ATP5H/KCTD2 locus is associated with Alzheimer’s disease risk

M Boada1,2,15, C Antúnez3,15, R Ramírez-Lorca4,15, AL DeStefano5,6, A González-Pérez4, J Gayán4, J López-Arrieta7, MA Ikrar8, I Hernández1, J Marin5, JJ Galán5, JC Bis9, A Mauleón1, M Rosende-Roca1, C Moreno-Rey5, V Gudnason10, FJ Morón4, J Velasco4, JM Carrasco5, M Alegret1, A Espinosa1, G Vinyes1, A Lafuente1, L Vargas1, AL Fitzpatrick9, for the Alzheimer’s Disease Neuroimaging Initiative16, LJ Launer11, ME Sáez4, EV Azáquez4, JT Becker12, OL López12, M Serrano-Ríos13, L Tarraga1, CM van Duijn8, LM Real4, S Seshadri1,6,11 and A Ruiz1,4

To identify loci associated with Alzheimer disease, we conducted a three-stage analysis using existing genome-wide association studies (GWAS) and genotyping in a new sample. In Stage I, all suggestive single-nucleotide polymorphisms (at \( P < 0.001 \)) in a previously reported GWAS of seven independent studies (8082 Alzheimer’s disease (AD) cases; 12 040 controls) were selected, and in Stage II these were examined in an in silico analysis within the Cohorts for Heart and Aging Research in Genomic Epidemiology consortium GWAS (1367 cases and 12904 controls). Six novel signals reaching \( P < 5 \times 10^{-6} \) were genotyped in an independent Stage III sample (the Fundació ACE data set) of 2200 sporadic AD patients and 2301 controls. We identified a novel association with AD in the adenosine triphosphate (ATP) synthase, H+ transporting, mitochondrial F0 (ATP5H)/Potassium channel tetramerization domain-containing protein 2 (KCTD2) locus, which reached genome-wide significance in the combined discovery and genotyping sample (rs11870474, odds ratio (OR) = 1.58, \( P = 2.6 \times 10^{-7} \) in discovery and OR = 1.43, \( P = 0.004 \) in Fundació ACE data set; combined OR = 1.53, \( P = 4.7 \times 10^{-5} \)). This ATP5H/KCTD2 locus has an important function in mitochondrial energy production and neuronal hyperpolarization during cellular stress conditions, such as hypoxia or glucose deprivation.

Molecular Psychiatry (2014) 19, 682–687; doi:10.1038/mp.2013.86; published online 16 July 2013

Keywords: Alzheimer’s disease; genomics; GWAS; molecular epidemiology; SNP

INTRODUCTION

Alzheimer’s disease (AD) is the most common cause of dementia. It is expected that AD prevalence will be quadrupled by 2040, reaching a worldwide number of 81.1 million affected individuals.\(^1\) In spite of the knowledge that genetic factors may account for about 60–80% of AD susceptibility,\(^2\) the APOE epsilon 4 allele was, until very recently, the only accepted risk factor for late-onset AD (LOAD).\(^3\) Fortunately, genome-wide association study (GWAS) technologies are rapidly transforming our knowledge of susceptibility factors related to LOAD. Specifically, in the past three years, nine additional loci located in or adjacent to clusterin (CLU), PICALM, CR1, BIN1, the MS4A gene cluster, ABCA7, EPHA1, CD33 and CD2AP, have been identified.\(^4–9\) There is no obvious relationship between the most of these novel loci and the current models of the pathogenesis of AD (that is, the amyloid and tau hypotheses), rather the novel genes identified point to immune system function, cholesterol metabolism and synaptic cell membrane processes as important in determining the risk of LOAD.\(^10\) However, researchers are intensively looking for direct relationships between these novel loci and amyloid deposition speculating that new genes might have effects on amyloid metabolism or through previously unsuspected pathophysiological pathways, and indeed preliminary evidence for relationships between the amyloid hypothesis and some of the novel loci is rapidly emerging.\(^11\)

Specifically, it has been reported that PICALM has a role in beta-amyloid membrane trafficking in yeast models.\(^12\) Furthermore using highly sensitive single-molecule fluorescence methods, Narayan et al.\(^12\) have established a direct link between the CLU protein and beta-amyloid toxicity, observing that beta amyloid forms a heterogeneous group of small oligomers (from dimers to 50-mers), all of which interact with the sequestering clu protein to form long-lived complexes.\(^12\)

Overall, the known loci explain only a small fraction of the known heritability of polygenic AD. In this paper, we present the...
Results of a collaborative effort to identify additional AD genes. We followed up on all suggestive (P < 0.001) results in our previously published GWAS (Stage I) with in silico analysis using unpublished data from a previously reported GWAS in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium (Stage II). Six novel single-nucleotide polymorphisms (SNPs) that reached a \( P < 5 \times 10^{-6} \) were genotyped in an independent data set (Stage III). This sequential analysis and novel genotyping allowed us to identify a new AD locus (adrenosine triphosphate (ATP) synthase, H+ transporting, mitochon- drial F0/Potassium channel tetramerization domain-contain- ing protein 2 (ATP5H/KCTD2) at 17q25.1.

