MS, johannes.roehr@fu-berlin.de
M. Mattheissen, PhD, mm@biomed.au.dk
A. R. Parrado, PhD, aparrado@its.jnj.com
L. Bertram, MD, lars.bertram@uni-luebeck.de
C. Lange, PhD, clange@hsph.harvard.edu
R. E. Tanzi, PhD, tanzi@helix.MGH.Harvard.edu

[3] Department of Biostatistics, Harvard T.H Chan School of Public Health, Boston, 02115 Massachusetts, USA

[4] Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin, Germany

[5] Department of Biomedicine, Bartholins Allé 6, 8000 Aarhus Denmark

[6] Computational Sciences, Janssen Research & Development, Pharmaceutical Company of Johnson & Johnson, Spring House, Pennsylvania, 19477

[7] School of Public Health, Faculty of Medicine, The Imperial College of Science, Technology, and Medicine, London, W6 8RP, UK

[8] Platform for Genome Analytics, Institutes of Neurogenetics & Integrative and Experimental Genomics, University of Lübeck, 23552 Lübeck, Germany

Abstract

The genetic basis of Alzheimer's disease (AD) is complex and heterogeneous. Over 200 highly penetrant pathogenic variants in the genes *APP*, *PSEN1* and *PSEN2* cause a subset of early-onset familial Alzheimer's disease (EOFAD). On the other hand, susceptibility to late-onset forms of AD (LOAD) is indisputably associated to the ε4 allele in the gene APOE, and more recently to variants in more than two-dozen additional genes identified in the large-scale genome-wide association studies (GWAS) and meta-analyses reports. Taken together however, although the heritability in AD is estimated to be as high as 80%, a large proportion of the underlying genetic factors still remain to be elucidated. In this study we performed a systematic family-based genome-wide association and meta-analysis on close to 15 million imputed variants from three large collections of AD families (~3,500 subjects from 1,070 families). Using a multivariate phenotype combining affection status and onset age, meta-analysis of the association results revealed three single nucleotide polymorphisms (SNPs) that achieved genome-wide significance for association with AD risk: rs7609954 in the gene *PTPRG* (P-value = $3.98 \cdot 10^{-08}$), rs1347297 in the gene *OSBPL6* (P-value = $4.53 \cdot 10^{-08}$), and rs1513625 near *PDCL3* (P-value = $4.28 \cdot 10^{-08}$). In addition, rs72953347 in *OSBPL6* (P-value = $6.36 \cdot 10^{-07}$) and two SNPs in the gene *CDKAL1* showed marginally significant association with LOAD (rs10456232, P-value: $4.76 \cdot 10^{-07}$; rs62400067, P-value: $3.54 \cdot 10^{-07}$). In summary, family-based GWAS meta-analysis of imputed SNPs revealed novel genomic variants in (or near) *PTPRG, OSBPL6,* and *PDCL3* that influence risk for AD with genome-wide significance.

Introduction

Alzheimer disease (AD) is the most common form of senile dementia. AD is the sixth leading cause of death in the US with the healthcare costs surpassing $200 billion in the year 2013, and anticipated to increase exponentially with aging population (1-3). Clinical symptoms are broadly characterized by a slowly progressing loss of memory and cognitive functions, dementia, and ultimately death. Neuropathologically, deposition of β-amyloid (Aβ) peptide in the form of senile 'plaques' and oligomers (crucial to initiating AD pathogenesis), and accumulation of hyperphosphorlylated tau protein in the form of

intracellular neurofibrillary 'tangles' (NFTs), along with inflammation and neurodegeneration, are the hallmark characteristics in post-mortem AD brains.

Following advancing age, family history is the second strongest risk factor in AD. Traditionally, AD is classified into two dichotomous forms based on the age of onset and the associated genetic factors. The relatively rare early-onset familial form of AD (EOFAD, onset age <65 years of age, <5% of diagnosed AD cases) is caused by highly penetrant, autosomal-dominant mutations in the three genes *APP*, *PSEN1* and *PSEN2*. On the other hand, the more predominantly diagnosed late-onset form of AD (LOAD, onset >65 years of age, >95% of AD cases) shows less obvious familial aggregation. *APOE-ε4* still remains the strongest risk factor in LOAD, where the *ε4* allele confers between 3.7- and 14-fold increases in risk, in heterozygotes and homozygotes, respectively. Importantly, the identification of the four above genes were key to understanding the underlying molecular mechanism leading to AD - driven by Aβ oligomers, leading to the tangles formation, loss of neurons, neuroinflammation and dementia ("amyloidosis", (4)).

