Published in final edited form as:
Hum Mutat. 2012 December ; 33(12): 1630–1634. doi:10.1002/humu.22167.

Exploratory Data from Complete Genomes of Familial Alzheimer Disease Age-at-Onset Outliers

Matthew A. Lalli[1,2], **Gloria Garcia**[3], **Lucia Madrigal**[3], **Mauricio Arcos-Burgos**[3,4], **Mary Luz Arcila**[1,5], **Kenneth S. Kosik**[1,2,5,*], and **Francisco Lopera**[3]

[1]University of California at Santa Barbara, Neuroscience Research Institute, Santa Barbara, California, United States

[2]University of California at Santa Barbara, Program in Biomolecular Science and Engineering,Santa Barbara, California, United States

[3]Grupo de Neurociencias de Antioquia, Universidad de Antioquia, Medellín, Colombia

[4]ANU College of Medicine, Biology & Environment, John Curtin School of Medical Research, The Australian National University, Translational Genomics Group, Translational Medicine Department, Canberra, ACT, Australia

[5]University of California at Santa Barbara, Department of Molecular, Cellular, and Developmental Biology, Santa Barbara, California, United States

Abstract

Identifying genes that modify the age-at-onset (AAO) of Alzheimer disease and targeting them pharmacologically represent a potential treatment strategy. In this exploratory study, we sequenced the complete genomes of six individuals with familial Alzheimer disease due to the autosomal dominant mutation p.Glu280Ala in *PSEN1* (MIM# 104311; NM_000021.3:c.839A>C). The disease and its age-at-onset are highly heritable, motivating our search for genetic variants that modulate AAO. The median AAO of dementia in carriers of the mutant allele is 49 years. Extreme phenotypic outliers for AAO in this genetically isolated population with limited environmental variance are likely to harbor onset-modifying genetic variants. A narrow distribution of AAO in this kindred suggests large effect sizes of genetic determinants of AAO in these outliers. Identity by Descent (IBD) analysis and a combination of bioinformatics filters have suggested several candidate variants for AAO modifiers. Future work and replication studies on these variants may provide mechanistic insights into the etiopathology of Alzheimer disease.

Keywords

Alzheimer disease; whole-genome sequencing; age-at-onset; Identity by Descent

Here we describe exploratory data on the complete genomic sequences of six individuals from a well-characterized kindred with familial Alzheimer disease caused by a fully penetrant autosomal dominant missense mutation encoding p.Glu280Ala in *PSEN1* on chromosome 14q24.3 [Lopera et al., 1997]. The gene product of *PSEN1,* presenilin 1, is a component of the gamma-secretase complex that cleaves amyloid beta (A4) precursor protein (*APP*; MIM# 104760) into amyloid beta (Aβ), the main component of Alzheimer

*address correspondence to: Kenneth S. Kosik, University of California at Santa Barbara, Neuroscience Research Institute, Santa Barbara, California, United States, kosik@lifesci.ucsb.edu.

Supporting Information for this preprint is available from the *Human Mutation* editorial office upon request (humu@wiley.com)

disease (AD) senile plaques [Glenner and Wong, 1984]. The median age at onset (AAO) of dementia in mutation carriers is 49 years [95% CI 49–50] [Acosta-Baena et al., 2011]. AAO in sporadic AD has a high heritability with a predicted genetic contribution of 42% [Li et al., 2002]. A study in a monogenic early-onset AD population predicted a heritability of AAO of 93% [Marchani et al., 2010]. *APOE* (MIM# 107741), the major known modifier gene of AAO, can also shift the AAO in carriers of the p.E280A mutation [Pastor et al., 2003], suggesting shared mechanisms in the monogenetic and sporadic forms of the disease.

