

This Provisional PDF corresponds to the article as it appeared upon acceptance. Copyedited and fully formatted PDF and full text (HTML) versions will be made available soon.

The membrane-spanning 4-domains, subfamily A (MS4A) gene cluster contains a common variant associated with Alzheimer's Disease

Genome Medicine 2011, **3**:33 doi:10.1186/gm249

Carmen Antunez (mcarmen.antunez@carm.es)
Merce Boada (mbr@fundacioace.com)
Antonio Gonzalez-Perez (agonzalez@neocodex.es)
Javier Gayan (jgayan@neocodex.es)
Reposo Ramirez-Lorca (rrl@neocodex.es)
Juan Marin (juan.marin5@carm.es)
Isabel Hernandez (ihernandez@fundacioace.com)
Concha Moreno-Rey (cmoreno@neocodex.es)
Francisco Jesus Moron (fran@neocodex.es)
Jesus Lopez-Arrieta (jlopezar.hulp@salud.madrid.org)
Ana Mauleon (amauleon@fundacioace.com)
Maitee Rosende-Roca (mrosende@fundacioace.com)
Fuensanta Noguera-Perea (mariaf.noguera@carm.es)
Agustina Legaz-Garcia (agustina.legaz@carm.es)
Laura Vivancos-Moreau (laura.vivancos@carm.es)
Juan Velasco (jvelasco@neocodex.es)
Jose Miguel Carrasco (jmcarrasco@neocodex.es)
Montserrat Alegret (malegret@fundacioace.com)
Martirio Antequera-Torres (mariam.antequera@carm.es)
Salvadora Manzanares (salvi.manzanares@gmail.com)
Alejandro Romo (aromo@neocodex.es)
Irene Blanca (iblanca@neocodex.es)
Susana Ruiz (sruiz@fundacioace.com)
Anna Espinosa (aespিনosa@fundacioace.com)
Sandra Castano (sclorete@gmail.com)
Blanca Garcia (blagarto@hotmail.com)
Begona Martinez-Herrada (begona.martinez@carm.es)
Georgina Vinyes (gvinyas@fundacioace.com)
Asuncion Lafuente (alafuente@fundacioace.com)
James T Becker (BeckerJT@upmc.edu)
Jose Jorge Galan (jjgalan@neocodex.es)
Manuel Serrano-Rios (mserrano.hcsc@salud.madrid.org)
Alzheimer's Disease Neuroimaging Initiative (adni) (aruiz@neocodex.es)
Enrique Vazquez (evt@neocodex.es)
Lluís Tarraga (ltarraga@fundacioace.com)
Maria Eugenia Saez (mesaez@neocodex.es)
Oscar L Lopez (lopezOL@upmc.edu)
Luis Miguel Real (lreal@neocodex.es)
Agustin Ruiz (aruiz@neocodex.es)

© 2011 Antunez *et al.*; licensee BioMed Central Ltd.

This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ISSN 1756-994X
Article type Research
Submission date 23 April 2011
Acceptance date 31 May 2011
Publication date 31 May 2011
Article URL <http://genomemedicine.com/content/3/5/33>

This peer-reviewed article was published immediately upon acceptance. It can be downloaded, printed and distributed freely for any purposes (see copyright notice below).

Articles in *Genome Medicine* are listed in PubMed and archived at PubMed Central.

For information about publishing your research in *Genome Medicine* go to

<http://genomemedicine.com/info/instructions/>

The membrane-spanning 4-domains, subfamily A (*MS4A*) gene cluster contains a common variant associated with Alzheimer's Disease

Carmen Antúnez^{1,2*}, Mercè Boada^{3,4*}, Antonio González-Pérez^{5*}, Javier Gayán^{5*}, Reposo Ramírez-Lorca⁵, Juan Marín¹, Isabel Hernández³, Concha Moreno-Rey⁵, Francisco Jesús Morón⁵, Jesús López-Arrieta⁶, Ana Mauleón³, Maitée Rosende-Roca³, Fuensanta Noguera-Perea¹, Agustina Legaz-García¹, Laura Vivancos-Moreau¹, Juan Velasco⁵, José Miguel Carrasco⁵, Montserrat Alegret³, Martirio Antequera-Torres¹, Salvadora Manzanares¹, Alejandro Romo⁵, Irene Blanca⁵, Susana Ruiz³, Anna Espinosa³, Sandra Castaño¹, Blanca García¹, Begoña Martínez-Herrada¹, Georgina Vinyes³, Asunción Lafuente³, James T Becker⁷, José Jorge Galán⁵, Manuel Serrano-Ríos⁸, Alzheimer's Disease Neuroimaging Initiative, Enrique Vázquez⁵, Lluís Tárraga³, María Eugenia Sáez⁵, Oscar L López⁷, Luis Miguel Real⁵ and Agustín Ruiz^{5#}

1. Dementia Unit, University Hospital Virgen de la Arrixaca, Ctra. Madrid-Cartagena, Murcia, s/n - 30120 El Palmar, Spain
2. Alzheimer Foundation, Avda. Juan Carlos, Building Cajamurcia, Murcia, 30100, Spain
3. Memory Clinic of Fundació ACE. Institut Català de Neurociències Aplicades, Calle del Marqués de Sentmenat, Barcelona, 35-3708029, Spain
4. Hospital Universitari Vall d'Hebron - Institut de Recerca, Universitat Autònoma de Barcelona. (VHIR-UAB), Carretera bellaterra, Barcelona, S/N 08290 Cerdanyola del Vallès, Spain
5. Department of Structural Genomics, Neocodex, Avda. Charles Darwin, Sevilla, s/n 41092, Spain
6. Memory Unit, University Hospital La Paz-Cantoblanco, Paseo Castellana, 261, Madrid, 28046, Spain
7. Alzheimer's Disease Research Center. Departments of Neurology, Psychiatry and Psychology. University of Pittsburgh School of Medicine, 200 Lothrop Street, Pittsburgh PA, PA 15213-2536, U.S.A
8. Diabetes Research Laboratory, Biomedical Research Foundation, University Hospital Clínico San Carlos, E – 28040, Madrid, Spain

*contributed equally

#correspondence: aruiz@neocodex.es

Abstract

Background

In order to identify novel loci associated to Alzheimer's Disease (AD) we conducted a genome-wide association study (GWAS) in the Spanish population.

Methods

We genotyped 1128 individuals using the Affymetrix Nsp I 250K chip. A sample of 327 sporadic AD patients and 801 controls with unknown cognitive status from the Spanish general population were included in our initial study. To increase the power of the study, we combined our results with those of four other public GWAS datasets by applying identical quality control filters, and the same imputation methods which were then analyzed with a global meta-GWAS. A replication sample with 2200 sporadic AD patients and 2301 controls was genotyped to confirm our GWAS findings.