Since the first wave of genome-wide association studies beginning in 2007, more than a dozen GWAS and meta-analysis have been published to date, revealing several novel genetic loci in LOAD. Some of the genes that either encompass AD GWA SNPs or present in close proximity include, Triggering Receptor Expressed On Myeloid Cells 2 (TREM2), Bridging Integrator 1 (BIN1), Sialic Acid–binding Immunoglobulin (Ig)–like Lectin (CD33), Clusterin (CLU), ATP-binding Cassette Transporter (ABCA7), Complement Receptor 1 (CR1), Phosphatidylinositol Binding Clathrin Assembly Protein (PICALM), to name a few. Overall, although twin and population studies estimate heritability in AD as high as 80% (5), all the above genetic factors taken together explain less than 50% of heritability in AD. Identification of the remaining genetic factors in AD will not only explain the missing heritability but will be vital to fully understanding the disease pathogenesis and developing treatment strategies.

In this study we performed a systematic meta-analysis of the family-based association test results using imputed genotypes (limited to MAF >0.05) generated in the subjects from three large Alzheimer's family collections; National Institute of Mental Health (NIMH), National Institute of Aging Genetics Initiative for Late Onset Alzheimer's Disease (NIA-LOAD) and National Repository of Research on Alzheimer's Disease (NCRAD). A total of 3,500 subjects from 1,070 families were assessed in this study. To maximize statistical power to detect disease associated variants, we implemented a novel approach combining AD affection status and age of onset jointly using the multivariate extension of the FBAT-approach (6, 7). We performed family-within component analyses and family-within and family-between component analyses (FBAT-GEE method), and finally the results from the three family-based samples were combined via meta-analysis. Meta-analysis results were computed separately for the two statistical approaches, (1) based on the family-within component analyses, and, (2) for the family-within component and family-between component analyses.

Material and Methods

Study families

We utilized three large family-based AD samples in the association tests and the meta-analyses: National Institute of Mental Health (NIMH), National Repository of Research on Alzheimer's Disease (NCRAD) and National Institute on Aging (NIA). The National Institute of Mental Health (NIMH) Alzheimer Disease Genetics Initiative Study (8), was originally ascertained for the study of genetic risk factors in AD in a family-based setting. The complete NIMH sample contains a total of 1,536 subjects from 457 families. For the purposes for our study 1,376 participants (941 affected and 404 unaffected) from 410 families were included. The complete National Institute of Aging Genetics Initiative for Late Onset Alzheimer's Disease Family sample and the National Cell Repository for Alzheimer's Disease (NIA-LOAD) sample contains 4,006 subjects from 1,653 families. Here, we included 1,040 subjects (748 affected and 282 unaffected) from 329 multiplex families. The families originally ascertained as part of the National Repository of Research on Alzheimer's Disease (NCRAD) subset of families includes 1,108 subjects from 331 families, with 799 affected and 293 unaffected siblings. Affection status was based on clinical dementia diagnosis documentation according to NINCDA-ADRDA criteria. Patients diagnosed with mild cognitive impairment, unknown dementia, or unconfirmed family reports of dementia were excluded from our analysis. The initial age of detection of cognitive impairment in the patients was used for the age of onset phenotype. The basis for each cohort was the presence of at least two affected individuals within a family, typically siblings. All subjects are of self-reported European ancestry.

Genotyping and Imputation

DNA samples from study subjects belonging to the NIMH and NCRAD AD families were processed on Affymetrix Human Genome Wide SNP 6.0 arrays. Samples that failed to pass Affymetrix quality control (cQC), showed conflicting gender or carried large chromosomal abnormalities and were excluded from the study, as described in detail elsewhere (9, 10). The Human610-Quad array genotypes of the NIA-LOAD study samples were obtained from dbGAP (Genetic Consortium for Late Onset Alzheimer's Disease 6K, ID: phs000160.v1.p1). The quality of the array genotype data is available in the original report (11).

Before performing an association analysis of our three family cohorts, we applied standard GWAS quality control procedures for all three samples (NCRAD, NIA-LOAD and NIMH) as described here (12). SNPs and individuals were filtered for a call rate of at least 99%. In addition, SNPs with a minor allele frequency (MAF) of <5% were excluded. Population stratification within and between the samples was also checked by performing a multi-dimensional scaling (MDS, identification of population outliers) implemented in PLINK (13). Duplicated DNA samples were identified by consideration of Genome-wide genotype identity-by-state (IBS) status (IBS > 1.98). From each pair the individual with the lower genotyping rate was removed. In a second step, we used IMPUTE2 (14) to impute the QCed datasets NCRAD and NIMH into the May 2013 release of the 1,000 Genomes Project and NIA into the Sep 2013 release of the 1,000 Genomes Project (15). SNPs with an info score

smaller than 0.4 were removed. Next, we "called" individual genotypes in the family studies by assigning the genotype with the highest imputation probability.

