

# Integrating Genome-Wide Association Study and Brain Expression Data Highlights Cell Adhesion Molecules and Purine Metabolism in Alzheimer's Disease

Zimin Xiang · Meiling Xu · Mingzhi Liao · Yongshuai Jiang · Qinghua Jiang ·  
Rennan Feng · Liangcai Zhang · Guoda Ma · Guangyu Wang · Zugen Chen · Bin Zhao ·  
Tiansheng Sun · Keshen Li · Guiyou Liu

Received: 29 June 2014 / Accepted: 27 August 2014 / Published online: 10 September 2014  
© Springer Science+Business Media New York 2014

**Abstract** Alzheimer's disease (AD) is the most common neurodegenerative disease in the elderly. Recently, genome-wide association studies (GWAS) have been used to investigate AD pathogenesis. However, a large proportion of AD heritability has yet to be explained. We previously identified the cell adhesion molecule (CAM) pathway as a consistent signal in two AD GWAS. However, it is unclear whether CAM is present in the Genetic and Environmental Risk for Alzheimer's Disease Consortium (GERAD) GWAS and brain expression GWAS. Meanwhile, we think integrating AD GWAS and AD brain expression datasets may provide

complementary information to identify important pathways involved in AD. Here, we conducted a systems analysis using (1) KEGG pathways, (2) large-scale AD GWAS from GERAD ( $n=11,789$ ), (3) two brain expression GWAS datasets ( $n=399$ ) from the AD cerebellum and temporal cortex, and (4) previous results from pathway analysis of AD GWAS. Our results indicate that (1) CAM is a consistent signal in five AD GWAS; (2) CAM is the most significant signal in AD; (3) we confirmed previous AD risk pathways related to immune system and diseases, and cardiovascular disease, etc.; and (4) we highlighted the purine metabolism

Zimin Xiang, Meiling Xu, and Mingzhi Liao contributed equally to this work.

Z. Xiang · T. Sun (✉)  
Department of Orthopedic Surgery, Beijing Army General Hospital,  
Clinical Medicine College of Third Military Medical University, No.  
5 Nanmencang, Dongcheng District, Beijing 100700, China  
e-mail: dr\_suntiansheng@163.com

M. Xu  
Department of Neurology, The Fourth Affiliated Hospital of Harbin  
Medical University, Harbin, Heilongjiang, China

M. Liao  
College of Life Science, Northwest A&F University, Yangling,  
Shaanxi, China

Y. Jiang  
College of Bioinformatics Science and Technology, Harbin Medical  
University, Harbin, China

Q. Jiang  
School of Life Science and Technology, Harbin Institute of  
Technology, Harbin, China

R. Feng  
Department of Nutrition and Food Hygiene, School of Public Health,  
Harbin Medical University, Harbin, China

L. Zhang  
Department of Statistics, Rice University, Houston, TX, USA

G. Ma · B. Zhao · K. Li (✉)  
Institute of Neurology, Guangdong Medical College,  
Zhanjiang 524001, China  
e-mail: keshenli2012@163.com

Z. Chen  
Department of Human Genetics, University of California at Los  
Angeles, Los Angeles, CA, USA

G. Wang  
Department of Oncology, The First Hospital of Harbin, Harbin,  
China

G. Liu (✉)  
Genome Analysis Laboratory, Tianjin Institute of Industrial  
Biotechnology, Chinese Academy of Sciences, Xiqi Dao 32, Tianjin  
Airport Economic Area, Tianjin 300308, China  
e-mail: liuguiyou1981@163.com

pathway in AD for the first time. We believe that our results may advance our understanding of AD mechanisms and will be very informative for future genetic studies in AD.

**Keywords** Genome-wide association study · Alzheimer’s disease · Brain expression · Pathway analysis

## Introduction

Alzheimer’s disease (AD) is the most common neurodegenerative disease in the elderly [1]. Recently, genome-wide association studies (GWAS) have been used to investigate AD pathogenesis. GWAS have yielded important new insights into the genetic mechanisms of AD [2]. However, newly identified AD susceptibility loci exert very small risk effects and cannot fully explain the underlying genetic risk [3]. A large proportion of AD heritability has yet to be explained. Fortunately, the existing large-scale GWAS datasets provide strong support for the investigation of AD mechanisms using pathway analysis methods [4–9].

Lambert et al. [4] used two different pathway analysis tools to analyze a large GWAS in a French population (French GWAS). They identified significant pathways related to the immune system [4]. Jones et al. [6] performed multiple pathway analyses of two large GWAS datasets including the French GWAS and the Genetic and Environmental Risk for Alzheimer’s Disease Consortium (GERAD) GWAS. They observed significantly enriched pathways related to metabolism and the immune system [6]. We performed multiple pathway analyses of two publicly available AD GWAS datasets (French and Pfizer GWAS datasets) [7]. We identified cell adhesion molecules (CAM) as a consistent signal in AD. Recently, Ramanan et al. [8] performed a pathway enrichment analysis on GWAS data from the 742 Alzheimer’s Disease Neuroimaging Initiative (ADNI) participants. They confirmed the involvement of CAM in AD ( $P=5.60E-04$ ) [8].

