

RESEARCH

Open Access

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

Carmen Antúñez^{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, for Alzheimer's Disease Neuroimaging Initiative⁵, Enrique Vázquez⁵, Lluís Tàrraga³, María Eugenia Sáez⁵, Oscar L López⁷, Luis Miguel Real⁵ and Agustín Ruiz^{5*}

Abstract

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

Methods: We genotyped 1,128 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 2,200 sporadic AD patients and 2,301 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 the membrane-spanning 4-domains subfamily A (*MS4A*) gene cluster (rs1562990, $P = 4.40E-11$, odds ratio = 0.88, 95% confidence interval 0.85 to 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.

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 that causes a progressive neuronal loss in the brain, resulting

in a devastating cognitive phenotype, which ends with the death of the patient.

Although its etiology is poorly understood, genetic factors seem to play a pivotal role in AD. In fact, three genes containing multiple full penetrance mutations, *APP* (amyloid precursor protein), *PSEN1* (presenilin 1) and *PSEN2* (presenilin 2), have been described for Mendelian AD [2-4]. A non-necessary, non-sufficient common allele near the *APOE* (apolipoprotein E) transcript is almost universally associated with non-Mendelian AD [5]. In spite of research efforts in AD genetics, until very

* Correspondence: aruiz@neocodex.es

† Contributed equally

⁵Department of Structural Genomics, Neocodex, Avda. Charles Darwin, Sevilla, s/n 41092, Spain

Full list of author information is available at the end of the article

recently no other genetic risk factor has been consistently associated with the AD phenotype. However, recent advances in genome wide association study (GWAS) techniques have permitted the isolation of four uncontroversial meta-GWAS-significant ($P < 5 \times E-8$) genetic markers associated with AD, which are located near the *CLU* (clusterin), *PICALM* (phosphatidylinositol binding clathrin assembly protein), *CRI* (complement component (3b/4b) receptor 1) and *BINI* (bridging integrator 1) genes [6-8]. No other result derived from genetic studies has been consistently validated for AD other than these loci.

Materials and 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 1,128 individuals using the Affymetrix Nsp I 250 K chip as previously described [9]. A sample of 327 sporadic AD patients diagnosed as possible or probable AD in accordance with the criteria of the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) [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 (standard deviation (SD)) age at recruitment was 79.1 (6.8) years in cases and 52.0 (8.9) in controls. The corresponding number (percentage) of female samples was 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. Institutional review board 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 GWASs, including the Alzheimer's Disease Neuroimaging Initiative (ADNI) longitudinal study, the GenADA study, the National Institute of Aging (NIA) Genetic Consortium for Late Onset Alzheimer's Disease study, and the Translational Genomics Research Institute (TGEN) GWAS [11-14]. The ADNI longitudinal study, which is aimed at identifying biomarkers of AD using the Illumina 610 Quad 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 ADNI database [16]. The ADNI was launched in 2003 by the 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 and early 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 reduce 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 US and Canada. The initial goal of ADNI was to recruit 800 adults aged 55 to 90 years to participate in the research - approximately 200 cognitively normal older individuals to be followed for 3 years, 400 people with mild cognitive impairment 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]. The GenADA study 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 the Affymetrix 500 K GeneChip Array set [12,17]. The NIA Genetic Consortium for Late Onset Alzheimer's Disease study originally included 1,985 cases and 2,059 controls genotyped with the Illumina Human 610 Quad 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. The TGEN GWAS study included 643 late onset AD cases and 404 controls from a neuropathological cohort and 197 late onset AD cases and 114 controls from a clinical cohort all genotyped with the Affymetrix 500 K GeneChip Array set [11].

Aggregated data 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 odds ratio (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 the supplementary material of Hu *et al.* These data correspond 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 was 82.0 (7.7) years in cases and 54.7 (12.4) in controls. The corresponding number (percentage) of female samples was 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 analyses

We performed 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 mix-ups, 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 -value $< 10^{-4}$) in control samples were also removed. We also removed SNPs with a significantly different rate of missingness (P -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 the opposite strand) or removed (if the 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 $< 10^{-6}$ in a test for association.

Principal components 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 linkage disequilibrium in the analysis,

we also excluded markers in high linkage disequilibrium (with the indep-pairwise option in Plink) and long-range linkage disequilibrium regions reported previously or detected in our population. Individuals identified as outliers (six SDs or more along one of the top ten principal components) were removed from all subsequent analyses. Principal component 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 GWASs analyzed, we imputed genotypes using HapMap phase 2 CEU founders ($n = 60$) as a 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 GWASs whereas 1,098,485 and 1,951,797 SNPs were common to at least four and two studies, respectively.

