

ORIGINAL ARTICLE

Family-based association analyses of imputed genotypes reveal genome-wide significant association of Alzheimer's disease with *OSBPL6*, *PTPRG*, and *PDCL3*C Herold^{1,2,9}, BV Hooli^{3,9}, K Mullin³, T Liu⁴, JT Roehr⁴, M Mattheisen⁵, AR Parrado⁶, L Bertram^{4,7,8}, C Lange² and RE Tanzi³

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 AD. On the other hand, susceptibility to late-onset forms of AD (LOAD) is indisputably associated to the $\epsilon 4$ 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 (~3500 subjects from 1070 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 = 3.98×10^{-8}), rs1347297 in the gene *OSBPL6* (P -value = 4.53×10^{-8}), and rs1513625 near *PDCL3* (P -value = 4.28×10^{-8}). In addition, rs72953347 in *OSBPL6* (P -value = 6.36×10^{-7}) and two SNPs in the gene *CDKAL1* showed marginally significant association with LOAD (rs10456232, P -value = 4.76×10^{-7} ; rs62400067, P -value = 3.54×10^{-7}). 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.

Molecular Psychiatry (2016) **21**, 1608–1612; doi:10.1038/mp.2015.218; published online 2 February 2016

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 peptide in the form of senile 'plaques' and oligomers (crucial to initiating AD pathogenesis), and accumulation of hyperphosphorylated tau protein in the form of intracellular neurofibrillary 'tangles', 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 AD (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- $\epsilon 4$* still remains the strongest risk factor in LOAD, where the $\epsilon 4$ allele confers between 3.7- and 14-fold increases in risk, in heterozygotes and homozygotes, respectively. Importantly, the identification of the above four genes was the key to understand the underlying molecular mechanism leading to AD driven by β -amyloid 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 genome-wide association studies (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 single nucleotide polymorphisms (SNPs) or present in close proximity include, triggering receptor expressed on myeloid cells 2, bridging integrator 1, sialic acid-binding immunoglobulin-like lectin, clusterin, ATP-binding cassette transporter (ABCA7), complement receptor 1, phosphatidylinositol binding clathrin assembly protein, 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

¹German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany; ²Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA; ³Genetics and Aging Research Unit, Department of Neurology, Massachusetts General Hospital, Boston, MA, USA; ⁴Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin, Germany; ⁵Department of Biomedicine, Bartholins Allé 6, Aarhus, Denmark; ⁶Computational Sciences, Janssen Research and Development, Pharmaceutical Company of Johnson and Johnson, Spring House, PA, USA; ⁷School of Public Health, Faculty of Medicine, The Imperial College of Science, Technology, and Medicine, London, UK and ⁸Platform for Genome Analytics, Institutes of Neurogenetics and Integrative and Experimental Genomics, University of Lübeck, Lübeck, Germany. Correspondence: Professor C Lange, Department of Biostatistics, Harvard T.H. Chan School of Public Health, 665 Huntington Avenue, Boston, MA 02115, USA or Dr RE Tanzi, Genetics and Aging Research Unit, Department of Neurology, Massachusetts General Hospital, Charlestown, MA 02129, USA or Department of Neurology, Harvard Medical School, Boston, MA 02114, USA. E-mail: clange@hsph.harvard.edu or tanzi@helix.mgh.harvard.edu

⁹These authors contributed equally to the work.

Received 14 October 2015; revised 25 November 2015; accepted 7 December 2015; published online 2 February 2016

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 for fully understanding the disease pathogenesis and developing the treatment strategies.

In this study we performed a systematic meta-analysis of the family-based association test (FBAT) results using imputed genotypes (limited to minor allele frequency > 0.05) generated in the subjects from three large Alzheimer's family collections: National Institute of Mental Health (NIMH), National Institute on Aging Genetics Initiative for Late Onset Alzheimer's Disease (NIA-LOAD), and National Repository of Research on Alzheimer's Disease (NCRAD). A total of 3500 subjects from 1070 families were assessed in this study. To maximize the 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 (family-based association test using generalized estimating equations (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.

MATERIALS AND METHODS

Study families

We utilized three large family-based AD samples in the association tests and the meta-analyses: NIMH, NCRAD and NIA. The NIMH Alzheimer disease genetics initiative study⁸ 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 1536 subjects from 457 families. For the purposes for our study 1376 participants (941 affected and 404 unaffected) from 410 families were included. The complete NIA-LOAD family sample and the National Cell Repository for Alzheimer's Disease sample contains 4006 subjects from 1653 families. Here, we included 1040 subjects (748 affected and 282 unaffected) from 329 multiplex families. The families originally ascertained as part of the NCRAD subset of families include 1108 subjects from 331 families, with 799 affected and 293 unaffected siblings. Affection status was based on clinical dementia diagnosis documentation according to the 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, 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 of < 5% were excluded. Population stratification within and between the samples was also checked by performing a multi-dimensional scaling (identification of population outliers) implemented in PLINK.¹³ Duplicated DNA samples were identified by consideration of genome-wide genotype identity-by-state status (identity-by-state > 1.98). From each pair the individual with the lower genotyping rate was removed. In a second step, we used IMPUTE2 (ref. 14) to impute the QCed datasets NCRAD and NIMH into the May 2013 release of the 1000 genomes