Based on the above findings, we think that CAM may be a consistent signal in multiple AD GWAS. However, it was unclear whether CAM involvement was present in the GERAD GWAS dataset. Evidence also shows that genetic variants that modify gene expression in the brain may influence AD risk [10]. In order to verify CAM to be a consistent signal in multiple AD GWAS, we conducted a pathway analysis of the GERAD GWAS dataset and brain expression GWAS dataset. Meanwhile, we think that integrating the AD GWAS and AD brain expression datasets may provide complementary information to identify important pathways involved in AD. Here, we conducted a systems analysis using (1) KEGG pathways, (2) large-scale AD GWAS from GERAD ( $n=11,789$ ), (3) two brain expression GWAS datasets ( $n=399$ ) using quantitative expression trait loci from

the human cerebellum and temporal cortex, and (4) previous results from pathway analysis of AD GWAS.

## Materials and Methods

### AD GWAS Dataset

The GERAD GWAS dataset included 11,789 samples from individuals of European ancestry (3941 AD cases and 7848 controls) [11]. A total of 529,205 autosomal single nucleotide polymorphisms (SNPs) passed quality control checks. SNPs were tested for association with AD using logistic regression under an additive model. Here, we selected 761 significant SNPs with  $P\leq 1.00E-03$ . For more detailed information, please refer to the original study [11].

### Two AD Brain Expression GWAS Datasets

The brain expression GWAS datasets were originally analyzed by Zou et al. [10]. A total of 773 brain samples from the cerebellum and temporal cortex were available for analysis. The samples were divided into four datasets: 177 non-AD cerebellar samples, 197 non-AD temporal cortex samples, 197 AD cerebellar case samples, and 202 AD temporal cortex case samples. Zou et al. analyzed 213,528 *cis*SNPs within  $\pm 100$  kb of the 24,526 tested transcripts. Levels of 24,526 transcripts for 18,401 genes were measured using WG-DASL assays. False discovery rate (FDR)-based  $P$  values were used to correct for multiple testing. More detailed results are described in Supplementary Tables [10].

### Previous Results from Pathway Analysis of AD GWAS

Prior to this study, six pathway analyses of AD GWAS have been reported with five studies using the KEGG database [4–9]. We compared our findings with the previous pathway analyses of AD GWAS. All of the pathway analysis results were publicly available from the original studies [4–9].

### AD GWAS Dataset Preprocessing

ProxyGeneLD was used to assign SNPs to specific genes [12]. This software flexibly takes into consideration the complex linkage disequilibrium (LD) patterns in the human genome and corrects for the inflation of significance caused by gene length. In brief, ProxyGeneLD begins with the retrieval of LD structures in the HapMap genotyping data (Utah residents with ancestry from northern and western Europe (CEU) samples of HapMap phase II, release 22) [12]. If a group of markers is in high LD in HapMap ( $r^2>0.8$ ), they are tied to a “proxy cluster” and taken as a single signal [5]. Next, each marker in the AD GWAS with statistically significant

evidence of association is evaluated to see whether (a) it belongs to any proxy cluster and (b) whether the marker itself or any marker in the cluster is located in a genetic region [5]. If a marker or cluster overlaps a region extending across a gene, then it is assigned as showing possible association with that gene. For more detailed algorithms, please refer to the original study [12].

#### AD Brain Expression GWAS Datasets Preprocessing

Here, we selected the 197 AD cerebellar case samples and 202 AD temporal cortex case samples from Zou et al. [10]. For each dataset, we selected *cis*SNPs associated with gene expression with  $P < 1.0E-4$  and with AD risk with  $P < 1.0E-3$  in the meta-analysis of the Alzheimer's Disease Genetic Consortium (ADGC) dataset [10]. In the end, we selected 3660 *cis*-SNPs associated with the expression of 298 genes in the cerebellum and 3659 *cis*SNPs associated with the expression of 298 genes in the temporal cortex. More detailed results are described in Supplementary Tables from the original study [10].

#### Pathway-Based Testing for AD Genes

The KEGG pathways in WebGestalt were used in our research (May 10, 2014) [13]. For a given pathway, a hypergeometric test was used to detect the overrepresentation of AD-related genes among all of the genes in the pathway [13]. The  $P$  value of  $K$  AD-related genes in the pathway was calculated using

$$P = 1 - \sum_{i=0}^K \frac{\binom{S}{i} \binom{N-S}{m-i}}{\binom{N}{m}}$$

where  $N$  is the total number of genes of interest,  $S$  is the number of all of the AD-related genes,  $m$  is the number of genes in the pathway, and  $K$  is the number of AD-related genes in the pathway. The FDR method was used to correct for multiple testing. Any pathway with an adjusted  $P < 0.05$  and at least five upregulated or downregulated AD genes was considered significant. In order to reduce the multiple-testing issue and to avoid testing overly narrow or broad pathways, we selected pathways that contained at least 20 and at most 300 genes for subsequent analysis.

