npg

Genome-wide association study of Alzheimer's disease

MI Kamboh[1,2,3], FY Demirci[1], X Wang[1], RL Minster[1], MM Carrasquillo[4], VS Pankratz[4], SG Younkin[4], AJ Saykin[5], for the Alzheimer's Disease Neuroimaging Initiative[17], G Jun[6,7,8], C Baldwin[6,9], MW Logue[6,7], J Buros[6], L Farrer[6,7,8,10,11], MA Pericak-Vance[12], JL Haines[13], RA Sweet[3,14,15], M Ganguli[3], E Feingold[1], ST DeKosky[16], OL Lopez[2,14] and MM Barmada[1]

In addition to apolipoprotein E (*APOE*), recent large genome-wide association studies (GWASs) have identified nine other genes/ loci (*CR1, BIN1, CLU, PICALM, MS4A4/MS4A6E, CD2AP, CD33, EPHA1* and *ABCA7*) for late-onset Alzheimer's disease (LOAD). However, the genetic effect attributable to known loci is about 50%, indicating that additional risk genes for LOAD remain to be identified. In this study, we have used a new GWAS data set from the University of Pittsburgh (1291 cases and 938 controls) to examine in detail the recently implicated nine new regions with Alzheimer's disease (AD) risk, and also performed a meta-analysis utilizing the top 1% GWAS single-nucleotide polymorphisms (SNPs) with $P < 0.01$ along with four independent data sets (2727 cases and 3336 controls) for these SNPs in an effort to identify new AD loci. The new GWAS data were generated on the Illumina Omni1-Quad chip and imputed at $\sim$2.5 million markers. As expected, several markers in the *APOE* regions showed genome-wide significant associations in the Pittsburg sample. While we observed nominal significant associations ($P < 0.05$) either within or adjacent to five genes (*PICALM, BIN1, ABCA7, MS4A4/MS4A6E* and *EPHA1*), significant signals were observed 69–180 kb outside of the remaining four genes (*CD33, CLU, CD2AP* and *CR1*). Meta-analysis on the top 1% SNPs revealed a suggestive novel association in the *PPP1R3B* gene (top SNP rs3848140 with $P = 3.05E-07$). The association of this SNP with AD risk was consistent in all five samples with a meta-analysis odds ratio of 2.43. This is a potential candidate gene for AD as this is expressed in the brain and is involved in lipid metabolism. These findings need to be confirmed in additional samples.

Introduction

Alzheimer's disease (AD), especially the late-onset form (LOAD), is a complex multifactorial neurodegenerative disease and the leading cause of dementia among the elderly people.[1] Currently, there are $\sim$5 million AD cases in the United States and about 35 million cases worldwide.[2] Due to its long clinical course, AD is a major public health problem. Genetic susceptibility at multiple genes and interactions among them and/or environmental factors likely influence the risk of AD. AD has a strong genetic basis with heritability estimates up to 80%.[3] However, genetic variants in the four well-established genes for AD; amyloid precursor protein and presenilin 1 and 2 (*PSEN1* and *PSEN2*) for the rare early-

onset, and apolipoprotein E (*APOE*) for the common LOAD explain less than half of this heritability.

To identify the remaining genes for the common LOAD, efforts have been focused on conducting genome-wide association studies (GWASs)[4–11] because this approach is hypothesis free and conceptually would identify all known and unknown genes. However, with the exception of the *APOE* region, no other significant associations were replicated across these initial GWAS. This highlights the difficulties in identifying the remaining LOAD genes that are thought to make a relatively small contribution to the overall risk of disease and thus would require much large sample size than used in earlier GWASs.[4–11] Recent large GWASs have identified nine additional genes/loci for LOAD, including *CR1,*

BIN1, CLU (a.k.a. APOJ), PICALM, MS4A4/MS4A6E, CD2AP, CD33, EPHA1 and ABCA7.[12–16] Although up to 80% of the AD risk is attributable to genetic factors,[3] all of the known LOAD genes (including APOE and new ones) account for about 50% of the total genetic variance. This indicates that additional risk genes for LOAD remain to be identified.

In this study, we used 2229 samples from the University of Pittsburgh Alzheimer's Disease Research Center (ADRC) that were genotyped using the Illumina Omni1-Quad chip to examine the extent of associations of the recently implicated nine non-APOE risk loci for LOAD in this sample. In addition, we performed a GWAS analysis in the Pittsburgh sample and conducted a meta-analysis on the top 1% single-nucleotide polymorphisms (SNPs) by incorporating additional 2727 cases and 3336 controls from prior studies to identify new loci for AD.

Subjects and methods

Stage 1 sample. Genomic DNA from 1440 AD cases and 1000 controls were genotyped using the Illumina Human Omni1-Quad as part of the stage 1 discovery sample. All subjects were European Americans. AD cases (mean age-at-onset (AAO) 72.6 ± 6.4 years; 65.6% women; 23.5% autopsy-confirmed) were derived from the University of Pittsburgh ADRC, all of whom met the National Institute of Neurological and Communication Disorders and Stroke (NINCDS)/Alzheimer's Disease and Related Disorders Association (ADRDA) criteria for probable or definite AD. The University of Pittsburgh ADRC follows a standard evaluation protocol, including medical history, general medical and neurological examinations, psychiatric interview, neuropsychological testing and magnetic resonance imaging scan. Controls were cognitively normal subjects that were recruited from the same geographical area as the cases. They were 60 years or older (mean age 74.07 ± 6.20 years; 59.8% women; 0.2% autopsy-confirmed), had no psychiatric or neurological disorders, and did not meet criteria for mild cognitive impairment or dementia. All subjects were recruited with informed consent, and the study was approved by the University of Pittsburgh Institutional Review Board.