Replication genotyping

The *MS4A* (membrane-spanning 4-domains, subfamily A) 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. The protocols were performed in the LightCycler[®] 480 System instrument (Roche Diagnostics, Indianapolis, IN, USA). 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). We used an initial denaturation step of 95°C for 5 minutes, followed by 45 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 30 s. Melting curves were 95°C for 2 minutes (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 degree celsius. 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 degree of freedom) 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 the 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 the Logistic option in Plink. A final meta-analysis and Forest plot for the marker rs1562990, including the five original GWASs plus the two '*in silico*' replicas and the final replica, was done with the Stata 10.0 (College Station, TX, USA) *metan* command.

Results and discussion

The meta-analysis of the five GWASs (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. 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). The only GWAS-significant result ($P = 4.71 \times 10^{-15}$) corresponded to rs10402271 in chromosome 19, a marker located 78 kb upstream of the *APOE* locus. Other suggestive signals were located in chromosome 2 (rs7561528, located 25 kb downstream of the *BINI* locus), chromosome 22 (rs7561528 and rs13447284), and multiple regions within chromosome 11. In fact, among the top 100 markers, 45 were located on chromosome 11 (Additional file 3). Chromosome 11 contains several independent suggestive association signals, including the *HBG2* (hemoglobin, gamma G) locus (peak association at rs10838245, $P = 1.04E-5$), *MSE4A* gene family cluster (peak association at rs7626344, $P = 5.48E-6$), *GAB2* (GRB2-associated binding protein 2;

rs450128, $P = 2.79E-6$), downstream *PICALM* (rs4944558, $P = 1.50E-4$), and putative downstream gene BC038205 (rs7935502, $P = 7.47E-5$).

We then conducted an '*in silico*' replication of our results using aggregated data from Harold *et al.* [7] (which includes the top 731 signals from their study, many of them also located in chromosome 11) and Hu *et al.* [18] (a comprehensive rank of 451,001 SNPs genotyped in their GWAS). 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). Several signals belonged to known AD loci: *APOE* with eight SNPs, *PICALM* (three SNPs, the most significant being rs536841, $P = 2.96E-9$), *CLU* (rs569214, $P = 4.11E-8$), and *BINI* (rs744373, $P = 2.13E-9$). 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 SNP was rs1562990 (OR 0.87; $P = 3.01E-10$).

Since we have previously published replication studies of *APOE*, *CLU*, *PICALM* and *BINI* signals in the Spanish population [8,24], we decided to replicate only rs1562990 in 2,200 cases and 2,301 controls from the this population. Importantly, the result of this new independent replica was fully consistent, yielding a significant OR of 0.90 (95% confidence interval (CI) 0.83 to 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. 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 to 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 GWASs and the two '*in silico*' replications yields an OR of 0.88 (95% CI 0.85 to 0.91; $P = 4.4E-11$), which exceeds the accepted threshold for testing multiple comparisons (that is, $P < 5E-8$). 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 (Figure 2).

Our results point to the existence of a new AD locus located within the *MS4A* cluster at 11q12. Coincidentally, during the drafting of this manuscript two independent articles 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