#### Pathway-Based Meta-analysis of AD GWAS and Brain Expression GWAS

We used Fisher's method to combine the  $P$  values for each pathway identified in AD GWAS and AD brain expression

GWAS [14]. For a given pathway, the formula for the statistic is

$$x^2 = -2 \sum_{i=1}^k \ln(p_i)$$

where  $p_i$  is the  $P$  value of the pathway in the  $i$ th study and  $k$  is the total number of studies.  $x^2$  follows a chi-square distribution with  $2k$  degrees of freedom [14]. The pathway-based meta-analysis was carried out using the program R (<http://www.r-project.org/>).

## Results

#### Pathway Analysis of GERAD GWAS

Using the AD GWAS, we found 320 AD genes using ProxyGeneLD. After FDR correction for multiple testing, we identified 14 significant KEGG pathways ( $P < 0.05$ ) with at least five AD genes, among which CAM (hsa04514) was the most significant pathway (Table 1). Nine of the 133 genes in the CAM pathway were identified. We also observed a significant overrepresentation of KEGG pathways related to metabolism (metabolic pathways and purine metabolism), cardiovascular diseases (dilated cardiomyopathy, arrhythmogenic right ventricular cardiomyopathy (ARVC), and hypertrophic cardiomyopathy (HCM)), and neurological disorders (Alzheimer's disease and Huntington's disease) (Table 1).

#### Pathway Analysis of Brain Expression GWAS Datasets

We got 425 AD genes regulated by *cis*SNPs in the cerebellar or temporal cortex. After FDR correction, we identified 20 significant KEGG pathways ( $P < 0.05$ ), which included at least five AD genes (Table 2). We found a significant overrepresentation of KEGG pathways related to the immune system and diseases, neurodegenerative diseases and cardiovascular diseases, metabolism, and genetic and environmental information processing. The antigen processing and presentation and CAM pathways were the most and the second significant signals (Table 2).

#### Pathway-Based Meta-analysis of AD GWAS and Brain Expression GWAS

We combined the findings from the AD GWAS and the AD brain expression datasets. Using pathway-based meta-analysis and FDR correction for multiple testing, we identified CAM (hsa04514) and antigen processing and presentation to be the

**Table 1** Significant KEGG pathways with  $P < 0.05$  in the AD GWAS dataset

Pathway ID	Pathway name	C	O	E	R	rawP	adjP	Gene ID
hsa04514	Cell adhesion molecules (CAMs)	133	9	0.99	9.12	7.41E-07	1.04E-05	3122 1002 9369 26285 3118 5817 5819 1462 4756
hsa05414	Dilated cardiomyopathy	90	6	0.67	8.98	5.87E-05	3.00E-04	6262 22801 108 3674 10369 6444
hsa05010	Alzheimer's disease	167	8	1.24	6.46	3.84E-05	3.00E-04	348 4722 2776 6868 3709 841 25825 1340
hsa05416	Viral myocarditis	70	5	0.52	9.63	2.00E-04	4.00E-04	3122 3118 841 4626 6444
hsa05412	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	74	5	0.55	9.11	2.00E-04	4.00E-04	6262 22801 3674 10369 6444
hsa04080	Neuroactive ligand-receptor interaction	272	9	2.02	4.46	2.00E-04	4.00E-04	2890 9340 1138 2898 9934 53829 8001 1136 1143
hsa00230	Purine metabolism	162	7	1.2	5.82	2.00E-04	4.00E-04	5144 29922 2977 5315 5152 108 5142
hsa05410	Hypertrophic cardiomyopathy (HCM)	83	5	0.62	8.12	4.00E-04	7.00E-04	6262 22801 3674 10369 6444
hsa04270	Vascular smooth muscle contraction	116	5	0.86	5.81	1.80E-03	2.50E-03	5581 2776 3709 2977 108
hsa04020	Calcium signaling pathway	177	6	1.31	4.57	2.20E-03	2.80E-03	6262 814 2185 2776 3709 108
hsa04360	Axon guidance	129	5	0.96	5.22	2.80E-03	3.30E-03	8829 57144 10725 84448 57522
hsa05016	Huntington's disease	183	5	1.36	3.68	1.21E-02	1.30E-02	23186 4722 2776 841 1340
hsa04060	Cytokine-cytokine receptor interaction	265	6	1.97	3.05	1.47E-02	1.47E-02	3626 8809 91 3556 8807 83729

*C* the number of reference genes in the category, *O* the number of genes in the gene set and also in the category, *E* expected number in the category, *R* the ratio of enrichment, *rawP* the *P* value from hypergeometric test, *adjP* the *P* value adjusted by the multiple test adjustment

most and the second significant signals (Table 3). Here, we list the top 20 significant pathways (Table 3). Based on the KEGG classifications, these pathways can be divided into six main classes: immune system and diseases ( $n=8$ ), cardiovascular disease ( $n=4$ ), environmental information processing ( $n=2$ ), metabolism ( $n=2$ ), neurodegenerative system and disease ( $n=2$ ), cellular processes ( $n=1$ ), and infectious diseases ( $n=1$ ) (Table 3).