Stage 2 samples. A total of 2727 cases and 3336 controls were derived from four prior studies and they are briefly described below.

Mayo sample. The Mayo sample comprised 844 AD cases (mean AAO: 74 ± 4.8 years; 57.2% women) and 1255 controls (mean age: 73.2 ± 4.4 years; 51.7% women) that were previously genotyped using the Illumina HumanHap300 BeadChip.[11] AD diagnosis was established using the NINCDS/ADRDA criteria.

Alzheimer's Disease Neuroimaging Initiative (ADNI) sample. This data set consisted of 188 AD cases with a clinical diagnosis of AD at baseline visit and 193 controls (mean age: 78.6 ± 5.3 years; 46.3% women) that were previously genotyped using the Illumina 610-Quad BeadChip.[17]

AD cases were between the ages of 55 and 90 (mean AAO: 71.9 ± 8.1 years; 44.6% women) and met NINCDS/ADRDA criteria for AD. Details of the clinical evaluation and sample characterization are described elsewhere.[18,19] The ADNI data used in this report were obtained from the ADNI database (adni.loni.ucla.edu). The initial goal of ADNI was to recruit 800 adults, aged 55–90, to participate in research on the sensitivity and specificity of neuroimaging and other biomarkers for detecting and monitoring AD pathology *in vivo*. In ADNI, ~200 cognitively normal older individuals and 400 people with amnestic mild cognitive impairment were followed for 3 years and 200 people with mild early-stage AD who were followed for 2 years. For up-to-date information, see http://www.adni-info.org.

University of Miami/Vanderbilt University/Mt. Sinai School of Medicine sample. The University of Miami/Vanderbilt University/Mt. Sinai School of Medicine data set contains 1186 cases (mean AAO: 74.1 ± 7.8 years; 64% women) and 1135 controls (mean age: 74.0 ± 8.3; 61% women) ascertained at the University of Miami, Vanderbilt University and Mt. Sinai School of Medicine. Cases met NINCDS/ADRDA criteria for probable or definite AD with AAO >60 years. Cognitively healthy controls were unrelated individuals from the same catchment areas.[16]

MIRAGE study sample. MIRAGE is a family-based genetic epidemiological study of AD that enrolled AD cases and unaffected sibling controls at 17 clinical centers in the United States, Canada, Germany and Greece. Briefly, families were ascertained through a proband meeting the NINCDS/ADRDA criteria for definite or probable AD. Unaffected sibling controls were verified as cognitively healthy based on a Modified Telephone Interview of Cognitive Status score of ≥86. For this study, we used 1262 subjects (509 cases and 753 controls)[16] from the MIRAGE Caucasian data set.

Genotyping and quality control of genotype data in the Pittsburgh sample. The stage 1 Pittsburgh sample was genotyped using the Illumina Omni1-Quad chip (containing probes for 1 016 423 SNPs and/or copy-number variations) at the Feinstein Institute of Medical Research (Manhasset, NY). Genotypes for two APOE SNPs rs429358 (E*4) and rs7412 (E*2) were determined either as previously described[20] or using TaqMan SNP genotyping assays.

Exclusion criteria for individual samples included consent and diagnostic criteria (five controls were excluded because of updated diagnostic criteria, or refusal to participate further in this study), high genotype failure rate (141 individuals were removed because of a genotype failure rate >2%) and cryptic relatedness (65 individuals were removed because they displayed an average degree of sharing (identity by state or IBS) >0.4 with other members of the data set). Exclusion criteria for markers included minor allele frequency (189 727 SNPs were removed because of minor allele frequency <1%), deviation from Hardy–Weinberg expectations in controls (2239 SNPs gave a Hardy–Weinberg expectations test P-value ≤1E–06), and high genotype failure rates (22 385 SNPs were removed because of genotype failure rates >2%). The final

sample after all exclusions was applied consisted of 1291 cases and 938 controls.

Imputation. Genotype posterior probabilities were imputed with MACH v.1.0 (http://www.sph.umich.edu/csg/abecasis/MACH/), on all data sets using haplotypes from the HapMap CEU v3 data release as a reference sample. The imputation generated data for >3 million SNPs that were subsequently filtered to exclude SNPs with $r2<0.3$ and 2 543 888 were included in the analysis. Additional details of the imputation procedure were published previously.[16]

Population stratification. Population stratification was examined using a multi-dimensional scaling method as implemented in PLINK.[21] Four components were conservatively determined to be relevant to the determination of population origin based on the visual examination of principle component plots.