After imputing, we had a total of 43,207,737 markers from the three study cohorts, NIMH (n=14,129,045 markers), NCRAD (n=13,971,550), and NIA-LOAD (n=15,107,142), for association analysis. 273 families from NCRAD with 470 affected and 279 unaffected siblings, 401 families from NIMH with 905 affected and 318 unaffected siblings and 618 families from NIA-LOAD with 1464 affected and 1096 unaffected siblings were analyzed.

Family-based Association Analyses

SNPs showing Mendelian errors were excluded from all the following analyses. In the presence of markers showing Mendelian errors in a pedigree, all genotypes for those markers with the Mendelian inconsistencies were set to zero (missing) in those pedigrees. In other words, markers showing Mendelian inconsistencies in a family were set to 'missing' in that specific pedigree only prior to performing association tests. To maximize statistical power and avoid multiple comparison problems, we used for our analyses a multivariate extension of the FBAT-approach (6), the FBAT-GEE (7) statistic and Van-Steen-like testing strategy (16, 17). FBAT-GEE, as the original FBAT, does not require any distributional assumptions for the phenotypes and it tests different trait types simultaneously. Assuming that m traits for each offspring that we want to test simultaneously by the FBAT approach, we denote the vector containing all m observations for each offspring by $Y_{ij} = (Y_{ij1}, \ldots, Y_{ijm})$, where Y_{ijk} is the kth phenotype for the jth offspring in the family. The multivariate FBAT-GEE statistic is constructed by replacing the univariate coding variable T_{ij} in C by the coding vector defined by (18)

$$T_{ij} \quad = \quad Y_{ij} - \widehat{Y_{\imath\jmath}} \quad = \quad \begin{pmatrix} Y_{ij1} \\ \ldots \\ Y_{ijk} \\ \ldots \\ Y_{ijm} \end{pmatrix} - \begin{pmatrix} \widehat{Y_{\imath\jmath1}} \\ \ldots \\ \widehat{Y_{\imath\jmath k}} \\ \ldots \\ \widehat{Y_{\imath\jmath m}} \end{pmatrix}$$

where, the $\hat{Y}_{ijk}$'s are the predicted trait values based on the regression model for covariates. Replacing univariate coding variable T_{ij} in FBAT statistic by the vector $\boldsymbol{T_{ij}}$ results in the FBAT-GEE statistic

$$T_{FBAT-GEE} \quad = \quad C^T V^{-1} C$$

Under the null hypothesis, the FBAT-GEE statistic has a X^2 distribution with m degrees of freedom.

In our case the FBAT-GEE statistic contains affection status and time to onset as phenotypes, coded as Wilcoxon statistic. A more detailed description can be found in the original article (7).

The analyses were divided in two steps: the family-within components and the family-within and family-between components approach (17, 19). The family-within component is a genetic association tests that is based on Mendelian transmissions, where as the family-between component is a population-based association analysis in which the genotypes are replaced by the expected genotyped conditional on the sufficient statistic, i.e. the conditional mean model (19). The advantage of the within-family component is that it is robust against population stratification. However, the between-family component remains sensitive against population stratification and requires further adjustment (20). Therefore, several statistical techniques proposed to use both approaches (16). After performing within-family analysis (FBAT-GEE) and between-family analysis (conditional mean model), the results from the two family analyses are combined via meta-analysis, in which the FBAT P-value is used for the within family-analysis and a rank-based P-value for the between-family-analysis (16). The rank-based P-values for the conditional mean model ensure maximal efficiency and, at the same time, maintain the robustness against population stratification of the overall approach.

In the meta-analyses that was performed with METAL (21), the P-values across our studies NCRAD (n=7,432,385 variants), NIMH (n=7,346,118) and NIA-LOAD (n=7,556,673) were summarized and also the sample size and direction of effect were taking into account. First, a meta-analysis for the family-within component analysis (1) was performed for our three samples, second for the family-within component and family-between component analysis (2).

To check if our top SNPs are in linkage disequilibrium (LD) with a gene we used the software package epigwas (22). For each SNP, the SNPs in a 1M window (upstream 500K, and downstream 500K) are included in the calculation. The tool calculates LD using 1000 genome data for the EUR population. Only SNPs with r^2 greater than 0.5 are shown in the results section.

Results

The results of our two meta-analyses (1): FBAT-GEE results for the family-within component analysis and (2): FBAT-GEE for the family-within and family-between component analysis) are shown in Table 1. We present SNPs exhibiting family-based association with AD with P-value $<10^{-06}$, a minor allele frequency (MAF) >0.05, and with the same effect direction in each family sample. The complete list of SNPs with P-values $>10^{-05}$ are listed in the supplementary tables; Supplementary Table 3a and Supplementary Table 3b.