Results

Meta-analysis of our data and independent replication datasets allowed us to confirm a novel genome-wide significant association of AD with membrane-spanning 4-domains subfamily A (*MS4A*) gene cluster (rs1562990, $P=4.40E-11$, odds ratio (OR)=0.88, c.i.95%[0.85-0.91], $n=10,181$ cases and 14,341 controls).

Conclusions

Our results underscore the importance of international efforts combining GWAS datasets to isolate genetic loci for complex diseases.

Keywords: GWAS, SNP, association studies, Alzheimer's Disease, molecular epidemiology, neurogenetics

Background

Alzheimer's disease (AD) is the most common neurodegenerative pathology afflicting humans. The prevalence of AD is rapidly growing due to a continuous increase in life expectancy in developed countries[1]. AD is considered a complex neurodegenerative disorder which renders a progressive neuronal loss in the brain, causing a devastating cognitive phenotype, which ends with the death of the patient.

Although AD etiology is poorly understood, genetic factors seem to play a pivotal role in AD. In fact, three genes containing multiple full penetrance mutations, *APP*, *PSEN1* and *PSEN2* have been described for mendelian AD[2-4]. A non-necessary non-sufficient common allele near *APOE* transcript is almost universally associated with non-mendelian AD[5]. In spite of research efforts in AD genetics, until very recently, no other genetic risk factor has been consistently associated to AD phenotype. However, recent advances in genome wide association study (GWAS) techniques have permitted the isolation of four uncontroversial meta-GWAS significant ($p < 5 \times 10^{-8}$) genetic markers associated to AD which are located near *CLU*, *PICALM*, *CR1* and *BIN1* genes[6-8]. No other result derived from genetic studies has been consistently validated for AD other than these loci.

Methods

Samples and datasets

In order to identify new AD associated SNPs, we designed a new case-control GWAS in the Spanish population. We genotyped 1128 individuals using Affymetrix Nsp I 250K chip as previously described[9]. A sample of 327 sporadic AD patients diagnosed as possible or probable AD in accordance with NINCDS-ADRDA criteria[10] by neurologists at the Virgen de Arrixaca University Hospital in Murcia (Spain) and 801 controls with unknown cognitive status from the Spanish general population were included in our initial study. Mean (SD) age at recruitment were 79.1 (6.8) years in cases and 52.0 (8.9) in controls. The corresponding number (%) of female samples were 228 (71.5%), and 348 (45.4%) respectively. Mean (SD) age at AD diagnosis in cases was 76.2 (6.9) years. Informed consent was obtained from each blood donor. IRB approval for this research was obtained from the regional Ministry of Health (Comunidad Autónoma de Murcia) and conforms to the World Medical Association's Declaration of Helsinki.

To increase the power of our study to detect small genetic effects, we combined our results with those of four other public GWAS[11-14] GWAS datasets in the analysis include: a) the Alzheimer's Disease Neuroimaging Initiative (ADNI) longitudinal study which is aimed at identifying biomarkers of AD by using Illumina 610Quad platform and extensive Neuroimaging techniques. A total of 187 early AD cases and 229 elderly controls were initially identified to be included in this study[15]. ADNI Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database[16]. The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a \$60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), and other biological markers are related to the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California - San Francisco. ADNI is the result of efforts of many co- investigators from a broad range of academic institutions

and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 adults, ages 55 to 90, to participate in the research - approximately 200 cognitively normal older individuals to be followed for 3 years, 400 people with MCI to be followed for 3 years and 200 people with early AD to be followed for 2 years." For up-to-date information, visit ADNI's webpage[16] . b) GenADA study which includes 801 cases meeting the NINCDS-ADRDA and DSM-IV criteria for probable AD and 776 control subjects without family history of dementia that were genotyped using Affymetrix 500K GeneChip Array set [12, 17]. c) The National Institute of Aging (NIA) Genetic Consortium for Late Onset Alzheimer's Disease Study. Data from this study originally included 1,985 cases and 2,059 controls genotyped with Illumina Human 610Quad platform [13]. However using family trees provided we excluded all related controls and kept only one case per family. A total of 1,077 cases and 876 unrelated controls were eligible for our study. d) TGEN GWAS study included 643 LOAD cases and 404 controls from a neuropathological cohort and 197 LOAD cases and 114 controls from a clinical cohort all genotyped with Affymetrix 500K GeneChip Array set [11].

Aggregated data available from Harold et al[7] and Hu et al.[18] were also used as "in silico" replication studies. Available data from Harold et al include allelic OR estimates and p-values for the 731 top signals from their study of 3,941 cases and 7,848 controls. A comprehensive list of allelic OR estimates and p-values for 451,001 SNPs was obtained from Hu et al. supplementary material. These data corresponded to the GWAS described in their manuscript that includes 1,034 cases and 1,186 controls.

Finally a replication sample with 2,200 sporadic AD patients diagnosed as possible or probable AD in accordance with NINCDS-ADRDA criteria by neurologists at Fundació ACE in Barcelona (Spain) and Hospital de Cantoblanco (Madrid), along with 2,301 general population controls was used. Mean (SD) age at recruitment in this sample were 82.0 (7.7) years in cases and 54.7 (12.4) in controls. The corresponding number (%) of female samples were 1,559 (71.0%), and 1,540 (67.1%) respectively. Mean (SD) age at AD diagnosis was 77.9 (7.6) years.

GWAS Quality Control (QC) analyses.

We performed an extensive quality control on the five datasets with individual genotypes included in the analysis (Murcia, ADNI, GenADA, NIA, TGEN) using Affymetrix Genotyping Console software and Plink[19]. For our genotyped samples, only individuals with a sample call rate above 93% were later re-called with the Bayesian Robust Linear Model with Malalanobis

(BRLMM) distance algorithm, ran with default parameters, which improves call rates in most samples. Self-reported sex was compared to sex assigned by chromosome X genotypes, and discrepancies were resolved or samples removed. For all datasets, the program Graphical Representation of Relationships (GRR)[20] was used to check sample relatedness and to correct potential sample mixups, duplications, or contaminations. SNPs were selected to have a call rate above 95% (in each case, control, and combined group, within each dataset), and a minor allele frequency above 1% (again in each case, control, and combined group, within each dataset). SNPs that deviated grossly from Hardy-Weinberg equilibrium ($p\text{-value} < 10^{-4}$) in control samples were also removed. We also removed SNPs with a significantly different rate of missingness ($p\text{-value} < 5 \times 10^{-4}$) between case and control samples within each dataset.