Extreme outliers in AAO provide an opportunity to identify disease modifying factors. In this exploratory study, four extreme outliers for AAO (two a decade on either side of the mean) were selected for whole genome re-sequencing. Two individuals with the average AAO were also sequenced as controls. This study was approved by Institutional Review Boards at the Universidad de Antioquia where the samples were obtained and informed consent was obtained from all subjects. Additional information about the study participants is provided in Supp. Table S1. The narrow distribution of AAO in carriers of the p.E280A allele suggests the presence of determinants of AAO with large effect sizes in the outliers. Due to common founders and a high occurrence of endogamy and consanguinity, this kindred contains limited genomic variability and consequently a narrowed scope of search for variants that affect phenotype. Similarities in education, socio-economic status, and lifestyle among population members reduce environmental variance thereby increasing the likelihood of modifier genes among the factors that control AAO.

Although insightful, genome-wide association studies (GWAS) remain underpowered to detect association with variants of low frequency or small effect sizes. Many statistically associated variants reside in noncoding regions and lack an obvious biological mechanism of action. Associated single nucleotide polymorphisms (SNPs) may also serve as proxies for the causal SNP due to strong linkage disequilibrium (LD) and thus themselves make no contribution to phenotype. Arrays used in GWAS are typically unable to detect structural variants, which can contribute significantly to disease phenotype. Using whole-genome sequencing, we obtained a comprehensive assessment of genomic variation in each individual and address some of these shortcomings. By combining a variety of bioinformatics analyses, our exploratory investigation highlights a variety of biologically plausible mechanisms of how variants might modulate the Alzheimer AAO. While this approach can capture the scope of genetic variation, it lacks statistical significance and therefore serves as an initial step toward identifying candidate loci.

Sequencing was performed by Complete Genomics using unchained combinatorial probe anchor (cPAL) chemistry [Drmanac et al., 2010]. An average of 192 Gb of sequencing reads was generated for each genome covering over 96% of both the genome and exome at an average depth of >60X for all samples. Reads were aligned to the reference genome (NCBI Build 37) and variant calling was performed by Complete Genomics. Each genome averaged 3.46±.07 million SNPs with a genome-wide average transition to transversion ratio of 2.13, consistent with current estimates from the 1000 Genomes Project Consortium. The union of all genomic variants in the 6 genomes comprised 8.2 million variants including 6.93 million SNPs, 505,000 insertions, 567,000 deletions, and 198,000 substitutions. 1.05 million novel variants absent from dbSNP 132 were observed. 1.16 million of the variants were shared across all 6 genomes and 450±90 thousand variants were private to each genome.

The intersection of variants shared in the early or late onset outlier phenotype pairs will contain the AAO modifier genes common to the outliers. A filter was established to find variants that were 1) homozygous (two copies of the same variant) in both genomes of the extreme phenotype pair and absent or heterozygous in the remaining 4 genomes, or 2) present in the phenotype pair and absent from the other 4 genomes. Variants falling into

these categories are hereafter classified as tracking with the disease AAO. To determine which variants are likely to contribute to disease phenotype, we focus primarily on non-synonymous variants (NSVs) that track with the disease AAO. Future work will assess contributions from non protein coding variants such as the rare homozygous variant NR_030326.1: n.29T>C located in the mature sequence of *MIR596* and present only in the late-onset phenotype pair. The results of our screening process are summarized in Supp. Table S2. 275 and 151 distinct genes contained NSVs tracking with early and late onset respectively. Seven genes contained both potential risk and protective variants, similar to APOE in which the ε4 allele is a risk factor and the ε2 allele is protective [Corder et al., 1993; Corder et al., 1994], potentially implicating these genes as molecular switches for the onset of the disease (Supp. Table S3).

We used Identity by Descent (IBD) mapping to identify regions of these genomes inherited from a recent common ancestor, which are likely to contain causal alleles shared between individuals with the same phenotype. We expect casual alleles to be rare because common variants tend to have smaller effect sizes [Kryukov et al., 2007] which likely cannot account for our extreme AAO phenotypes. Some casual variants are also likely to have arisen recently or else they would have been removed by purifying selection [Browning and Thompson, 2012]. Variants may have arisen through drift during a recent bottleneck in the founder population. Rare, recent, or founder introduced causal variants are readily detected through IBD mapping.

