Genetic Associations with Age at Dementia Onset in the PSEN1 E280A Colombian Kindred

J. Nicholas Cochran[1]*, Juliana Acosta-Uribe[2,3]*, Lucia Madrigal[3], David Aguillón[3], Gloria P. Garcia[3], Margarita M. Giraldo[3], Natalia Acosta-Baena[3], Francisco Piedrahita[3], Diana Alzate[3], Hugo E. Lopez[3], Kevin Roberts[1], Devin Absher[1], Richard M. Myers[1], Francisco Lopera[3**], Kenneth S. Kosik[2**]

[1] HudsonAlpha Institute for Biotechnology. Huntsville, Alabama, USA

[2] Neuroscience Research Institute, University of California, Santa Barbara. California. USA and Department of Molecular Cellular and Developmental Biology University of California, Santa Barbara. California. USA

[3] Grupo de Neurociencias de Antioquia. School of Medicine. Universidad de Antioquia, Medellín, Antioquia, Colombia

*,**These authors contributed equally to this work

Abstract

Background: Genetic associations with Alzheimer's disease (AD) age at onset (AAO) have revealed genetic variants with the potential for therapeutic application. Such studies in late onset AD are limited by population heterogeneity and co-morbidities associated with aging. Studies to date in ADAD (autosomal dominant AD) have been limited by their small sample size. The large Colombian kindred presented here provides a unique opportunity to discover AAO genetic associations.

Methods: A genetic association study was conducted for AAO in 344 individuals with the PSEN1 E280A mutation via array imputation and whole genome sequencing in a subset of 80 individuals. Scrambled age conditions were used as a control, and replication was assessed using three studies on AD age of onset, age of onset survival, and meta-analysis.

Results: The proportion of variance explained (+/- 95% CI) was 56% (+/-15%). 29 loci reached genome-wide significance, of which 23 had a GWAS hit from the NHGRI-EBI catalog within 500 kb relevant to neurodegeneration. Hits included a new variant in the *CLU* locus, rs35980966, (which replicated), a variant in a locus on a separate chromosome previously associated with plasma clusterin (rs138295139), and a missense variant in *SORBS3* (rs34059820). Former GWAS index variants at the *CLU* locus also replicated in this cohort (rs4236673, rs9331896). *APOE* ε4 (rs429358) and *APOE* ε2 (rs7412)-associated variants exhibited modest (less than 3 years) associations with earlier and later age of dementia onset, respectively. Other nominal associations included coding variants in *IL34* (rs4985556), *TSPAN10* (rs6565617, rs7210026), *STIM2* (rs1457401458), *HTT* (rs1473464204), and *KCNT1* (rs557219607).

Interpretation: A large proportion of the variability in AAO was explained by genetic variation at numerous loci. The unique demography of this population as a tri-continental admixture that passed through a bottleneck about 500 years ago might predict that drift would uncover rare variants with a large effect size on AAO. Indeed, candidates for large contributions from rare alleles were observed. Common variation, presumably ancestral to the bottleneck, was also associated with AAO. The detection of these effects in the presence of a strong mutation for ADAD reinforce the potentially impactful role of the identified variants.

Introduction

Complex genetic, environmental, and lifestyle risk factors confounded by the aging process underlie risk for late onset Alzheimer's disease (LOAD). Autosomal dominant Alzheimer's disease (ADAD) closely resembles the clinical and neuropathological features of LOAD, but without the confound of aging, and thus provides a less heterogeneous view of underlying AD-associated processes. ADAD accounts for less than 1% of all cases of AD and mutations in *PSEN1* account for 80% of this monogenic group (reviewed in[1]).

The age of onset of all other individuals with a given ADAD mutation correlates with an individual's age of onset ($r^2 = 0.52$)[2], but there remains substantial unexplained variability in age at onset. Large ADAD families such as the kindred harboring the Colombian *PSEN1* NM_000021:c.839A>C, p.(Glu280Ala) (canonically known as PSEN1 E280A) mutation, the world's largest ADAD founder population with a comprehensive family tree of thousands of individuals[3], provide an opportunity to assess the contribution of genetic variation to unexplained variability in age of dementia onset. PSEN1 E280A mutation carriers typically develop MCI at a median age of 44 (95% CI, 43–45) and dementia at age of 49 (95% CI, 49–50)[4]. The value of this family for the nomination of genetic variants that delay the onset of Alzheimer's disease was recently affirmed by the report of a PSEN1 E280A carrier who developed mild cognitive impairment nearly three decades after the kindred's median age at clinical onset[5]. This individual was homozygous for the rare *APOE* ε3 Christchurch variant (*APOE* NM_000041:c.460C>A, p.(Arg154Ser)) and had an exceptionally high amyloid-β plaque burden, but limited neurofibrillary tau burden. In addition to this case report, several studies have explored genetic associations with age at onset in this population of patients (see supplemental references), but all with substantially lower numbers of cases (at most 72 individuals)[6]. To expand on the valuable insights gained from these previous studies, we conducted the most comprehensive search to date for genetic variants associated with age of dementia onset in this founder population by assessing 344 individuals, which is the current snapshot of all individuals from this cohort that currently have high quality genotypic and phenotypic information available.

Methods

Patient Recruitment

A cohort of 368 patients was selected from the Neuroscience Group of Antioquia (GNA) database of the PSEN1 E280A family. Selection criteria included being a PSEN1 E280A carrier with diagnosis of dementia, having adequate medical and neuropsychological evaluations and follow-up for a confident age determination of clinical age of dementia onset, and having a DNA sample. Participants were evaluated following a standard protocol including physical and neurological examination, as well as population-validated neuropsychological assessment. The clinician who performed these evaluations was blinded to the genetic status of the patient. Dementia was diagnosed according to most recent DSM criteria at the time of diagnosis. Collected data were stored in medical records software (SISNE v2.0). Family history was obtained from the patients and their relatives, and genealogical data from baptism and death certificates was gathered from local parishes and was incorporated into the pedigree reconstruction. Blood samples from each individual were obtained through standard phlebotomy and collected in EDTA tubes. Genomic DNA was purified from peripheral blood leukocytes using a modified salting-out technique (Gentra Puregene Blood Kit, Qiagen). All individuals were genotyped for *PSEN1* NM_000021:c.839A>C, p.(Glu280Ala) using a restriction length fragment polymorphism assay.

Genotyping Arrays

1,923,394 variants were genotyped using the Illumina Multi-Ethnic Genotyping Array plus Neuro consortium content (catalog #WG-316-1014, beadchip #20028352). Data were annotated with build hg38 and analyzed using PLINK v1.90b5.2, PLINK v2.00aLM, and GEMMA. Relatedness was assessed using KING 2.2.

Genome Sequencing and Annotation

Genome sequencing and preparatory processing of genome data prior to analysis is described in the supplemental methods.

Variant Measurement, Imputation, and Quality Control

Imputation was conducted using the TOPMed Imputation Panel and Server (version 1.3.3), which includes 97,256 references samples and 308,107,085 variants and uses Minimac4 for imputation. Pre-imputation scripts (version 4.3.0 from William Rayner at the University of Oxford) were run using default settings, which filtered out palindromic single nucleotide variants (SNVs) with minor allele frequency (MAF) > 0.4 or variants with > 0.2 MAF difference from the TOPMed reference panel. For samples with both array and genome sequencing data, imputed genotypes were checked for concordance with genome sequencing data using SnpEff 4.3s.

Details of Variant measurement, imputation, and quality control are described in the supplemental methods. Briefly, 540,753 high quality variants from array measurements for all individuals that overlapped with measured genome sequencing variants for 80 individuals were used for imputation using the TOPMed imputation server, and imputed variants were filtered for presence in genome sequencing data resulting in a final set of 9,451,360 variants. Median concordance between imputed genotypes with these filters and genome sequencing genotypes was 99.4% for the 80 samples with both array and genome sequencing data available, indicating high quality imputation.

Replication sets

Three cohorts were selected for replication: an AD age of onset study (9,162 cases)[7], a study of AD age at onset survival (14,406 cases and 25,849 controls)[8], and a recent GWAS meta-analysis for AD (21,982 AD vs. 41,944 controls)[9]. See supplemental methods for discussion of International Genomics of Alzheimer's Project (IGAP) replication data.

Role of the funding source

The study sponsors were not involved in study design, the collection, analysis, and interpretation of data, the writing of the report, or the decision to submit the paper for publication.

Supplemental methods and references

Additional methods and references relevant to methods are provided in the supplemental material.

