Biochemical and Biophysical Research Communications 463 (2015) 490–495

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

# Functional enrichment analysis of three Alzheimer's disease genome-wide association studies identities *DAB1* as a novel candidate liability/protective gene

Hui Gao [a, b, 2], Yu Tao [a, b, 2], Qin He [a, b, 1], Fan Song [a], David Saffen [a, b, c, *]

[a] Department of Cellular and Genetic Medicine, School of Basic Medical Sciences, China
[b] Institutes of Brain Science, Fudan University, Shanghai, China
[c] State Key Laboratory for Medical Neurobiology Fudan University, 130 Dong'an Road, Shanghai 200032, China

ARTICLE INFO

*Article history:*
Received 26 April 2015
Accepted 2 May 2015
Available online 29 May 2015

*Keywords:*
Alzheimer's disease
Genome-wide association study
Database for Annotation
Visualization and Integrated Discovery

ABSTRACT

To explore genetic contributions of Alzheimer's disease (AD) at the level of biological terms and pathways, we analyzed three Caucasian population-based genome-wide association study datasets (TGEN_ND, GeneADA and NIA_LOAD) using the Database for Annotation, Visualization and Integrated Discovery (DAVID). This analysis identified 4 annotation terms ("Fibronectin type III-like fold," "Cell adhesion," "Cell motion" and "Ig-like-C2-type 3") and 17 genes that associated with AD susceptibility in two or more of the GWAS datasets. Ten of these genes, have previously been identified as candidate AD liability genes in genetic association studies (*AGT*, *COL11A1*) or encode proteins that function in biological systems or pathways previously implicated in AD (*BARHL2*, *CSF3R*, *DAB1*, *HMCN1*, *LEPR*, *PTPRF*, *PXDN*, *TNR*). Among these, *DAB1* (*Dab*, *reelin signal transducer*, *homolog 1*) was of particular interest, since it encodes a protein that functions downstream from reelin, a signaling pathway previously identified as protective in AD. Multiple linear regression analysis of correlations between brain *DAB1* mRNA expression and SNP genotype using data from the "BrainCloud" database identified five SNPs within the *DAB1* locus that correlated with mRNA expression in human dorsolateral prefrontal cortex. Analysis of predicted levels of *DAB1* mRNA expression based on genotype combinations present in AD cases and controls vs. the $\log_{10}$-transformed odds ratios for AD diagnosis, revealed statistically significant correlations in one of the GWAS datasets (GenADA), with high *DAB1* mRNA expression correlating with AD protection. Multidimensional scaling (MDS) analysis of cases and controls in the three GWAS, revealed genetic differences between GenADA and TGEN_ND/NIA_LOAD, which were similar to each other. To our knowledge, this study is the first to provide genetic evidence for *DAB1* as a candidate AD liability/protection gene, although the strength of the contribution of *DAB1* may differ among populations.



## 1. Introduction

Alzheimer's disease (AD) is a progressive and ultimately fatal neurodegenerative disorder, that affected an estimated 36.5 million individuals worldwide in 2010 and is projected to affect as many as 115 million in 2050 [1]. Family- and case/control-based association studies have demonstrated a large genetic contribution to AD, and, to date, genome-wide association studies (GWAS) have identified more than twenty candidate AD liability genes that associate with AD diagnosis at genome-wide level of significance [2,3]. Genetic markers that associate with AD liability (usually single nucleotide polymorphisms: SNPs) in GWAS, however, often lack supporting biological mechanisms.

To gain insights into the range of biological systems and pathways that contribute to AD liability and protection, researchers have employed various kinds of gene set or pathway enrichment analyses. For example, Liu GY and Guiyou Liu et al. used multiple pathway analysis programs to study two French population-based AD GWAS datasets, producing evidence that cell adhesion molecules (CAMs) contribute to AD [4]. Likewise, Lee YH and colleagues

* Corresponding author. Department of Cellular and Genetic Medicine, School of Basic Medical Sciences, Fudan University, 130 Dong'an Road, Shanghai 200032, China.
*E-mail address:* saffen@fudan.edu.cn (D. Saffen).
[1] Current address: Centre de recherche, Institut universitaire en santé mentale de Montréal, Canada.
[2] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.bbrc.2015.05.044
0006-291X/

used Identify Candidate Causal SNPs and Pathway (ICSNPathway) analysis with a European-descent population-based AD GWAS dataset to obtain evidence that two pathways, one that negatively regulates programmed cell death and another related to the M phase of the mitotic cell cycle, contribute to AD susceptibility [5]. Additional gene- or pathway-based analyses of AD GWAS datasets have provided support for contributions of genes related to the immune system [6,7], lipid pathways [8], cardiovascular disease-related pathways [9] and glutamate-signaling [10].