Association analyses. Initial association analysis was performed in the University of Pittsburgh ADRC's GWAS data set. After standard quality control filters for both genotypes and samples and imputing for unobserved genotypes, a total of ~2.5 million SNPs were examined in 2229 subjects (1291 cases and 938 controls) for association analysis. Association of SNPs with AD risk was tested using logistic regression under an additive model that included age, sex and the first four principal components as covariates. All SNPs with *P*-values of <0.01 were carried forward to the meta-analysis. Significance values from logistic regression analyses were used for ranking purposes only, and so were not adjusted for multiple testing. All analyses were done in R and/or PLINK. Meta-analysis was done using a fixed-effect methodology, as implemented in PLINK. Heterogeneity testing was accomplished using Cochran's Q statistic, summarized as the I^2 statistic (the percentage of total variation across studies that is due to heterogeneity rather than chance).[22,23] Values closer to 0 indicate no heterogeneity, whereas larger numbers indicate increasing degrees of heterogeneity between studies.

Results

Association analysis in nine non-APOE AD loci.
We first examined the association signals in the Pittsburgh GWAS data in recently reported nine non-*APOE* AD gene regions (*CR1, BIN1, CLU, PICALM, MS4A4/MS4A6E, CD2AP, CD33, EPHA1* and *ABCA7*)[12–16] in 2229 cases and controls that passed quality control for both genotypes and samples and included imputed data for unobserved genotypes. Although this sample was included as one of several replication samples in one GWAS that reported the latter five gene region,[16] the extent of replications in these five and the other four regions in this sample has previously not been examined closely. We observed significant associations ($P<0.05$) either within or adjacent to five genes (*PICALM, BIN1, ABCA7, MS4A4/MS4A6E* and *EPHA1*). On the other hand, significant signals were observed 69–180 kb

outside of the remaining four genes (*CD33, CLU, CD2AP* and *CR1*). The regional association plots containing SNPs within ~500 kb on either side of these nine genes are presented in Supplementary Figures 1–9.

In the *PICALM* gene and its surrounding sequence, 69 SNPs showed significant associations (Supplementary Figure 1). The top 11 SNPs with $P<E-03$ were located in four different introns of the *PICALM* gene. The most significant SNP, rs17817992 ($P=1.37E-04$), was located in intron 2 followed by rs12790526 ($P=1.43E-04$) in intron 2, rs12795381 ($P=1.68E-04$) in intron 12 and rs12802399 ($P=1.72E-04$) in intron 2. The originally reported genome-wide significant SNP/rs3851179,[13] which is present about 89 kb 5′ to *PICALM*, was observed at $P=4.6E-02$ in our sample. In the *BIN1* gene, 17 SNPs showed significant associations (Supplementary Figure 2), mostly in the 5′ region, including the 5 top significant SNPs/rs11680911 ($P=2.9E-04$), rs7561528 ($P=3.0E-04$), rs6743470 ($P=1.18E-03$), rs17014923 ($P=1.64E-03$) and rs9394826 ($P=2.15E-03$). Among our top five significant SNPs, rs7561528 was genome-wide significant in the discovery sample of the Alzheimer Disease Genetics Consortium.[16] The originally reported genome-wide significant SNP/rs744373 in the initial GWAS after stage 3 meta-analysis[14] was observed at $P=1.02E-02$ in our sample. In the *ABCA7* gene, seven SNPs showed significant association (Supplementary Figure 3), including the two most significant SNPs, rs4147916 in intron 34 ($P=3.18E-03$) and rs4147918 in exon 37 ($P=3.19E-03$) that is associated with an amino-acid change (Gln → Arg). The reported genome-wide significant SNP (rs3764650)[15] gave a non-significant but similar trend of association in our sample ($P=1.68E-01$).

There are multiple genes in the *MS4A* gene clusters on chromosome 11. While one GWAS reported genome-wide significant SNPs in the *MS4A4A* gene,[16] another GWAS found genome-wide significant SNPs in the *MS4A6A* and *MS4A4E* genes.[15] In our data set, we found a total of 39 SNPs with $P<0.05$ in the *MS4A* gene cluster (Supplementary Figure 4) and the 5 most significant SNPs were located in the *MS4A4A* gene and its surrounding sequence: rs11827324 ($P=2.58E-03$), rs4939338 ($P=3.39E-03$), rs10792263 ($P=4.85E-03$), rs11824773 ($P=8.06E-03$) and rs11824734 ($P=6.93E-03$). The genome-wide significant SNP in the Alzheimer Disease Genetics Consortium (rs4938933) gave the same trend of association in our sample ($P=8.80E-02$). There were eight significant SNPs in the *EPHA1* gene (Supplementary Figure 5), including the top significant SNPs: rs10241042 ($P=2.27E-03$), rs1525119 ($P=1.02E-02$) and rs10233030 ($P=1.29E-02$). The reported genome-wide significant SNP (rs11767557)[15,16] was observed with a similar trend of association in our sample ($P=1.11E-01$).