Results

Cohort demographics

The final cohort had a mean age of dementia onset of 49.2 years [range: 37–75, 10th–90th percentile: 43–56]. 200 of the patients were genetically female (58.1%). Alcohol abuse was reported for 46 patients. The patients had extensive follow up data; the mean number of medical evaluations was 6.7 [1–27], and 4.8 [1–18] for neuropsychological evaluations. A partial pedigree of enrolled individuals annotated with age of dementia onset is presented in **Supplemental Figure 1**.

Association analysis

Association analysis was conducted using age at dementia onset as a quantitative outcome for 344 individuals passing QC. We employed PLINK with allelic, dominant, genotypic, and recessive models as well as GEMMA, a package that performs a likelihood ratio test using a linear mixed model to adjust for relatedness between individuals. For both packages, we adjusted for genetic sex, the first two principal components (calculated from the set of 540,753 high quality variants used as imputation input using PLINK), reported alcohol abuse, and batch.

Assessment of biological relevance of identified signals

We first assessed the dataset as a whole. To assess if the number of hits observed was greater than random chance, we generated QQ plots (**Supplemental Figure 2**). QQ plots for the age of dementia onset did not deviate detectably from randomly scrambled (10 scrambles) ages of dementia onset across individuals, but this was not surprising given the small size of the cohort. We therefore pursued alternative approaches to assess the biological relevance of the observed associations. First, we compared the number variants with $p<1x10^{-5}$ that exhibit nominal replication in one of the three replication cohorts ($p<0.05$) for the observed age of dementia

onset compared to scrambles (**Figure 1A**), which revealed a definitive enrichment for replication with the observed ages of onset vs. scrambles. Second, we identified the regions surrounding (+/-500 kb) genome-wide significant hits ($p<5x10^{-8}$), where multiple hits within 5 MB of one another were considered a part of the same continuous block. Concurrently, we identified these regions in scrambled conditions. We then searched (blinded to whether the source was from the observed ages or scrambled ages) the NHGRI-EBI GWAS catalog for GWAS hits within these regions with either a permissive relationship to AD (**Figure 1B**) (i.e. cognition phenotypes, other neurodegenerative diseases, stroke, AD and AD-related phenotypes, etc., full list in **Supplemental Table 2**), or a strict relationship to AD (**Figure 1C**) (i.e. AD itself or related endophenotypes such as CSF tau and Aβ, full list in **Supplemental Table 2**). This revealed an enrichment (observed > 95% CI of scramble) of GWAS hits related to AD or other related phenotypes with the observed age distribution for allelic, dominant, and LRT models, and enrichment of GWAS hits with a strict relationship to AD in allelic, dominant, LRT, and genotypic models. A final test for biological relevance comes from the proportion of variance explained ("chip heritability") calculated by GEMMA in the observed model at 56.4%, well above the median of 2.9% for scrambled conditions (**Figure 1D**).

Genome-wide significant loci

Genome-wide significant loci ($p<5x10^{-8}$) are shown in **Table 1** and included two variants at different loci associated with clusterin biology, rs35980966 and rs138295139. In addition, the age of each individual harboring each variant in **Table 1** is illustrated in **Figure 2** by variant. NHGRI-EBI GWAS hits near genome-wide significant hits from this study are detailed in **Supplemental Table 3**. Manhattan plots for all models are shown in **Supplemental Figure 3** and correlation plots for all models tested are shown in **Supplemental Figure 4**. All models were correlated (average Pearson's r=0.54). Allelic, dominant, and LRT models were highly correlated (average Pearson's r=0.76) and genotypic and recessive models were highly correlated (Pearson's r=0.83). Summary statistics for all variants and models are provided in **Supplemental Table 1**. LocusZoom plots of all regions in **Table 1** are presented in **Supplemental File 1**. Because this is a founder population, we present *p* values from the LRT model (which accounts for family structure most stringently) for

all variants listed, although we note that the LRT model *p* value may be overly conservative when compared to genotypic and recessive models because the LRT model does not assign additional weight to multiple alleles present in the same individual.

Results at key APOE variants

Effects of previously established *APOE* variants important for AD association in LOAD are in the expected direction based on previous studies, but modest in magnitude (**Table 2**). Overall, the observations are consistent with previously reported observations including a protective effect of *APOE* ε2 in the Colombian E280A population[10], a deleterious effect of *APOE* ε4 in the Colombian E280A population in one study[11] but an inability to detect an effect of *APOE* ε4 in three other studies in this population (see supplemental material for references), and a non-significant trend towards an *APOE* ε2 > *APOE* ε3 > *APOE* ε4 age-of-onset in dominant AD families with a variety of mutations[2].

A recent case report implicated the *APOE* Christchurch variant (rs121918393)[5]. That individual is also enrolled in this study, and while we do observe a significant effect on age at onset of this variant, the effect is driven by this one individual with homozygosity at this variant. In a dominant model, which weights heterozygous and homozygous variants equally, there is not a detectable effect of the Christchurch variant.

Replication at known AD-associated loci

We evaluated 15 AD genome-wide association studies (GWAS), including the largest case/control studies for AD in European populations, studies in non-Europeans, age of onset modifier studies, and endophenotype studies (see supplemental references). These studies identified 55 loci (100 index variants) with the highest confidence associations for AD and endophenotypes (**Supplemental Table 4**). Replication of hits with genome-wide significance for AD-associated phenotypes with nominal significance ($p<0.05$) in at least one model and with consistent effect direction in this cohort are shown in **Table 3** and included two variants at the *CLU* locus (rs9331896[12] and rs4236673[13]) and a stop gain variant in *IL34* (rs4985556[13]).

Coding variants of interest

We next asked if any coding variants associate with age of dementia onset (**Table 4**). Two missense variants reached genome-wide significance ($p<5x10^{-8}$) which were in Linkage Disequilibrium (LD) with one another ($r^2=1$ and D'=1), rs34059820 in *SORBS3* and rs143895597 in *CCAR2*, both associated with a later age of onset. To capture hits less stringently, we allowed a more permissive threshold of $p<1x10^{-5}$, but required replication, which identified two independent signals: a missense variant in *ZBTB3* that replicates in one previous AD AAO study[7] and was associated with a later age of onset, and a pair of missense variants in *TSPAN10* (rs6565617 and rs7210026) in LD with one another ($r^2=1$ and D'=1) that replicated in both previous AD AAO studies[7,8] and were associated with an earlier age of onset. It is possible that rare variants not imputed by TOPMed are associated, so we assessed variants present in genomes but absent in the imputed dataset with $p<0.05$ and also in perfect LD ($r^2=1$ and D'=1) with a variant ($p<1x10^{-5}$) in the same model from the imputation set at a population allele frequency > 0.01%, and CADD > 20. This analysis revealed one canonical loss-of-function and two missense variants including a variant in *STIM2* (rs1457401458) associated with a later age of onset. Next we searched for variants in 765 genes previously implicated in neurodegeneration (**Supplemental Table 5**) with $p<0.05$, a population allele frequency > 0.01%, and CADD > 20, which uncovered seven missense variants including variants in *HTT* (rs1473464204), *KCNT1* (rs557219607), and the *APOE* Christchurch mutation (rs121918393)[5], all associated with a later age of onset. Finally, we used permissive filters for eight key genes important for AD including *APP*, *PSEN1*, *PSEN2*, *MAPT*, *APOE*, *ABCA7*, *SORL1*, or *TREM2* with $p<0.05$, a population allele frequency > 1%, and CADD > 10, which identified one canonical loss-of-function variant in *ABCA7* (rs547447016) and a missense variant in *APP* (rs112263157), both associated with an earlier age of onset.

Discussion

Genetic association studies for LOAD are limited by heterogeneity of cases and unknown levels of contribution from environmental sources. This study addresses these limitations by employing a well-described

phenotype in a geographically isolated population with a monogenic form of AD[3]. While environmental influences will always be present, this population has a relatively homogeneous set of environmental influences.

We identified 29 loci reaching genome-wide significance, 27 of which were associated with a later age of onset (median β=17.6 years), implying they may be protective. Two loci directly replicated in previously conducted AD-relevant GWAS studies, and 23 of the 29 loci had a neurodegeneration-associated GWAS hits within 500 kb, indicating the general relevance of identified loci for neurodegeneration. These observations will serve as the foundation for additional studies to nominate particularly important pathways. For example, one variant that exhibited replication, rs35980966, is a rare variant that tags the *CLU* locus on chromosome 8, a locus for which our cohort also provides replication for two common variants identified as genome-wide significant index variants in previous studies, rs9331896[12] and rs4236673[13]. In addition, rs138295139 on chromosome 1 is only 4.4kb from a variant previously associated with plasma clusterin, rs4428865[14]. Taken together, these observations suggest that further investigation on the role of clusterin in EOAD is warranted. Only two coding variants reached genome-wide significance, including a variant in *SORBS3* (rs34059820). A previous study for age of onset modifiers for a different *PSEN1* mutation implicated a related gene, *SORBS2*[15].