In this study, we used the web-accessible, gene annotation term-based Database for Annotation, Visualization and Integrated Discovery (DAVID, http://david.niaid.nih.gov) [11], for functional enrichment analysis of three independent AD GWAS datasets. This analysis identified four annotation terms and 10 genes with plausible links to AD. Among these genes, the reelin pathway gene *DAB1* (*Dab*, *reelin signal transducer*, *homolog 1*) showed statistically significant correlations between mRNA expression in human brain and odds ratios (ORs) for AD liability or protection in the GenADA GWAS, providing the first genetic evidence linking *DAB1* to AD.

## 2. Materials and methods

### 2.1. Study populations

Three independent, Caucasian population-based AD GWAS datasets were used for genome-wide enrichment analysis: 1) Translational Genomics Research Institute Neuropathology Discovery study (TGEN_ND: www.tgen.org), 2) Canadian-Based Genetics of Alzheimer's Disease Study (GenADA: http://www.ncbi.nlm.nih.gov/gap/?term=phs000219.v1.p1) and 3) National Institute of Aging Late-Onset Alzheimer's Disease Family Study (NIA_LOAD: http://www.ncbi.nlm.nih.gov/gap/?term=phs000168.v1.p1). AD cases and controls in each study were independent from the other studies. AD cases and controls selected from TGEN_ND and GenADA datasets self-identified as Caucasian. AD probands and controls selected from NIA_LOAD dataset (one from each family) self-identified as "White, non-Hispanic," The three datasets had been filtered to exclude individuals with missing genotype rates >10% and SNPs with minor allele frequencies (MAF) < 0.02, Hardy–Weinberg Equilibrium (HWE) $p > 0.01$ or missing rates >10%.

### 2.2. Genotype imputation

Imputation of genotypes for autosomal SNPs into the three "cleaned" datasets were performed using IMPUTE version 2.2.2 [12] based on 1000 Genomes Phase I integrated haplotypes, which were produced using SHAPEIT2 [13] and released in June 2014 (http://mathgen.stats.ox.ac.uk/impute/impute_v2.html). For each case or control, imputed SNP genotypes with posterior probabilities >0.9 were accepted, otherwise the genotypes were classified as missing. Quality control for imputed SNPs also included removing SNPs with info scores <0.3 or MAF <0.01 or HWE $p > 0.0001$.

### 2.3. Genome-wide association study

Differences in SNP allele frequencies between cases and controls in the TGEN_ND, GenADA and NIA_LOAD were quantified by association analysis using the basic association test (allelic $\chi^2$ test) in PLINK [14], which produced an asymptotic p-value for each SNP. SNPs with $p < 10^{-2}$ were extracted using PLINK.

### 2.4. DAVID-based tools for functional enrichment analysis

To identify biological processes implicated in AD based on GWAS data, we performed functional enrichment analysis of the TGEN_ND, GeneADA and NIA_LOAD datasets using DAVID. Because this program requires lists of genes as input, we used Scandb software (www.scandb.org) to map SNPs that correlated ($p < 10^{-2}$) with AD diagnosis in three GWAS into specific genes. Scandb performs "physical" annotations, where SNPs are categorized according to their physical locations within genes or, in the case of intergenic SNPs, strength of linkage to nearby genes. Intergenic SNPs are assigned to genes with which they are in linkage disequilibrium (LD) at the level of $r^2 > 0.20$.

Two DAVID analytic tools were used to analyze gene lists generated for each of the three GWAS: 1) "Functional Annotation Chart" and 2) "Functional Annotation Clustering" [15]. The "Functional Annotation Chart" tool identifies gene annotation terms that are enriched in the input gene list, with genes with similar functions linked to the same term, and ranks the terms according to their enrichment *P*-values. *P*-values lower than 0.05 are considered to be nominally significant. Analysis of each GWAS produces a list of enriched terms (based on the input list of genes), allowing terms enriched in two or more GWAS to be identified. For each of the common enriched terms, there is an associated list of genes (a subset of the input gene list) for each GWAS. These lists can be compared to identify genes common to two or more lists. The "Functional Annotation Clustering" tool identifies "clusters" of related gene annotation terms that are enriched in the input gene list. Each cluster of terms receives an "enrichment score" (ES), defined as the geometric mean of the *P*-values (on the $-\log_{10}$ scale) for all single terms in the cluster. The Benjamini-Hochberg-false discovery rate (BH-FDR) is calculated as a method to correct for multiple testing, with nominal significance assigned for BH-FDR $\leq 0.25$.