#### Comparison with Previous Studies Using Single GERAD GWAS and Other GWAS

Jones et al. [6] analyzed the GERAD GWAS and reported six immune system-related pathways including asthma (hsa05310), hematopoietic cell lineage (hsa04640), graft-versus-host disease (hsa05332), allograft rejection (hsa05330), autoimmune thyroid disease (hsa05320), and type I diabetes mellitus (hsa04940). Here, we confirmed these six pathways, which were all ranked to be top 20 significant signals (Table 3). Meanwhile, we identified another two immune pathways including antigen processing and presentation (hsa04612) and systemic lupus erythematosus (hsa05322). Previous pathway analysis using French AD GWAS reported weak association between antigen processing and presentation (hsa04612) and AD with  $P=0.02$  [4]. Here, we identified it to be the second and significant signals with  $P=1.87E-07$  (Table 3).

We previously integrated three previous large-scale AD GWAS including GERAD GWAS using a gene-based meta-analysis and subsequently conducted a pathway analysis [9]. We highlighted, for the first time, the involvement of cardiovascular disease-related pathways in AD [9]. There are four pathways related to cardiovascular disease in the KEGG database including viral myocarditis (hsa05416), dilated cardiomyopathy (hsa05414), hypertrophic cardiomyopathy (hsa05410), and arrhythmogenic right ventricular cardiomyopathy (hsa05412). In our research, we identified all the four pathways to be significantly associated with AD, among which viral myocarditis was the third significant signal with  $P=7.79E-07$  (Table 3). After careful comparison with previous pathway analyses of AD GWAS [4–9], we reported, for the first time, the involvement of the purine metabolism (hsa00230) pathway in AD, which is the eighth significant signal in Table 3. More detailed comparison results are described in Table 3.

#### Discussion

Recently, multiple AD GWAS have been conducted. In this research, we consider CAM to be a consistent signal in multiple AD GWAS. We analyzed the GERAD GWAS and AD brain expression GWAS. We identified the involvement of CAM in AD by pathway analysis of the AD GWAS and AD brain expression GWAS. Based on previous and our results

**Table 2** Significant KEGG pathways with  $P < 0.05$  in AD brain expression GWAS datasets

Pathway ID	Pathway name	C	O	E	R	rawP	adjP	Gene ID
hsa04612	Antigen processing and presentation	76	10	0.75	13.35	4.16E-09	9.98E-08	3821 3122 3308 3119 3823 6891 3107 3112 3822 3127
hsa04514	Cell adhesion molecules (CAMs)	133	11	1.31	8.39	9.63E-08	1.16E-06	3122 80380 5819 3119 923 3385 29126 3107 3112 3127 1366
hsa05332	Graft-versus-host disease	41	6	0.4	14.85	2.97E-06	2.38E-05	3821 3122 3107 3119 3112 3127
hsa05416	Viral myocarditis	70	7	0.69	10.15	6.03E-06	3.62E-05	3122 3107 8735 637 3119 3112 3127
hsa04145	Phagosome	153	9	1.51	5.97	2.30E-05	1.00E-04	3122 23480 3119 6891 3107 78989 1311 3112 3127
hsa05330	Allograft rejection	37	5	0.36	13.71	3.05E-05	1.00E-04	3122 3107 3119 3112 3127
hsa04940	Type I diabetes mellitus	43	5	0.42	11.8	6.42E-05	2.00E-04	3122 3107 3119 3112 3127
hsa05322	Systemic lupus erythematosus	136	8	1.34	5.97	6.54E-05	2.00E-04	3122 8361 3119 8364 8343 89 3127 3112
hsa05320	Autoimmune thyroid disease	52	5	0.51	9.76	2.00E-04	5.00E-04	3122 3107 3119 3112 3127
hsa05010	Alzheimer's disease	167	8	1.65	4.86	3.00E-04	7.00E-04	6390 840 4707 637 517 4137 4311 2081
hsa03018	RNA degradation	71	5	0.7	7.15	7.00E-04	1.30E-03	22803 118460 1656 167227 2027
hsa05140	Leishmaniasis	72	5	0.71	7.05	7.00E-04	1.30E-03	3122 3119 3112 1378 3127
hsa04640	Hematopoietic cell lineage	88	5	0.87	5.77	1.80E-03	3.10E-03	3122 952 1378 3127 4311
hsa04650	Natural killer cell mediated cytotoxicity	136	6	1.34	4.48	2.40E-03	3.60E-03	3821 10451 3107 637 3823 3822
hsa05016	Huntington's disease	183	7	1.8	3.88	2.40E-03	3.60E-03	1742 6390 84699 4707 5441 7802 517
hsa00240	Pyrimidine metabolism	99	5	0.98	5.13	3.00E-03	4.20E-03	5441 50484 51082 30834 5557
hsa04916	Melanogenesis	101	5	1	5.02	3.30E-03	4.40E-03	109 7479 84699 1638 8325
hsa00230	Purine metabolism	162	6	1.6	3.76	5.60E-03	7.10E-03	109 5441 50484 51082 30834 5557
hsa04510	Focal adhesion	200	6	1.97	3.04	1.48E-02	1.78E-02	10451 50509 5906 1290 1311 89
hsa04020	Calcium signaling pathway	177	5	1.74	2.87	3.15E-02	3.60E-02	109 952 2769 777 148