In addition to genome-wide significant loci, we investigated previously implicated GWAS loci. We confirmed the modest effect of variants tagging *APOE* ε2 and of *APOE* ε4 in this population, consistent with either modest or undetectable effects of these alleles for ADAD in other studies (supplemental references and[2,10,11]). We also observed replication of 13 index variants from 10 previously identified loci linked to AD in this cohort. In addition to the variants at the *CLU* locus that we previously discussed, another previous GWAS hit of interest that replicates in this study was a stop gain variant in *IL34*[13] associated with earlier age of dementia onset. IL-34 has protective roles such as protection from excitotoxicity[16] and promoting clearance of oligomeric Aβ[17], so it is consistent that a stop gain variant in *IL34* would associate with an earlier age of onset by preventing these protective effects.

Beyond currently or previously implicated genome-wide significant loci, we conducted genome-wide searches for coding variation using more permissive criteria. Two variants in LD in *TSPAN10* were nominally (p=4.8x10^{-7}) associated with an earlier age of dementia onset in this cohort, and this association replicated in

both AD AAO studies queried[7,8]. Tetraspanin-10 is a part of the TspanC8 subgroup of tetraspanins which promote ADAM10 maturation[18]. Given ADAM10's established role as an α-secretase promoting non-amyloidogenic processing of Aβ[19] and the recent association of genetic variation in or near *ADAM10* with AD risk by GWAS (supplemental references and[9,13]), the basis of the observed association between earlier age of dementia onset and these variants in *TSPAN10* may result from disruption of a protective role of ADAM10. An additional coding variant of interest was a rare missense variant in *STIM2* was nominally associated with later age of dementia onset (LRT model p=6.3x10^{-3} in genome-only analysis and in perfect LD (r^2=1 and D'=1) with a variant with p=9.0x10^{-6} in the imputation analysis). STIM2 is protective in AD models[20-22], implying that this variant may be gain of function (or could block STIM2 loss of function).

In addition to variants identified through more permissive genome-wide criteria, we also explored variation in genes previously associated with neurodegeneration. One example was a variant in *HTT*. Repeat expansions in *HTT* are associated with Huntington's disease[23], but a recent study has implicated huntingtin in regulation of presynaptic APP levels[24], which may explain the observation of a rare missense variant in *HTT* associated with later age of dementia onset. A second example was a variant in *KCNT1*. Variants in *KCNT1* can cause seizures[25] and were also identified in the discovery phase of an ALS rare variant study[26]. Therefore, the observation of a missense variant in *KCNT1* with an association with later age of dementia onset (p=3.7x10^{-5}) in the genome-only analysis is interesting in light of the growing evidence for excitability abnormalities in AD and AD models[27,28], including abnormalities in other potassium channels[29]

As the largest age of onset modifier study in dominant AD to date (to our knowledge), this study has nominated several new candidate genetic associations with age of dementia onset in dominant AD. However, we acknowledge that this study is limited by the small size of the dominant AD population. Recruitment of more patients with early onset dementias from South American countries will help to overcome this limitation in future studies.

An important overarching theme from this analysis is that while age at dementia onset in ADAD has a strong heritable component, it is likely that, as with LOAD, there are many different genetic contributors that sum to determine an individual's age at dementia onset for ADAD. Based on the unique demography of this

population as a tri-continental admixture that passed through a narrow bottleneck[30], we conducted this study with the hypothesis that rare variants with a large effect size, i.e., the *APOE* Christchurch mutation[5], could account for much of the difference in age at dementia onset. Indeed, we identified many genetic variants of a similar rarity and damage level and with similar effect sizes and association significance levels to *APOE* Christchurch in this study (**Table 4**). For example, *SORBS3* (rs34059820) exhibited a similar effect size even in the heterozygous state, and several others such as variants discussed in *STIM2* (rs1457401458) and *KCNT1* (rs557219607) occurred in the heterozygous state but could have a similar effect size to *APOE* Christchurch if they occurred in the homozygous state, which was observed for *ZBTB3* (rs544641) and *PPM1N* (hg38 19:45498605:C>CT). We note that due to the nature of the models employed, it is possible for the presence of alleles in a small number of individuals with a particularly late age of onset can result in a low *p* value. However, as a class these variants could confer a larger effect. Nevertheless, observations driven by a small number of alleles should be considered with caution. Importantly, we also detected more common and/or lower effect size variation associated with age of dementia onset. Some of the variants identified are associated with the amyloid processing pathway, others with pathways that include immune response and neuronal signaling. Because many of these variants replicate or were identified in non-admixed European populations, it suggests that the associations for many of these variants are robust to ancestral background. The identified variants in this study occur in the presence of a very strong causative mutation for ADAD, emphasizing the importance of the association signals observed for these variants and the need for more investigation of these variants in future studies.

Funding

Funding was provided by the HudsonAlpha Memory and Mobility Program, the Larry L. Hillblom Foundation, the Rainwater Foundation, and NIH grants 1K99AG068271-01 (Cochran), 1R56AG062479-01 (Kosik), and 5U54NS100717-04 (Kosik). Sustainability Program of the CODI University of Antioquia (Lopera).

Acknowledgements

We thank the patients and their caregivers for their generous contributions to this study. We thank other members of the GNA for their contributions to study of this cohort which made this study possible. We thank Nithesh Perumal for assisting with pedigree reconstruction. Replication statistics were obtained from publicly available data for NIAGADS project NG00075[9] and controlled access data from NIAGADS projects NG00048[7] and NG00058[8]. More extensive acknowledgement details for IGAP are provided in the supplemental information.

References

1. Neuner SM, Tcw J, Goate AM. Genetic architecture of Alzheimer's disease. *Neurobiol Dis* 2020; **143**: 104976.

2. Ryman DC, Acosta-Baena N, Aisen PS, et al. Symptom onset in autosomal dominant Alzheimer disease: a systematic review and meta-analysis. *Neurology* 2014; **83**(3): 253–60.

3. Lopera F, Ardilla A, Martinez A, et al. Clinical features of early-onset Alzheimer disease in a large kindred with an E280A presenilin-1 mutation. *JAMA* 1997; **277**(10): 793–9.

4. Acosta-Baena N, Sepulveda-Falla D, Lopera-Gomez CM, et al. Pre-dementia clinical stages in presenilin 1 E280A familial early-onset Alzheimer's disease: a retrospective cohort study. *Lancet neurology* 2011; **10**(3): 213–20.

5. Arboleda-Velasquez JF, Lopera F, O'Hare M, et al. Resistance to autosomal dominant Alzheimer's disease in an APOE3 Christchurch homozygote: a case report. *Nat Med* 2019; **25**(11): 1680–3.

6. Lalli MA, Bettcher BM, Arcila ML, et al. Whole-genome sequencing suggests a chemokine gene cluster that modifies age at onset in familial Alzheimer's disease. *Mol Psychiatry* 2015; **20**(11): 1294–300.

7. Naj AC, Jun G, Reitz C, et al. Effects of multiple genetic loci on age at onset in late-onset Alzheimer disease: a genome-wide association study. *JAMA Neurol* 2014; **71**(11): 1394–404.

8. Huang KL, Marcora E, Pimenova AA, et al. A common haplotype lowers PU.1 expression in myeloid cells and delays onset of Alzheimer's disease. *Nat Neurosci* 2017; **20**(8): 1052–61.

9. Kunkle BW, Grenier-Boley B, Sims R, et al. Genetic meta-analysis of diagnosed Alzheimer’s disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat Genet* 2019; **51**(3): 414–30.

10. Velez JI, Lopera F, Sepulveda-Falla D, et al. APOE*E2 allele delays age of onset in PSEN1 E280A Alzheimer's disease. *Mol Psychiatry* 2016; **21**(7): 916–24.

11. Pastor P, Roe CM, Villegas A, et al. Apolipoprotein Eepsilon4 modifies Alzheimer's disease onset in an E280A PS1 kindred. *Ann Neurol* 2003; **54**(2): 163–9.

12. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 2013; **45**(12): 1452–8.

13. Marioni RE, Harris SE, Zhang Q, et al. GWAS on family history of Alzheimer's disease. *Transl Psychiatry* 2018; **8**(1): 99.