### 2.5. Candidate gene identification

To identify genes related to AD, a list of genes identified using both the "Functional Annotation Chart" tool and "Functional Annotation Clustering" tool and common to two or three GWAS was compared to the most recent list of AD candidate genes in the AlzGene database (http://www.alzgene.org) [16]. In addition, we searched PubMed (http://www.ncbi.nlm.nih.gov/pubmed) for the terms: 1) "official gene name" and "Alzheimer's disease" or "Alzheimer disease." 2) "Official gene name" and "neurodegeneration," 3) "Official gene" and a specific "biological pathway," associated with the gene in KEGG database (http://www.kegg.jp). For each gene, we also consulted the Gene database (http://www.ncbi.nlm.nih.gov) to identify published studies relevant to AD or neurodegeneration that used non-official names for the gene.

### 2.6. Analysis of DAB1 mRNA expression and association with AD

Data for *DAB1* mRNA expression obtained from 109 independent samples of dorsolateral prefrontal cortex (DLPFC) of Caucasian origin was downloaded from BrainCloud Web site (http://braincloud.jhmi.edu/downloads.htm). The normalized $\log_2$-transformed expression values were adjusted for the covariates age, sex, postmortem interval, RNA integrity number (RIN), smoking status, sample source and batch number by multiple linear regression. Genotype data for SNPs in the region of the *DAB1* gene for each brain sample were obtained from database for Genotypes and Phenotypes (dbGAP): http://www.ncbi.nlm.nih.gov/gap, and imputed using IMPUTE version 2.2.2 as described above.

Correlations between mRNA expression and genotypes of *DAB1*-region SNPs were analyzed by single-variable linear regression analysis using the –assoc command in PLINK, providing each SNP with a P-value and coefficient of determination ($R^2$), the later serving as a measure of the contribution of each SNP to the variance

of *DAB1* mRNA expression. Pairwise linkage $r^2$ disequilibrium LD constants (designated "$\Delta^2$" to avoid confusion with "$R^2$") for each SNP were calculated using the–ld command in PLINK. Because the *DAB1*-region extends over an approximately 1252 kb segment of chromosome 1 (Chr1p32-p31), we subdivided this region into 2 non-overlapping sub-regions, which were examined independently using "$R^2$–$\Delta^2$" analysis, a method that we have recently developed to estimate the number of independent or semi-independent SNP families that correlate with mRNA expression [Wang, J et al., *Annals of Human Genetics*, 2015; in press]. A brief summary of this method is provided in the Legend to Fig. S1.

Multiple stepwise linear regression (implemented in SPSS) of *DAB1* mRNA expression vs. genotypes of "index" SNPs identified in the "$R^2$–$\Delta^2$" analysis, was used to identify a set of SNPs with the highest adjusted-$R^2$ contribution to the variance of *DAB1* mRNA expression. The regression equation from this analysis (mRNA expression level $= \beta_0 + \sum_{j=1}^{k} \beta_j X_j$ (k = 1 through 5), where $X_j$ = genotype of the jth SNP and $\beta_j$ = regression coefficients] was used to estimate brain *DAB1* mRNA expression levels in AD cases and controls, based on the specific combinations of genotypes for these SNPs harbored by each AD case or control. Correlations between *DAB1* mRNA expression and AD liability were investigated using weighted linear regression analysis (implemented in R) of estimated *DAB1* mRNA expression vs. $\log_{10}$-transformed odds ratios ($\log_{10}$ORs, AD cases vs controls) for specific genotype combinations in cases and controls (calculated using RANDOMPAT (http://www.jurgott.org/linkage/randompat.html; [17]) in the TGEN_ND, GenADA or NIA_LOAD GWAS datasets.

## 3. Results

### 3.1. Imputation of genotypes and GWAS and generation of lists of genes linked to SNPs that associated with AD diagnosis

Imputation of genotypes into the three GWAS datasets using IMPUTE 2.2.2 with 1000 Genomes Phase I integrated haplotypes as a reference, followed by selection of SNPs that passed quality control, produced ~14-, 10- and 11-fold increases in the number of SNPs available for analysis in the three GWAS (Table 1, Table S1). Because we were interested in identifying SNPs that associate with AD diagnosis that are unlinked to the known AD liability locus *APOE* [18], SNPs linked to *APOE* with $r^2$ LD > 0.1 were removed from the three datasets. Association analysis between SNP allele frequencies and AD diagnosis was carried out using the basic association test in PLINK [14]. With the exclusion of SNPs linked to *APOE*, no SNPs reached genome-wide significant level in any of the three GWAS. As described in Methods, we used Scandb software to identify genes that contain or are in LD with SNPs that associated with AD diagnosis in the three GWAS (Table 1, Table S2).

### 3.2. Functional enrichment analysis using the DAVID "Functional Annotation Chart" tool

This analysis yielded 125 annotation terms in TGEN_ND, 309 in GenADA and 120 in NIA_LOAD that associated with AD ($P < 0.05$, Table S3). Six of these terms were found to be enriched ($P < 0.03$) in each of the three GWAS (Table 2). The terms "alterative splicing" and "splice variants" showed the highest level of enrichment. Two of the remaining four terms are related to cell adhesion ("IPR008957: Fibronectin type III-like fold" and "cell adhesion") and one term each to cell mobility ("cell motion") and Ig-like domains ("Domain: Ig-like C2-type 3").