The APOE region (rs56131196) shows highly significant results (P-values of $3.09 \cdot 10^{-24}$, and $3.96 \cdot 10^{-24}$ for approaches (1) and (2), respectively; 187 SNPs <0.05 for approach (1) and 283 SNPs for approach (2). The three EOFAD genes (*APP, PSEN1,* and *PSEN2*) harbor several SNPs with nominally significant association with AD. In the FBAT-GEE results for the family-within component analysis, the most strongly associated SNP in the *APP* region (among 47 SNPs with P-value <0.05) is rs190685835 (P-value = $3.74 \cdot 10^{-03}$). For the *PSEN1* region, there are 61 SNPs with P-value <0.05 (rs3025774: P-value = 0.02775); for the

PSEN2 region, 15 SNPs <0.05, (rs182226938 has a P-value = $2.23 \cdot 10^{-03}$). Looking at the FBAT-GEE results for the family-within component and the family-between components the number of nominal significant SNPs is increased. The *APP* region has 452 SNPs <0.05 (rs141145244 P-value = $1.9 \cdot 10^{-04}$; the *PSEN1* region has 311 SNPs <0.05 (rs214277 P-value = $1.07 \cdot 10^{-04}$) and the *PSEN2* region with 87 SNPs <0.05 (rs149734051 P-value = $1.06 \cdot 10^{-03}$). This is in concordance with previous GWAS in AD, where variants in the three early-onset familial AD genes failed to show consistent genome-wide significant association with AD.

The results of our two meta-analysis approaches are shown in Table 1 (Panel A & B). We found 32 novel variants showing genome-wide significant association with AD and fulfilling the criteria described above, that is, P-value <10^{-06}, MAF >5% and same effect direction across all samples tested. Detailed information can be found in Table Supp-1. 15 variants were either in a gene (Table 1, Panel A) or in LD with SNPs in a gene, while 18 other SNPs (Table 1, Panel B) were in proximity of a known gene. Three variants reached genome-wide significance in at least one of the four meta-analysis: rs7609954 (*PTPRG*): P-value = $3.98 \cdot 10^{-08}$; rs1513625 (*PDCL3*): P-value = $4.28 \cdot 10^{-08}$; rs1347297 (*OSBPL6*): P-value = $4.53 \cdot 10^{-08}$. A second SNP, rs72953347 in *OSBPL6* also showed marginally significant evidence for association with AD using the other meta-analysis approach (approach 2): P = $6.36 \cdot 10^{-07}$). In addition, two SNPs in the gene, *CDKAL1,* showed marginally significant evidence for association with AD in the two different testing meta-analysis approaches (rs62400067: P-value = $3.54 \cdot 10^{-07}$; rs10456232: P-value = $9.60 \cdot 10^{-07}$).

We next tested the 32 SNPs from Table 1 (Both, Panel A and Panel B) showing GW-significant association with AD using family-based methods in the IGAP case-control dataset (Supplementary Table 4). SNPs showing association with AD in the family-based studies do not consistently replicate in case-control data, and vice-versa. As seen in previous reports (reviewed elsewhere (23) none of our top 32 SNPs from Table 1 showed genome-wide significance (with the exception of APOE SNPs) in the IGAP case-control GWAS dataset (24).

Discussion

We carried out a family-based genome-wide association and meta-analysis on roughly 15 million imputed variants using three large AD family samples (~3,500 subjects from 1,070 families). We employed a multivariate phenotype combining affection status and onset age and then performed meta-analysis of the association results. Three SNPs: one in *PTPRG* (rs7609954), one in *OSBPL6* (rs1347297) and another near *PDCL3* revealed genome-wide association with AD in the meta-analysis. Additionally, another SNP, rs72953347 in OSBPL6 (P-value = $6.36 \cdot 10^{-07}$) and two SNPs (rs10456232, rs62400067) in the gene *CDKAL1* showed marginally significant association with AD.