**PATIENTS AND METHODS**

**Setting and participants**

**Stage I meta-GWAS.** We followed up the results obtained in our initial meta-analysis described previously. Briefly, we undertook GWAS on a sample of 319 sporadic AD patients diagnosed with possible or probable AD and 801 population-based controls. Due to the limited power of our sample to detect small genetic effects, we combined our data with the individual level data from four other publicly available GWAS: TGEN (Translational Genomics Research Institute; 757 cases and 468 controls), ADNI (Alzheimer Disease Neuroimaging Initiative; 164 cases and 194 controls),4 and the GERAD (Genetic and Environmental Risk in silico analyses and selection of SNPs for genotyping follow-up. The procedure to select candidate SNPs is detailed in Supplementary Figure S1. Briefly, we designed a multi-stage strategy to prioritize SNPs for further de novo genotyping in the Fundacio´ ACE data set. In the first stage, we selected a relatively large number of SNPs by establishing a permissive cutoff in our original meta-GWAS (P < 0.001). A total of 1202 SNPs met this threshold, and results for these SNPs were meta-analyzed with results from the CHARGE GWAS. Thirty-five SNPs with \( P < 5 \times 10^{-6} \) in the joint analysis were selected for follow-up and mapped in the UCSC genome browser. Twenty-eight SNPs located within known AD loci were excluded. Seven novel SNPs reaching a predetermined suggestive \( P \)-value \( (P < 5 \times 10^{-5}) \) but outside known loci were selected for the final genotyping step in the Fundacio´ ACE data set. Of note, based on 1000 genotypes, we observed that two markers (rs2896209 and rs4406992) were physically close and displayed strong linkage disequilibrium (LD; \( r^2 = 0.950 \)). We decided to analyze only one of them. Consequently, the rs2896209 SNP within SLC24A4 locus was excluded due to strong LD with and close proximity to the rs4406992 marker (Supplementary Figure S1).

So, we finally selected six SNPs within new candidate regions for further follow-up. The sample size, the effective sample size, the data sets that were informative and the genotype status (imputed or genotyped) for each selected marker are detailed in Supplementary Table S2.

**Genotyping.** Selected candidate SNPs were genotyped in the Fundacio´ ACE data set using real-time PCR coupled to fluorescence resonance energy transfer (FRET) technology, as previously described using MagnapletTM technology (MegaTide Diagnostics, Mannheim, Germany). Of note, all samples were centralized and processed in the same location (Neocodex DNA Laboratory, Seville, Spain). Identical DNA extraction methods, quality controls, equipment and personnel were applied for the entire genotyping project. Primers and probes designed for genotyping protocols are summarized in Supplementary Table S3. The protocols were performed in the LightCycler 480 System instrument (Roche Diagnostics). PCR reactions were performed in a final volume of 20 \( \mu l \) using 20 ng of genomic DNA. 0.5 \( \mu l \) of each amplification primer, 0.2 \( \mu l \) of each detection probe and 4 \( \mu l \) of LC480 Genotyping Master 5X (Roche Diagnostics). We used an initial denaturation step of 95 °C for 5 min, followed by 45 cycles of 95 °C for 30 s, 56 °C for 30 s and 72 °C for 30 s. Melting curves were 95 °C for 2 min (ramping rate 4.4 °C \cdot s^{-1} \), 45 °C for 30 s (ramping rate of 1 °C \cdot s^{-1} \) and 70 °C for 0 s (ramping rate of 0.15 °C \cdot s^{-1} \). In the last step of each melting curve, a continuous fluorimetric register was performed by the system at one acquisition register per each degree Celsius. Melting peaks and genotype calls were obtained by using the LightCycler480 software (Roche). In order to confirm SNPs, selected PCR amplicons were bi-directionally sequenced using standard capillary electrophoresis techniques.

**Statistical analysis.** Association analyses in Stage I were carried out using an allelic association test model with no covariates, as implemented in the software Plink (http://pngu.mgh.harvard.edu/~purcell/plink), to obtain univariate \( \chi^2 \) tests of the effect size and \( \beta \)-values. SNPs were selected only from the autosomal chromosomes. X, Y and mitochondrial SNPs were excluded. The filters for genotyping completeness, imputation quality, minor allele frequency have been described previously. Briefly, SNPs were selected to have a call rate > 95% in each case, control and combined group, within each data set, and a minor allele frequency > 1% (again in each case, control and combined group, within each data set). SNPs that deviated grossly from Hardy–Weinberg equilibrium (\( P \)-value < \( 10^{-4} \)) in control samples were 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 data set.