To ensure all SNPs across all datasets were typed according to the same DNA strand, each dataset was normalized using HapMap phase 2 data as the reference set. We merged each study with the HapMap CEU samples and compared genotype calls. SNP calls were flipped (if typed on opposite strand) or removed (if strand could not be undoubtedly assigned) as necessary. We also removed SNPs that were significantly associated with “study status”. That is, we labeled control individuals from each study as cases and HapMap CEU individuals as controls, and removed SNPs with p -values smaller than 10^{-6} in a test for association.

Principal Components (PC) analysis.

Principal components analysis was carried out with EIGENSOFT [21, 22] to evaluate population admixture within each population, and to identify individuals as outliers. We ran the SMARTPCA program with default parameters, excluding chromosome X markers. To minimize the effect of LD in the analysis, we also excluded markers in high LD (with the independent option in Plink) and long-range LD regions reported previously or detected in our population. Individuals identified as outliers (6 standard deviations or more along one of the top ten principal components) were removed from all subsequent analyses. PC analysis was run within each dataset, and also together with other HapMap European and worldwide populations to detect individuals of different ethnicities.

Imputation

Since different platforms were used in the five GWAS analyzed, we imputed genotypes using HapMap phase 2 CEU founders ($n=60$) as reference panel using two different methodologies: Plink [19] and Mach[23]. Genome-wide imputation was carried out with plink, and genotype

calls with high quality scores were used in subsequent association analyses. Best association results were also imputed with Mach 1.0, to confirm the consistency of imputed genotypes.

After all these quality control and preparatory steps, the Murcia study kept 1,034,239 SNPs for 319 cases and 769 controls; the ADNI dataset kept 1,794,894 SNPs for 164 cases and 194 controls; the GenADA dataset kept 1,436,577 SNPs for 782 cases and 773 controls; the NIA dataset kept 1,738,663 SNPs for 987 cases and 802 controls.; and the TGEN dataset contained 1,237,568 SNPs in 757 cases and 468 controls. A total of 696,707 SNPs were common to all GWAS studies whereas 1,098,485 and 1,951,797 SNPs were common to at least four and two studies respectively.

Replication genotyping

The MS4A cluster polymorphism rs1562990 was genotyped in 2,200 cases and 2,301 controls from the Spanish population using Real-time PCR coupled to Fluorescence Resonance Energy Transfer (FRET). Primers and probes employed for these genotyping protocols are summarized in Additional file 1, Table S1. The protocols were performed in the LightCycler[®] 480 System instrument (Roche Diagnostics). Briefly, PCR reactions were performed in a final volume of 20 μ l using 20 ng of genomic DNA, 0.5 μ M of each amplification primer, 0.20 μ M each detection probe, and 4 μ L of LC480 Genotyping Master (5X, Roche Diagnostics, Germany). We used an initial denaturation step of 95°C for 5 min, followed by 45 cycles of 95°C for 30 sec, 55°C for 30 sec, and 72°C for 30 sec. Melting curves were 95°C for 2 min s (ramping rate 4.4 °C/s), 62°C for 30 s (ramping rate of 1°C/s), 40°C for 30 s (ramping rate of 1°C/s), and 68°C for 0 s (ramping rate of 0.15°C/s). In the last step of each melting curve, a continuous fluorimetric register was performed by the system at one acquisition register per °C. Melting peaks and genotype calls were obtained by using the LightCycler[®]480 software (Roche). In order to confirm genotypes selected PCR amplicons were bi-directionally sequenced using standard capillary electrophoresis techniques.

Association analysis

Unadjusted single-locus allelic (1 df) association analysis within each independent sample, and of the combined sample, was carried out using Plink. We combined data from these five GWAS datasets using meta-analysis tool in Plink selecting only those markers common to at least four of these studies (1,098,485 SNPs). The most promising and consistent results from these single-locus analyses were compared to the aggregated estimates available from Harold et al

[7]and Hu et al[18]. Finally a replication sample of 2,200 cases and 2,301 controls from the Spanish population was used to validate rs1562990. Although the main results of the study are unadjusted estimates and p-values from the allelic test, multivariate logistic regression models were also used to adjust estimates for the combined Spanish samples (Murcia GWAS and the Replica) by age, sex, and APOE E+ status using Logistic option in Plink. A final meta-analysis and Forest plot for marker rs1562990 including the five original GWAS, plus the two “in silico” replica, and the final replica, was done with Stata 10.0 (College Station, TX) metan command.

Results and Discussion

The meta-analysis of the five GWAS (Murcia, ADNI, GenADA, NIA, and TGEN) included a total of 3,009 cases and 3,006 controls. A total of 696,707 SNPs were common to all GWAS whereas 1,098,485 SNPs were common to at least four studies. Figure 1 shows a Manhattan plot with the results of this GWAS Meta-analysis. We identified several signals, most of them found in previously reported AD loci (Additional file 2, Table S2). The only GWAS significant result ($p=4.71 \times 10^{-15}$) corresponded to rs10402271 in chromosome 19, a marker located 78kb upstream from the *APOE* locus. Other suggestive signals were located in chromosome 2 (rs7561528 located 25kb downstream from the *BIN1* locus), chromosome 22 (rs7561528 and rs13447284), and multiple regions within chromosome 11. In fact, among the top 100 markers, 45 were located on chr. 11 (Additional file 3, Figure S1). Chr. 11 contains several independent suggestive association signals including the *HBG2* locus (peak association at rs10838245, $p=1.04 \times 10^{-5}$), *MSE4A* gene family cluster (peak association at rs7626344, $p=5.48 \times 10^{-6}$), *GAB2* (rs450128, $p=2.79 \times 10^{-6}$), downstream *PICALM* (rs4944558, $p=1.50 \times 10^{-4}$), and downstream putative gene mRNA BC038205 (rs7935502, $p=7.47 \times 10^{-5}$).

We then conducted an “in silico” replication of our results using aggregated data from Harold et al (which includes the top 731 signals from their study, many of them also located in chromosome 11) and Hu et al (a comprehensive rank of 451,001 SNPs genotyped in their GWAS)[7, 18]. Although limited by the number of SNPs available from these studies, the new meta-analysis yielded quite interesting results with a total of 17 markers above the GWAS significance level (Additional file 4, Table S3). Several signals belonged to known AD loci: *APOE* with eight SNPs, *PICALM* (three SNPs, the most significant being rs536841, $p=2.96 \times 10^{-9}$), *CLU* (rs569214, $p=4.11 \times 10^{-8}$), and *BIN1* (rs744373, $p=2.13 \times 10^{-9}$). Most important, we found four SNPs that belong to a region in chromosome 11q12 not previously reported as GWAS significant for AD. The new peak for AD is located within the *MS4A* cluster and the most significant SNPs was rs1562990 (OR: 0.87; $p=3.01 \times 10^{-10}$).