OSBPL6 encodes a member of the oxysterol-binding protein (OSBP) family, a group of intracellular lipid receptors. This gene adds to a growing number of other cholesterol-related genes implicated in AD genetics, e.g. *APOE* and *ABCA7*. Differential gene expression studies have previously implicated OSBPL6 in Niemann-Pick Type C Disease, Parkinson

disease and AD (25-27). The precise pathogenic mechanism still remains unclear but it is speculated that OSBPL6 may affect cognition decline through cholesterol mediated pathways (28). *PTPRG* encodes a member of the protein tyrosine phosphatase (PTP) family, known to function as signaling molecules that regulate cell growth, differentiation, mitotic cycle, and oncogenic transformation. PDCL3 encodes phosphoducin-like 3, which is believed to modulate heterotrimeric G-proteins via binding to beta and gamma subunits of G-proteins. It has also been proposed to play a role in angiogenesis by serving as a chaperone for the VEGF receptor, KDR/VEGFR2 and regulating its ubiquitination and degradation (29). PDCL3 has also been proposed to modulate caspase activation by interacting with the inhibitor of apoptosis (IAP) (30). CDKAL1 encodes the methylthiotransferase family member, cyclin-dependent kinase 5 (CDK5) Regulatory Subunit Associated protein-Like 1, and has been previously associated with noninsulin-dependent diabetes mellitus. Interestingly, CDK5 has also been implicated in AD tangle pathology (31).

The most significant results of our meta-analysis in family-based GWAS studies differ from those of the published meta-analyses of large-scale case/control studies. In addition, while many of the top hits from IGAP (24) are also significantly associated with AD and age-at-onset of AD in our family-based meta-analysis (Supplementary Table 4), they do not achieve genome significance in our family-based association analyses. This observation can most likely be attributed to the fact that different types of association tests were used across these studies, i.e. population-based association tests vs family-based association tests, which require the presence of both linkage and association. Given that the family-based tests combine the evidence of both, linkage and association, the p-values in both meta-analyses may vary for each SNP and the same p-value ranking cannot be expected. This lack of replication is not uncommon in the GWAS of complex human traits and often attributed to several other factors, including, insufficient statistical power, population stratification, differences in genetic ancestry and age-dependent genetic effects, to name a few (32). While the case-control method is the most common study design due to ease of sample ascertainment, the main concern is the population stratification effects, most notable in the SNPs present in the region involved with natural selection (33). On the other hand, family-based studies are more robust against population admixture and stratification that allows both linkage and association testing (34, 35), but may lack power due to small number of families present in the studies. The analytic approaches used in most studies address these pitfalls of the two study designs, and allowing for these caveats, both types of designs yield useful and complementary information (36). In this study, another important factor is that in order to maximize power, our family-based meta-analysis used a multi-variate phenotype combining AD and age-at-onset of AD, while the meta-analysis of case/control design tested for AD without taking age-at-onset into account (24).

The use of a multi-variate phenotype may also explain why our top meta-analysis association findings do not replicate in IGAP (24), as the meta-analysis of case/control studies does not incorporate the age-at-onset information. However, the most important factor that contributes to the non-replication in IGAP may be the adjustment for population substructure. The case/control studies use principal component approaches which works well to adjust for global genetic stratification, but cannot account for local genetic

stratification. The FBAT-based meta-analysis is robust against any genetic confounding, global and local. Any type of locus-specific stratification therefore could bias a principal-component based association analysis and therefore result in undetected, true genetic association, which could be the case here.

In summary, using close to 15 million imputed variants we performed a systematic family-based genome-wide association and meta-analyses using a multivariate phenotype combining affection status and onset age in three large collections of AD families. The meta-analysis of the association results revealed three SNPs that show genome-wide significance for association with AD risk in the genes *PTPRG* and *OSBPL6*, and near *PDCL3* gene. One of our top genes, *OSBPL6* has previously been implicated in AD in the transcriptomic studies of the post-mortem brains. Further studies will be required to replicate these novel findings and to elucidate the pathophysiologic role of these AD-associated genes and variants in the etiology and pathogenesis of AD.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Canadian study of health and aging: study methods and prevalence of dementia. CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne. 1994; 150(6):899–913. Epub 1994/03/15.

2. Cure Alzheimer's Fund. 2015. Available from: http://curealz.org/alzheimers-disease

3. Alzheimer's Assoc. Facts & Figures. 2015. Available from: http://www.alz.org/facts/overview.asp

4. Glenner GG, Wong CW. Alzheimer's disease and Down's syndrome: sharing of a unique cerebrovascular amyloid fibril protein. Biochemical and biophysical research communications. 1984; 122(3):1131–5. Epub 1984/08/16. [PubMed: 6236805]

5. Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S, et al. Role of genes and environments for explaining Alzheimer disease. Archives of general psychiatry. 2006; 63(2):168–74. Epub 2006/02/08. [PubMed: 16461860]

6. Laird NM, Horvath S, Xu X. Implementing a unified approach to family-based tests of association. Genetic epidemiology. 2000; 19(Suppl 1):S36–42. Epub 2000/10/31. [PubMed: 11055368]