### 3.3. Functional enrichment analysis using the DAVID "Functional Annotation Clustering" tool

To further validate terms replicated in the three datasets, we used the DAVID "Functional Annotation Clustering" tool, which clusters gene annotation terms based on related biological functions and pathways and assesses the enrichment of individual clusters within a given dataset. In this analysis, input gene lists from the TGEN_ND study yielded 10 clusters, GeneADA 28 clusters and NIA_LOAD 9 clusters of annotation terms with enrichment scores (ES) greater than 1.3 (the consensus threshold for statistical significance). Among those terms, five clusters were common to all three GWAS datasets, with 4 clusters showing statistically significant enrichment in two or more of the GWAS datasets (Table 3).

Cluster A, which included the terms "alternative splicing" and "splice variant," did not reach statistically significant level of enrichment in any of the datasets. By contrast, cluster E, which included the term "Domain: Ig-like C2-type 3" obtained an ES larger than 1.3 in all three datasets. Three additional clusters (B, C and D, containing the terms "IPR008957: Fibronectin type III-like fold," "GO: 0007155~cell adhesion" and "GO: 0006928~cell motion," respectively) achieved ES larger than 1.3 in both the TGEN_DN and GenADA datasets, but not in NIA_LOAD dataset.

### 3.4. Identification of candidate AD liability/protective genes

Because cluster A did not reach statistical significance in the DAVID "cluster" analysis, we decided to focus on genes with annotation terms identified as enriched using the "Functional Annotation Chart" tool and also listed within clusters B, C, D or E. Altogether, 17 genes met these criteria for two or more GWAS (Table 4). Among these genes, *AGT* (*angiotensinogen*) and *COL11A1* (*collagen, type XI, alpha 1*) have previously been identified as candidate AD liability genes in the AlzGene database (http://www.alzgene.org/). To investigate possible contributions of the remaining genes to AD, we conducted an extensive search of the scientific literature using the NCBI Gene and PubMed databases, with the search terms described in Methods. These searches identified an additional eight genes with links to biological pathways implicated in AD (Table 4 and Table S4). In total, 10 of the 17 genes (58.8%) listed in Table 4 have direct or indirect associations with AD.

### 3.5. DAB1 mRNA expression in human brain correlates with AD protection

Among the 10 genes with plausible associations with AD, our attention was drawn to *DAB1*, which functions within the reelin signaling pathway, immediately downstream from the reelin receptors VLDLR (very low density lipid receptor) and APOER2 (Apolipoprotein E receptor 2) [19,20]. Although reelin is best known for its role in brain development [21], synaptic functions [22,23], and as a candidate liability gene for neurodevelopmental disorders, it has also been implicated in Alzheimer's disease [24,25], with higher expression levels possibly correlated with AD protection [26,27].

**Table 1**
Sample information for three GWAS.

| | TGEN_ND | GenADA | NIA_LOAD |
|---|---|---|---|
| Genotyped SNPs | 299404 | 358608 | 531691 |
| Imputed SNPs | 4142083 | 3681871 | 5892223 |
| AD Cases | 857 | 799 | 662 |
| Controls | 551 | 778 | 408 |
| Mapped genes | 813 | 1496 | 569 |

**Table 2**
Enriched annotation terms replicated in three GWAS datasets.

| Annotation term | Source[d] | Number of genes included in each annotation term[a] | *P*-values[a,b] | BH-FDR[a,c] |
|---|---|---|---|---|
| "Alternative splicing" | SP_PIR_KEYWORDS | 269/439/191 | 1.34E-05/3.11E-08/7.92E-05 | 0.006/1.68E-05/0.009 |
| "Splice variant" | UP_SEQ_FEATURE | 268/439/190 | 2.02E-05/3.470E-08/1.158E-04 | 0.034/8.98E-05/0.045 |
| "IPR008957:Fibronectin type III-like fold" | INTERPRO | 16/23/8 | 4.04E-04/8.18E-05/0.018 | 0.166/0.102/0.817 |
| "GO:0007155~cell adhesion" | GOTERM_BP_FAT | 32/57/0 | 0.009/3.10E-04/- | 0.998/0.091/- |
| "GO:0006928~cell motion" | GOTERM_BP_FAT | 22/41/18 | 0.029/8.27E-04/0.021 | 0.983/0.156/0.911 |
| "Domain:Ig-like-C2-type 3" | UP_SEQ_FEATURE | 11/15/3 | 0.003/0.002/0.003 | 0.499/0.512/0.099 |

Note:
[a] Values ordered: TGEN_ND/GenADA/NIA_LOAD.
[b] Defined as nominally significant for *P*-value < 0.05.
[c] Defined as nominally significant for Benjamini-Hochberg-False Discovery Rate (BH-FDR) ≤ 0.25.
[d] As listed in DAVID Knowledgebase: http://david.abcc.ncifcrf.gov/helps/knowledgebase/DAVID_gene.html#assign.