Since we have previously published replication studies of *APOE*, *CLU*, *PICALM* and *BIN1* signals in the Spanish population[8, 24], we decided to replicate only rs1562990 in 2,200 cases and 2,301 controls from the Spanish population. Importantly, the result of this new independent replica was fully consistent yielding a significant OR of 0.90 (95%CI:0.83-0.98; $p=.01$). Detailed results for the original Spanish GWAS dataset, Spanish replica sample, and the combined Spanish dataset are described in Additional file 5, Table S4. We fitted a multivariate logistic

regression model for the combined Spanish sample in which we adjusted for age, sex and *APOE*. The adjusted OR estimate was virtually unchanged (OR 0.87; 95%CI: 0.74-1.04; $p=0.12$) suggesting that the observed effect is not influenced by age, sex or *APOE* in our series.

Finally, combining this new replication in a final meta-analysis together with the five original GWAS and the two “in silico” replications yields an OR of 0.88 (95%CI 0.85-0.91; $p=4.4E-11$) which exceeds the accepted threshold for testing multiple comparisons (i.e., $p<5E-8$). A total of 10,181 cases and 14,341 controls are included in this combined analysis. The magnitude of effect is consistent across studies, with all ten estimates between 0.74 and 0.91 (See Figure 2).

Our results point to the existence of a new AD locus located within *MS4A* cluster at 11q12. Coincidentally, during the drafting of this manuscript two independent articles have emerged reaching similar conclusions regarding *MS4A* cluster involvement in AD[25, 26]. Certainly, the SNP markers described in the three studies are different, but they are only 83,871 bp apart. However, our signal is closer to rs4938933 (reported by Naj et al., 2011)[27] which is only 9 kb centromeric to rs1562990. In any case, peak markers observed in these studies are located in the same haplotypic block and have identical effect size and direction which strongly suggest that they are tracking the same functional variant.

It is important to mention that there exists sampling overlapping between these studies. Nonetheless, at least three full datasets contained in our study (comprising 7809 individuals, 31%) do not overlap with previous published works. Importantly, meta-analysis using only these non-overlapping samples also rendered a significant association to the *MS4A* region OR=0,897 (95%CI 0,838- 0,961; $p=0.0018$). Therefore, our study could be considered an independent replication of the involvement of the *MS4A* gene cluster on AD. The concurrence of three independent studies reaching the same conclusion by employing different SNP platforms, imputation methods and datasets underscores the strength and consistency of this new AD locus, at least in European populations. Further studies will be necessary to corroborate its involvement in AD etiology in other ethnic groups.

The *MS4A* family includes at least 16 paralogues. Each gene has been probably generated by an ancestral cascade of intrachromosomal duplications during vertebrate evolution. Unfortunately, this gene family is poorly characterized although a role in immunity has already been shown for several members this cluster, including *MS4A1* (CD20), *MS4A2* and *MS4A4B*[28]. However, the function in humans of many other members remains obscure and

a more general involvement of *MS4A* family members as ion channel adaptor proteins in non-immune tissues is suspected[28].

The rs1562990 marker maps between *MS4A4E* and *MS4A4A* members of the cluster. However, we detected a critical LD haplotype block spanning 163kb that comprises three members of the family (*MS4A2*, *MS4A6A*, *MS4A4*) and the top four meta-GWAS significant markers (Additional file 4, Table S3) . With the available data it is difficult to determine the precise location of the functional variant associated with AD, or even which gene could be the best candidate for AD etiology. Furthermore, it may be the case that a functional *non-coding* variant within the cluster might be altering, by cis-regulation, the function of other members of the cluster simultaneously. Re-sequencing and functional studies of candidate mutations could help resolve this question.

The most centromeric gene within the critical block, *MS4A2*, encodes a protein that binds to the Fc region of immunoglobulins epsilon. *MS4A2* seems responsible for initiating the allergic response by binding of allergen to receptor-bound IgE that leads to cell activation and the release of mediators (such as histamine). This signal cascade is responsible for the manifestations of allergy[29]. Indeed, polymorphisms within the *MS4A2* gene have been associated with susceptibility to aspirin-intolerant asthma[30], and some epidemiological studies suggest a link between asthma and AD [31]. Consequently, a hypothetical link between *MS4A2* and AD would add new evidence in favor of the AD neuroinflammatory hypothesis, suggesting a role for the immune system in the pathogenesis of AD. The rest of genes within the candidate block are poorly characterized and it is not easy to delineate a plausible hypothesis for them yet.

Conclusions

In summary, we report a new genetic locus associated with AD. Our work underscores the importance of the combination of new GWAS data with existing datasets in order to identify novel signals that can only emerge through meta-analysis. We are confident that the increasing sample size of GWAS, the growing number of publicly available GWAS datasets, the higher marker density and the development of novel strategies for GWAS data analysis will help isolate novel genetic signals related to AD in the future and might contribute to decreasing the missing piece of heritability in neurodegenerative disorders.