7. Lange C, Silverman EK, Xu X, Weiss ST, Laird NM. A multivariate family-based association test using generalized estimating equations: FBAT-GEE. Biostatistics. 2003; 4(2):195–206. Epub 2003/08/20. [PubMed: 12925516]

8. Blacker D, Haines JL, Rodes L, Terwedow H, Go RC, Harrell LE, et al. ApoE-4 and age at onset of Alzheimer's disease: the NIMH genetics initiative. Neurology. 1997; 48(1):139–47. Epub 1997/01/01. [PubMed: 9008509]

9. Ku CS, Pawitan Y, Sim X, Ong RT, Seielstad M, Lee EJ, et al. Genomic copy number variations in three Southeast Asian populations. Human mutation. 2010; 31(7):851–7. Epub 2010/05/28. [PubMed: 20506136]

10. de Andrade M, Atkinson EJ, Bamlet WR, Matsumoto ME, Maharjan S, Slager SL, et al. Evaluating the influence of quality control decisions and software algorithms on SNP calling for the affymetrix 6.0 SNP array platform. Human heredity. 2011; 71(4):221–33. Epub 2011/07/08. [PubMed: 21734406]

11. Lee JH, Cheng R, Graff-Radford N, Foroud T, Mayeux R, National Institute on Aging Late-Onset Alzheimer's Disease Family Study G. Analyses of the National Institute on Aging Late-Onset Alzheimer's Disease Family Study: implication of additional loci. Archives of neurology. 2008; 65(11):1518–26. Epub 2008/11/13. [PubMed: 19001172]

12. Bertram L, Lange C, Mullin K, Parkinson M, Hsiao M, Hogan MF, et al. Genome-wide association analysis reveals putative Alzheimer's disease susceptibility loci in addition to APOE. American journal of human genetics. 2008; 83(5):623–32. Epub 2008/11/04. [PubMed: 18976728]

13. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics. 2007; 81(3):559–75. Epub 2007/08/19. [PubMed: 17701901]

14. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS genetics. 2009; 5(6):e1000529. Epub 2009/06/23. [PubMed: 19543373]

15. 1,000 Genome Project Reference data. 2013.

16. Won S, Wilk JB, Mathias RA, O'Donnell CJ, Silverman EK, Barnes K, et al. On the analysis of genome-wide association studies in family-based designs: a universal, robust analysis approach and an application to four genome-wide association studies. PLoS genetics. 2009; 5(11):e1000741. Epub 2009/12/04. [PubMed: 19956679]

17. Van Steen K, McQueen MB, Herbert A, Raby B, Lyon H, Demeo DL, et al. Genomic screening and replication using the same data set in family-based association testing. Nature genetics. 2005; 37(7):683–91. Epub 2005/06/07. [PubMed: 15937480]

18. Laird NM, Lange C. Family-based methods for linkage and association analysis. Advances in genetics. 2008; 60:219–52. Epub 2008/03/25. [PubMed: 18358323]

19. Lange C, DeMeo D, Silverman EK, Weiss ST, Laird NM. PBAT: tools for family-based association studies. American journal of human genetics. 2004; 74(2):367–9. Epub 2004/01/24. [PubMed: 14740322]

20. Magi R, Asimit JL, Day-Williams AG, Zeggini E, Morris AP. Genome-wide association analysis of imputed rare variants: application to seven common complex diseases. Genetic epidemiology. 2012; 36(8):785–96. Epub 2012/09/07. [PubMed: 22951892]

21. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010; 26(17):2190–1. Epub 2010/07/10. [PubMed: 20616382]

22. Trynka G, Sandor C, Han B, Xu H, Stranger BE, Liu XS, et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. Nature genetics. 2013; 45(2):124–30. Epub 2012/12/25. [PubMed: 23263488]

23. Bertram L, Tanzi RE. Genome-wide association studies in Alzheimer's disease. Human molecular genetics. 2009; 18(R2):R137–45. Epub 2009/10/08. [PubMed: 19808789]

24. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nature genetics. 2013; 45(12):1452–8. Epub 2013/10/29. [PubMed: 24162737]

25. Reddy JV, Ganley IG, Pfeffer SR. Clues to neuro-degeneration in Niemann-Pick type C disease from global gene expression profiling. PloS one. 2006; 1:e19. Epub 2006/12/22. [PubMed: 17183645]

26. Yeger-Lotem E, Riva L, Su LJ, Gitler AD, Cashikar AG, King OD, et al. Bridging high-throughput genetic and transcriptional data reveals cellular responses to alpha-synuclein toxicity. Nature genetics. 2009; 41(3):316–23. Epub 2009/02/24. [PubMed: 19234470]