**Meta analyses in Stages I, II and III were conducted using inverse variance method (fixed effects model) and random effects model in PLINK’s ‘meta’ option.** We presented random effects meta-analysis results only when heterogeneity was observed (Q-test was statistically significant). The original GWAS in Stage I were treated as separate studies, CHARGE GWAS results were treated as a single additional study. The weighting of each study was calculated using the estimated \( \tau \)-value. Genome-wide significant and highly suggestive \( P \)-value thresholds were established at \( P < 5 \times 10^{-8} \) and \( P < 5 \times 10^{-5} \), respectively. Final meta-analysis results and Forest plot for rs11870474 showing association results in the original meta-GWAS, the CHARGE data and the Fundacio´ ACE data set were derived using the Stata 10.0 (College Station, TX, USA) ‘metan’ command. Global \( P \)-values were calculated in different ways using PLINK or Episheet software (academic software non-commercial).
Multivariate logistic regression models were used to adjust the effect estimates for our top SNP, rs11870474, using age, sex and/or principal components (PCs) and the presence of APOE E4 as covariates in data sets wherein these data were available. These analyses were conducted in SPSS 18.0 software (IBM, Armonk, NY, USA) evaluating a dominant model for the minor allele (CC vs CA + AA genotypes).

Power calculations were done with EpiSheet spreadsheet (http://www.drugepi.org/links/downloads/episheet.xls). The basic idea was to calculate minimum sample size necessary to have 80% power to detect a moderate effect of rs11870474 SNP assuming z-alpha = 1.96, case/control ratio = 1, exposure prevalence 3% and different odds ratio (OR) effects (1.43 and 1.53).

Graphical Representation of Relationships (GRR) software was used to estimate identity by state (IBS) mean values in all individual pairs and visualize the resulting relationships. Any potential duplication or cryptic relatedness across samples was also explored using PLINK or GRR. Individual IBS mean values were calculated to identify samples with common ancestors. In Murcia data set, we found two possible sibling pairs (IBS 1.63–1.67), and one possible second-degree pair (IBS 1.50), while the remaining individuals represented a cluster of diverse ranges of relatedness. Therefore, three individuals were removed to eliminate these relationships. We also found two possible pairs of first-degree relatives (IBS = 1.63–1.70) in the ADNI data set, who were also removed from the analysis. We do not need to remove any subject from the NIA, TGEN or GENA data sets (IBS < 1.50 in all individuals). However, when relatedness was explored across these databases, we detected 15 samples that might be duplicated and 6 related individuals. These were patients from the Mayo clinic (TGEN), GENA and ADNI data sets who had also been included in the NIA GWAS (Supplementary Table S12). The potential impact of this unexpected finding, undetected in our previous report, was evaluated separately. In fact, after removing detected duplications we re-calculated effect sizes and P-values that varied only at the thousandth and millionth levels, respectively (data not shown). We concluded that undetected sample redundancy in our original GWAS has little impact on our results.

LD analyses and proxy searches were conducted using SNAP software (academic software non-commercial)35 (CEU 1000 genomes information, 500 kb window and r2 > 0.7).

Graphic software and other informatics tools. Regional plots containing meta-GWAS results were generated using LocusZoom (academic software non-commercial).23 Stage II Manhattan plot was generated using Haploview software36 (academic software non-commercial; Supplementary Figure S2). Q-Q plots and inflation factor were calculated using SPSS 18.0. A minimal inflation of statistics was observed using fixed effects model meta-analysis (lambda = 1.05). In contrast, a clear deflation was observed for the random effects model (lambda = 0.84), demonstrating that this strategy is over-conservative (Supplementary Figure S3). PC scatterplots were generated using SPSS 18.0.

RESULTS

The top 1202 SNP (P < 0.001) signals obtained in our meta-analysis previously published9 (Supplementary Figure S1 and Supplementary Table S4) were submitted to the CHARGE consortium for further in silico analysis. We received ORs with 95% confidence interval (CI) and risk allele information for 1142/1202 (95%) of the requested markers. The rest of markers (58 SNPs, 5%) were not available in the CHARGE consortium GWAS results, because they did not meet pre-specified quality control criteria. Hence they were excluded from our Stage II meta-analysis. Of note, effect estimates were calculated without any modification of published methodologies.5,9 Then, we combined data from each data set using the meta-analysis tool in PLINK generating a novel SNP list with top signals in terms of effect size and direction. Finally, we mapped the signals using the UCSC genome gateway (http://genome.ucsc.edu/) and measured physical distance and LD with known loci with SNAP software. (Supplementary Figure S1 and Supplementary Table S5).