14. Deming Y, Xia J, Cai Y, et al. A potential endophenotype for Alzheimer's disease: cerebrospinal fluid clusterin. *Neurobiol Aging* 2016; **37**: 208 e1– e9.

15. Lee JH, Cheng R, Vardarajan B, et al. Genetic Modifiers of Age at Onset in Carriers of the G206A Mutation in PSEN1 With Familial Alzheimer Disease Among Caribbean Hispanics. *JAMA Neurol* 2015; **72**(9): 1043–51.

16. Luo J, Elwood F, Britschgi M, et al. Colony-stimulating factor 1 receptor (CSF1R) signaling in injured neurons facilitates protection and survival. *J Exp Med* 2013; **210**(1): 157–72.

17. Mizuno T, Doi Y, Mizoguchi H, et al. Interleukin-34 selectively enhances the neuroprotective effects of microglia to attenuate oligomeric amyloid-beta neurotoxicity. *Am J Pathol* 2011; **179**(4): 2016–27.

18. Haining EJ, Yang J, Bailey RL, et al. The TspanC8 subgroup of tetraspanins interacts with A disintegrin and metalloprotease 10 (ADAM10) and regulates its maturation and cell surface expression. *J Biol Chem* 2012; **287**(47): 39753–65.

19. Peron R, Vatanabe IP, Manzine PR, Camins A, Cominetti MR. Alpha-Secretase ADAM10 Regulation: Insights into Alzheimer's Disease Treatment. *Pharmaceuticals (Basel, Switzerland)* 2018; **11**(1).

20. Deng M, Zhang Q, Wu Z, et al. Mossy cell synaptic dysfunction causes memory imprecision via miR-128 inhibition of STIM2 in Alzheimer's disease mouse model. *Aging Cell* 2020; **19**(5): e13144.

21. Popugaeva E, Pchitskaya E, Speshilova A, et al. STIM2 protects hippocampal mushroom spines from amyloid synaptotoxicity. *Mol Neurodegener* 2015; **10**: 37.

22. Sun S, Zhang H, Liu J, et al. Reduced synaptic STIM2 expression and impaired store-operated calcium entry cause destabilization of mature spines in mutant presenilin mice. *Neuron* 2014; **82**(1): 79–93.

23. The Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 1993; **72**(6): 971–83.

24. Bruyere J, Abada YS, Vitet H, et al. Presynaptic APP levels and synaptic homeostasis are regulated by Akt phosphorylation of huntingtin. *Elife* 2020; **9**.

25. Barcia G, Fleming MR, Deligniere A, et al. De novo gain-of-function KCNT1 channel mutations cause malignant migrating partial seizures of infancy. *Nat Genet* 2012; **44**(11): 1255–9.

26. Cirulli ET, Lasseigne BN, Petrovski S, et al. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science* 2015; **347**(6229): 1436–41.

27. Vossel KA, Tartaglia MC, Nygaard HB, Zeman AZ, Miller BL. Epileptic activity in Alzheimer's disease: causes and clinical relevance. *Lancet neurology* 2017; **16**(4): 311–22.

28. Dunn AR, Kaczorowski CC. Regulation of intrinsic excitability: Roles for learning and memory, aging and Alzheimer's disease, and genetic diversity. *Neurobiol Learn Mem* 2019; **164**: 107069.

29. Hall AM, Throesch BT, Buckingham SC, et al. Tau-dependent Kv4.2 depletion and dendritic hyperexcitability in a mouse model of Alzheimer's disease. *J Neurosci* 2015; **35**(15): 6221–30.

30. Mooney JA, Huber CD, Service S, et al. Understanding the Hidden Complexity of Latin American Population Isolates. *Am J Hum Genet* 2018; **103**(5): 707–26.

Table 1: Genome-wide significant hits ($p<5x10^{-8}$). All models are from PLINK except LRT (likelihood ratio rest) which is calculated using GEMMA. The model with the lowest *p* value is shown. For all variants, the LRT *p* value is also shown, which accounts for family structure using a linear mixed model that adjusts for kinship. β is shown for each model (with the β from the Allelic model shown if LRT is the best model, as GEMMA does not output β values), and is derived from the basis of age of dementia onset. Replication indicates nominal significance ($p<0.05$) in the specified study with consistent effect direction. A full list of NHGRI-EBI hits along with the variants associated is provided in **Supplemental Table 3**.

Chr	Pos (MB)	Variant	Best Model	*p*	LRT *p*	β (SE)	MAF	Replication	Nearby (+/- 500 kb) NHGRI-EBI Hits
1	22	rs149331601	Allelic	$2.5x10^{-8}$	$4.2x10^{-8}$	21.8 (3.8)	0.003		AD and age of onset, ALS
1	80.5	rs77107089	Allelic	$7.7x10^{-9}$	$4.7x10^{-7}$	6.1 (1)	0.044	Huang, 2017	aging, stroke
1	112.8	rs115679813	Dominant	$1.9x10^{-8}$	$2.5x10^{-6}$	5.7 (1)	0.049		cognitive performance, *APOE* ε4- amyloid deposition
1	194.4	rs138295139	Allelic	$3.1x10^{-10}$	$2.3x10^{-9}$	17.2 (2.7)	0.006		WM hyperintensity burden, plasma clusterin
1	237.9	rs79319775	Allelic	$2.5x10^{-8}$	$4.2x10^{-8}$	21.8 (3.8)	0.003		WM microstructure, neurofibrillary tangles
3	23.9	rs11718580	Allelic	$2.5x10^{-8}$	$4.2x10^{-8}$	21.8 (3.8)	0.003		AD and age of onset x2, cognitive function, LOAD x2, PD, AD, cognitive decline rate in MCI
4	82.1	rs181746127	Allelic	$7.7x10^{-9}$	$2.5x10^{-8}$	18.3 (3.1)	0.004		cognitive performance
5	95.7	rs75966684	Allelic	$2.5x10^{-8}$	$4.2x10^{-8}$	21.8 (3.8)	0.003		lg. artery stroke, age-rel. cog. decline, ALS, educational attainment x3, AD and age of onset, hippocampal atrophy, AD biomarkers, plasma Aβ42, AD cognitive decline
5	166.3	rs951483	Allelic	$3.4x10^{-8}$	$9.3x10^{-8}$	17.6 (3.1	0.004		cognitive decline in MCI
5	176.3	rs56389738	Allelic	$8.6x10^{-10}$	$2.2x10^{-9}$	16.9 (2.7)	0.006		
6	51.5	rs34506826	Genotypic	$5.0x10^{-9}$	0.02	11.4 (1.9)	0.112		
6	151.6	rs62444327	Recessive	$1.3x10^{-9}$	0.12	19.2 (3.1)	0.105		educational attainment
7	16	rs17352534	Recessive	$3.7x10^{-8}$	0.02	9.4 (1.7)	0.151		stroke
7	65.1	rs34848990	Genotypic	$4.2x10^{-10}$	$2.4x10^{-3}$	10 (1.5)	0.129		AD and age of onset
8	27.6	rs35980966	Allelic	$2.6x10^{-8}$	$4.6x10^{-8}$	18 (3.2)	0.004	Kunkle, 2019	cognitive function, verb. decl. mem., WM integrity, Aβ load, MSA, cognitive decline rate in late MCI, LOAD x3, AD x2, AD in *APOE* ε4- x2, AD in *APOE* ε4+, plasma clusterin
9	19.7	rs4977232	Genotypic	$3.8x10^{-8}$	$5.7x10^{-5}$	-3.6 (0.6)	0.740		cognitive performance, LOAD
10	0.5	rs75435570	Allelic	$3.4x10^{-8}$	$9.3x10^{-8}$	17.6 (3.1)	0.004		
10	35.6	rs673438	Dominant	$5.9x10^{-9}$	0.32	-22.7 (3.8)	0.923		
12	71.3	rs77075460	Allelic	$3.4x10^{-8}$	$9.3x10^{-8}$	17.6 (3.1)	0.004		ALS age of onset
12	120.6	rs79248899	Allelic	$2.5x10^{-8}$	$4.2x10^{-8}$	21.8 (3.8)	0.003		longevity, PD x2, educational attainment x3, LOAD, cognitive performance, ALS
13	40.8	rs71427424	Recessive	$5.9x10^{-9}$	0.09	22.7 (3.8)	0.125		
13	93.6	rs1887660	Recessive	$5.0x10^{-9}$	$2.0x10^{-3}$	7.2 (1.2)	0.224		AD cognitive decline
13	113	rs138705540	Allelic	$2.5x10^{-8}$	$4.2x10^{-8}$	21.8 (3.8)	0.003		AD and age of onset
16	15.7	rs7195365	Genotypic	$1.3x10^{-8}$	0.03	8 (1.4)	0.142		cerebral amyloid angiopathy, cog. function, isch. stroke
19	3.2	rs4807426	Recessive	$6.4x10^{-9}$	$1.4x10^{-3}$	16 (2.7)	0.180		CSF t-tau, CSF t-tau:Aβ42 ratio, dementia and core AD neuropathologic changes
20	36.5	rs35609661	Allelic	$3.4x10^{-8}$	$9.3x10^{-8}$	17.6 (3.1)	0.004		hippocampal atrophy x2
21	22.3	rs17592663	LRT	$3.0x10^{-8}$	$3.0x10^{-8}$	4.6 (1.1)	0.039		
21	44.3	rs1800525	Allelic	$1.7x10^{-10}$	$2.1x10^{-8}$	4.2 (0.6)	0.110		corticobasal degeneration
22	48.5	rs11705431	Allelic	$7.7x10^{-9}$	$2.5x10^{-8}$	18.3 (3.1)	0.004		AD and age of onset, LOAD