Using the GERMLINE algorithm [Gusev et al., 2008], we identified IBD segments between all pairs of individuals and visualized these regions using Circos [Krzywinski et al., 2009] (Fig. 1). The parameter used for IBD determination was a minimum region of 3 MB containing fewer than 4 mismatches between two genomes. With about 1 SNP/KB, in the human genome, we expect around 3000 SNPs in the minimum region. The length and number of IBD segments decline exponentially after every generation and are largely undetectable by current algorithms after ~25 generations. Interestingly, the only region of IBD (excluding centromeres) shared in all 6 genomes was located on chromosome 14 and contained the causal *PSEN1* p.E280A mutation (Fig. 1A). This result demonstrates the utility of this method in identifying regions likely to contain causal variants. We next identified regions of IBD between the pairs of extreme outliers and the NSVs within these regions that track with AAO. Relationship inference analysis, although underestimating kinship coefficients, suggests that cryptic relatedness did not contribute to IBD sharing between our extreme phenotype pairs. (Supp. Table S4).

In the delayed onset pair, a unique stretch of IBD on chromosome 16 spanned 5.3 MB and contained 3 NSVs that track with the disease onset (Fig 1B). One of the variants in *CCL22* (Supp. Table S5) was very rare. The other two variants were rare, homozygous and affect the genes *GPR56* (MIM# 604110; NM_001145770.1: c.918A>C, p.Gln306His) and *CCDC135* (NM_032269.5: c.1298C>T, p.Pro433Leu). The protein encoded by *GPR56* is involved in brain cortical patterning. The gene product of *CCDC135* is involved in motile cilia and could possibly affect cerebrospinal fluid flow [Yang et al., 2011]. Another region of IBD shared in the late-onset pair fell on chromosome 6 and contained a rare variant in *MYO6* (MIM# 600970; NM_004999.3:c.1120T>C, p.Tyr374His) encoding a protein which is highly expressed in the brain and enriched at the postsynaptic density [Osterweil et al., 2005].

In the early-onset phenotype pair, a stretch of IBD on chromosome 1 contained 6 distinct NSVs including several rare variants (Fig. 1C). The six variants affected 5 genes: *EPHA10* (Table 1), *INPP5B* (MIM# 147264; NM_005540.2: c.136G>A, p.Gly46Ser), *RLF* (MIM# 180610; NM_012421.3: c.5053C>G, p.Gln1685Glu), *DEM1* (NM_022774.1: c.515G>T,

p.Gly172Val and NM_022774.1: c.343G>A, p.Asp115Asn), and *KCNQ4* (MIM# 603537; NM_004700.3: c.1365T>G, p.His455Gln). Many of these genes are functional in neurons. INPP5B is a type 2 5-phosphatase that localizes to the Golgi and is co-localized with the gene product of *RABEP1* (MIM# 603616) [Shin et al., 2005]. RABEP1-positive endosomes are one of the earliest pathological hallmarks of AD [Cataldo et al., 2000]. *KCNQ4* encodes a potassium voltage gated channel critical in regulating neuronal excitability. Another region of IBD on chromosome 1 contained 3 NSVs in the genes *SWT1* (NM_017673.6: c.442A>G, p.Ile148Val), *PRG4* (Supp. Table S5), and *HMCN1* (MIM# 608548; NM_031935.2: c. 7253T>C, p.Ile2418Thr). An IBD segment on chromosome 6 contained two nonsynonymous variants in the early-onset pair. These variants reside within the carnitine transporter *SLC22A16* (MIM# 608276; NM_033125.2: c.755T>C. p.Val252Ala) and glutaminyl-tRNA synthase *QRSL1* (NM_018292.4: c.32C>T, p.Ala11Val). Carnitine may protect against the toxicity of Aβ [Virmani et al., 2001] thus defects in this transporter could accelerate the onset of the disease.