A total of 35 SNPs reached our pre-established threshold for being labeled highly suggestive (P < 5 x 10−8) in the GWAS; these are listed in Supplementary Table S5. Of note, 16 of these 35 SNPs reached our pre-established threshold for GW significance (P < 5 x 10−8), but all signals belonged to known AD loci, including eight SNPs at the APOE locus, MS4A gene cluster (four SNPs, the most significant being rs1562990, P = 5.05 x 10−10), PICALM (three SNPs, the most significant being rs536841, P = 9.67 x 10−10) and BIN1 (rs744373, P = 2.13 x 10−9).

The remaining 19 SNPs that reached pre-established highly suggestive P-value thresholds also included 12 markers near known AD loci such as PICALM (rs2077815), CR1/ APOJ (rs569214), CR1 (rs3818361), BIN1 (rs11685593 and rs7561528) and, again, the APOE chromosomal region (7 markers). Of note, another SNP, rs16871253, was located within the NEDD9 gene. This locus has been previously proposed for AD (Supplementary Table S5). The list also included seven SNPs, comprising five different chromosomal regions not previously detached with AD (Supplementary Table S5). Specifically, we detected the strongest signal at the ATPSH/KCCT2 locus (rs11870474, P = 2.65 x 10−7) in a region previously associated with the information processing speed cognitive phenotype.28 The next strongest signal corresponded to two markers located within the SLC24A4 gene (rs4406992, P = 9.54 x 10−7; rs2896209, P = 4.47 x 10−7). The third signal was a synonymous cSNP in exon 2 of the cholinergic receptor, nicotinic, alpha 9 gene (rs10022491, P = 2.51 x 10−5). The fourth was an intragenic marker in the utrophin gene (rs2473130, P = 3.50 x 10−4). The last SNP, rs11151137 (P = 4.05 x 10−5), is located in an 800-Kb gene desert at 18q22.1.21

We have previously confirmed, using the Fundacio´ ACE data set, all the GWAS-significant loci detected in this study (APOE, MS4A, BIN1 and PICALM) and also previously known loci classified as highly suggestive in our SNP list (CR1 and CLU).5,29 Consequently, we decided to genotype in the Fundacio´ ACE data set only the five new SNPs (KCTD2/ATP5H and ADN1). Hence, we decided to genotype in the Fundacio´ ACE data set only the five new suggestive loci (CHRNA9, UTRN, SLC24A4, ATPSH/KCCT2 and the within the gene desert at chromosome 18) and NEDD9. We decided to follow the NEDD9 marker because that gene, while previously described as probably associated with AD in one study, has not been confirmed at a genome-wide significance level to date. We achieved a nominally significant signal only for rs11870474 at the ATPSH/KCCT2 locus (OR = 1.43, 95% CI (1.12–1.83), P = 0.0038). The involvement of the other signals in determining risk of AD remains uncertain after our attempted validation and will require further research (Supplementary Table S6).

The rs11870474 signal remained statistically significant even after applying Bonferroni’s multiple testing correction for the six markers genotyped during the last stage of our study (P = 0.0083) and its effect size remained almost unchanged after covariate adjustments (OR = 1.40, 95% CI (1.08–1.82), P = 0.01, dominant model). Of note, the magnitude of the effect is quite large and consistent across studies, with all six estimates ranging between 1.31 and 1.85 (see Figure 1). Fixed effects model meta-analysis with all available data sets confirmed this as a novel GWAS

![Figure 1](image-url)
significant locus for AD (Figure 1, OR = 1.533, 95% CI (1.329–1.770), \( P = 5.07 \times 10^{-5} \)).

**DISCUSSION**

We present additional results from a GWAS generated by our group,\(^9\) this time following up on selected top markers in a large independent GWAS generated by the CHARGE consortium.\(^5\) After adding CHARGE information, we confirmed seven different SNPs identical to those previously reported (Supplementary Table S5). However, we cannot consider these independent replications, because some data sets used here overlap with previous studies. We also detected 21 unreported SNPs within previously described loci. Overall, most of the newly identified SNPs are physically close to previous detected signals (300 kb window) (Supplementary Table S7). These new markers might help to refine the association at the previously identified loci and aid the search for functional variants.