Table 2: Assessment of previously implicated variants in *APOE*. Dominant (Dom.) and recessive (Rec.) models are from PLINK, while LRT (likelihood ratio rest) is calculated using GEMMA. The LRT *p* value accounts for family structure using a linear mixed model that adjusts for kinship. β is shown for dominant and recessive models (but not for LRT, as GEMMA does not output β values), and is derived from the basis of age of dementia onset. NA = not applicable, because no individuals in this cohort were homozygous for rs7412. MAF indicates minor allele frequency in this cohort. Hom. Indicates the number of individuals homozygous for the variant. PopAF indicates the population allele frequency (TOPMed freeze 5).

Variant	Description	Dom. β (SE)	Dom. *p*	Rec. β (SE)	Rec. *p*	LRT *p*	MAF	Hom.	PopAF
rs429358 T>C	*APOE* ε4 (w/ rs7412 ref.)	-2.3 (0.7)	9.7×10^{-4}	-2.2 (2.0)	0.27	1.2×10^{-3}	0.137	8	0.156
rs121918393 C>A	*APOE* Christchurch	1.8 (1.8)	0.31	25.6 (5.4)	2.8×10^{-6}	0.027	0.019	1	8.0×10^{-6}
rs7412 C>T	*APOE* ε2 (w/ rs429358 ref.)	2.6 (0.9)	3.0×10^{-3}	NA	NA	0.014	0.067	0	0.082

Table 3: Replication of hits with genome-wide significance for AD-associated phenotypes from previous studies in this cohort. All models are from PLINK except LRT (likelihood ratio rest) which is calculated using GEMMA. Only variants with nominal significance ($p<0.05$) in at least one model and with consistent effect direction are shown, and the model with the lowest *p* value is shown. For all variants, the LRT *p* value is also shown, which accounts for family structure using a linear mixed model that adjusts for kinship. β is shown for each model (with the β from the Allelic model shown if LRT is the best model, as GEMMA does not output β values), and is derived from the basis of age of dementia onset. For previous studies, all effects are OR (95% CI) except Beecham et al., 2014 and Deming et al., 2017, which are β (SE). MAF indicates minor allele frequency in this cohort. PopAF indicates the population allele frequency (TOPMed freeze 5).

Variant	Locus	Nearest Gene	Best Model	*p*	LRT *p*	β (SE)	MAF	PopAF	Previous Study	Previous Study *p*	Effect (Error)	Effect
rs6448453	4p16.1	*CLNK*	Allelic	0.002	0.008	1.6 (0.5)	0.75	0.77	Jansen, 2019	1.9×10^{-09}	0.99 (0.98–0.99)	AD Risk
rs62341097	4q34.1	*GALNT7*	LRT	0.036	0.036	1.2 (1.0)	0.05	0.03	Beecham, 2014	6.0×10^{-09}	-1.147 (0.198)	Neuritic Plaque
rs316341	6p25	*SERPINB1*	Recessive	0.004	0.015	-1.8 (0.6)	0.68	0.71	Deming, 2017	1.8×10^{-08}	-0.025 (0.004)	CSF Aβ42
rs9271058	6p21.32	*HLA-DRB1*	Dominant	0.017	0.125	3.0 (1.3)	0.76	0.75	Kunkle, 2019	1.4×10^{-11}	0.91 (0.88–0.93)	AD Risk
rs9271192	6p21.32	*HLA-DQA1*	Dominant	0.017	0.125	3.0 (1.3)	0.76	0.75	Lambert, 2013	2.9×10^{-12}	0.90 (0.87–0.93)	AD Risk
rs10808026	7q34–q35	*EPHA1*	Dominant	0.035	0.410	1.3 (0.6)	0.24	0.19	Marioni, 2018	1.1×10^{-14}	0.91 (0.89–0.93)	AD Risk
rs4236673	8p21.1	*CLU*	Dominant	0.017	0.138	-2.3 (1.0)	0.69	0.72	Marioni, 2018	1.1×10^{-28}	1.12 (1.09–1.14)	AD Risk
rs9331896	8p21.1	*CLU*	Dominant	0.043	0.174	-1.9 (0.9)	0.64	0.57	Lambert, 2013	2.8×10^{-25}	1.16 (1.13–1.19)	AD Risk
rs983392	11q12.2	*MS4A6A*	Recessive	0.046	0.331	1.6 (0.8)	0.41	0.27	Lambert, 2013	6.1×10^{-16}	0.90 (0.87–0.92)	AD Risk
rs4985556	16q22.1	*IL34*	Allelic	0.049	0.170	-2.9 (1.5)	0.02	0.08	Marioni, 2018	3.7×10^{-08}	1.09 (1.05–1.12)	AD Risk
rs138190086	17q23.3	*CYB561*	LRT	0.015	0.015	-2.7 (1.2)	0.03	0.01	Marioni, 2018	1.9×10^{-09}	1.25 (1.16–1.35)	AD Risk
rs6069736	20q13.2–31	*CSTF1*	Recessive	0.003	0.178	7.5 (2.5)	0.14	0.13	Marioni, 2018	2.0×10^{-10}	0.89 (0.86–0.93)	AD Risk
rs6014724	20q13.2–31	*CASS4*	Recessive	0.003	0.362	7.5 (2.5)	0.15	0.12	Jansen, 2019	6.6×10^{-10}	0.98 (0.97–0.98)	AD Risk

Table 4: Coding variants of interest. Variants were required to meet one of five categories, all of which required observation of alleles in at least two individuals: (1) genome-wide (GW) significant ($p<5x10^{-8}$), (2) $p<1x10^{-5}$ with replication in at least one of the aforementioned replication cohorts, (3) a variant observed in the genome only analysis (but not observed in the imputation set) with $p<0.05$ in perfect Linkage Disequilibrium (LD) ($r^2=1$ and D'=1) with a variant with $p<1x10^{-5}$ with the same model in the imputation set, a population allele frequency > 0.01%, and CADD > 20, (4) A variant in a gene previously implicated in neurodegeneration (**Supplemental Table 5**) with $p<0.05$, a population allele frequency > 0.01%, and CADD > 20, or (5) A variant in *APP*, *PSEN1*, *PSEN2*, *MAPT*, *APOE*, *ABCA7*, *SORL1*, or *TREM2* with $p<0.05$, a population allele frequency > 1%, and CADD > 10. Source indicates if the analysis was on the imputation plus genome call set, or genome only set (implying a rare variant not imputed by TOPMed). Pos indicates hg38 position in megabases. MAC (Total) indicates minor allele count and total allele count. PopAF is the maximum population allele frequency from four databases: ExAC non-psychiatric disease, gnomAD exomes, gnomAD genomes, and TOPMed Bravo genomes.