project and NIA into the September 2013 release of the 1000 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 ($n = 141\,29\,045$ markers), NCRAD ($n = 13\,971\,550$), and NIA-LOAD ($n = 15\,107\,142$), for association analysis. A total of 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-GEE⁷ 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 $\mathbf{Y}_{ij} = (\mathbf{Y}_{ij1}, \dots, \mathbf{Y}_{ijm})$, where \mathbf{Y}_{ijk} is the k th phenotype for the j th offspring in the family. The multivariate FBAT-GEE statistic is constructed by replacing the univariate coding variable \mathbf{T}_{ij} in C by the coding vector defined by¹⁸

$$\mathbf{T}_{ij} = \mathbf{Y}_{ij} - \hat{\mathbf{Y}}_{ij} = \begin{pmatrix} \mathbf{Y}_{ij1} \\ \dots \\ \mathbf{Y}_{ijk} \\ \dots \\ \mathbf{Y}_{ijm} \end{pmatrix} - \begin{pmatrix} \hat{\mathbf{Y}}_{ij1} \\ \dots \\ \hat{\mathbf{Y}}_{ijk} \\ \dots \\ \hat{\mathbf{Y}}_{ijm} \end{pmatrix}$$

where, the $\hat{\mathbf{Y}}_{ijk}$'s are the predicted trait values based on the regression model for covariates. Replacing univariate coding variable \mathbf{T}_{ij} in FBAT statistic by the vector \mathbf{T}_{ij} results in the FBAT-GEE statistic

$$\mathbf{T}_{FBAT-GEE} = \mathbf{C}^T \mathbf{V}^{-1} \mathbf{C}$$

Under the null hypothesis, the FBAT-GEE statistic has a χ^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 test 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, that is, 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 analysis¹ 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 with a gene we used the software package epigwas.²² For each SNP, the SNPs in a 1M window (upstream 500 K and downstream 500 K) are included in the calculation. The tool calculates linkage disequilibrium using 1000 genome

Table 1. Results of genome-wide family-based association analysis of imputed genotypes using three large collections of multiplex AD study families

SNP	Chr	Position	Fams (MA)	Zscore (MA)	P-value (MA)	Effect direction (MA)	Gene	NIMH	NCRAD	NIA	Approach
<i>Panel A: functional variants; SNPs that are either in a gene or in LD with the gene</i>											
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
SNP	Chr	Position	Fams (MA)	Zscore (MA)	P-value (MA)	Effect direction (MA)	Gene (proximity)	NIMH	NCRAD	NIA	Approach
<i>Panel B: intergenic variants; genes close to the associated markers are listed with proximal distance in base-pairs</i>											
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
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

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 5×10^{-8} . Effect direction (NCRAD, NIA, and 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 Materials and Methods section.

data for the European population. Only SNPs with $r^2 > 0.5$ 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 $< 10^{-6}$, a minor allele frequency > 0.05 , and with the same effect direction in each family sample. The complete list of SNPs with *P*-values $> 10^{-5}$ are listed in the Supplementary Table 3a and b.

The APOE region (rs56131196) shows highly significant results (*P*-values of 3.09×10^{-24} and 3.96×10^{-24} for approaches (1) and (2), respectively; 187 SNPs < 0.05 for approach (1) and 283 SNPs for approach (2). The three early-onset familial AD 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 = 3.74×10^{-3}). 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 = 2.23×10^{-3}). 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 = 1.9×10^{-4} ; the PSEN1 region has 311 SNPs < 0.05 (rs214277 *P*-value = 1.07×10^{-4}) and the PSEN2 region with 87 SNPs < 0.05 (rs149734051 *P*-value = 1.06×10^{-3}). 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 and B). We found 32 novel variants showing

genome-wide significant association with AD and fulfilling the criteria described above, that is, P -value $< 10^{-6}$, minor allele frequency $> 5\%$ and same effect direction across all samples tested. Detailed information can be found in Supplementary Table 1. A total of 15 variants were either in a gene (Table 1) or in linkage disequilibrium with SNPs in a gene, while 18 other SNPs (Table 1) 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 = 3.98×10^{-8} ; rs1513625 (*PDCL3*): P -value = 4.28×10^{-8} ; rs1347297 (*OSBPL6*): P -value = 4.53×10^{-8} . A second SNP, rs72953347 in *OSBPL6* also showed marginally significant evidence for association with AD using the other meta-analysis approach (approach 2): $P = 6.36 \times 10^{-7}$). 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 = 3.54×10^{-7} ; rs10456232: P -value = 9.60×10^{-7}).