*C* the number of reference genes in the category, *O* the number of genes in the gene set and also in the category, *E* expected number in the category, *R* the ratio of enrichment, *rawP* the *P* value from hypergeometric test, *adjP* the *P* value adjusted by the multiple test adjustment

from pathway analysis of AD GWAS, CAM is a consistent signal in the French [4], Pfizer [4], ADNI [8], GERAD [11], and AD brain expression GWAS [10]. Meanwhile, we combined the findings from the AD GWAS and the AD brain expression GWAS datasets by a pathway-based meta-analysis method. We further identified CAM (hsa04514) to be the most significant signal (Table 3). In addition to the CAM, we identified significant pathways related to immune system and diseases, cardiovascular disease, metabolism, neurodegenerative system and disease, cellular processes, and infectious diseases. We confirmed previous findings and highlighted the purine metabolism pathway (hsa00230) in AD for the first time, which is the eighth significant signal in Table 3.

Until now, the exact pathogenetic role of purine metabolism in AD is still unknown. Kaddurah-Daouk et al. [15] used targeted metabolomics platform to profile cerebrospinal fluid from 40 AD, 36 mild cognitive impairment (MCI), and 38 control subjects. Levels of 71 metabolites, including 24 known compounds quantified by the liquid chromatography electrochemical array (LCECA) platform, were measured. There are six known compounds in purine metabolism

including guanosine, hypoxanthine, uric acid, xanthine, xanthosine, and paraxanthine. The results showed that AD subjects had elevated xanthosine versus controls [15]. MCI subjects had elevated hypoxanthine and uric acid versus controls. Metabolite ratios revealed changes of uric acid/xanthine, xanthine/hypoxanthine, and xanthine/xanthosine within the purine pathway. A partial correlation network showed total tau most directly related to purine pathway [15]. These findings indicate that AD is associated with an overlapping pattern of perturbations in the purine pathway [15].

In addition to the AD, the involvement of purine metabolism in other neurodegenerative diseases has been reported. Johansen et al. [16] compared plasma profiles from people with idiopathic Parkinson's disease (PD) and those whose disease was due to a mutation in LRRK2, as well as nonrelated control subjects. Although the two PD categories shared much in common, the profile of a dozen mostly unknown metabolites was sufficient to distinguish them. Levels and ratios of some purine metabolites, such as uric acid, hypoxanthine, and xanthine, were significantly decreased in both kinds of PD patients, when compared to control subjects [16].