27. Berchtold NC, Coleman PD, Cribbs DH, Rogers J, Gillen DL, Cotman CW. Synaptic genes are extensively downregulated across multiple brain regions in normal human aging and Alzheimer's disease. Neurobiology of aging. 2013; 34(6):1653–61. Epub 2013/01/01. [PubMed: 23273601]

28. Olkkonen VM, Li S. Oxysterol-binding proteins: sterol and phosphoinositide sensors coordinating transport, signaling and metabolism. Progress in lipid research. 2013; 52(4):529–38. Epub 2013/07/09. [PubMed: 23830809]

29. Srinivasan S, Meyer RD, Lugo R, Rahimi N. Identification of PDCL3 as a novel chaperone protein involved in the generation of functional VEGF receptor 2. The Journal of biological chemistry. 2013; 288(32):23171–81. Epub 2013/06/25. [PubMed: 23792958]

30. Wilkinson JC, Richter BW, Wilkinson AS, Burstein E, Rumble JM, Balliu B, et al. VIAF, a conserved inhibitor of apoptosis (IAP)-interacting factor that modulates caspase activation. The Journal of biological chemistry. 2004; 279(49):51091–9. Epub 2004/09/17. [PubMed: 15371430]

31. Baumann K, Mandelkow EM, Biernat J, Piwnica-Worms H, Mandelkow E. Abnormal Alzheimer-like phosphorylation of tau-protein by cyclin-dependent kinases cdk2 and cdk5. FEBS letters. 1993; 336(3):417–24. Epub 1993/12/28. [PubMed: 8282104]

32. Li A, Meyre D. Challenges in reproducibility of genetic association studies: lessons learned from the obesity field. International journal of obesity. 2013; 37(4):559–67. Epub 2012/05/16. [PubMed: 22584455]

33. Lange EM, Sun J, Lange LA, Zheng SL, Duggan D, Carpten JD, et al. Family-based samples can play an important role in genetic association studies. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2008; 17(9):2208–14. Epub 2008/09/05.

34. Laird NM, Lange C. Family-based methods for linkage and association analysis. Advances in genetics. 2008; 60:219–52. Epub 2008/03/25. [PubMed: 18358323]

35. Ott J, Kamatani Y, Lathrop M. Family-based designs for genome-wide association studies. Nature reviews Genetics. 2011; 12(7):465–74. Epub 2011/06/02.

36. Evangelou E, Trikalinos TA, Salanti G, Ioannidis JP. Family-based versus unrelated case-control designs for genetic associations. PLoS genetics. 2006; 2(8):e123. Epub 2006/08/10. [PubMed: 16895437]

Table 1

Results of Genome-wide Family-based Association Analysis of Imputed Genotypes using Three Large Collections of Multiplex AD Study Families.

Panel A: Functional variants; SNPs that are either in a gene or in LD with the gene.

SNP	Chr	Position	Fams (MA)	Zscore (MA)	P-value (MA)	Effect Direction (MA)	Gene	NIMH	NCRAD	NIA	Approach
rs7609954	3	61636156	696	5.492	3.98E-08	++++++	PTPRG	8.73E-03	7.57E-03	1.73E-03	2
rs1347297	2	179244986	640	5.469	4.53E-08	+++	OSBPL6	3.65E-02	5.85E-02	1.24E-06	1
rs11749176	5	44145931	840	5.218	1.81E-07	++++++	FGF10	5.95E-02	3.74E-02	3.12E-02	2
rs12378800	9	100631820	438	5.218	1.81E-07	++++++	FOXE1	6.33E-02	1.92E-02	1.05E-01	2
rs185968827	6	56708510	844	5.158	2.49E-07	++++++	DST	1.12E-02	4.09E-02	5.25E-02	2
rs62400067	6	20592984	502	5.092	3.54E-07	++++++	CDKAL1	2.51E-02	6.81E-02	1.65E-04	2
rs9304861	19	35271888	622	5.082	3.74E-07	++++++	ZNF599	2.19E-02	5.92E-03	3.97E-02	2
rs9994441	4	170094562	622	4.99	6.05E-07	++++++	SH3RF1	5.92E-01	1.52E-02	1.11E-01	2
rs3931397	4	149079497	524	4.984	6.22E-07	++++++	NR3C2	1.62E-02	1.85E-02	2.73E-01	2
rs72953347	2	179274829	670	4.98	6.36E-07	+++	OSBPL6	3.27E-02	2.72E-01	3.39E-06	2
rs115500410	5	76852235	922	4.963	6.94E-07	++++++	WDR41	1.85E-01	1.88E-01	3.06E-04	2
rs56146971	14	91920101	904	4.941	7.76E-07	++++++	SMEK1	7.24E-03	5.18E-02	7.94E-03	2
rs77220498	9	100700832	440	4.946	7.59E-07	++++++	HEMGN	2.16E-01	1.31E-02	1.15E-01	2
rs6491411	13	98904568	872	4.911	9.08E-07	++++++	FARP1	1.91E-01	2.04E-02	1.36E-04	2
rs10456232	6	20579123	258	4.9	9.60E-07	+++	CDKAL1	2.73E-02	2.75E-02	1.49E-04	1

Panel B: Intergenic variants; genes close to the associated markers are listed with proximal distance in base-pairs.