Abbreviations

ACE: Alzheimer Centre Educacional; AD: Alzheimer's Disease; ADNI: Alzheimer's Disease Neuroimaging Initiative; APOE: Apolipoprotein E gene; APP: Amyloid Precursor Protein gene; BIN1: Bridging integrator 1 gene; Bp: base pair; BRLMM: Bayesian Robust Linear Model with Malalanobis distance algorithm; CEU HapMap: Utah residents with ancestry from northern and western Europe; Chr: Chromosome; CI: Confidence Interval; CLU: Clusterin gene; CR1: complement component (3b/4b) receptor 1 gene; Df: Degree of Freedom; Fc: Constant fraction of immunoglobulins; FDA: The Food and Drug Administration; FRET: Fluorescence Resonance Energy Transfer; GAB2: GRB2-associated binding protein 2 gene; GenADA: Multi-Site Collaborative Study for Genotype-Phenotype Associations in Alzheimer's disease and Longitudinal follow-up of Genotype-Phenotype Associations in Alzheimer's disease and Neuroimaging component of Genotype-Phenotype Associations in Alzheimer's disease; GRR: Graphical Representation of Relationships; GWAS: Genome Wide Association Study; *HBG2*: hemoglobin, gamma G gene; IRB: Institutional Review Board; Kb: Kilobase; LD: Linkage Disequilibrium; LOAD: Late Onset Alzheimer's Disease; Mb: Megabase; MCI: Mild cognitive impairment; MD: Medical Doctor; MRI: Magnetic resonance imaging; mRNA: messenger ribonucleic acid; MS4A: membrane-spanning 4-domains, subfamily A; MS4A1: membrane-spanning 4-domains, subfamily A, member 1 gene; MS4A2: membrane-spanning 4-domains, subfamily A, member 2 gene; MS4A4A: membrane-spanning 4-domains, subfamily A, member 4 gene; MS4A4B: membrane-spanning 4-domains, subfamily A, member 4B gene; MS4A4E: membrane-spanning 4-domains, subfamily A, member 4E gene; NIA: National Institute on Aging; NIBIB: National Institute of Biomedical Imaging and Bioengineering; NINCDS-ADRDA: National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association; OR: Odds ratio; PC: Principal Component; PCR: Polymerase Chain Reaction; PET: Positron emission tomography; PICALM: Phosphatidylinositol binding clathrin assembly protein gene; PSEN1: Presenilin 1 gene; PSEN2: Presenilin 2 gene; SD: Standard Deviation; Sec: second; SNPs: Single Nucleotide Polymorphisms; TGEN: Translational Genomics Research Institute.

Competing interests

Some authors declare competing interests: RR-L, FJM, JV, JMC, LMR, AG-P, JG, CM-R, ARo, IB, JJG, MES, EV, and AR are employees in Neocodex SL. LMR, EV and AR are shareholders in Neocodex SL. The remaining authors declare that they have no competing interests.

Authors' contributions

Phenome characterization, database and Biobank construction: CA, MB, JM, IH, CMR, JL-A, AM, MR-R, FN-P, AL-G, LV-M, MA, MA-T, SM, SR, AE, SC, BG, BM-H, GV, AL, JTB, OLL, MS-R, LT, EV, ARo, LMR, AR.

Clinical research oversee (Spanish series): CA, MB, IH, JM, OLL, JTB.

DNA management & Genome analysis: RR-L, FJM, JV, JMC, JJG, MES, LMR, AR.

Bioinformatics, Statistical analysis & IT support: AG-P, JG, RRL, CM-R, ARo, IB, JJG, MES, AR.

Writers: AG-P, JG, JTB & AR with contributions of all authors.

Project design & Funding: CA, MB, LT, EV, LMR, AR.

Project oversee: CA, MB, AR.

All authors read and approved the final manuscript.

Acknowledgements

We would like to thank patients and controls who participated in this project. This work has been funded by the Fundación Alzheimer (Murcia), the Ministerio de Educación y Ciencia (Gobierno de España), Corporación Tecnológica de Andalucía and Agencia IDEA (Consejería de Innovación, Junta de Andalucía). The Diabetes Research Laboratory, Biomedical Research Foundation. University Hospital Clínico San Carlos has been supported by CIBER de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM); CIBERDEM is an ISCIII Project. We also are indebted to TGEN investigators who provided a free access to genotype data to other researchers via Coriell Biorepositories [32]. The genotypic and associated phenotypic data used in the study, "Multi-Site Collaborative Study for Genotype-Phenotype Associations in Alzheimer's Disease (GenADA)" were provided by the GlaxoSmithKline, R&D Limited. The datasets used for analyses described in this manuscript were obtained from dbGaP [33] through dbGaP accession number phs000219.v1.p1. Funding support for the "Genetic Consortium for Late Onset Alzheimer's Disease" was provided through the Division of Neuroscience, NIA. The Genetic Consortium for Late Onset Alzheimer's Disease includes a genome-wide association study funded as part of the Division of Neuroscience, NIA. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by Genetic Consortium for Late Onset Alzheimer's Disease. The

datasets used for analyses described in this manuscript were obtained from dbGaP [33] through dbGaP accession number phs000168.v1.p1. Furthermore, Part of data collection and sharing for Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson, Eli Lilly and Co., Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer Inc, F. Hoffman-La Roche, Schering-Plough, Synarc, Inc., as well as non-profit partners the Alzheimer's Association and Alzheimer's Drug Discovery Foundation, with participation from the U.S. Food and Drug Administration. Private sector contributions to ADNI are facilitated by the Foundation for the National Institutes of Health[34]. The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129, K01 AG030514, and the Dana Foundation. The investigators within ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. ADNI investigators include (complete listing available in Additional file 6, File S1).

Data access

GWAS data from Spanish patients is available for qualified researchers after IRB approval by the Comunidad Autónoma de la Región de Murcia (Spain). Request to Dr. Carmen Antúnez Almagro (mcarmen.antunez@carm.es).