**Table 3** Top 20 significant KEGG pathways by meta-analysis of AD GWAS and AD brain expression GWAS datasets

Study	Pathway ID	Pathway name	AD GWAS				AD brain expression GWAS				Meta-analysis			
			C	O	E	R	rawP	C	O	E	R	rawP	rawP	adjP
[7–9]	hsa04514	Cell adhesion molecules (CAMs)	133	9	0.99	9.12	7.41E-07	133	11	1.31	8.39	9.63E-08	2.23E-12	2.01E-10
[4, 9]	hsa04612	Antigen processing and presentation	76	N	N	N	N	76	10	0.75	13.35	4.16E-09	4.16E-09	1.87E-07
[9]	hsa05416	Viral myocarditis	70	5	0.52	9.63	2.00E-04	70	7	0.69	10.15	6.03E-06	2.60E-08	7.79E-07
[4, 9]	hsa05010	Alzheimer's disease	167	8	1.24	6.46	3.84E-05	167	8	1.65	4.86	3.00E-04	2.22E-07	5.00E-06
[6, 9]	hsa05332	Graft-versus-host disease	41	2	0.3	6.57	3.72E-02	41	6	0.4	14.85	2.97E-06	1.88E-06	3.38E-05
[6, 9]	hsa05310	Asthma	30	3	0.22	13.48	1.40E-03	30	4	0.3	13.53	2.00E-04	4.50E-06	5.79E-05
[6, 8, 9]	hsa05330	Allograft rejection	37	2	0.27	7.28	3.08E-02	37	5	0.36	13.71	3.05E-05	1.40E-05	1.57E-04
	hsa00230	Purine metabolism	162	7	1.2	5.82	2.00E-04	162	6	1.6	3.76	5.60E-03	1.65E-05	1.65E-04
[9]	hsa04145	Phagosome	153	N	N	N	N	153	9	1.51	5.97	2.30E-05	2.30E-05	2.07E-04
[6, 8, 9]	hsa04940	Type 1 diabetes mellitus	43	2	0.32	6.27	4.06E-02	43	5	0.42	11.8	6.42E-05	3.61E-05	2.96E-04
[9]	hsa05414	Dilated cardiomyopathy	90	6	0.67	8.98	5.87E-05	90	N	N	N	N	5.87E-05	4.40E-04
[9]	hsa05322	Systemic lupus erythematosus	136	N	N	N	N	136	8	1.34	5.97	6.54E-05	6.54E-05	4.53E-04
[6, 9]	hsa04640	Hematopoietic cell lineage	88	4	0.65	6.13	4.30E-03	88	5	0.87	5.77	1.80E-03	9.88E-05	6.35E-04
[9]	hsa04080	Neuroactive ligand-receptor interaction	272	9	2.02	4.46	2.00E-04	272	6	2.68	2.24	5.37E-02	1.34E-04	8.02E-04
[6, 9]	hsa05320	Autoimmune thyroid disease	52	2	0.39	5.18	5.71E-02	52	5	0.51	9.76	2.00E-04	1.41E-04	7.95E-04
[9]	hsa05140	Leishmaniasis	72	3	0.53	5.62	1.66E-02	72	5	0.71	7.05	7.00E-04	1.44E-04	7.61E-04
[8]	hsa05412	Arrhythmogenic right ventricular cardiomyopathy	74	5	0.55	9.11	2.00E-04	74	N	N	N	N	2.00E-04	1.00E-03
[9]	hsa00860	Porphyryn and chlorophyll metabolism	43	4	0.32	12.54	3.00E-04	43	2	0.42	4.72	6.72E-02	2.38E-04	1.13E-03
[9]	hsa05016	Huntington's disease	183	5	1.36	3.68	1.21E-02	183	7	1.8	3.88	2.40E-03	3.32E-04	1.50E-03
[9]	hsa05410	Hypertrophic cardiomyopathy (HCM)	83	5	0.62	8.12	4.00E-04	83	N	N	N	N	4.00E-04	1.71E-03

C the number of reference genes in the category, O the number of genes in the gene set and also in the category, E expected number in the category, R the ratio of enrichment, rawP the P value from hypergeometric test, adjP the P value adjusted by the multiple test adjustment

LeWitt et al. [17] analyzed cerebrospinal fluid (CSF) concentrations of homovanillic acid (the major catabolite of dopamine) and the purine compound xanthine for a comparison of 217 unmedicated PD subjects and 26 healthy controls. The xanthine/homovanillic acid ratio is different between PD cases and controls. The mean xanthine/homovanillic acid quotient from controls was  $13.1 \pm 5.5$  as compared to the PD value of  $17.4 \pm 6.7$  at an initial lumbar CSF collection ( $P=0.0017$ ), and  $19.7 \pm 8.7$  ( $P<0.001$ ) at a second CSF collection up to 24 months later. These observations further provide neurochemical evidence that links purine metabolism to PD [17].

Here, a hypergeometric test was used for pathway analysis. We set a high threshold ( $P \leq 1.00E-03$ ) for the inclusion of SNPs or *cis*SNPs, similar to Lambert and Jones et al. [4, 6]. Evidence suggests that using a stringent  $P$  value ( $P \leq 1.00E-03$ ) to define associated SNPs will accurately test whether highly associated SNPs are enriched in a pathway [18]. Jia et al. [19] suggested that a hypergeometric test performs better and with a higher power than gene set enrichment analysis (GSEA) or a SNP ratio test (SRT) for gene sets consisting of markers that are highly associated with the disease ( $P \leq 1.00E-03$ ).

In this study, we used the pathway-based meta-analysis method to integrate AD GWAS and expression datasets, as this method is widely used in previous studies. Arasappan et al. [20] applied a pathway-based meta-analysis into four independent gene expression datasets to identify gene expression signatures for systemic lupus erythematosus, and identified a 37-gene expression signature for systemic lupus erythematosus in human peripheral blood mononuclear cells. In order to combine the results from different analyses, Kaeffer et al. [21] introduce a methodical framework for the meta-analysis of  $P$  values obtained from pathway enrichment analysis (set enrichment analysis based on pathways) of multiple dependent or independent data sets from different omics platforms. Shen et al. [22] proposed two approaches of meta-analysis for pathway enrichment by combining statistical significance across studies at the gene level or at the pathway level. They applied these methods into real data on drug response of breast cancer cell lines and lung cancer tissues [22].

Here, we selected the pathways from the KEGG database, but not the GO database, for pathway analysis based on the following considerations. KEGG database is manually compiled on the basis of biological evidence and does not have a hierarchical structure [4, 6]. The GO database is based mainly on computer predictions as well as human annotation [4, 6]. The GO database has a hierarchical structure. GO analysis typically assumes that each functional category is independent, and less than 1 % of the GO annotations have been confirmed experimentally [4, 6].