Reelin and *DAB1* are coexpressed in mature human cortical neurons [28], and studies using cultured cells or genetically engineered mice provide support for contributions of *DAB1* to AD. For example, *DAB1* is coexpressed with amyloid-beta precursor protein (APP) in hippocampal neurons and binds to the cytoplasmic C-terminal domain of APP in transfected cells [29]. Over-expression of *DAB1* in cultured COS7 cells stimulates the translocation of APP to the cell surface, facilitating α-cleavage of APP and the reduction of toxic $A\beta_{40-42}$ levels [27,30]. By contrast, primary cultures of cortical neurons derived from *dab1* (−/−) homozygous knockout mice secrete significantly more $A\beta_{40-42}$, compared to *dab1* (−/+) heterozygous controls [20,27]. In addition, tyrosine phosphorylation of *DAB1* downstream of reelin-mediated activation VLDLR and APOER2, an event required for DAB1 activation [20], is inversely correlated with phosphorylation of the microtubule binding protein tau [31], and deletion of *dab1* in mice increases levels of hyperphosphorylated tau in a strain-dependent manner [31].

Based on these observations, we hypothesized that common regulatory variants of *DAB1* may contribute to AD, with low-expression increasing AD liability and high-expression providing protection. To test this hypothesis we examined possible correlations between *DAB1* mRNA expression in human brain and AD risk.

Stepwise linear regression analysis of mRNA expression vs genotypes of 16 *DAB1* "index" SNPs (selected by "$R^2$-$\Delta^2$" analysis and shown in Fig. S1), identified 5 SNPs within or near the *DAB1* gene that together account for up to 27% (adjusted $R^2 = 0.272$) of the variance in *DAB1* mRNA expression in human dorsolateral prefrontal cortex (Fig 1A). To investigate possible associations between *DAB1* mRNA expression and AD risk, we used weighted linear regression to analyze correlations between predicted brain *DAB1* mRNA levels (calculated using regression equations from the above analyses) and $\log_{10}$ORs for specific combinations of genotypes observed among AD cases and controls. For the GenADA GWAS, correlations between predicted *DAB1* mRNA expression and $\log_{10}$-ORs for individual genotype combinations revealed that higher predicted *DAB1* mRNA expression correlated with lower AD risk, with nominally significant correlations observed for genotype combinations comprising two, three, four or five SNPs (Fig. 1B).

By contrast, statistically significant correlations between *DAB1* mRNA expression and AD risk were not observed for the TGEN and NIA_LOAD GWAS (data not shown). Analysis of possible genetic differences between the three GWAS populations by multi-dimensional scaling (MDS) implemented using the "–extract prune1.prune.in –genome –out ibs1" commands in PLINK, revealed differences between the GenADA population vs. TGEN_ND/NIA_-LOAD populations, which mapped more closely to each other and to the HapMap CEU sample [32], a well-studied Caucasian population (Fig. S2). Together, these results support our hypothesis that AD risk varies inversely with levels of *DAB1* mRNA expression in human brain, although the strength of this association may vary among populations.

## 4. Discussion

In this study, we used the DAVID bioinformatics webserver as a tool to explore three AD GWAS datasets. Compared with other enrichment analysis tools, DAVID has extended annotation coverage, providing not only SNP-gene-term enrichment analysis, but also clustering analysis that allows multiple genes or terms to be linked to specific biological systems and pathways. In our analysis, six annotation terms were identified as enriched in each of the three GWAS studies analyzed using the DAVID "Functional Annotation Chart" tool. Four of these terms, IPR008957: Fibronectin type III-like fold, GO: 0007155~cell adhesion, GO: 0006928~cell motion and Domain: Ig-like-C2-type 3 were also found within annotation term clusters enriched in two or more GWAS studies.

Focusing on genes linked to these four annotation terms, we identified 17 genes that were common to two or more GWAS, two of which (*AGT* and *COL11A1*) are listed in the AlzGene database as candidate AD susceptibility genes. A search of the scientific literature using the NCBI Gene and PubMed databases revealed that an additional eight genes have direct or indirect links to AD. In total, 10

**Table 3**
Functional annotation clustering enrichment scores (ES) for clusters.

| Cluster (terms included) | Enrichment scores (ES) | | |
|---|---|---|---|
| | TGEN_ND | GenADA | NIA_LOAD |
| Cluster A ("Alternative splicing," "splice variants") | 0.251 | 0.112 | 0.944 |
| Cluster B ("IPR008957: Fibronectin type III-like fold") | **2.85** | **3.003** | 1.273 |
| Cluster C ("GO: 0007155~cell adhesion") | **1.737** | **2.593** | 0.570 |
| Cluster D ("GO: 0006928~cell motion") | **1.638** | **3.941** | 1.130 |
| Cluster E ("Domain: Ig-like C2-type 3") | **1.926** | **1.418** | **1.369** |

Note: nominally significant ES indicated by **bold type**.