For the other four genes, significant SNPs ($P<0.05$) were observed 69, 125, 143 and 180 kb outside of the *CD33, CLU, CR1* and *CD2AP* genes, respectively. There were four SNPs with $P<0.05$ and another seven with *P*-values between 0.05 and 0.10 located in an intergenic region about 69 kb distal to the *CD33* gene. Likewise, there were four significant SNPs ($P=0.012–0.027$) about 125 kb proximal to the *CD33* gene (Supplementary Figure 6). The reported genome-wide
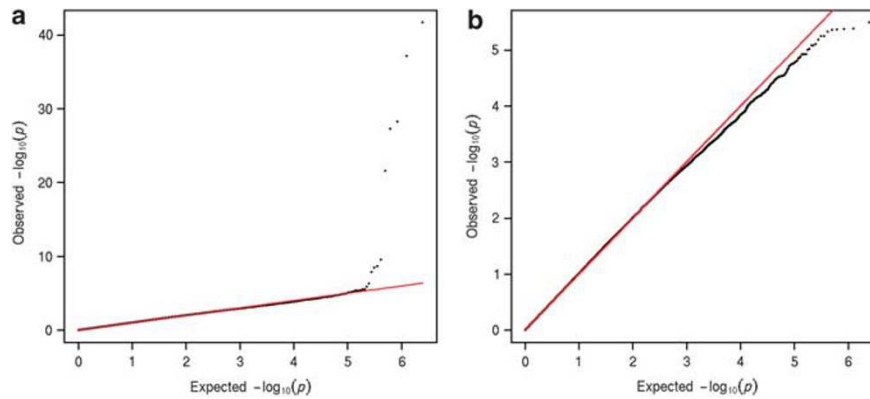
Figure 1 Quantile-quantile plots showing the observed versus the expected *P*-values in the University of Pittsburgh sample including all single-nucleotide polymorphisms (SNPs) (**a**) and after removing SNPs in the *APOE, CR1, BIN1, CLU, PICALM, MS4A4/MS4A6E, CD2AP, CD33, EPHA1* and *ABCA7* regions (**b**). The red line shows the distribution under the null hypothesis.
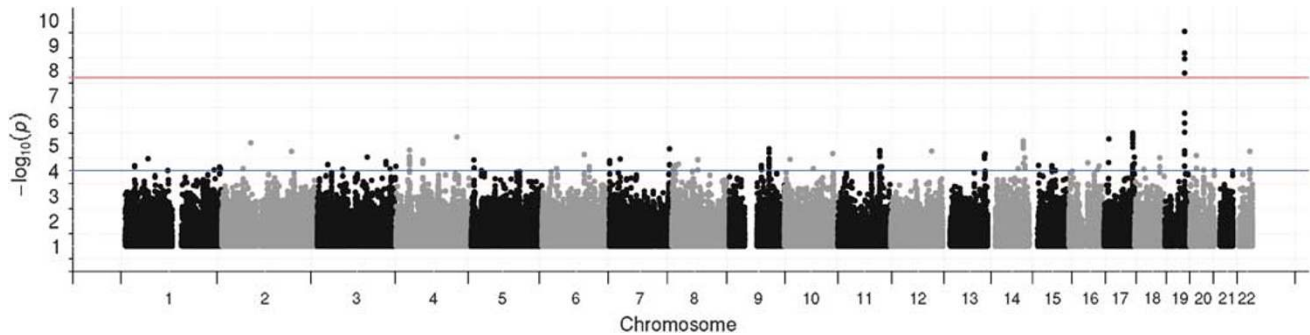

Figure 2 Manhattan plots showing the genome-wide *P*-values in the University of Pittsburgh sample after adjusting for age (age-at-onset in cases), sex and principle components. Red line indicates genome-wide significant level ($P < 5E–08$) and blue line indicates suggestive associations ($P < 1E–04$) in the Pittsburgh sample.
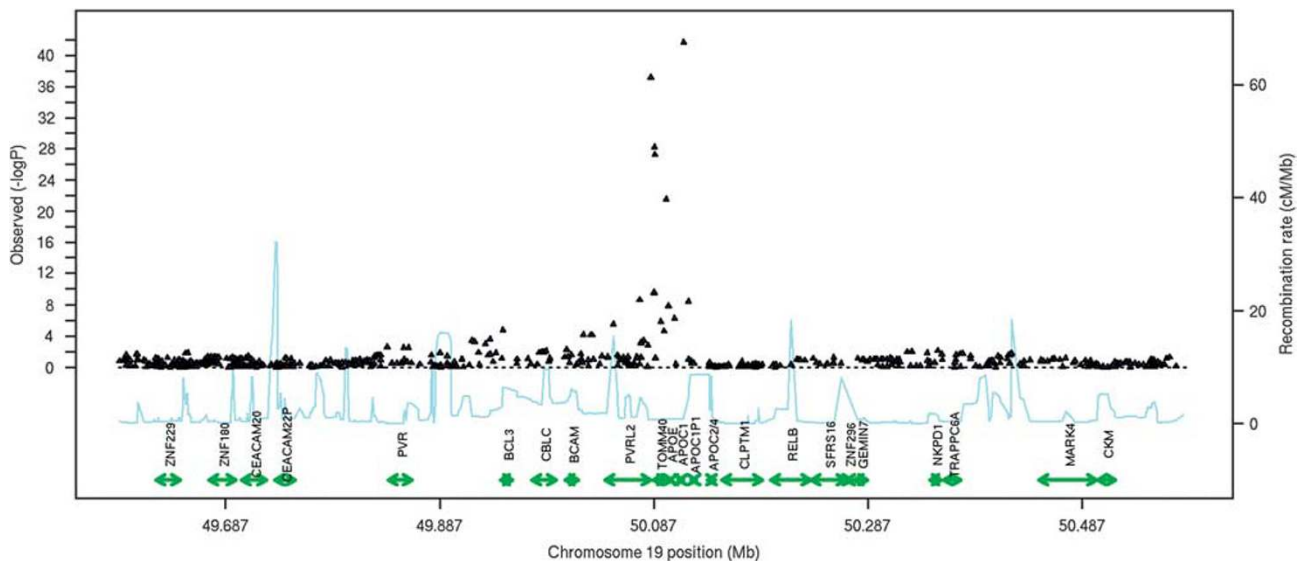