Despite these interesting results, we recognize some limitations in our study. First, we selected expression datasets

from the human cerebellum and temporal cortex. We will further verify our findings using expression data from other human brain tissues such as the frontal cortex, hippocampus, prefrontal cortex, and visual cortex in subsequent studies. Second, for many long genes in the CAM pathway, multiple-testing corrections may not be sufficient to account for all biases. The results from the GERAD GWAS should be adjusted using a permutation test. However, the original SNP genotype data are not available to us, so future replication studies using genotype data are required to replicate our findings.

Collectively, our integrated analysis shows that (1) CAM is a consistent signal in five AD GWAS, (2) CAM is the most significant signal in AD, (3) we confirmed previous findings, and (4) we highlighted the purine metabolism pathway (hsa00230) in AD for the first time. We believe that our results may advance our understanding of AD mechanisms and will be very informative for future genetic studies in AD.

**Acknowledgments** We thank the GERAD Consortium and Zou et al. for the AD GWAS and expression datasets. This work was supported by funding from the National Nature Science Foundation of China (grant numbers 81300945, 31200934, 31301938, 81471294, 31171219, 81271213, 81271214).

**Conflict of Interest** The authors declare no conflict of interests.

## References

- Pedersen NL (2010) Reaching the limits of genome-wide significance in Alzheimer disease: back to the environment. *JAMA* 303(18):1864–1865
- Borovecki F, Klepac N, Muck-Seler D, Hajnsek S, Mubrin Z, Pivac N (2011) Unraveling the biological mechanisms in Alzheimer's disease—lessons from genomics. *Prog Neuropsychopharmacol Biol Psychiatr* 35(2):340–347
- Bertram L, Lill CM, Tanzi RE (2010) The genetics of Alzheimer disease: back to the future. *Neuron* 68(2):270–281
- Lambert JC, Grenier-Boley B, Chouraki V, Heath S, Zelenika D, Fievet N, Hannequin D, Pasquier F, Hanon O, Brice A, Epelbaum J, Berr C, Dartigues JF, Tzourio C, Campion D, Lathrop M, Amouyel P (2010) Implication of the immune system in Alzheimer's disease: evidence from genome-wide pathway analysis. *J Alzheimers Dis* 20(4):1107–1118
- Hong MG, Alexeyenko A, Lambert JC, Amouyel P, Prince JA (2010) Genome-wide pathway analysis implicates intracellular transmembrane protein transport in Alzheimer disease. *J Hum Genet* 55(10):707–709
- Jones L, Holmans PA, Hamshere ML, Harold D, Moskva V, Ivanov D, Pocklington A, Abraham R, Hollingworth P, Sims R, Gerrish A, Pahwa JS, Jones N, Stretton A, Morgan AR, Lovestone S, Powell J, Proitsi P, Lupton MK, Brayne C, Rubinsztein DC, Gill M, Lawlor B, Lynch A, Morgan K, Brown KS, Passmore PA, Craig D, McGuinness B, Todd S, Holmes C, Mann D, Smith AD, Love S, Kehoe PG, Mead S, Fox N, Rossor M, Collinge J, Maier W, Jessen F, Schurmann B, Heun R, Kolsch H, van den Bussche H, Heuser I, Peters O, Kornhuber J, Wiltfang J, Dichgans M, Frolich L, Hampel H, Hull M, Rujescu D, Goate AM, Kauwe JS, Cruchaga C, Nowotny