Although we lack the statistical power to detect association between genomic variants and AAO, several bioinformatic tools were implemented to predict their effects on gene function. The most promising results of these combined analyses are presented in Table 1. Nonsense and frameshift variants are likely to cause a loss of protein function (Supp. Table S6). For all NSVs, loss of function in the resulting protein was predicted using SIFT [Ng and Henikoff, 2006] and PolyPhen2 [Adzhubei et al., 2010] which account for evolutionary conservation and structural features. Novel NSVs are likely to have deleterious effects. Only one novel NSV segregated exclusively in the early-onset phenotype pair. Located in *C5AR1* (MIM# 113995; NM_001736.3: c.524G>A, p.Arg175Gln), this variant was given the maximum deleterious prediction by PolyPhen2. This position is highly conserved across species and the gene resides in region of linkage with AD. This gene is expressed in hippocampal and cortical neurons and a role in Alzheimer neurodegeneration has been proposed [Farkas et al., 2003], making this variant an extremely promising candidate for follow up studies. Rare variants are more likely to have functional impact than common variants because deleterious variants will be eliminated through negative selection [Ford, 1965]. Variants were sorted by frequency observed in the 1000 Genomes Project Consortium data and rare variants were evaluated as candidate modifiers of AAO (Table 1, Supp. Table S5).

For each NSV tracking with AAO, we determined whether it was previously tested in association with AD. The AlzGene database contains the results from over a thousand candidate gene studies and GWAS related to AD [Bertram et al., 2007]. By identifying variants whose statistical association is already shown, we can see if the phenotypic outliers are enriched for risk genes. Linkage analysis has identified chromosomal regions likely to harbor the causal variants in AD, but the underlying variants often remain unknown. Having already identified almost all the genetic variation within our sequenced individuals, we can then determine the variants within known linkage regions. Putative AD linkage regions were obtained from genome-wide linkage meta-analyses [Butler et al., 2009] and allow us to take advantage of prior studies to evaluate the potential effects of our variants. Through this analysis, we found that both early-onset individuals are homozygous for a variant in *MTHFR* (MIM# 607093; NM_005957.4: c.665C>T) which encodes a p.Ala222V substitution in methylene tetrahydrofolate reductase that reduces the enzyme's activity by 70% and confers an Alzgene meta-analysis odds ratio of 1.13. This variant results in elevated homocysteine levels, which is a reported risk factor for AD [Frosst et al., 1995; Seshadri et al., 2002]. The delayed onset pair shared a variant in *APH1B* (MIM# 607630; NM_031301.3: c.651T>G) encoding a p.Phe217Leu substitution. The gene product of *APH1B* is a member of the gamma-secretase complex. Alzgene meta-analysis provides an

OR of 1.52 for carriers of this allele. In our sample, this allele seems to work in the opposite direction, possibly by mitigating the effects of the deleterious *PSEN1* mutation.

We next sought to identify whether individual genes contained multiple NSVs, which should increase the likelihood of a loss of function. Roughly half of the genes whose variants co-segregate with AAO contained multiple NSVs, with a third of these co-segregating with AAO. The number of NSVs that track with onset per gene are shown in Supp. Figure S1. From this analysis, we observed two missense mutations in the phospolipid-transporting ATPase encoded by *ATP10A* (MIM# 605855; NM_024490.3: c.2350G>A and c.3516G>C) that co-segregate exclusively in the early-onset genomes and have MAF < 0.10. The resulting substitutions p.Ala784Thr and p.Trp1172Cys both affect highly conserved amino acids.