Figure 3 Regional association plot on chromosome 19 including the best single-nucleotide polymorphism (SNP), rs429358 (*E*4*), for Alzheimer's disease (AD) risk in the University of Pittsburgh sample.

significant SNP (rs3865444)[15,16] gave a non-significant but similar trend of association in our sample ($P = 2.50E–01$). Overall, recombination was found to be low in this region. In the *CLU* region, 41 significant SNPs were present in the *PTK2B* and *CHRNA2* genes, both of which are within 100–

125 kb of the *CLU* gene (Supplementary Figure 7). Although no SNP showed an association at $P = 0.05$ in the *CLU* gene, several SNPs demonstrated the reported trend of association. The 14 most significant SNPs were present in the *PTK2B* gene with $P < 4.0E–03$. The top three SNPs in the *PTK2B*

gene were rs373625 in intron 25 ($P = 2.72E–03$) and rs9657295 and rs11135993 in intron 27 ($P = 3.06E–03$). The reported associations in the *CD2AP* and *CR1* genes were not replicated in our sample, as significant SNPs were located ~150 kb from these two genes (Supplementary Figures 8 and 9).

GWAS analysis. We next examined the entire GWAS data set (~2.5 million SNPs) in the Pittsburg sample. Figure 1a shows the comparison of observed *P*-values with the *P*-values for a null distribution. The observed *P*-value conformed to the null distribution until the tail of the distribution where it deviated, suggesting little evidence of stratification but compelling evidence of disease associations. As expected, several SNPs in the *APOE* region on chromosome 19 demonstrated genome-wide significant ($P < 5E–08$) associations (Figure 2). The most significant SNP was rs429358 (*E*4*) in the *APOE* gene ($P = 2.52E–53$) followed by rs4420638 ($P = 1.97E–42$), rs6857 ($P = 6.75E–38$), rs2075650 ($P = 5.67E–29$), rs157582 ($P = 5.56E–28$) and rs157580 ($P = 2.77E–10$) in the *APOE/APOC1/TOMM40* gene region. The regional association plot including all SNPs in the *APOE* region is given in Figure 3.

After removing SNPs in the *APOE, CR1, BIN1, CLU, PICALM, MS4A4/MS4A6E, CD2AP, CD33, EPHA1* and *ABCA7* regions, a deviation of *P*-values from the null distribution remained in the quantile-quantile plot, although within the 95% confidence interval of the expectation (Figure 1b). Since no other SNPs outside the *APOE* region were genome-wide significant in the Pittsburgh discovery sample, we performed a meta-analysis by combining the results of the top ~1% SNPs (~25 000) with $P < 0.01$ of the discovery sample with the four existing data sets (2727 AD cases and 3336 controls) for these SNPs. Of the known loci, *APOE, BIN1* and *PICALM* revealed genome-wide significant associations along with five suggestive new loci with $P < 1E–05$ in the meta-analysis (Table 1). All SNPs with $P < 1E–04$ in the meta-analysis are presented in Supplementary Table.

Four of the top five SNPs outside the *APOE* and nine other known loci were either directly genotyped or had proxy genotyped SNPs with $P < 1E–04$ on Illumina arrays (see Table 1 and Supplementary Table), thus eliminating potential spurious associations due to imputation artifact. The most significant SNP, rs3848140 ($P = 3.05E–07$), located in the *PPP1R3B* (encoding protein phosphatase 1, regulatory subunit 3B) region on chromosome 8 at position 9.04 Mb was actually genotyped and there were three additional significant SNPs in this gene at $P < 1E–06$. The association of this SNP with AD risk was consistent in all five samples with meta odds ratio of 2.43. Figure 4 shows the regional plot for SNPs within 500 kb on either side of the *PPP1R3B* index SNP and the meta *P*-values for markers which had $P < 1E–04$ in the meta-analysis. The regional plots for the remaining four top AD loci are shown in Supplementary Figures 10–13. To remove any variation in AD due to the established effect of *APOE*, we also analyzed the data after adjusting for the effect of *APOE*4*, but found no appreciable difference in *P*-values for the non-*APOE* loci (Table 1), indicating that their effects are independent of *APOE*.