Importantly, by merging results of a meta-GWAS conducted by our group, results in the CHARGE consortium data sets and an in vivo genotyping comprising 4501 individuals, we were able to detect a novel gene associated with AD risk. Compared with the previous meta-GWAS,\(^5,7,8\) our study had information on an additional 8000 individuals, including the Fundación ACE data set and data derived from Pfizer’s GWAS.\(^18\) The larger sample size of 24 227 persons might be one reason why we detected this novel signal, which was not observed in previous GWAS. The in vivo genotyping of over 4500 persons was an additional strength of our study design that permitted the signal to reach the pre-established GWAS significance threshold.

As the new locus reaches genome-wide significance only when including the final discovery sample (Stage III), it must still be considered a highly probable finding but not a replicated locus. Independent replications are still required to corroborate this signal. The low frequency of the rs11870474 marker must be taken into account for future replication efforts (Supplementary Table S11). We estimate that a sample size of more than 2450 cases and an equal number of controls will be necessary to reach 80% power to detect its observed effect on AD risk (\( \alpha=0.05 \), case/ control ratio = 1, exposure prevalence 3% and OR effect level = 1.53). If we consider adjustment for a possible winner’s curse effect,\(^10\) which in fact was observed in the Fundación ACE data set (decreasing observed effect size to 1.43), the number of cases necessary to detect this effect (80% power) could rise to 3660 AD cases. So, it is only by using very large case-control data sets that one can expect to have reasonable power to replicate this observation.

As age/sex data were not available for some data sets, we decided to apply homogeneous criteria to available data sets during Stage I. A potential criticism of our study design emerged from this decision based on our use of young, general population controls, as a proportion of these controls might develop AD as they age. However, although this misclassification might reduce our power to detect an association, it should not create a spurious association. Furthermore, age- and sex-adjusted logistic regression analyses in data sets with covariate data available demonstrated little difference in terms of effect size or statistical significance. (Supplementary Table S10a). Of note, association reported in the CHARGE data set did include age, sex and PC adjustments,\(^7\) and the effect size observed in this data set was remarkably consistent with our Stage I result. A second criticism might be that having controls that are younger than cases might lead to a spurious association with longevity-related genes. However, our discovery sample was largely age-matched for cases and controls and the observed association in the Fundación ACE data set was weaker rather than stronger making it unlikely that the detected locus represents a spurious association with longevity rather than AD. We used this same set of cases and general population controls to successfully replicate relatively ‘modest’ effects associated with uncontroversial SNPs located in PICALM, BIN1 and CLU loci.\(^5\) We also detected a consistent signal in the MS4A gene cluster previously reported by others.\(^7–9\) Notably for these known markers, the observed magnitude of the effect was virtually the same as that reported in the original studies. Furthermore, general population controls have some advantages over neurologically healthy elderly controls, as the latter represent a group of healthy survivors who escaped infectious, cardiovascular and neoplastic diseases. Using such ‘hypernormal’ controls might jeopardize the generalizability of the risk estimates observed and has been identified as a potential source of bias.\(^31,32\)

Another source of bias could be hidden population stratification affecting the rs11870474 results. We did not detect an adjusted effect on stratifying using two major eigenvectors in three data sets from Stage I with genomewide genotypic data available (Supplementary Table S10b). This analysis revealed little impact of PCs in terms of effect size (OR = 1.49, (1.05–2.13), \( P = 0.02 \)) with three data sets). Of note, the lack of impact of population stratification in our results was also re-enforced by the absence of correlation between PCs and rs11870474 A-allele carrying status (Pearson’s determination coefficient (\( r^2 < 1.5\% \)) and the homogenous distribution of carriers observed in multi-dimensional scatterplots representing PC1 and PC2 eigenvectors (Supplementary Figures S5b and S5c). A final model integrating population stratification PC4s, age and sex was also applied to the series with these data available (Murcia, ADNI and NIA). Again, little impact observed on OR estimates using PLINK software (data not shown), suggesting that imputation process have been successful for this marker. Furthermore, actual genotyping data were always preferentially selected when available (in the NIA, ADNI and Pfizer GWAS). Case-control differences in allele frequency comparing imputed and non-imputed data sets were almost identical (Supplementary Table S11). These observations suggest a lack of bias during the imputation process.

The potential significance of our finding is also reinforced by the independent previous observation that a locus at the same chromosomal region could also be related to information processing speed,\(^28\) which is an important cognitive function compromised in dementing disorders, and one that might share a genetic background with other complex cognitive traits, such as working memory or abstract reasoning.\(^28\) However, the marker previously associated with information processing speed (rs11077773) only reached a suggestive association \( P\)-value (8.33 \( \times 10^{-6} \)),\(^28\) Furthermore, in spite of its physical proximity to rs11870474 (it is only 29 kb away), LD among markers is null (\( r^2 \leq 1 \)) and SNAP calculations based on two matching simulations (genomes and CEU population). As rs11077773 is not in LD with rs11870474, it is less likely that there is a direct relationship among both observations. Rather, it is probable that these two markers could be tracking different alleles.