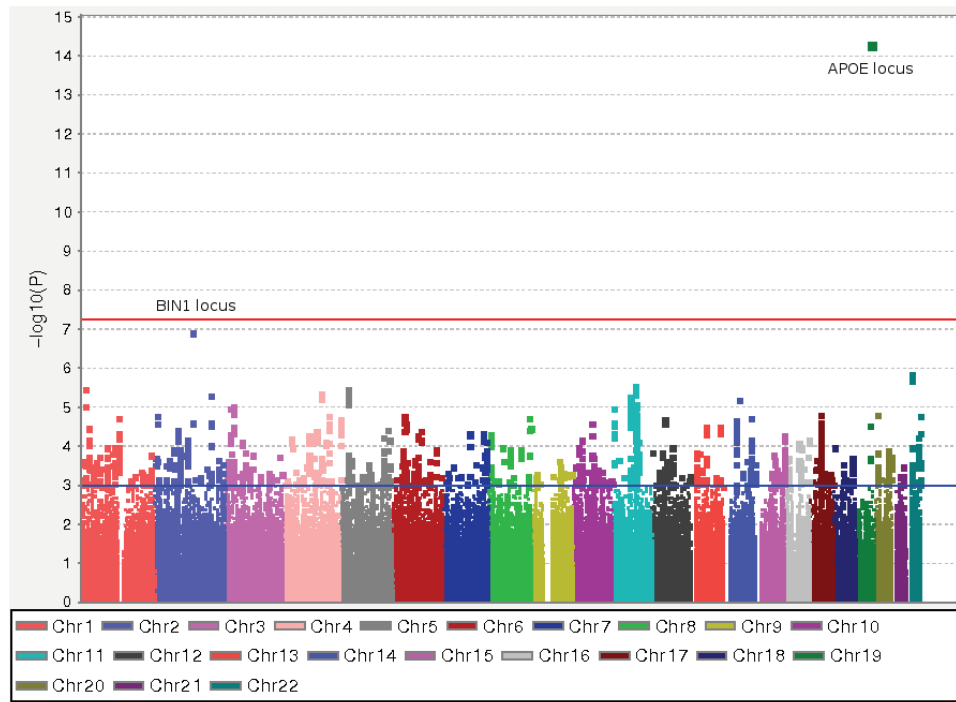


Figure 1 Manhattan plot of meta-analysis of five GWASs (Murcia, ADNI, GenADA, NIA, and TGEN), including a total of 3,009 cases and 3,006 controls. A total of 696,707 SNPs were common to all GWASs whereas 1,098,485 SNPs were common to at least four studies.

et al. [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 sample overlapping exists between these studies. Nonetheless, at least three full datasets contained in our study (comprising 7,809 individuals, 31%) do not overlap with previous published works. Importantly, meta-analysis using only these non-overlapping samples also rendered a significant association with the *MS4A* region (OR = 0.897; 95% CI 0.838 to 0.961; $P = 0.0018$). Therefore, our study could be considered an independent replication of the involvement of the *MS4A* gene cluster in 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 linkage disequilibrium haplotype block spanning 163 kb that comprises three members of the family (*MS4A2*, *MS4A6A*, and *MS4A4*) and the top four meta-GWAS-significant markers (Additional file 4). With the available data it is difficult to determine the precise location of the functional variant associated with AD, or