Table 1 Genetic loci associated with the risk of AD with $P < E–05$

Chr	Top SNP[a]	Position MB	Total SNPs[b]	Nearest gene	Pittsburgh OR	Pittsburgh P	Mayo OR	Mayo P	ADNI OR	ADNI P	MIRAGE OR	MIRAGE P	UM/VU/MSSM OR	UM/VU/MSSM P	Meta-analysis OR	Meta-analysis P	Meta APOE adjusted P	Heterogeneity test (I)
19	rs4420638	50.11	28 (14)	APOE/TOMM40/APOC1	0.32	1.97E–42	0.21	1.72E–51	0.28	2.11E–06	0.38	6.44E–22	0.25	7.38E–38	0.29	7.91E–149	1.99E–11	79.31
2	rs7561528	127.61	2 (1)	BIN1	1.27	2.96E–04	1.26	1.07E–03	1.34	9.05E–02	1.14	8.66E–02	1.29	9.94E–05	1.25	6.00E–11	4.80E–09	0
11	rs17817600	85.36	30 (7)	PICALM	1.41	3.67E–04	1.36	2.31E–03	0.94	8.48E–01	1.38	1.22E–02	1.25	2.12E–02	1.33	2.17E–08	1.40E–05	0
8	rs3748140	9.04	4 (2)	PPP1R3B	2.80	5.75E–04	3.47	6.57E–03	10.14	1.31E–01	1.62	1.82E–01	2.24	1.48E–02	2.43	3.05E–07	4.93E–06	0
11	rs12808148	102.24	1 (0)	MMP3/MMP12	0.76	1.38E–03	0.82	4.55E–02	0.62	5.60E–02	0.87	1.39E–01	0.82	1.55E–02	0.81	1.24E–06	1.02E–05	0
4	rs6856768	30.08	4 (1)	—	0.82	3.11E–03	0.81	3.05E–03	0.97	8.55E–01	0.88	1.21E–01	0.89	8.19E–02	0.85	4.63E–06	8.22E–05	0
5	rs11738335	60.99	9 (3)	FLJ37543	0.75	5.72E–04	0.86	6.83E–02	1.19	4.39E–01	0.84	5.62E–02	0.83	1.73E–02	0.83	4.66E–06	4.28E–06	3.43
2	rs1357692	107.04	11 (1)	—	0.82	4.63E–03	0.87	4.82E–02	1.05	7.89E–01	0.82	1.67E–02	0.88	5.17E–02	0.86	9.27E–06	3.28E–06	0

Abbreviations: ADNI, Alzheimer's Disease Neuroimaging Initiative; APOE, apolipoprotein E; MB, mega base; OR, odds ratio; SNP, single-nucleotide polymorphism; UM/VU/MSSM, University of Miami/Vanderbilt University/Mt. Sinai School of Medicine. [a]SNP with the lowest *P*-value in a given gene region after meta-analysis. [b]Total genotyped or imputed significant SNPs with $P < 1E–04$ in a given region after meta-analysis. The number of genotyped SNPs is given in parentheses.
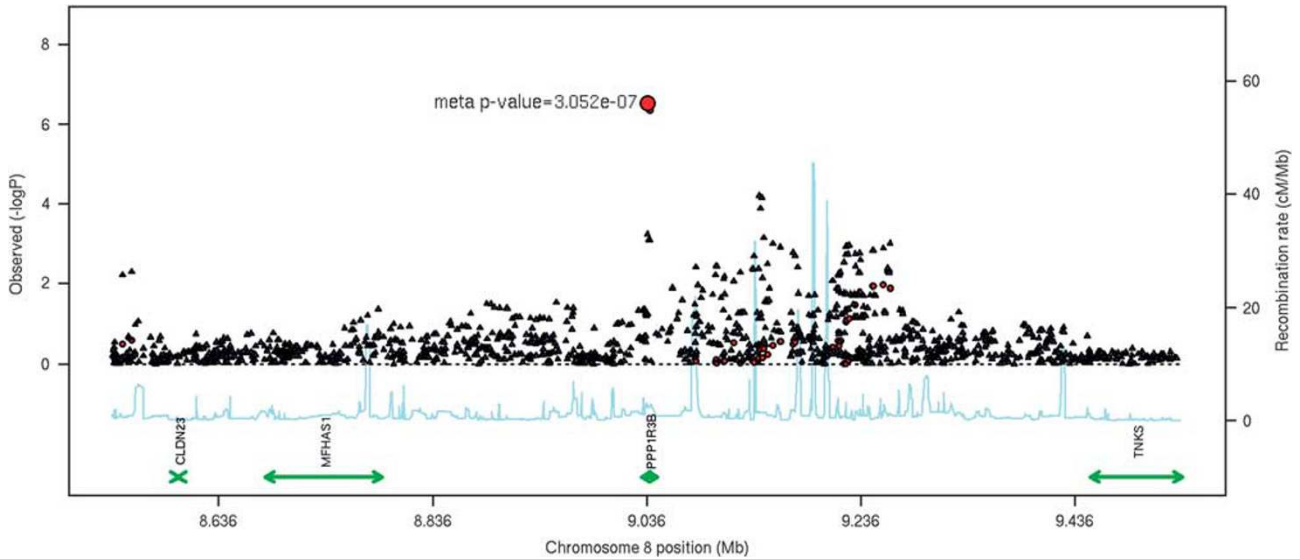
Figure 4 Regional association plot on chromosome 8, including the best single-nucleotide polymorphism (SNP) (rs3748140) for Alzheimer's disease (AD) risk in the *PPP1R3B* gene in the Pittsburgh sample (blue triangles), and meta-analysis (red dots).