- P, Morris JC, Mayo K, Livingston G, Bass NJ, Gurling H, McQuillin A, Gwilliam R, Deloukas P, Al-Chalabi A, Shaw CE, Singleton AB, Guerreiro R, Muhleisen TW, Nothen MM, Moebs S, Jockel KH, Klopp N, Wichmann HE, Ruther E, Carrasquillo MM, Pankratz VS, Younkin SG, Hardy J, O'Donovan MC, Owen MJ, Williams J (2010) Genetic evidence implicates the immune system and cholesterol metabolism in the aetiology of Alzheimer's disease. *PLoS One* 5(11):e13950
7. Liu G, Jiang Y, Wang P, Feng R, Jiang N, Chen X, Song H, Chen Z (2012) Cell adhesion molecules contribute to Alzheimer's disease: multiple pathway analyses of two genome-wide association studies. *J Neurochem* 120(1):190–198
  8. Ramanan VK, Kim S, Holohan K, Shen L, Nho K, Risacher SL, Foroud TM, Mukherjee S, Crane PK, Aisen PS, Petersen RC, Weiner MW, Saykin AJ (2012) Genome-wide pathway analysis of memory impairment in the Alzheimer's disease neuroimaging initiative (ADNI) cohort implicates gene candidates, canonical pathways, and networks. *Brain Imaging Behav* 6(4):634–648
  9. Liu G, Yao L, Liu J, Jiang Y, Ma G, Chen Z, Zhao B, Li K (2014) Cardiovascular disease contributes to Alzheimer's disease: evidence from large-scale genome-wide association studies. *Neurobiol Aging* 35(4):786–792
  10. Zou F, Chai HS, Younkin CS, Allen M, Crook J, Pankratz VS, Carrasquillo MM, Rowley CN, Nair AA, Middha S, Maharjan S, Nguyen T, Ma L, Malphrus KG, Palusak R, Lincoln S, Bisceglia G, Georgescu C, Kouri N, Kolbert CP, Jen J, Haines JL, Mayeux R, Pericak-Vance MA, Farrer LA, Schellenberg GD, Petersen RC, Graff-Radford NR, Dickson DW, Younkin SG, Ertekin-Taner N (2012) Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. *PLoS Genet* 8(6):e1002707
  11. Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, Pahwa JS, Moskvina V, Dowzell K, Williams A, Jones N, Thomas C, Stretton A, Morgan AR, Lovestone S, Powell J, Proitsi P, Lupton MK, Brayne C, Rubinsztein DC, Gill M, Lawlor B, Lynch A, Morgan K, Brown KS, Passmore PA, Craig D, McGuinness B, Todd S, Holmes C, Mann D, Smith AD, Love S, Kehoe PG, Hardy J, Mead S, Fox N, Rossor M, Collinge J, Maier W, Jessen F, Schurmann B, van den Bussche H, Heuser I, Kornhuber J, Wiltfang J, Dichgans M, Frolich L, Hampel H, Hull M, Rujescu D, Goate AM, Kauwe JS, Cruchaga C, Nowotny P, Morris JC, Mayo K, Sleegers K, Bettens K, Engelborghs S, De Deyn PP, Van Broeckhoven C, Livingston G, Bass NJ, Gurling H, McQuillin A, Gwilliam R, Deloukas P, Al-Chalabi A, Shaw CE, Tsolaki M, Singleton AB, Guerreiro R, Muhleisen TW, Nothen MM, Moebs S, Jockel KH, Klopp N, Wichmann HE, Carrasquillo MM, Pankratz VS, Younkin SG, Holmans PA, O'Donovan M, Owen MJ, Williams J (2009) Genome-wide association study identifies variants at *CLU* and *PICALM* associated with Alzheimer's disease. *Nat Genet* 41(10):1088–1093
  12. Hong MG, Pawitan Y, Magnusson PK, Prince JA (2009) Strategies and issues in the detection of pathway enrichment in genome-wide association studies. *Hum Genet* 126(2):289–301
  13. Zhang B, Kirov S, Snoddy J (2005) WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res* 33(Web Server issue):W741–W748
  14. Begum F, Ghosh D, Tseng GC, Feingold E (2012) Comprehensive literature review and statistical considerations for GWAS meta-analysis. *Nucleic Acids Res* 40(9):3777–3784
  15. Kaddurah-Daouk R, Zhu H, Sharma S, Bogdanov M, Rozen SG, Matson W, Oki NO, Motsinger-Reif AA, Churchill E, Lei Z, Appleby D, Kling MA, Trojanowski JQ, Doraiswamy PM, Arnold SE (2013) Alterations in metabolic pathways and networks in Alzheimer's disease. *Transl Psychiatry* 3:e244
  16. Johansen KK, Wang L, Aasly JO, White LR, Matson WR, Henchcliffe C, Beal MF, Bogdanov M (2009) Metabolomic profiling in *LRRK2*-related Parkinson's disease. *PLoS One* 4(10):e7551
  17. LeWitt P, Schultz L, Auinger P, Lu M (2011) CSF xanthine, homovanillic acid, and their ratio as biomarkers of Parkinson's disease. *Brain Res* 1408:88–97
  18. O'Dushlaine C, Kenny E, Heron EA, Segurado R, Gill M, Morris DW, Corvin A (2009) The SNP ratio test: pathway analysis of genome-wide association datasets. *Bioinformatics* 25(20):2762–2763
  19. Jia P, Wang L, Meltzer HY, Zhao Z (2011) Pathway-based analysis of GWAS datasets: effective but caution required. *Int J Neuropsychopharmacol* 14(4):567–572
  20. Arasappan D, Tong W, Mummaneni P, Fang H, Amur S (2011) Meta-analysis of microarray data using a pathway-based approach identifies a 37-gene expression signature for systemic lupus erythematosus in human peripheral blood mononuclear cells. *BMC Med* 9:65
  21. Kaefer A, Landesfeind M, Feussner K, Morgenstern B, Feussner I, Meinicke P (2014) Meta-analysis of pathway enrichment: combining independent and dependent omics data sets. *PLoS One* 9(2):e89297
  22. Shen K, Tseng GC (2010) Meta-analysis for pathway enrichment analysis when combining multiple genomic studies. *Bioinformatics* 26(10):1316–1323