The rs11870474 SNP is an intronic non-coding common variant physically located within the second intron of the KCTD2 gene at 17q25.1 (Figure 2; Supplementary Figure S5). In spite of its small length (33 kb), this locus is located within a region with low LD; there are no large LD blocks in this region and this remains true even when analyzing intragenic markers alone (Supplementary Table S8). This phenomenon makes it difficult to identify proxy signals around rs11870474. However, we did detect a proxy marker, rs12943281, just 716 bp away from rs11870474 (\( r^2=0.718, \ D'=1; \ P=0.008 \)).
The multi-subunit complexes: the soluble catalytic core, F1, and the mitochondrial ATP synthase catalyzes ATP synthesis, utilizing an mitochondrial F0 (ATP synthase complex V component). Mitochondrial functions are related to voltage-dependent potassium channel ubiquitinated proteins and proteasome physiology. Other KCTD rs12943281). Rs11870474 SNP is located within the KCTD2 gene intron 2. This gene is a member of the KCTD family, which is involved in diverse functions ranging from DNA transcription to degradation of ubiquitinated proteins and proteasome physiology. Other KCTD functions are related to voltage-dependent potassium channel function and GABA neurotransmitter receptor B heteromultimeric composition. KCTD2 expression is ubiquitous according to GeneNote. However, the highest levels of expression are noted in the cerebral cortex and cerebellum.

In spite of these suggestive data, it is important to mention that another transcription unit, named ATP5H, is embedded in the third intron of the KCTD2 gene (Supplementary Figure S4). So, the rs11870474 marker has an alternative candidate gene by position. This ATP5H gene encodes ATP synthase, H+ transporting, mitochondrial F0 (ATP synthase complex V component). Mitochondrial ATP synthase catalyzes ATP synthesis, utilizing an electrochemical gradient of protons across the inner membrane during oxidative phosphorylation. It is composed of two linked multi-subunit complexes: the soluble catalytic core, F1, and the membrane-spanning component, Fo, which comprises the proton channel. The Fo has nine subunits and the ATP5H gene encodes the d subunit of the Fo complex.21 Mutations in other members of the mitochondrial complex V, such as MTATP6, result in Leigh syndrome characterized by lactic acidemia, hypotonia, neurodegeneration and MRI (magnetic resonance imaging) brain lesions (OMIM 256000). So, the ATP5H gene is related to cell energy production via respiration and its expression is obviously pervasive. The oxidative stress hypothesis for AD, including mitochondrial disturbances, is well documented, and recent studies have confirmed that AD cases have significantly lower expression of the nuclear genes (including the ATP5H gene) encoding subunits of the mitochondrial electron transport chain in different regions of the brain.38,39

In any case, KCTD2, ATP5H or both together (as these are probably highly co-regulated loci) are very attractive candidates for AD risk as both are related to fundamental neuronal physiological processes associated with tolerance to hypoxia and other stressors. In fact, potassium conductance alteration and abolition of ATP synthesis are early events necessary to maintain neuronal survival during oxygen deprivation.40 However, taking into account available data, we cannot decide whether KCTD2, ATP5H, a common variant altering the transcription of either or even other adjacent genes could explain the observed associations. In fact, a close marker, rs9907177, has been described as an exon-quantitative trait loci (eQTL) for the growth factor receptor-bound protein 2 (GRB2) gene. Interestingly, the GRB2 gene has been proposed as a candidate for AD. However, this eQTL marker is not associated with LOAD in our series (P = 0.23, three studies) and its LD with rs11870474 is almost null (D = 1, r² = 0.037). Therefore, it seems unlikely that rs9907177 can explain the observed association.

We also looked for annotated functional variants around rs11870474 using SNAP software (Supplementary Table S9). We failed to identify any obvious candidate functional variant linked to rs11870474 within a 500-kb radius. Interestingly, this analysis identified ICT1, HN1 and SLC16A5 genes as potential candidates explaining our observations. In fact, we observed moderate LD between the detected signal and some intrinsic SNPs within these genes. The ICT1 gene was recently reported to be a component of the human mitoribosome essential for cell viability. HN1 encodes hemopoietic- and neurologically-expressed sequence-1 involved in neuronal regeneration. SLC16A5 gene is a mono-carboxylate transporter similar to MCT1, which had been involved in mediating axon damage. So, we conclude that other potentially interesting genes are present at this locus. To delineate a more precise hypothesis, further research using next-generation sequencing and detailed functional studies will be necessary.