Category	Source	Chr	Pos	ID	Gene	Protein Δ	Model	p	β (SE)	MAC (Total)	Hom	PopAF	CADD
GW significant ($p<5x10–8$)	Imputed	8	22.6	rs34059820	*SORBS3*	Arg550His	Allelic	$5.6x10^{-9}$	23.0 (3.8)	2 (688)	0	$2.3x10^{-3}$	8.2
	Imputed	8	22.6	rs143895597	*CCAR2*	Ile298Val	Allelic	$5.6x10^{-9}$	23.0 (3.8)	2 (688)	0	0.01	12.5
p<1x10-5 with replication	Imputed	11	62.8	rs544641	*ZBTB3*	Ile574Met	Genotypic	$5.5x10^{-6}$	12.7 (2.7)	22 (688)	1	0.09	11.6
	Imputed	17	81.6	rs6565617	*TSPAN10*	Ala118Thr	Genotypic	$4.8x10^{-7}$	-2.4 (0.5)	418 (688)	123	0.48	0
	Imputed	17	81.6	rs7210026	*TSPAN10*	Val134Ala	Genotypic	$4.8x10^{-7}$	-2.4 (0.5)	418 (688)	123	0.48	11.1
LD with imputed variant with p<1x10-5, Pop. AF < 0.01%, CADD > 20	Genome	4	27.0	rs1457401458	*STIM2*	Thr550Asn	LRT	$6.3x10^{-3}$	10.1 (5.7)	3 (156)	0	$4.1x10^{-6}$	24.6
	Imputed	4	27.1	rs780258210	NA	NA	LRT	$9.0x10^{-6}$	15.8 (3.6)	3 (688)	0	$2.6x10^{-4}$	NA
	Genome	10	32.0	10:32031212:T:A	*KIF5B*	Asn481Ile	Allelic	$5.6x10^{-3}$	13.0 (4.6)	3 (158)	0	Absent	26.3
	Imputed	10	32.0	rs115609415	NA	NA	Allelic	$4.4x10^{-6}$	8.2 (1.8)	11 (688)	0	$1.6x10^{-3}$	NA
	Genome	19	45.5	19:45498605:C:CT	*PPM1N*	Leu46fs	Genotypic	$1.4x10^{-3}$	11.8 (3.5)	6 (160)	1	Absent	22.9
	Imputed	19	45.5	rs561068037	NA	NA	Genotypic	$5.9x10^{-6}$	2.4 (1.3)	20 (688)	1	0.01	NA
Neurodegen. Association, Pop. AF < 0.01%, CADD > 20	Imputed	1	11.8	rs149322151	*CLCN6*	Glu779Lys	Allelic	$9.8x10^{-3}$	10.2 (3.9)	2 (688)	0	$1.1x10^{-5}$	28.2
	Imputed	4	3.2	rs1473464204	*HTT*	Thr1253Met	Allelic	$5.4x10^{-3}$	5.9 (2.1)	9 (688)	0	$4.0x10^{-5}$	22.9
	Imputed	9	77.3	rs572761766	*VPS13A*	Asn2162His	Allelic	0.016	4.2 (1.7)	11 (688)	0	$1.1x10^{-5}$	23.7
	Imputed	9	135.8	rs557219607	*KCNT1*	Arg573Cys	Allelic	$3.7x10^{-5}$	16.2 (3.9)	2 (688)	0	$4.4x10^{-5}$	25.4
	Imputed	16	16.2	rs200800189	*ABCC6*	Val711Met	Allelic	0.01	4.6 (1.8)	10 (688)	0	$7.7x10^{-5}$	25
	Imputed	19	44.9	rs121918393	*APOE*	Arg154Ser	Genotypic	$3.2x10^{-6}$	12.8 (2.7)	13 (674)	1	$8.0x10^{-6}$	31
	Genome	22	39.2	rs190122588	*PDGFB*	Arg66His	Genotypic	0.03	-6.7 (4.1)	4 (160)	0	$1.8x10^{-5}$	24.4
Key AD gene, Pop. AF < 1%, CADD > 10	Imputed	19	1	rs547447016	*ABCA7*	Glu709fs	Allelic	0.017	-4.6 (1.9)	9 (688)	0	$2.8x10^{-3}$	35
	Imputed	21	25.9	rs112263157	*APP*	Ser614Gly	LRT	$7.1x10^{-4}$	-4.2 (1.5)	15 (688)	0	$4.7x10^{-3}$	18.5

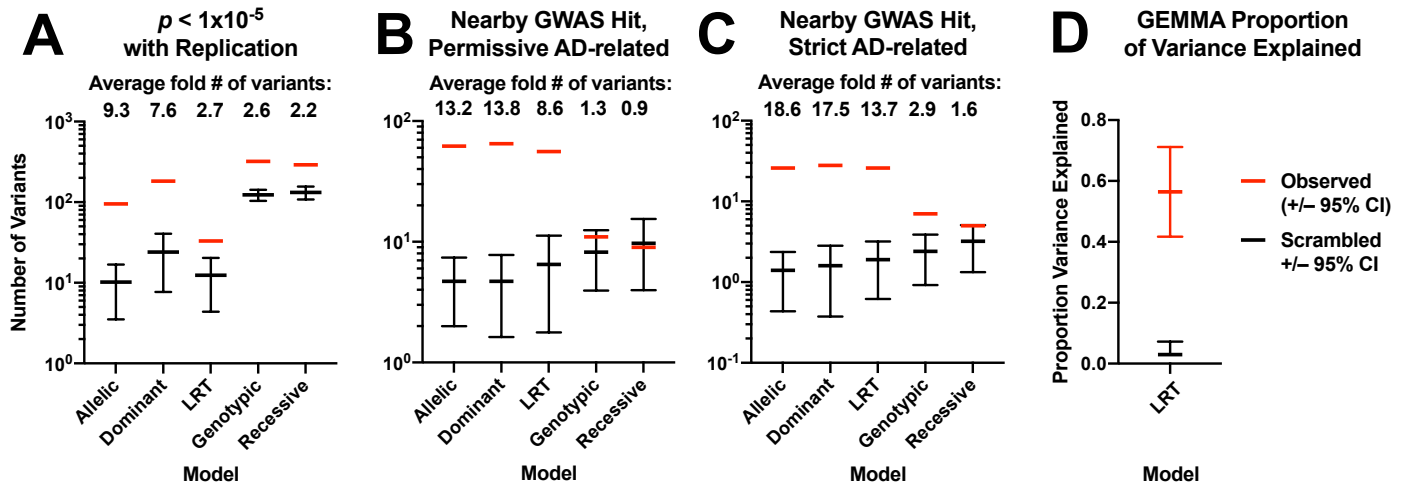


Figure 1: Comparisons to scrambled age distributions to assess biological relevance of signals observed. **(A)** Number of variants with $p < 1x10^{-5}$ in this cohort exhibiting nominal replication ($p < 0.05$) in at least 1 cohort (of three tested[7-9]) compared to the same cohort with age at dementia onset scrambled (10 scrambles) as a background control. **(B)** Number of variants +/–500 kb of $p < 5x10^{-8}$ variants (with variants considered as a continuous region from the first observed to last observed variant if multiple variants with $p < 5x10^{-8}$ were less than 5 MB from one another) in this cohort with a signal from the NHGRI-EBI GWAS catalog for permissive AD-related phenotypes (i.e. any aging, cognition, or neurodegeneration-associated phenotypes, full list in **Supplemental Table 2**) compared to the same cohort with age at dementia onset scrambled (10 scrambles) as a background control. **(C)** Number of variants +/–500 kb of $p < 5x10^{-8}$ variants (with variants considered as a continuous region from the first observed to last observed variant if multiple variants with $p < 5x10^{-8}$ were less than 5 MB from one another) in this cohort with a signal from the NHGRI-EBI GWAS catalog for strict AD-related phenotypes (i.e. AD-only phenotypes or endophenotypes, full list in **Supplemental Table 2**) compared to the same cohort with age at dementia onset scrambled (10 scrambles) as a background control. (**D**) Proportion of variance explained (PVE) from GEMMA. The 95% confidence interval provided from GEMMA is plotted for the Observed condition, while the 95% confidence interval was calculated based on the PVE values from all scrambles (10 scrambles) for clarity instead of plotting the provided 95% confidence interval outputted with each individual scramble condition. For all panels, errors bars indicated 95% confidence interval of the mean. For panels A–C, no error bars are available for the observed condition because it is a single condition. All models are from PLINK except LRT (likelihood ratio rest) which is calculated using GEMMA.