Discussion

In this study, we have used a new GWAS data set to (i) examine in detail the recently implicated nine new regions with AD risk, and to (ii) perform a meta-analysis utilizing the top 1% GWAS SNPs ($\sim$25 000) with $P<0.01$ along with four independent data sets for these SNPs in an effort to identify new AD loci.

Five of the nine loci are replicated in the Pittsburgh sample as we found multiple nominal significant SNPs in or around five genes (*PICALM, BIN1, ABCA7, MS4A4/MS4A6E* and *EPHA1*), suggesting that these are the relevant genes for AD in these regions. As associations in these genes have been validated in multiple studies,[13,16] it is acceptable to consider $P<0.05$ statistically significant in follow-up studies[24] like ours. However, the identity of the potential causal genes in the remaining four regions *(CD33, CLU, CD2AP* and *CR1*) is not clear in the Pittsburgh sample, as significant SNPs were observed 69–180 kb outside of these genes. The power in our Pittsburgh sample to detect these nine loci for the effect sizes reported in a previous GWAS[16] is 0.38, 0.43, 0.39, 0.31, 0.38, 0.30, 0.28, 0.28 and 0.47 for *PICALM, BIN1, ABCA7, MS4A4, EPHA1, CD33, CLU, CD2AP* and *CR1,* respectively. Although the overall power is low and this would reduce our confidence in our negative results, it enhances the interest in our positive results even further.

There were only 11 SNPs present in the *CD33* gene on the 1M chip. This may explain why we did not observe any significant SNP in this gene, although the reported genome-wide significant SNP (rs3865444) gave a *P*-value of 0.25 with the same effect size in our sample. Also, recombination is low in the *CD33* region; thus, it is likely that functional SNPs may be present in a broader region around *CD33*. Although none of the small number of 15 SNPs present in the *CLU* gene on the 1M chip showed significant association at $P<0.05$, a number of them showed the trend of association reported in earlier studies. However, several SNPs in a nearby *PTK2B* gene that is expressed in adult brain,[25] revealed significant

associations. Perhaps, genetic variations in the *CLU* and *PTK2B* genes, both of which are expressed in the brain, are relevant to AD risk and thus resequencing a broader region in and around these genes may reveal functional variants. However, a recent sequencing study found no functional SNPs in the coding region of the *CLU* gene and, likewise, no significant association was found between the previously associated *CLU* SNPs with AD risk with expression quantitative trait loci,[26] suggesting that the biological role of *CLU* is AD is not yet clear. While findings in our sample suggest the involvement of broader regions around the *CD33* and *CLU* genes, the evidence of association of the *CR1* and *CD2AP* with AD is weak or absent.

The GWAS analysis in the Pittsburgh data set found only the *APOE/TOMM40/APOC1* region to be genome-wide significant, which is due to the established association with the *APOE*4* SNP. Meta-analysis of the top 25 000 SNPs from the Pittsburgh GWAS data, in conjunction with four replication data sets revealed five suggestive non-*APOE* loci. The top SNP was located in the *PPP1R3B* gene, which is a potential candidate gene for AD as it is expressed in human brain.[27] Genetic variation in *PPP1R3B* has been shown to be associated with plasma high density lipoprotein and cholesterol levels and the risk of coronary artery disease.[28,29] Overexpression of the mouse ortholog *Ppp1r3b* in mouse liver results in significantly lower high density lipoprotein levels.[28] As high density lipoproteins levels and particle size affect the clearance of β-amyloid in mouse brain,[30] it is conceivable that the *PPP1R3B*-associated effect on AD risk is mediated through its impact on high density lipoproteins levels. Although the association of the top *PPP1R3B* SNP did not meet the strict criteria for genome-wide significance, it was consistent in direction in all samples as reflected in Odds ratio values (see Table 1). This suggests that this gene region should be followed up in additional samples. If these findings are confirmed, then this would be another cholesterol-related gene involved in AD along with *APOE*, *CLU* and *ABCA7*.

In conclusion, 6 of the 10 established AD loci, including the *APOE* locus, have been replicated in our new GWAS data set, while broader regions in the remaining four loci are suspected. Although we did not identify additional non-*APOE* SNPs meeting a conservative threshold of genome-wide significance, we have identified five suggestive loci that warrant follow-up in additional samples.

Conflict of interest

The authors declare no conflict of interest.