A problem with studying the genetics of complex diseases is the difficulty in identifying gene-gene interactions. Current statistical methods are severely underpowered to detect all but the strongest epistatic interactions. We employed a simple but novel paralog analysis to begin to address this issue. Through gene duplication, functionally similar paralogs may compensate for reduced expression of one another. By examining variation within paralogous gene sets, we may be able to enhance our predictions of the effects of a given variant. A list of paralogs for each of the genes containing NSVs that track with disease onset was generated from HomoloGene, Ensembl, and Pseudogene annotations. We then identified all the NSVs contained in these paralogs and in which genomes these variants were found. Our initial list of genes whose variants track with AAO contained 15 sets (2 or more genes per set) of paralogous genes. About half of the genes whose variants track with AAO have paralogs that also contain NSVs. Many genes located in regions in linkage with AD have paralogs that are also in regions of linkage or have variants that are extremely rare. For example, in the late onset genomes, *UNC13D* (MIM# 608897) only has only 1 paralog *BAIAP3* (MIM# 604009) encoding a brain-specific angiogenesis inhibitor. *UNC13D* lies in a region statistically linked with AD and *BAIAP3* contains a rare (MAF = 0.0544) homozygous variant (NM_001199096.1: c.1532A>C, p.Asp511Ala). In the early-onset pair, there were homozygous NSVs in *BICC1* (MIM# 614295; NM_001080512.1: c.2827T>C, p.Ser943Pro) and its only paralog *ANKS6* (NM_173551.3: c.1930G>A, p.Val644Ile). Both of these genes are in regions with high LOD scores for AD linkage. These observations suggest the hypothesis that variants in both paralogs might be required to have a functional effect, with neither variant significantly associated with AD on its own. Furthermore, deleterious variants in genes that lack redundant paralogs may be more likely to have a cellular phenotype (Supp. Table S7).

In conclusion, by combining several bioinformatic approaches and incorporating *a priori* biological information into our exploratory analysis of relatively few complete genomes, we can identify the scope of possible variants putatively related to the disease phenotype. Unlike GWAS, our analysis can highlight novel and rare variants and, by primarily focusing on NSVs, provide readily interpretable biological mechanisms as shown in the IBD analysis. Future work and replication studies on these variants may provide mechanistic insights into the etiopathology of Alzheimer disease. As the number of sequenced genomes increases, the regions of shared IBD between phenotypic groups will shrink, further narrowing our search scope for causal variants. Identification of modifier loci for phenotypic variation in an extended genetically isolated pedigree affected with a monogenic form of Alzheimer disease could apply to the more common sporadic form of the disease and point the way toward much needed novel pharmacologic targets.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Matt Huentelman for reading the manuscript and the Programa de Sostenibilidad, Universidad de Antioquia. Support was provided by the Errett Fisher Foundation and the NIH Fogarty Institute grants R21 AG024063 and R01 AG029802.

References

Aasheim HC, Delabie J, Finne EF. Ephrin-A1 binding to CD4+ T lymphocytes stimulates migration and induces tyrosine phosphorylation of PYK2. Blood. 2005; 105:2869–2876. [PubMed: 15585656]

Acosta-Baena N, Sepulveda-Falla D, Lopera-Gómez CM, Jaramillo-Elorza MC, Moreno S, Aguirre-Acevedo DC, Saldarriaga A, Lopera F. Pre-dementia clinical stages in presenilin 1 E280A familial early-onset Alzheimer's disease: a retrospective cohort study. Lancet Neurol. 2011; 10:213–220. [PubMed: 21296022]

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. Nat Methods. 2010; 7:248–249. [PubMed: 20354512]

Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. Systematic metaanalyses of Alzheimer disease genetic association studies: the AlzGene database. Nat Genet. 2007; 39:17–23. [PubMed: 17192785]

Browning BL, Browning SR. A fast, powerful method for detecting identity by descent. Am J Hum Genet. 2011; 88:173–182. [PubMed: 21310274]

Browning SR, Thompson EA. Detecting rare variant associations by identity by descent mapping in case-control studies. Genetics. 2012; 190:1521–1531. [PubMed: 22267498]