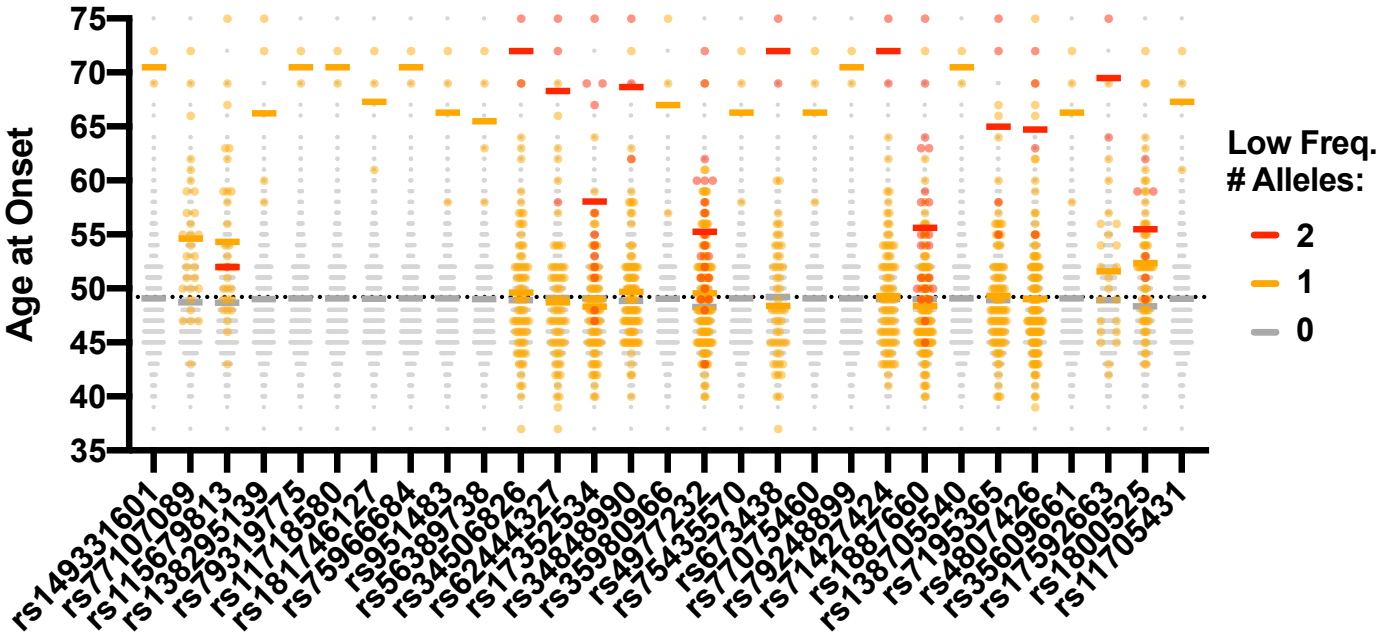


Figure 2: Genome-wide significant ($p < 5x10^{-8}$) variants by age at onset. Overall mean age is indicated by the black dashed line (49.2 years). For rs4977232 rs673438, the lower frequency allele is the reference allele. For all other variants, the minor allele is the lower frequency allele. Line indicates mean, points are individuals.

Supplemental Table 1: Summary statistics for all variants and models. (see supplemental zipped text file) [Note to reviewers: we will deposit this file to NIAGADS, but pre-publication, this file can be accessed via: http://mendel.hudsonalpha.org/Public/ncochran/Colombia/E280A-344-SuppTable1.tsv.gz]

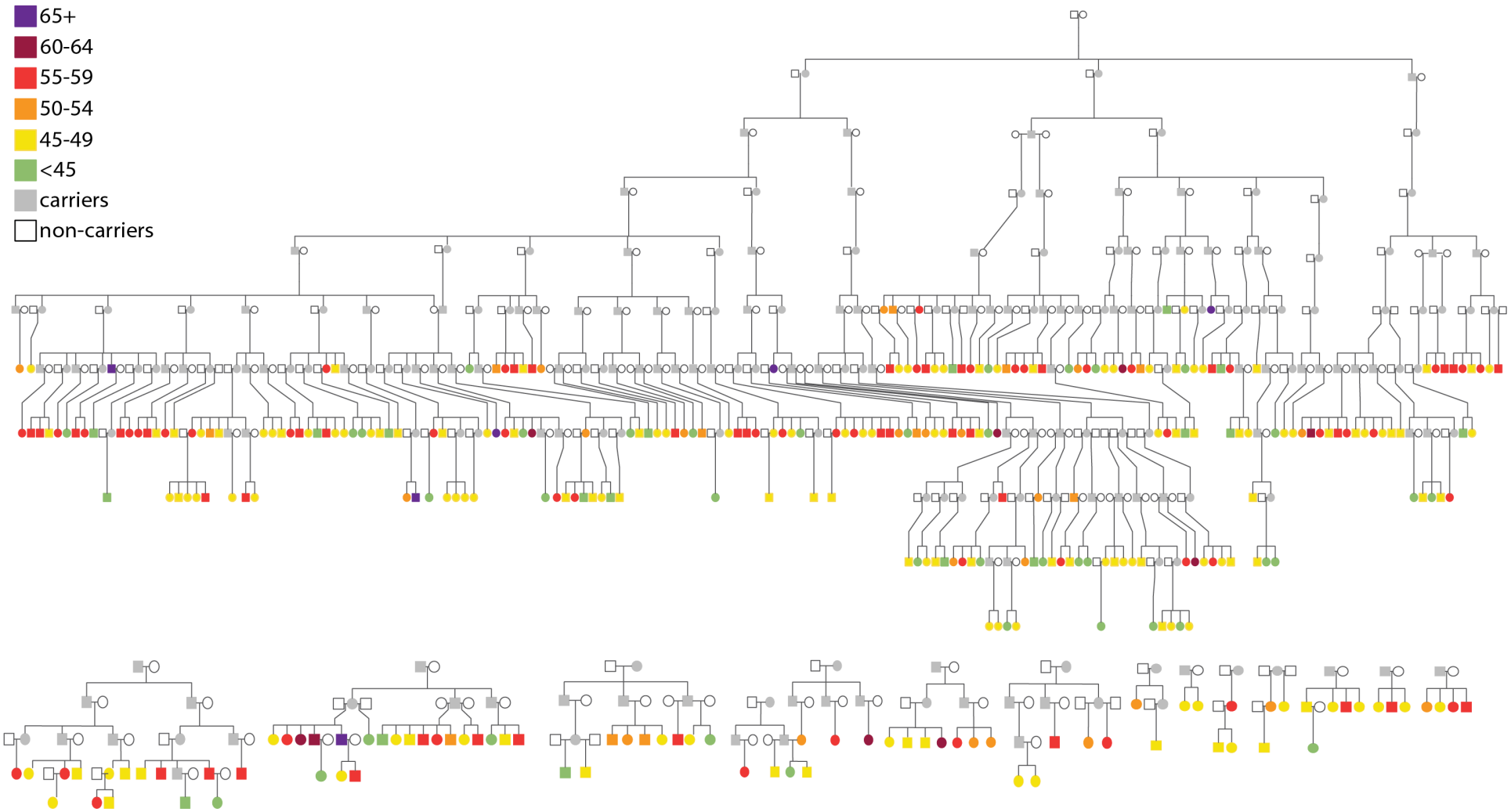
Supplemental Table 2: NHGRI-EBI terms present near genome-wide significant ($p < 5x10^{-8}$) hits in both observed and scrambled conditions. Terms selected as permissively related to AD or strictly related to AD are noted. (see supplemental Excel file)

Supplemental Table 3: NHGRI-EBI terms and associated variants present near genome-wide significant ($p < 5x10^{-8}$) hits.

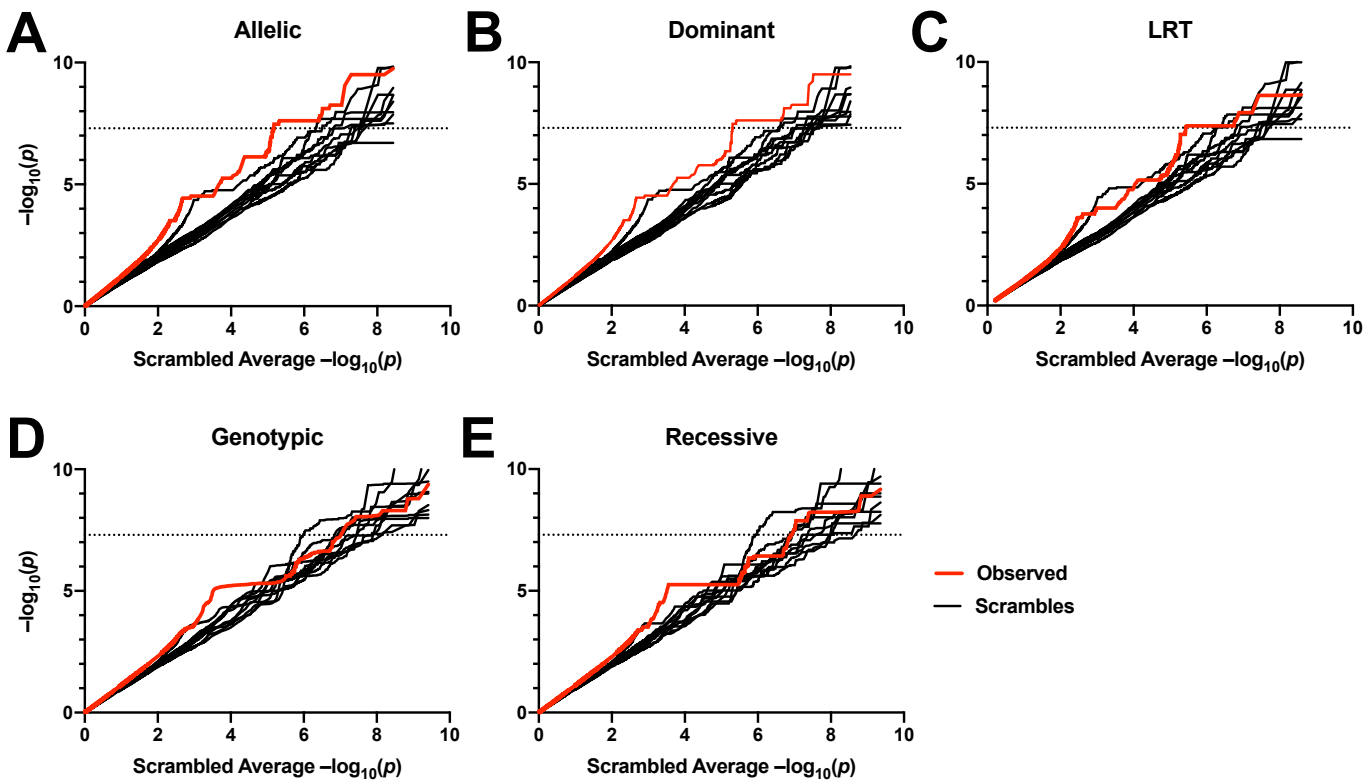
Supplemental Table 4: Full list of GWAS hits from previous studies evaluated. (see supplemental Excel file)