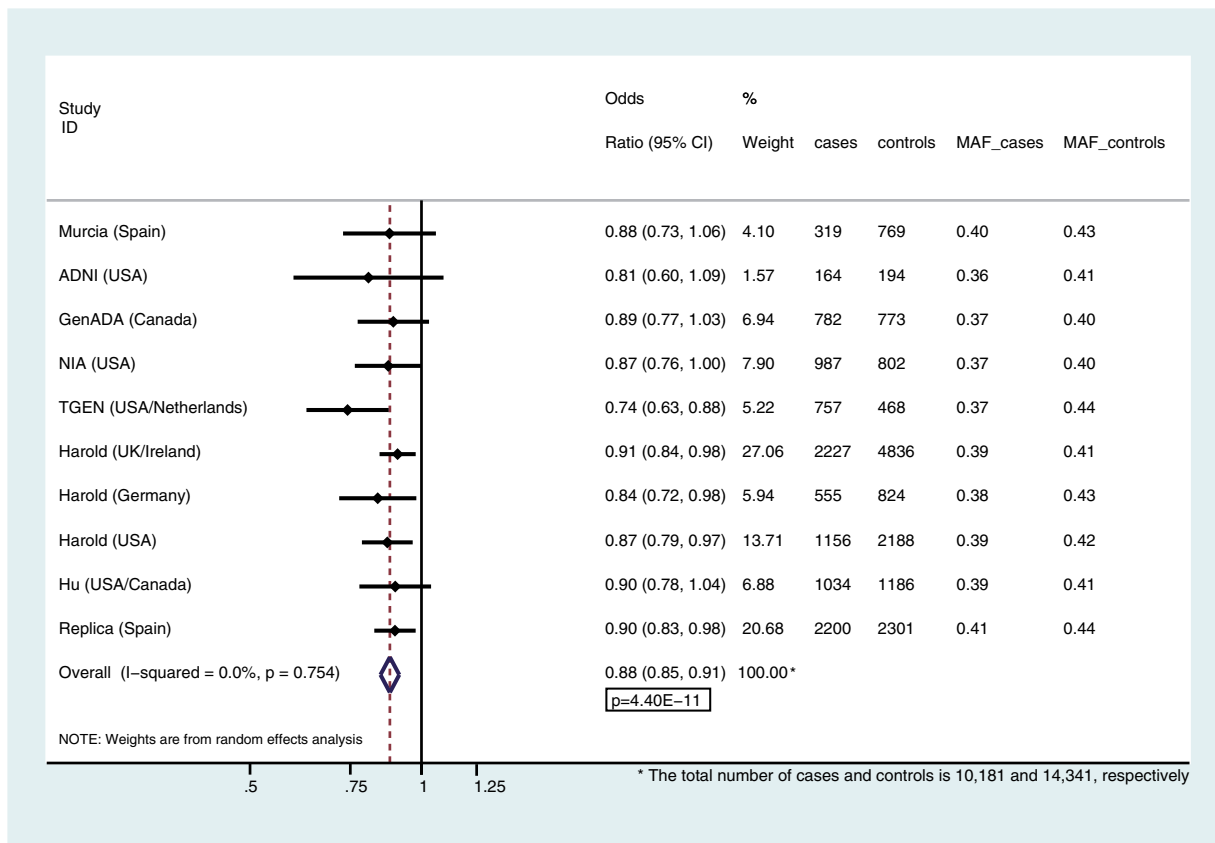


Figure 2 Meta-analysis and Forest plot of rs1562990, reporting odds ratio (OR) with 95% confidence interval (CI), study-specific weight, sample size and minor allele frequency (MAF) in cases and controls, for each study. The figure shows the remarkable consistency of the OR across studies.

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 alter, 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 immunoglobulin epsilon. *MS4A2* seems responsible for initiating the allergic response by binding of allergen to receptor-bound IgE, which 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 other genes within the candidate block are poorly characterized and it is not easy to delineate a plausible hypothesis for them yet.

Data access

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

Conclusions

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 GWASs, 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.

Additional material

Additional file 1: Table S1 - primers and probes employed for Real-time detection of MS4A cluster rs1562990 marker. Molecular Information for rs1562990 genotyping.

Additional file 2: Table S2 - top 100 results in the meta-analysis including five initial GWAS. Best results obtained in our study. 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, I² heterogeneity index.

Additional file 3: Table S3 - GWAS plus aggregated data from Harold *et al.* and Hu *et al.* GWAS-significant markers obtained after *in silico* replications. 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, I² heterogeneity index.

Additional file 4: Table S4 - MS4A rs1562990 minor allele frequency (MAF), Genotype distribution, effect estimates, and significance in the Spanish series. Table describing the results of MS4A cluster region in the Spanish population.

Additional file 5: Figure S1 - Manhattan plot with results from the meta-analysis of the five initial GWASs for markers in chromosome 11. MetaGWAS results obtained for chromosome 11.

Additional file 6: File S1 - Alzheimer's Disease Neuroimaging initiative (ADNI) active investigators. Full list of ADNI co-investigators (alphabetical order).

Abbreviations

AD: Alzheimer's disease; ADNI: Alzheimer's Disease Neuroimaging Initiative; bp: base pair; CI: confidence interval; GWAS: genome-wide association study; kb: kilobase; Mb: megabase; MS4A: membrane-spanning 4-domains, subfamily A; NIA: National Institute on Aging; NINCDS-ADRD: National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association; OR: odds ratio; PCR: polymerase chain reaction; SD: standard deviation; SNP: single nucleotide polymorphism; TGEN: Translational Genomics Research Institute.

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 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 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 GWAS 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, parts of data collection and sharing for this project was funded by the Alzheimer's Disease 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 US 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. A complete list of ADNI investigators is available in Additional file 6.

Author details

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

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 oversight (Spanish series): CA, MB, IH, JM, OLL, JTB. DNA management and genome analysis: RR-L, FJM, JV, JMC, JGG, MES, LMR, AR. Bioinformatics, statistical analysis and IT support: AG-P, JG, RRL, CM-R, ARo, IB, JGG, MES, AR. Writers: AG-P, JG, JTB and AR with contributions from all authors. Project design and funding: CA, MB, LT, EV, LMR, AR. Project oversight: CA, MB, AR. All authors read and approved the final manuscript.

Competing interests

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

Received: 23 April 2011 Revised: 19 May 2011 Accepted: 31 May 2011
Published: 31 May 2011

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, Rogaeve EI, Liang Y, Rogaeve 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, Moskva 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, Carrasquillo 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. Gayán J, Galán JJ, Gonzalez-Perez A, Saez ME, Martínez-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 JW, 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.** [http://adni.loni.ucla.edu/].
17. Filippini N, Rao A, Wetten S, Gibson RA, Borrie M, Guzman D, Kertesz A, Loy-English I, Williams J, Nichols T, Whitchee 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, Gayán 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, Moskva 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, Buros 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, Buros 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.
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 (FcepsilonR1beta) 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.** [http://www.fnih.org/].

doi:10.1186/gm249

Cite this article as: Antúñez et al: The membrane-spanning 4-domains, subfamily A (MS4A) gene cluster contains a common variant associated with Alzheimer's disease. *Genome Medicine* 2011 **3**:33.