**Table 4**
Genes shared among two or more GWAS datasets.

| IPR008957:Fibronectin type III-like fold" | "GO:0007155 ~cell adhesion" | "GO:0006928 ~cell motion" | "Domain: Ig-like C2-type 3" |
|---|---|---|---|
| ***LEPR**** | ***AGT**** | ***BARHL2**** | *HMCN1** |
| ***TNR**** | ***AJAP1*** | ***DAB1**** | ***KIRREL*** |
| | *CD34* | ***EDN2*** | *PTPRF** |
| | ***COL11A1**** | ***TNR**** | *PXDN** |
| | ***CSF3R**** | ***VAV3*** | |
| | ***DAB1**** | | |
| | ***NPHP4*** | | |
| | *TNN* | | |
| | ***TNR**** | | |

Notes: 1) Regular type = genes shared among two GWAS data sets; **Bold type** = genes shared among three data sets.
2) * Previously identified as an AD liability gene (AlzGene database).
3) **Directly or indirectly linked to AD based on search of scientific literature.

of the 17 genes identified using DAVID have plausible links to AD, a significant enrichment of candidate AD genes. Biological systems or processes that have been linked to one or more of the above 10 genes and to AD include: neurodegeneration, macular degeneration, reelin signaling, obesity, neurodevelopment, insulin resistance, and extracellular matrix (Table S4).

Among those 10 genes, *DAB1* functions within the reelin signaling pathway, which as mentioned above has been proposed to play a protective role in AD. Consistent with these reports, our study showed that AD risk varies inversely with levels of *DAB1* mRNA expression in human brain (Fig. 1). To our knowledge, these observations provide the first genetic evidence supporting *DAB1* as a candidate AD protection gene. Interestingly, the strength of this association varied among GWAS populations. The hypothesis that *DAB1* functions as a protective gene in a population-dependent manner is supported by multidimensional scaling (MDS) analysis, which revealed distinct genetic differences between GenADA population and the TGEN_ND/NIA_LOAD/CEU populations (Fig. S2). The observation that the effects of *DAB1* expression on levels of tau-phosphorylation in mice is strain dependent [31] and the identification several genetic loci in mice that modify the effects of *DAB1* deletions [31,33], suggest that genetic differences among human populations may also strengthen or weaken the contributions of *DAB1* to AD liability or protection.

Although our observations need to be confirmed in larger AD case–control studies based on different populations, the results