References

1. Hauw JJ, Duyckaerts C: **Dementia, the fate of brain? Neuropathological point of view.** *C R Biol* 2002, **325**:655-664.
2. Goate A, Chartier-Harlin MC, Mullan M, Brown J, Crawford F, Fidani L, Giuffra L, Haynes A, Irving N, James L, et al.: **Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease.** *Nature* 1991, **349**:704-706.
3. Sherrington R, Rogaev EI, Liang Y, Rogaeva EA, Levesque G, Ikeda M, Chi H, Lin C, Li G, Holman K, Tsuda T, Mar L, Foncin JF, Bruni AC, Montesi MP, Sorbi S, Rainero I, Pinessi L, Nee L, Chumakov I, Pollen D, Brookes A, Sanseau P, Polinsky RJ, Wasco W, Da Silva HA, Haines JL, Pericak-Vance MA, Tanzi RE, Roses AD, et al.: **Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease.** *Nature* 1995, **375**:754-760.
4. Levy-Lahad E, Wasco W, Poorkaj P, Romano DM, Oshima J, Pettingell WH, Yu CE, Jondro PD, Schmidt SD, Wang K, et al.: **Candidate gene for the chromosome 1 familial Alzheimer's disease locus.** *Science* 1995, **269**:973-977.
5. Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, Pericak-Vance MA: **Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families.** *Science* 1993, **261**:921-923.
6. Lambert JC, Heath S, Even G, Campion D, Sleegers K, Hiltunen M, Combarros O, Zelenika D, Bullido MJ, Tavernier B, Letenneur L, Bettens K, Berr C, Pasquier F, Fievet N, Barberger-Gateau P, Engelborghs S, De Deyn P, Mateo I, Franck A, Helisalmi S, Porcellini E, Hanon O, de Pancorbo MM, Lendon C, Dufouil C, Jaillard C, Leveillard T, Alvarez V, Bosco P, et al.: **Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease.** *Nat Genet* 2009, **41**:1094-1099.
7. 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, et al.: **Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease.** *Nat Genet* 2009, **41**:1088-1093.
8. Seshadri S, Fitzpatrick AL, Ikram MA, DeStefano AL, Gudnason V, Boada M, Bis JC, Smith AV, Carassquillo MM, Lambert JC, Harold D, Schrijvers EM, Ramirez-Lorca R, Debette S, Longstreth WT, Jr., Janssens AC, Pankratz VS, Dartigues JF, Hollingworth P, Aspelund T, Hernandez I, Beiser A, Kuller LH, Koudstaal PJ, Dickson DW, Tzourio C, Abraham R, Antunez C, Du Y, Rotter JJ, et al.: **Genome-wide analysis of genetic loci associated with Alzheimer disease.** *JAMA* 2010, **303**:1832-1840.
9. Gayan J, Galan JJ, Gonzalez-Perez A, Saez ME, Martinez-Larrad MT, Zabena C, Rivero MC, Salinas A, Ramirez-Lorca R, Moron FJ, Royo JL, Moreno-Rey C, Velasco J, Carrasco JM, Molero E, Ochoa C, Ochoa MD, Gutierrez M, Reina M, Pascual R, Romo-Astorga A, Susillo-Gonzalez JL, Vazquez E, Real LM, Ruiz A, Serrano-Rios M: **Genetic structure of the Spanish population.** *BMC Genomics* 2010, **11**:326.
10. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM: **Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease.** *Neurology* 1984, **34**:939-944.
11. Reiman EM, Webster JA, Myers AJ, Hardy J, Dunckley T, Zismann VL, Joshipura KD, Pearson JV, Hu-Lince D, Huentelman MJ, Craig DW, Coon KD, Liang WS, Herbert RH, Beach T, Rohrer KC, Zhao AS, Leung D, Bryden L, Marlowe L, Kaleem M, Mastroeni D, Grover A, Heward CB, Ravid R, Rogers J, Hutton ML, Melquist S, Petersen RC,

- Alexander GE, et al.: **GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers.** *Neuron* 2007, **54**:713-720.
12. Li H, Wetten S, Li L, St Jean PL, Upmanyu R, Surh L, Hosford D, Barnes MR, Briley JD, Borrie M, Coletta N, Delisle R, Dhalla D, Ehm MG, Feldman HH, Fornazzari L, Gauthier S, Goodgame N, Guzman D, Hammond S, Hollingworth P, Hsiung GY, Johnson J, Kelly DD, Keren R, Kertesz A, King KS, Lovestone S, Loy-English I, Matthews PM, et al.: **Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease.** *Arch Neurol* 2008, **65**:45-53.
 13. Wijsman EM, Pankratz ND, Choi Y, Rothstein JH, Faber KM, Cheng R, Lee JH, Bird TD, Bennett DA, Diaz-Arrastia R, Goate AM, Farlow M, Ghetti B, Sweet RA, Foroud TM, Mayeux R: **Genome-Wide Association of Familial Late-Onset Alzheimer's Disease Replicates BIN1 and CLU and Nominates CUGBP2 in Interaction with APOE.** *PLoS Genet* 2011, **7**:e1001308.
 14. Mueller SG, Weiner MW, Thal LJ, Petersen RC, Jack CR, Jagust W, Trojanowski JQ, Toga AW, Beckett L: **Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI).** *Alzheimers Dement* 2005, **1**:55-66.
 15. Mueller SG, Weiner MW, Thal LJ, Petersen RC, Jack C, Jagust W, Trojanowski JQ, Toga AW, Beckett L: **The Alzheimer's disease neuroimaging initiative.** *Neuroimaging Clin N Am* 2005, **15**:869-877, xi-xii.
 16. **ADNI web page** [www.loni.ucla.edu\ADNI]
 17. Filippini N, Rao A, Wetten S, Gibson RA, Borrie M, Guzman D, Kertesz A, Loy-English I, Williams J, Nichols T, Whitcer B, Matthews PM: **Anatomically-distinct genetic associations of APOE epsilon4 allele load with regional cortical atrophy in Alzheimer's disease.** *Neuroimage* 2009, **44**:724-728.
 18. Hu X, Pickering E, Liu YC, Hall S, Fournier H, Katz E, Dechairo B, John S, Van Eerdewegh P, Soares H: **Meta-Analysis for Genome-Wide Association Study Identifies Multiple Variants at the BIN1 Locus Associated with Late-Onset Alzheimer's Disease.** *PLoS One* 2011, **6**:e16616.
 19. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC: **PLINK: a tool set for whole-genome association and population-based linkage analyses.** *Am J Hum Genet* 2007, **81**:559-575.
 20. Abecasis GR, Cherny SS, Cookson WO, Cardon LR: **GRR: graphical representation of relationship errors.** *Bioinformatics* 2001, **17**:742-743.
 21. Patterson N, Price AL, Reich D: **Population structure and eigenanalysis.** *PLoS Genet* 2006, **2**:e190.
 22. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: **Principal components analysis corrects for stratification in genome-wide association studies.** *Nat Genet* 2006, **38**:904-909.
 23. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR: **MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes.** *Genet Epidemiol* 2010, **34**:816-834.
 24. Ramirez-Lorca R, Boada M, Saez ME, Hernandez I, Mauleon A, Rosende-Roca M, Martinez-Lage P, Gutierrez M, Real LM, Lopez-Arrieta J, Gayan J, Antunez C, Gonzalez-Perez A, Tarraga L, Ruiz A: **GAB2 gene does not modify the risk of Alzheimer's disease in Spanish APOE 4 carriers.** *J Nutr Health Aging* 2009, **13**:214-219.
 25. Hollingworth P, Harold D, Sims R, Gerrish A, Lambert JC, Carrasquillo MM, Abraham R, Hamshere ML, Pahwa JS, Moskvina V, Dowzell K, Jones N, Stretton A, Thomas C, Richards A, Ivanov D, Widdowson C, Chapman J, Lovestone S, Powell J, Proitsi P, Lupton MK, Brayne C, Rubinsztein DC, Gill M, Lawlor B, Lynch A, Brown KS, Passmore