Finally, it is important to mention that almost all the confirmed new loci identified to date have been unveiled using comprehensive meta-analyses of multiple GWAS and further genotyping on independent series. Given the large numbers of individuals needed to detect the associations, it seems like we will only be able to discover more markers with international cooperative efforts that incorporate larger GWAS data sets with an increased SNP density.

CONFLICT OF INTEREST

LMR, AR and EV are shareholders of Neocodex. RRL, AGP, JG, JIG, CMR, FJM, JV, JMC, MES, EV and LMR are employed by Neocodex. The other authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank the patients and controls who participated in this project. This work has been partially funded by the Fundación Alzheimer (Murcia), the Ministerio de Educación y Ciencia (FCT-010000-2007-18, DEE-5899000-2008-4; Gobierno de España), Corporación Tecnológica de Andalucía (08/211) and Agencia IDEA (841318) (Consejería de Innovación, Junta de Andalucía). The Diabetes Research Laboratory, Biomedical Research Foundation. University Hospital Clínicas San Carlos has been supported by CIBER Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM); CIBERDEM is an ISCIII Project. We are indebted to Trinitat Port-Carbo and her family who are supporting Fundación ACE research programs. We also are indebted to TGEN investigators who provided a free access to genotype data to other researchers via Coriell Biorepositories. 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 data sets used for analyses described in this manuscript were obtained from dbGaP 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 fund 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 data sets used for analyses described in this manuscript were obtained from dbGaP through dbGaP accession number phs000168.v1.p1. 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; Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Amanix Life Sciences Ltd; AstaZeneca; Bayer HealthCare; BioClinica, Inc.; Biogen Idec Inc; Bristol-Myers Squibb Company; Eisai Inc; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innomogenetics, N.V.; Janssen Alzheimer Immunotherapy Research & Development, LLC, Johnson & Johnson Pharmaceutical Research & Development LLC, Medpace, Inc; Merck & Co., Inc.; Meso...
Scale Diagnostics, LLC; Novartis Pharmaceuticals Corporation; Pfizer Inc; Servier; Synarc Inc; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education. The study is coordinated by the Alzheimer’s Disease Cooperative Study at the University of California, San Diego, CA, USA. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles, CA, USA. This research was also supported by NIH Grants P30 AG010129, K01 AG030514 and the Dana Foundation. Fundacíó ACE collaborates with CIBERED. The following have contributed to the papers as collaborators: Alejandro Romeo, Irene Blanca, Susana Ruiz, Mar Buendia, Fuensanta Noguera-Perea, Emma Rodríguez-Noriega, Evaristo Fernandez-Aranda, Agustina Legaz-García, Laura Vivancos-Moreau, América Morena, Mariona Guitar, Susana Lara, Martirio Antequera-Torres, Salvadora Manzanares, Sandra Castaño, Blanca García, Begónica Martínez-Herrada, Sergi Valero.