Supplemental Table 5: List of genes previously implicated in neurodegenerative diseases along with the source of data supporting their inclusion.

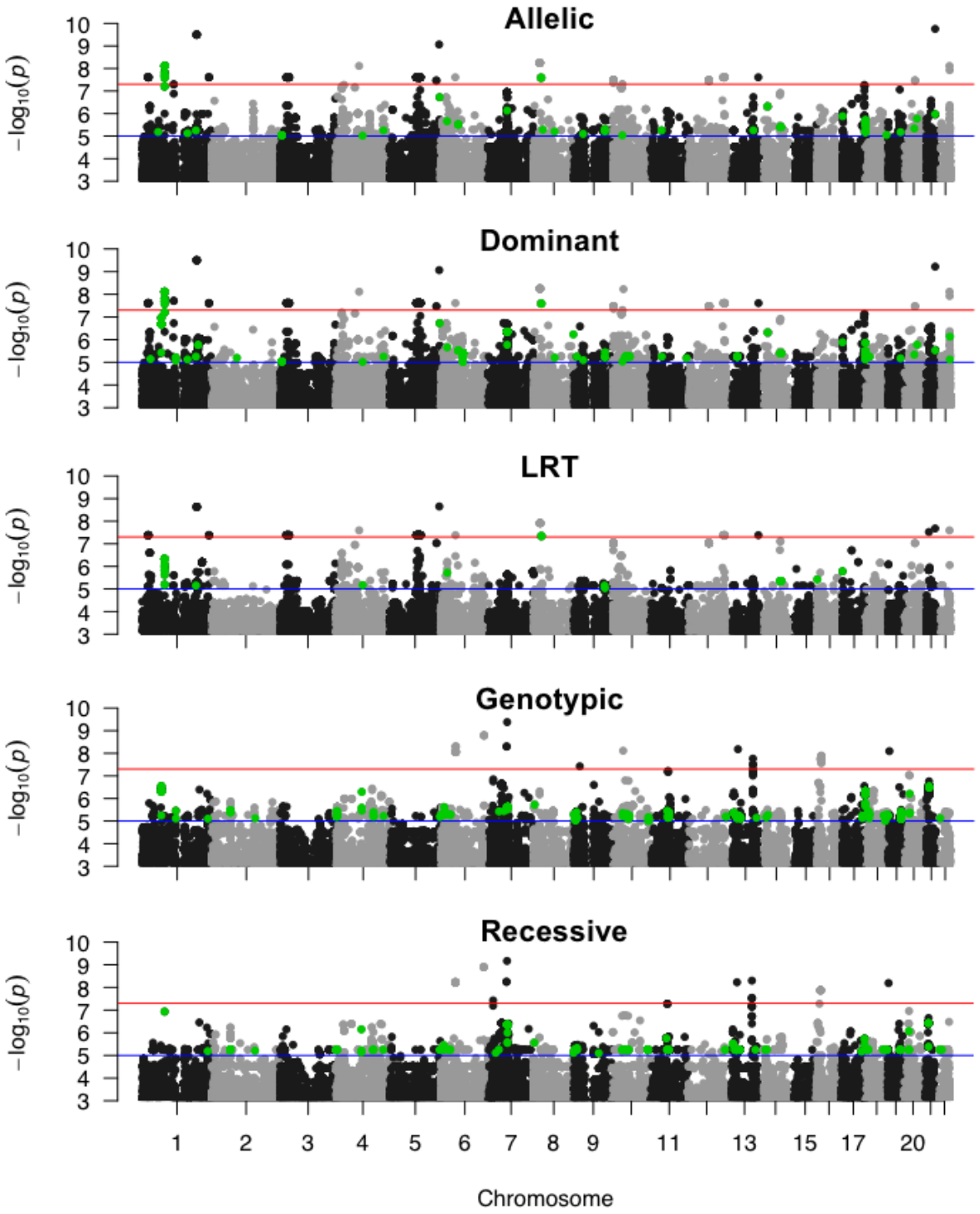
Supplemental File 1: LocusZoom plots of key regions. (see supplemental zipped file)



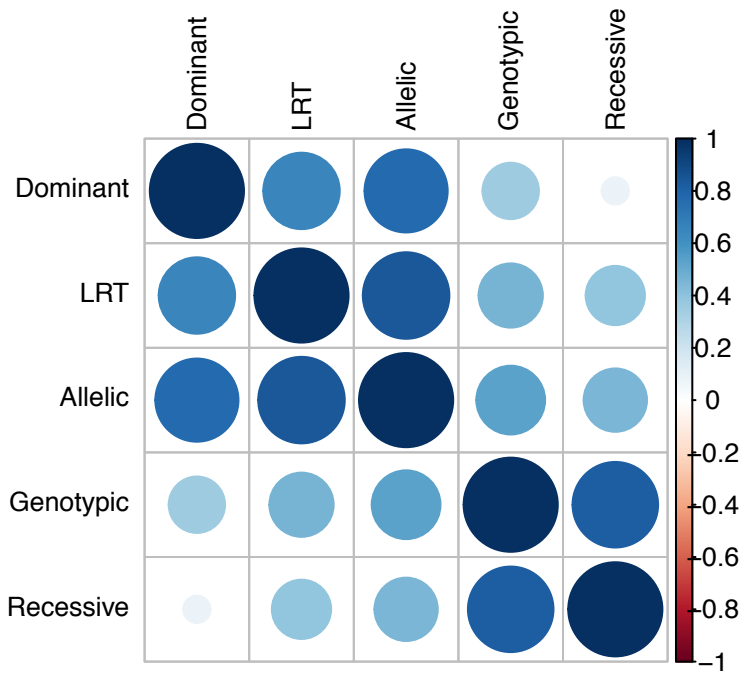
Supplemental Figure 1: Pedigrees of included patients colored according to age at dementia onset. The E280A mutation carriers (known and presumptive) that were not included in the present study are colored in gray. The large pedigree includes 252 of the sequenced participants, and their relatedness until the most recent common ancestors (which were born in 1743–1750). The small pedigrees in the lower row represent 87 additional participants. Healthy siblings and descendants of the participants were excluded for simplicity of the pedigree.



Supplemental Figure 2: QQ plots. Observed (actual age of dementia onset) $-\log_{10}(p)$ values are plotted vs. the average of scrambled age of dementia onset $-\log_{10}(p)$ values. All models are from PLINK except LRT (likelihood ratio rest) which is calculated using GEMMA.



Supplemental Figure 3: Manhattan plots for all models. Any variant with a discovery $p<1x10^{-5}$ that exhibited nominal replication ($p<0.05$) is highlighted in green. Blue line indicates $p<1x10^{-5}$, red line indicates $p<5x10^{-8}$.



Supplemental Figure 4: Pearson cross-correlations of $-\log_{10}(p)$ values for the 5 models employed. All models are from PLINK except LRT (likelihood ratio rest) which is calculated using GEMMA.