SNP	Chr	Position	Fams (MA)	Zscore (MA)	P-value (MA)	Effect Direction (MA)	Gene (proximity)	NIMH	NCRAD	NIA	Approach
rs1513625	2	101314473	992	5.479	4.28E-08	++++++	PDCL3 (+121276)	7.64E-03	1.12E-02	3.85E-01	2
rs543844	6	44424800	968	5.323	1.02E-07	++++++	CDC5L (+6637)	3.50E-03	2.07E-01	2.63E-02	2
rs186971130	2	104619458	576	5.311	1.09E-07	++++++	-	1.36E-03	8.69E-01	1.64E-02	2
rs7374058	3	26246575	794	5.137	2.80E-07	++++++	LINC00692 (331384)	5.00E-02	2.07E-02	4.40E-02	2
rs73310256	10	92438849	478	5.123	3.01E-07	++++++	HTR7 (−61729)	8.15E-02	1.25E-01	5.58E-02	2
rs6908580	6	92577371	1058	5.099	3.41E-07	++++++	-	5.79E-03	2.25E-01	3.30E-02	2
rs7047415	9	98541232	578	5.096	3.48E-07	++++++	DKFZP434H0512 (+4219)	3.73E-02	4.36E-01	3.33E-04	2

SNP	Chr	Position	Fams (MA)	Zscore (MA)	P-value (MA)	Effect Direction (MA)	Gene (proximity)	NIMH	NCRAD	NIA	Approach
rs982100	2	118475544	740	5.08	3.77E-07	++++++	DDX18 (−96682)	3.02E-02	5.88E-01	7.66E-03	2
rs75718659	4	187743086	460	5.052	4.37E-07	++++++	FAT1 (+95210)	2.12E-02	7.51E-01	5.57E-01	2
rs140633572	6	107279935	408	5.03	4.92E-07	++++++	C6orf203 (−69472)	1.66E-03	6.68E-01	3.68E-01	2
rs1443024	2	185376297	976	4.999	5.78E-07	++++++	ZNF804A (−86796)	6.57E-02	4.66E-02	5.23E-03	2
rs11773349	7	64407164	562	4.991	6.01E-07	++++++	ZNF273 (+15820)	4.69E-02	8.52E-01	1.73E-03	2
rs36060340	1	233622928	766	4.975	6.54E-07	++++++	MLK4 (+102034)	1.05E-03	2.89E-01	1.88E-01	2
rs1774093	9	104647156	766	4.956	7.19E-07	++++++	GRIN3A (+146294)	4.96E-03	6.29E-01	3.81E-01	2
rs857551	21	44829992	588	4.949	7.45E-07	++++++	SIK1 (−4403)	3.89E-03	5.87E-01	1.15E-01	2
rs9546312	13	83746951	658	4.92	8.68E-07	++++++	-	6.91E-02	1.16E-01	2.61E-01	2
rs4705644	5	113577164	862	4.905	9.33E-07	++++++	KCNN2 (−119478)	2.85E-05	6.82E-01	2.48E-01	2

Legend to Table 1: FBAT-GEE method was used in the analyses using affection status and age at onset as a multivariate phenotype. The p values are nominal and two-tailed for all the associated markers. Fams indicates the number of informative families contributing to the test statistic. Age of onset coding based on Wilcoxon statistic. Meta-analysis results including the imputed data sets NIMH, NIA and NCRAD. Zscore, Z-score of the test statistic (negative scores indicate under-transmission of minor allele to affected individuals). P-value = P-value derived from the meta-analysis. Thresholds to achieve genome-wide significance is $5 \cdot 10^{-08}$. Effect Direction (NCRAD, NIA, NIMH), the effect direction is positive if the involved studies have the same direction compared to the first study, otherwise it's negative. The FBAT-GEE p-values of the association signal from the three individual family cohort are listed in the columns, NIMH, NCRAD and NIA. These p-values were used in meta-analysis as described in the methods. Approach indicates the kind of analysis (1) FBAT GEE Within-Family and (2) FBAT GEE Between-Family as described in the material and method section.