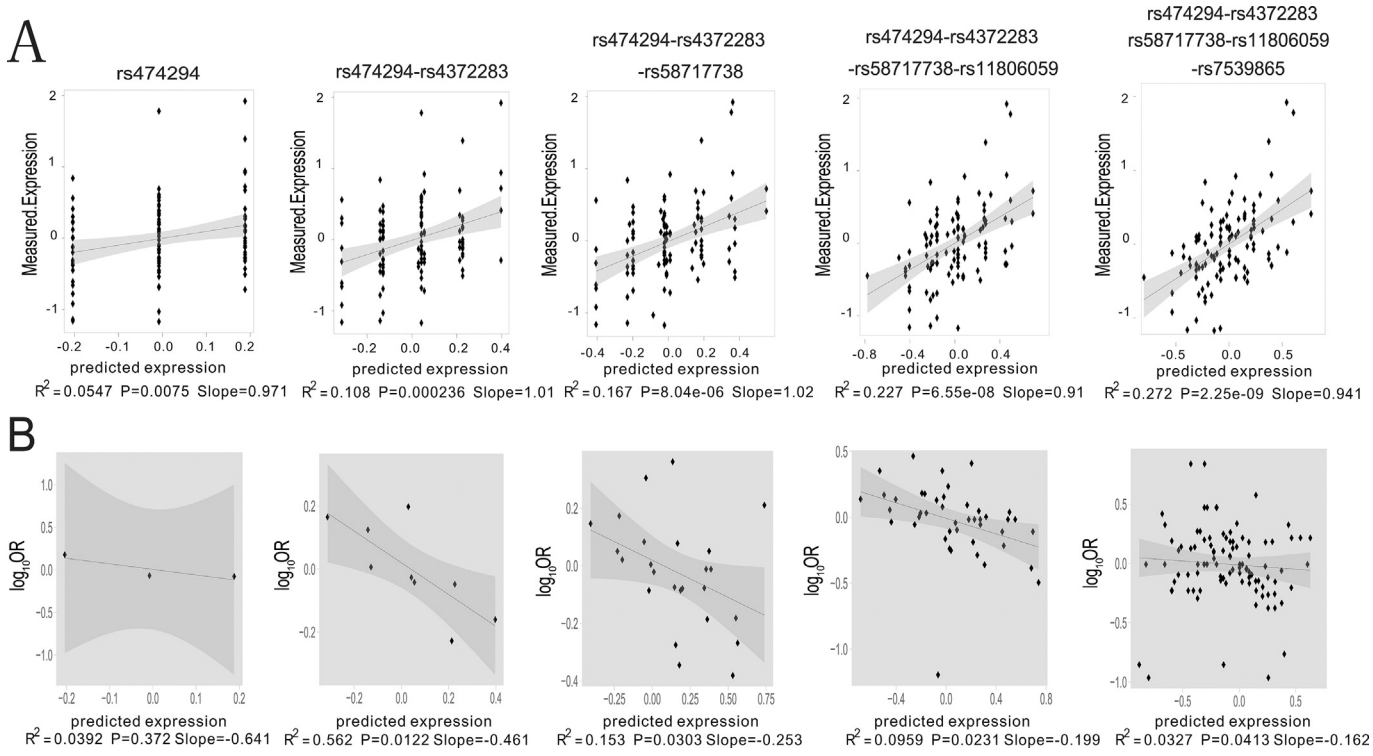


**Fig. 1.** Analysis of multiple SNPs reveals correlations between *DAB1* mRNA expression and AD risk. **A**. Plots of measured normalized *DAB1* mRNA levels in DLPFC (Y-axis) vs *DAB1* mRNA levels predicted from SNP genotypes (X-axis) based on regression equations derived from stepwise multiple linear regression analysis of normalized mRNA expression vs. SNP genotypes. Each point in the plots represents one individual (n = 109). **B**. Plots of $\log_{10}$-odds ratios (OR: AD case vs controls) for genotypes of a single SNP (left) or combinations of multiple SNPs (middle, right) (Y-axis) vs. predicted *DAB1* mRNA expression for each genotype or genotype combination (X-axis). Each point represents one genotype (left) or genotype combination (middle, right). The adjusted coefficients of determination ($R^2$) and P-values obtained from the weighted linear regression analysis of the data shown are listed above the respective plots. Shaded regions indicate the 95% confidence intervals (CI) for regression lines.

reported here provide, to the best of our knowledge, the first genetic evidence for contributions of *DAB1* to AD and suggest that DAVID-based annotation enrichment analyses combined with the analysis of correlations between mRNA expression and ORs for disorder risk are effective tools for identifying novel candidate disorder genes.

## Conflict of interest

None.

## Acknowledgements

This work was supported by a grant from the National Natural Science Foundation of China (81070908). We are also grateful for access to phenotype and genotype data from all of the datasets used in this study.

## Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.bbrc.2015.05.044.

## Transparency document

Transparency document related to this article can be found online at http://dx.doi.org/10.1016/j.bbrc.2015.05.044.