REFERENCES
1 Ballard C, Gauthier S, Corbett A, Brayne C, Aarsland D, Jones E. Alzheimer’s disease. Lancet 2011; 377: 1019–1031.
2 Wingo TS, Laj JJ, Levey AI, Cutler D. Autoimmune recessive causes likely in early-onset Alzheimer disease. Arch Neurol 2012; 69: 59–64.
3 Corser EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC. Genome-wide association study identifies variants at CLU and PICALM associated with late-onset Alzheimer’s disease. Nat Genet 2009; 41: 1088–1093.
4 Seshadri S, Fitzpatrick AL, Ikram MA, De Stefano AL, Gunawardana V, Boada M et al. Genome-wide analysis of genetic loci associated with Alzheimer disease. Jama 2010; 303: 1832–1840.
5 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. PLoS Genet 2010; 6: e1000888.
6 Hollingworth P, Harold D, Sims R, Gerrish A, Hamshere ML et al. Genome-wide association study identifies variants atCLU and PICALM associated with late-onset Alzheimer’s disease. Nat Genet 2009; 41: 1088–1093.
7 Antunez C, Boada M, Gonzalez-Perez A, Saez ME, Martinez-Larrad MT, Zabena C et al. Genetic structure of the Spanish population. BMC Genomics 2010; 11: 326.
8 Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buros J et al. Meta-analysis for genome-wide association of familial late-onset Alzheimer’s disease replicates BIN1 and CLU variants. PLoS Genet 2011; 7: e1002381.
9 Lee Y, Morrison BM, Li Y, Lengacher S, Farah MH, Hoffman PN. Essential for cell vitality. Proc Natl Acad Sci USA 2008; 105: 4441–4446.
10 Ramirez-Lorca R, Boada M, Saez ME, Hernandez I, Mauleon A, Rosende-Roca M et al. GAB2 gene does not modify the risk of Alzheimer’s disease in Spanish APOE 4 carriers. J Nutr Health Aging 2009; 13: 214–219.
11 Kent WJ, Sugnet CW, Fuery TS, Roskin KM, Pringle TH, Zahler AM et al. The human genome browser at UCSC. Genome Res 2002; 12: 996–1006.
12 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81: 559–575.
13 Abecasis GR, Chen H, Cookson WO, Cardon LR. GRR: graphical representation of relationship errors. Bioinformatics 2001; 17: 742–743.
14 Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O’Donnell CJ, de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics 2008; 24: 2398–2399.
15 Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 2010; 26: 2336–2337.
16 Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005; 21: 263–265.
17 Li Y, Grupe A, Rowland C, Holmans P, Segurado R, Abrahams R et al. Evidence that common variation in NEDD9 is associated with susceptibility to late-onset Alzheimer’s and Parkinson’s disease. Hum Mol Genet 2008; 17: 759–767.
18 Luciano M, Hansell NK, Lahti J, Davies G, Medland SE, Raikkonen K et al. Whole genome association scan for genetic polymorphisms influencing information processing speed. Biol Psychiatry 2011; 86: 193–202.
19 Antunez C, Boada M, Lopez-Arieta J, Moreno-Rey C, Hernandez I, Marin J et al. Genetic association of complement receptor 1 polymorphism rs3813631 in Alzheimer’s disease. Alzheimers Dement 2011; 7: e124–e129.
20 Zhong H, Prentice RL. Correcting ‘winner’s curse’ in odds ratios from genomewide association findings for major complex human diseases. Genet Epidemiol 2010; 34: 78–91.
21 Schwartz S, Link BG. The ‘well control’ artefact in case/control studies of specific psychiatric disorders. Psychol Med 1989; 19: 737–742.
22 Schwartz S, Suss E. Genome-wide association studies: does only size matter? Am J Epidemiol 2010; 167: 741–744.
23 Ding XF, Luo C, Ren KQ, Zhang J, Zhou JL, Hu X et al. Characterization and expression of a human KCTD1 gene containing the BTB domain, which mediates transcriptional repression and homeric interactions. DNA Cell Biol 2008; 27: 257–265.
24 Bayon Y, Trinidad AG, de la Puerta ML, Del Carmen RM, Boguet J, Rojas A et al. KCTD5, a putative substrate adaptor for cullin3 ubiquitin ligases. FEBS Lett 2008; 275: 3900–3910.
25 Schwenk J, Metz M, Zolles G, Turecek R, Fritzius T, Bihl W et al. Native GABA(B) receptors are heteromultimers with a family of auxiliary subunits. Nature 2010; 465: 231–235.
26 Shklar M, Strichman-Almashanu L, Shmueli O, Shmisho M, Safran M, Landon D. GeneTide—Terra Incognita Discovery Endeavor: a new transcriptome focused member of the GeneCards/GeneNote suite of databases. Nucleic Acids Res 2005; 33: (Database issue) D556–D561.
27 Hrál K, Áleš G, Nunnamora A, Fujikoa H, Russell RL, Atwood CS et al. Mitochondrial abnormalities in Alzheimer’s disease. J Neurosci 2001; 21: 3017–3023.
28 Terzi B, Boada J, Portero-Otín M, Pampolona R, Ferrer I. Mitochondrial ATP-synthase in the entorhinal cortex is a target of oxidative stress at stages I/II of Alzheimer’s disease pathology. Brain Pathol 2010; 20: 222–233.
29 Liang WS, Reiman EM, Valla J, Duncley T, Beach TG, Grober A et al. Alzheimer’s disease is associated with reduced expression of energy metabolism genes in posterior cortical neurons. Proc Natl Acad Sci USA 2008; 105: 4441–4446.
30 Milton SL, Prentice HM. Beyond axonin: the physiology of metabolic down-regulation and recovery in the axo-synaptic tolerant turtle. Comp Biochem Physiol A Mol 2007; 147: 277–290.
31 Hnda Y, Hikawa Y, Tochio N, Kogure H, Inoue M, Koshiba S et al. Solution structure of the catalytic domain of the mitochondrial protein ICT1 that is essential for cell vitality. J Mol Biol 2010; 404: 260–273.
32 Zujovic V, Luo D, Baker HV, Lopez MC, Miller KR, Streit WJ et al. The facial motor nuclei transcriptional program in response to peripheral nerve injury identifies Hn1 as a regeneration-associated gene. J Neurosci Res 2005; 82: 581–591.
33 Lee Y, Morrison BM, Li Y, Lengacher S, Farah MH, Hoffman PN et al. Oligodendroglia metabolically support axons and contribute to neurodegeneration. Nature 2012; 487: 443–448.