Butler AW, Ng MY, Hamshere ML, Forabosco P, Wroe R, Al-Chalabi A, Lewis CM, Powell JF. Meta-analysis of linkage studies for Alzheimer's disease--a web resource. Neurobiol Aging. 2009; 30:1037–1047. [PubMed: 19362756]

Cataldo AM, Peterhoff CM, Troncoso JC, Gomez-Isla T, Hyman BT, Nixon RA. Endocytic pathway abnormalities precede amyloid β deposition in sporadic Alzheimer's disease and Down syndrome. Am J Pathol. 2000; 157:277–286. [PubMed: 10880397]

Cheng JS, Dubal DB, Kim DH, Legleiter J, Cheng IH, Yu GQ, Tesseur I, Wyss-Coray T, Bonaldo P, Mucke L. Collagen VI protects neurons against Abeta toxicity. Nat Neurosci. 2009; 12:119–121. [PubMed: 19122666]

Corder EH, Saunders AM, Risch NJ, Strittmatter WJ, Schmechel DE, Gaskell PC Jr, Rimmler JB, Locke PA, Conneally PM, Schmader KE, Small GW, Roses AD, et al. Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. Nat Genet. 1994; 7:180–184. [PubMed: 7920638]

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. [PubMed: 8346443]

Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, Carnevali P, Nazarenko I, Nilsen GB, Yeung G, Dahl F, Fernandez A, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science. 2010; 327:78–81. [PubMed: 19892942]

Farkas I, Takahashi M, Fukuda A, Yamamoto N, Akatsu H, Baranyi L, Tateyama H, Yamamoto T, Okada N, Okada H. Complement C5a receptor-mediated signaling may be involved in neurodegeneration in Alzheimer's disease. J Immunol. 2003; 170:5764–5771. [PubMed: 12759460]

Ford EB. Genetic Polymorphism. MIT Press, Cambridge, MA. Proc R Soc Lond B Biol Sci. 1965; 164:350–351. [PubMed: 4379524]

Frosst P, Blom HJ, Milos R, Goyette P, Sheppard CA, Matthews RG, Boers GJ, den Heijer M, Kluijtmans LA, van den Heuvel LP, Rozen R. A candidate genetic risk factor for vascular disease:

a common mutation in methylenetetrahydrofolate reductase. Nat Genet. 1995; 10:111–113. [PubMed: 7647779]

Glenner GG, Wong CW. Alzheimer's disease: Initial report of the purification and characterization of a novel cerebrovascular amyloid protein. Biochem Biophys Res Commun. 1984; 120:885–890. [PubMed: 6375662]

Grupe A, Li Y, Rowland C, Nowotny P, Hinrichs AL, Smemo S, Kauwe JS, Maxwell TJ, Cherny S, Doil L, Tacey K, van Luchene R, et al. A scan of chromosome 10 identifies a novel locus showing strong association with late-onset Alzheimer disease. Am J Hum Genet. 2006; 78:78–88. [PubMed: 16385451]

Gusev A, Lowe JK, Stoffel M, Daly MJ, Altshuler D, Breslow JL, Friedman JM, Pe'er I. Whole population, genomewide mapping of hidden relatedness. Genome Res. 2008; 19:318–326. [PubMed: 18971310]

Hu J, Igarashi A, Kamata M, Nakagawa H. Angiotensin-converting enzyme degrades Alzheimer amyloid beta-peptide (A beta); retards A beta aggregation, deposition, fibril formation; and inhibits cytotoxicity. J Biol Chem. 2001; 276:47863–47868. [PubMed: 11604391]

Kryukov GV, Pennacchio LA , Sunyaev SR. Most Rare Missense Alleles Are Deleterious in Humans: Implications for Complex Disease and Association Studies. Am J Hum Genet. 2007; 80:727–739. [PubMed: 17357078]

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an Information Aesthetic for Comparative genomics. Genome Res. 2009; 19:1639–1645. [PubMed: 19541911]