1. Evans DA, Funkenstein HH, Albert MS, Scherr PA, Cook NR, Chown MJ et al. Prevalence of Alzheimer's disease in a community population of older persons. Higher than previously reported. *JAMA* 1989; **262**: 2551–2556.
2. Alzheimer's Disease International. *World Alzheimer Report* 2009. www.alz.co.uk/research/world-report.
3. 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.
4. Grupe A, Abraham R, Li Y, Rowland C, Hollingworth P, Morgan A et al. Evidence for novel susceptibility genes for late-onset Alzheimer's disease from a genome-wide association study of putative functional variants. *Hum Mol Genet* 2007; **16**: 865–873.
5. Coon KD, Myers AJ, Craig DW, Webster JA, Pearson JV, Lince DH et al. A high-density whole-genome association study reveals that APOE is the major susceptibility gene for sporadic late-onset Alzheimer's disease. *J Clin Psychiatry* 2007; **68**: 613–618.
6. Reiman EM, Webster JA, Myers AJ, Hardy J, Dunckley T, Zismann VL et al. GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers. *Neuron* 2007; **54**: 713–720.
7. Li H, Wetten S, Li L, St Jean PL, Upmanyu R, Surh L et al. Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. *Arch Neurol* 2008; **65**: 45–53.
8. 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.
9. Beecham GW, Martin ER, Li YJ, Slifer MA, Gilbert JR, Haines JL et al. Genome-wide association study implicates a chromosome 12 risk locus for late-onset Alzheimer disease. *Am J Hum Genet* 2009; **84**: 35–43.
10. Abraham R, Moskvina V, Sims R, Hollingworth P, Morgan A, Georgieva L et al. A genome-wide association study for late-onset Alzheimer's disease using DNA pooling. *BMC Med Genomics* 2008; **1**: 44.
11. Carrasquillo MM, Zou F, Pankratz VS, Wilcox SL, Ma L, Walker LP et al. Genetic variation in PCDH11X is associated with susceptibility to late-onset Alzheimer's disease. *Nat Genet* 2009; **41**: 192–198.
12. Lambert JC, Heath S, Even G, Campion D, Sleegers K, Hiltunen M et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet* 2009; **41**: 1094–1099.
13. Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* 2009; **41**: 1088–1093.
14. Seshadri S, Fitzpatrick AL, Ikram MA, DeStefano AL, Gudnason V, Boada M et al. Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA* 2010; **303**: 1832–1840.
15. Hollingworth P, Harold D, Sims R, Gerrish A, Lambert JC, Carrasquillo MM et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet* 2011; **43**: 429–435.
16. Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buros J et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet* 2011; **43**: 436–441.
17. Saykin AJ, Shen L, Foroud TM, Potkin SG, Swaminathan S, Kim S et al. Alzheimer's Disease Neuroimaging Initiative biomarkers as quantitative phenotypes: genetics core aims, progress, and plans. *Alzheimers Dement* 2010; **6**: 265–273.
18. Aisen PS, Petersen RC, Donohue MC, Gamst A, Raman R, Thomas RG et al. Clinical core of the Alzheimer's Disease Neuroimaging Initiative: progress and plans. *Alzheimers Dement* 2010; **6**: 239–246.
19. Petersen RC, Aisen PS, Beckett LA, Donohue MC, Gamst AC, Harvey DJ et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology* 2010; **74**: 201–209.
20. Kamboh MI, Aston CE, Hamman RF. The relationship of APOE polymorphism and cholesterol levels in normoglycemic and diabetic subjects in a biethnic population from the San Luis Valley, Colorado. *Atherosclerosis* 1995; **112**: 145–159.
21. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D et al. PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet* 2007; **81**: 559–575.
22. Brockwell SE, Gordon IR. A comparison of statistical methods for meta-analysis. *Stat Med* 2001; **20**: 825–840.
23. Higgins JC, Ivanov NV, Hodor P, Xia M, Wei N, Blevins R. et al. Measuring inconsistency in meta-analyses. *BMJ* 2003; **327**: 557–560.
24. Plenge RM, Bridges SL, Huizinga WJ, Criswell LA, Gregersen PK. Recommendation of genetic association studies in *Arthritis & Rheumatism*. *Arthritis Rheum* 2001; **63**: 2839–2847.
25. Lev S, Moreno H, Martinez R, Canoll P, Peles P, Musacchio JM et al. Proteintyrosine kinase PYK2 involved in Ca(2+)-induced regulation of ion channel and MAP kinase functions. *Nature* 1995; **376**: 737–745.
26. Guerreiro RJ, Beck J, Gibbs JR, Santana I, Rossor MN, Schott JM et al. Genetic variability in CLU and its association with Alzheimer's diseases. *PLoS One* 2010; **5**: e9510.
27. Glorioso C, Oh S, Douillard GG, Sibille E. Brain molecular aging, promotion of neurological disease and modulation by sirtuin 5 longevity gene polymorphism. *Neurobiol Dis* 2011; **41**: 279–290.
28. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010; **466**: 707–713.
29. Waterworth DM, Ricketts SL, Song K, Chen L, Zhao JH, Ripatto S et al. Genetic variants influencing circulating lipid levels and risk of coronary artery disease. *Arterioscler Thromb Vasc Biol* 2010; **30**: 2264–2276.
30. Cramer PE, Cirrito JR, Wesson DW, Lee CYD, Karlo C, Zinn AE et al. ApoE-directed therapeutics rapidly clear β-amyloid and reverse deficits in AD mouse models. *Science* 2012; **335**: 1503–1506.

Supplementary Information accompanies the paper on the Translational Psychiatry website (http://www.nature.com/tp)