## References

[1] M. Wortmann, Dementia: a global health priority – highlights from an ADI and World Health Organization report, Alzheimers Res. Ther. 4 (2012) 40.
[2] V. Escott-Price, C. Bellenguez, L.S. Wang, et al., Gene-wide analysis detects two new susceptibility genes for Alzheimer's disease, PLoS One 9 (2014) e94661.
[3] G. Jun, C.A. Ibrahim-Verbaas, M. Vronskaya, et al., A novel Alzheimer disease locus located near the gene encoding tau protein, 2015.
[4] G. Liu, Y. Jiang, P. Wang, et al., Cell adhesion molecules contribute to Alzheimer's disease: multiple pathway analyses of two genome-wide association studies, J. Neurochem. 120 (2012) 190–198.
[5] Y.H. Lee, G.G. Song, Genome-wide pathway analysis of a genome-wide association study on Alzheimer's disease, Neurol. Sci. 36 (2015) 53–59.
[6] J.C. Lambert, K. Sleegers, A. Gonzalez-Perez, et al., The CALHM1 P86L polymorphism is a genetic modifier of age at onset in Alzheimer's disease: a meta-analysis study, J. Alzheimers Dis. 22 (2010) 247–255.
[7] Convergent genetic and expression data implicate immunity in Alzheimer's disease, Alzheimers Dement. (2014) 1552–5260.
[8] C.A. Reynolds, M.G. Hong, U.K. Eriksson, et al., Analysis of lipid pathway genes indicates association of sequence variation near SREBF1/TOM1L2/ATPAF2 with dementia risk, Hum. Mol. Genet. 19 (2010) 2068–2078.
[9] G. Liu, L. Yao, J. Liu, et al., Cardiovascular disease contributes to Alzheimer's disease: evidence from large-scale genome-wide association studies, Neurobiol. Aging 35 (2014) 786–792.
[10] E. Perez-Palma, B.I. Bustos, C.F. Villaman, et al., Overrepresentation of glutamate signaling in Alzheimer's disease: network-based pathway enrichment using meta-analysis of genome-wide association studies, PLoS One 9 (2014) e95413.
[11] W. Huang da, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, Nat. Protoc. 4 (2009) 44–57.
[12] B. Howie, J. Marchini, M. Stephens, Genotype imputation with thousands of genomes, G3 (Bethesda) 1 (2011) 457–470.
[13] O. Delaneau, J. Marchini, J.F. Zagury, A linear complexity phasing method for thousands of genomes, Nat. Methods 9 (2012) 179–181.
[14] S. Purcell, B. Neale, K. Todd-Brown, et al., PLINK: a tool set for whole-genome association and population-based linkage analyses, Am. J. Hum. Genet. 81 (2007) 559–575.
[15] D.W. Huang, B.T. Sherman, Q. Tan, et al., DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists, Nucleic Acids Res. 35 (2007) W169–W175.
[16] L. Bertram, M.B. McQueen, K. Mullin, et al., Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database, Nat. Genet. 39 (2007) 17–23.
[17] Q. Long, Q. Zhang, J. Ott, Detecting disease-associated genotype patterns, BMC Bioinforma. 10 (Suppl. 1) (2009) S75.
[18] G. Bu, Apolipoprotein E and its receptors in Alzheimer's disease: pathways, pathogenesis and therapy, Nat. Rev. Neurosci. 10 (2009) 333–344.
[19] D.S. Rice, M. Sheldon, G. D'Arcangelo, et al., Disabled-1 acts downstream of Reelin in a signaling pathway that controls laminar organization in the mammalian brain, Development 125 (1998) 3719–3729.
[20] T. Hiesberger, M. Trommsdorff, B.W. Howell, et al., Direct binding of Reelin to VLDL receptor and ApoE receptor 2 induces tyrosine phosphorylation of disabled-1 and modulates tau phosphorylation, Neuron 24 (1999) 481–489.
[21] F. Tissir, A.M. Goffinet, Reelin and brain development, Nat. Rev. Neurosci. 4 (2003) 496–505.
[22] J. Trotter, G.H. Lee, T.M. Kazdoba, et al., Dab1 is required for synaptic plasticity and associative learning, J. Neurosci. 33 (2013) 15652–15668.
[23] J. Herz, Y. Chen, Reelin, lipoprotein receptors and synaptic plasticity, Nat. Rev. Neurosci. 7 (2006) 850–859.
[24] D. Seripa, M.G. Matera, M. Franceschi, et al., The RELN locus in Alzheimer's disease, J. Alzheimers Dis. 14 (2008) 335–344.
[25] T.D. Folsom, S.H. Fatemi, The involvement of Reelin in neurodevelopmental disorders, Neuropharmacology 68 (2013) 122–135.
[26] L. Pujadas, D. Rossi, R. Andres, et al., Reelin delays amyloid-beta fibril formation and rescues cognitive deficits in a model of Alzheimer's disease, Nat. Commun. 5 (2014) 3443.
[27] H.S. Hoe, T.S. Tran, Y. Matsuoka, et al., DAB1 and Reelin effects on amyloid precursor protein and ApoE receptor 2 trafficking and processing, J. Biol. Chem. 281 (2006) 35176–35185.
[28] K. Deguchi, K. Inoue, W.E. Avila, et al., Reelin and disabled-1 expression in developing and mature human cortical neurons, J. Neuropathol. Exp. Neurol. 62 (2003) 676–684.
[29] B.W. Howell, L.M. Lanier, R. Frank, et al., The disabled 1 phosphotyrosine-binding domain binds to the internalization signals of transmembrane glycoproteins and to phospholipids, Mol. Cell Biol. 19 (1999) 5179–5188.
[30] H.S. Hoe, S.S. Minami, A. Makarova, et al., Fyn modulation of Dab1 effects on amyloid precursor protein and ApoE receptor 2 processing, J. Biol. Chem. 283 (2008) 6288–6299.
[31] J. Brich, F.S. Shie, B.W. Howell, et al., Genetic modulation of tau phosphorylation in the mouse, J. Neurosci. 23 (2003) 187–192.
[32] The International HapMap Project, Nature 426 (2003) 789–796.
[33] T. Matsuki, M. Zaka, R. Guerreiro, et al., Identification of Stk25 as a genetic modifier of Tau phosphorylation in Dab1-mutant mice, PLoS One 7 (2012) e31152.