- PA, Craig D, et al.: **Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease.** *Nat Genet* 2011, **43**:429-435.
26. Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buross J, Gallins PJ, Buxbaum JD, Jarvik GP, Crane PK, Larson EB, Bird TD, Boeve BF, Graff-Radford NR, De Jager PL, Evans D, Schneider JA, Carrasquillo MM, Ertekin-Taner N, Younkin SG, Cruchaga C, Kauwe JS, Nowotny P, Kramer P, Hardy J, Huentelman MJ, Myers AJ, Barmada MM, Demirci FY, Baldwin CT, 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.
27. Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buross J, Gallins PJ, Buxbaum JD, Jarvik GP, Crane PK, Larson EB, Bird TD, Boeve BF, Graff-Radford NR, De Jager PL, Evans D, Schneider JA, Carrasquillo MM, Ertekin-Taner N, Younkin SG, Cruchaga C, Kauwe JS, Nowotny P, Kramer P, Hardy J, Huentelman MJ, Myers AJ, Barmada MM, Demirci FY, Baldwin CT, et al.: **Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease.** *Nat Genet* 2011.
28. Zuccolo J, Bau J, Childs SJ, Goss GG, Sensen CW, Deans JP: **Phylogenetic analysis of the MS4A and TMEM176 gene families.** *PLoS One* 2010, **5**:e9369.
29. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D: **The human genome browser at UCSC.** *Genome Res* 2002, **12**:996-1006.
30. Kim SH, Bae JS, Holloway JW, Lee JT, Suh CH, Nahm DH, Park HS: **A polymorphism of MS4A2 (- 109T > C) encoding the beta-chain of the high-affinity immunoglobulin E receptor (FepsilonR1beta) is associated with a susceptibility to aspirin-intolerant asthma.** *Clin Exp Allergy* 2006, **36**:877-883.
31. Eriksson UK, Gatz M, Dickman PW, Fratiglioni L, Pedersen NL: **Asthma, eczema, rhinitis and the risk for dementia.** *Dement Geriatr Cogn Disord* 2008, **25**:148-156.
32. **Coriell Biorepositories** [<http://www.coriell.org/>]
33. **dbGAP** [<http://www.ncbi.nlm.nih.gov/gap>]
34. **Foundation for the National Institutes of Health** [www.fnih.org]

Figure Legends

Figure 1. Manhattan plot of meta-analysis of five GWAS (Murcia, ADNI, GenADA, NIA, and TGEN), including a total of 3,009 cases and 3,006 controls. A total of 696,707 SNPs were common to all GWAS whereas 1,098,485 SNPs were common to at least four studies.

Figure 2. Meta-analysis and Forest plot of rs1562990, reporting Odds-Ratio (OR) with 95% confidence interval, study-specific weight, sample size and minor allele frequency (MAF) in cases and controls, for each study. The figure shows the remarkable consistency of the OR across studies.

Additional files

Additional file 1

Title: Table S1. Primers and probes employed for Real-time detection of MS4A cluster rs1562990 marker

Description: Molecular Information for rs1562990 genotyping.

Additional file 2

Title: Table S2. Top 100 results in the meta-analysis including 5 initial GWAS

Description: Best results obtained in our study.

Table legends: CHR:Chromosome; A1:Allele 1; A2:Allele 2; N:Number of studies in the meta-analysis contributing to the overall estimate of the marker; P:p-value from fixed effects model; P(R):p-value from random effects model; OR:Pooled Odds Ratio estimate from fixed effects model; OR(R):Pooled Odds Ratio estimate from random effects model; Q:p-value for Cochran's Q statistic; I:² heterogeneity index.

Additional file 3

Title: Table S3. GWAS plus aggregated data from Harold et al. and Hu et al.

Description: GWAS significant markers obtained after *in silico* replications.

Table legends: CHR:Chromosome; A1:Allele 1; A2:Allele 2; N:Number of studies in the meta-analysis contributing to the overall estimate of the marker; P:p-value from fixed effects model; P(Random):p-value from random effects model; OR:Pooled Odds Ratio estimate from fixed effects model; OR(Random):Pooled Odds Ratio estimate from random effects model; Q:p-value for Cochran's Q statistic; I:² heterogeneity index.

Additional file 4

Title: Table S4. *MS4A* rs1562990 MAF, Genotype distribution, effect estimates, and significance in the Spanish series.

Description: Table describing the results of *MS4A* cluster region in the Spanish Population.

Additional file 5

Title: Figure S1. Manhattan plot with results from the meta-analysis of the 5 initial GWAS for markers in chromosome 11.

Description: MetaGWAS results obtained for chromosome 11.

Additional file 6

Title: File S1. Alzheimer's Disease Neuroimaging initiative (ADNI) active investigators.

Description: Full list of ADNI co-investigators (alphabetic order).