We next tested the 32 SNPs from Table 1 (Panel A and B) showing GW-significant association with AD using family-based methods in the International Genomics of Alzheimer's project (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 elsewhere²³ 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 (~3500 subjects from 1070 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. In addition, another SNP, rs72953347 in *OSBPL6* (P -value = 6.36×10^{-7}) and two SNPs (rs10456232 and 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, for example, *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 vascular endothelial growth factor (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.³⁰ *CDKAL1* encodes the methylthiotransferase family member, cyclin-dependent kinase 5 Regulatory Subunit Associated protein-Like 1, and has been previously associated with noninsulin-dependent diabetes mellitus. Interestingly, cyclin-dependent kinase 5 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 IGAP²⁴ 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, that is, 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,^{18,34} 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.³⁵ 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.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Herold performed pre-GWAS QC, the imputation of the GWAS datasets, statistical analysis of family-based cohorts and drafted the manuscript. Hooli processed DNA samples on human microarray, SNP genotype data generation and quality control, and drafted manuscript. Mullin performed genotyping, data

generation and quality assessment. Liu assisted in the imputation of the GWAS datasets, statistical analysis of case-control datasets. Roehr performed the imputation of the various GWAS datasets. Mattheissen performed pre-GWAS QC, statistical analysis and revised the manuscript for intellectual content. Parrado performed pre-GWAS QC and statistical analysis. Bertram helped design, conceptualization, and planning of the study, and manuscript revision. Lange designed, conceptualized, planned, and oversaw the study, and helped with the meta-analysis of the GWAS studies, the GWAS approaches and plan the study. Tanzi designed, conceptualized, planned, and oversaw the study, and helped in drafting and revising the manuscript.

REFERENCES

- Canadian study of health and aging: study methods and prevalence of dementia. *CMAJ* 1994; **150**: 899–913.
- Cure Alzheimer's Fund. 2015; Available from <http://curealz.org/alzheimers-disease>.
- Alzheimer's Assoc. Facts & Figures. 2015; Available from <http://www.alz.org/facts/overview.asp>.
- Glennier GG, Wong CW. Alzheimer's disease and Down's syndrome: sharing of a unique cerebrovascular amyloid fibril protein. *Biochem Biophys Res Commun* 1984; **122**: 1131–1135.
- Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S et al. Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry* 2006; **63**: 168–174.
- Laird NM, Horvath S, Xu X. Implementing a unified approach to family-based tests of association. *Genet Epidemiol* 2000; **19**: S36–S42.
- 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**: 195–206.
- Blackler 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**: 139–147.
- Ku CS, Pawitan Y, Sim X, Ong RT, Seielstad M, Lee EJ et al. Genomic copy number variations in three Southeast Asian populations. *Hum Mutat* 2010; **31**: 851–857.
- 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. *Hum Hered* 2011; **71**: 221–233.
- 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. *Arch Neurol* 2008; **65**: 1518–1526.
- 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. *Am J Hum Genet* 2008; **83**: 623–632.
- 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. *Am J Hum Genet* 2007; **81**: 559–575.
- Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009; **5**: e1000529.
- 1,000 Genome Project Reference data. 2013.
- 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 Genet* 2009; **5**: e1000741.
- 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. *Nat Genet* 2005; **37**: 683–691.
- Laird NM, Lange C. Family-based methods for linkage and association analysis. *Adv Genet* 2008; **60**: 219–252.
- Lange C, DeMeo D, Silverman EK, Weiss ST, Laird NM. PBAT: tools for family-based association studies. *Am J Hum Genet* 2004; **74**: 367–369.
- 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. *Genet Epidemiol* 2012; **36**: 785–796.
- Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinformatics* 2010; **26**: 2190–2191.
- 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. *Nat Genet* 2013; **45**: 124–130.
- Bertram L, Tanzi RE. Genome-wide association studies in Alzheimer's disease. *Hum Mol Genet* 2009; **18**: R137–R145.
- 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. *Nat Genet* 2013; **45**: 1452–1458.
- 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.
- 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. *Nat Genet* 2009; **41**: 316–323.
- 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. *Neurobiol Aging* 2013; **34**: 1653–1661.
- Olkkonen VM, Li S. Oxysterol-binding proteins: sterol and phosphoinositide sensors coordinating transport, signaling and metabolism. *Prog Lipid Res* 2013; **52**: 529–538.
- 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. *J Biol Chem* 2013; **288**: 23171–23181.
- 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. *J Biol Chem* 2004; **279**: 51091–51099.
- Baumann K, Mandelkow EM, Biernat J, Piwnicka-Worms H, Mandelkow E. Abnormal Alzheimer-like phosphorylation of tau-protein by cyclin-dependent kinases cdk2 and cdk5. *FEBS Lett* 1993; **336**: 417–424.
- Li A, Meyre D. Challenges in reproducibility of genetic association studies: lessons learned from the obesity field. *Int J Obes* 2013; **37**: 559–567.
- 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 Epidemiol Biomarkers Prev* 2008; **17**: 2208–2214.
- Ott J, Kamatani Y, Lathrop M. Family-based designs for genome-wide association studies. *Nat Rev Genet*. 2011; **12**: 465–474.
- Evangelou E, Trikalinos TA, Salanti G, Ioannidis JP. Family-based versus unrelated case-control designs for genetic associations. *PLoS Genet* 2006; **2**: e123.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (<http://www.nature.com/mp>)