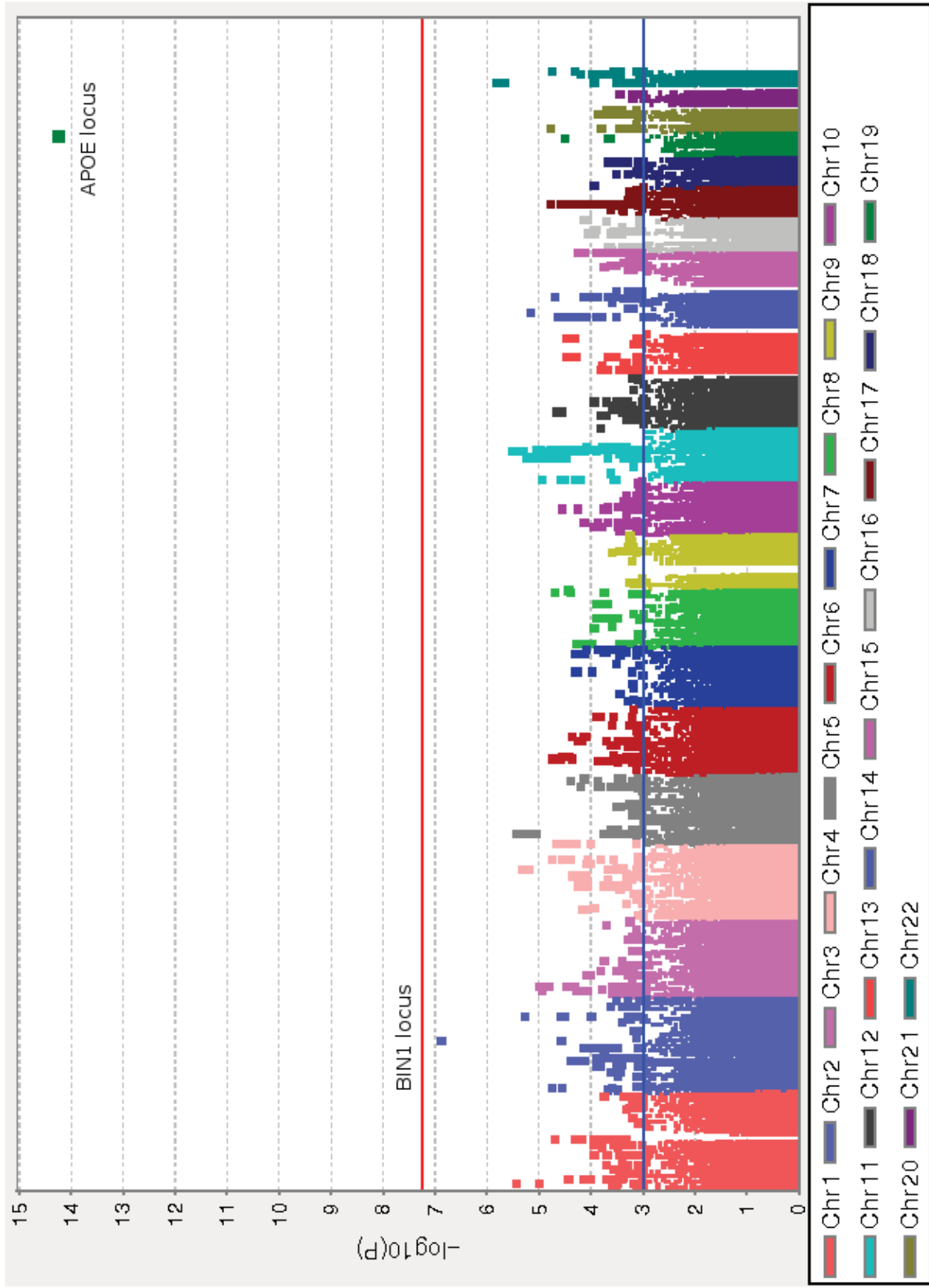


Figure 1

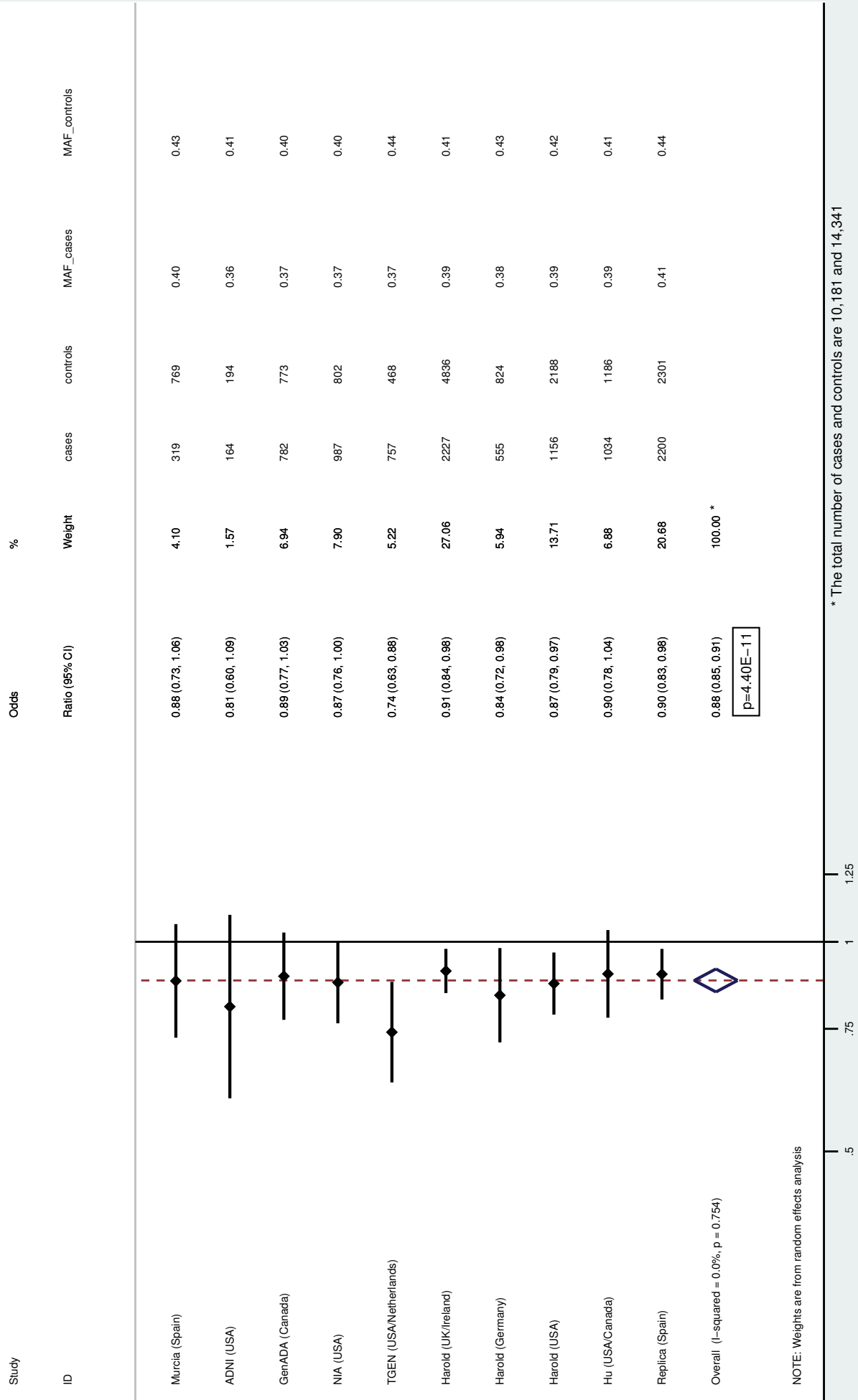


Figure 2

Additional files provided with this submission:

Additional file 1: Additional file 1.doc, 27K

<http://genomemedicine.com/imedia/1412785082555351/supp1.doc>

Additional file 2: Additional file 2.xls, 55K

<http://genomemedicine.com/imedia/2057619586555355/supp2.xls>

Additional file 3: AF3.eps, 246K

<http://genomemedicine.com/imedia/2812112615553808/supp3.eps>

Additional file 4: AF4.xls, 23K

<http://genomemedicine.com/imedia/2129478125553809/supp4.xls>

Additional file 5: AF5.doc, 30K

<http://genomemedicine.com/imedia/1825467996555380/supp5.doc>

Additional file 6: Additional file 6.doc, 41K

<http://genomemedicine.com/imedia/2064822234555356/supp6.doc>

Additional file 7: Additional file 7.doc, 24K

<http://genomemedicine.com/imedia/6427276945553560/supp7.doc>