Li YJ, Scott WK, Hedges DJ, Zhang F, Gaskell PC, Nance MA, Watts RL, Hubble JP, Koller WC, Pahwa R, Stern MB, Hiner BC, et al. Age at onset in two common neurodegenerative diseases is genetically controlled. Am J Hum Genet. 2002; 70:985–993. [PubMed: 11875758]

Lopera F, Ardilla A, Martínez A, Madrigal L, Arango-Viana JC, Lemere CA, Arango-Lasprilla JC, Hincapíe L, Arcos-Burgos M, Ossa JE, Behrens IM, Norton J, et al. Clinical features of early-onset Alzheimer disease in a large kindred with an E280A presenilin-1 mutation. JAMA. 1997; 277:793–799. [PubMed: 9052708]

Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. Bioinformatics. 2010; 26:2867–2873. [PubMed: 20926424]

Marchani EE, Bird TD, Steinbart EJ, Rosenthal E, Yu CE, Schellenberg GD, Wijsman EM. Evidence for three loci modifying age-at-onset of Alzheimer's disease in early-onset PSEN2 families. Am J Med Genet B Neuropsychiatr Genet. 2010; 153B:1031–1041. [PubMed: 20333730]

Miller KG, Alfonso A, Nguyen M, Crowell JA, Johnson CD, Rand JB. A genetic selection for Caenorhabditis elegans synaptic transmission mutants. Proc Nat Acad Sci. 1996; 93:12593–12598. [PubMed: 8901627]

Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. Nucl Acids Res. 2006; 31:3812–3814. [PubMed: 12824425]

Osterweil E, Wells DG, Mooseker MS. A role for myosin VI in postsynaptic structure and glutamate receptor endocytosis. J Cell Biol. 2005; 168:329–338. [PubMed: 15657400]

Pastor P, Roe CM, Villegas A, Bedoya G, Chakraverty S, García G, Tirado V, Norton J, Ríos S, Martínez M, Kosik KS, Lopera F, et al. Apolipoprotein E epsilon4 modifies Alzheimer's disease onset in an E280A PS1 kindred. Ann Neurol. 2003; 54:163–169. [PubMed: 12891668]

Perl DP, Olanow CW, Calne D. Alzheimer's disease and Parkinson's disease: distinct entities or extremes of a spectrum of neurodegeneration? Ann Neurol. 1998; 44:S19–S31. [PubMed: 9749570]

Roach JC, Glusman G, Smit AF, Huff CD, Hubley R, Shannon PT, Rowen L, Pant KP, Goodman N, Bamshad M, Shendure J, Drmanac R, et al. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science. 2010; 328:636–639. [PubMed: 20220176]

Satake W, Nakabayashi Y, Mizuta I, Hirota Y, Ito C, Kubo M, Kawaguchi T, Tsunoda T, Watanabe M, Takeda A, Tomiyama H, Nakashima K, et al. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. Nat Genet. 2009; 41:1303–1307. [PubMed: 19915576]

Seshadri S, Beiser A, Selhub J, Jacques PF, Rosenberg IH, D'Agostino RB, Wilson PW, Wolf PA. Plasma homocysteine as a risk factor for dementia and Alzheimer's disease. N Engl J Med. 2002; 346:476–483. [PubMed: 11844848]

Shin HW, Hayashi M, Christoforidis S, Lacas-Gervais S, Hoepfner S, Wenk MR, Modregger J, Uttenweiler-Joseph S, Wilm M, Nystuen A, Frankel WN, Solimena M, et al. An enzymatic cascade of Rab5 effectors regulates phosphoinositide turnover in the endocytic pathway. J Cell Bio. 2005; 170:607–618. [PubMed: 16103228]

Virmani MA, Caso V, Spadoni A, Rossi S, Russo F, Gaetani F. The action of acetyl-L-carnitine on the neurotoxicity evoked by amyloid fragments and peroxide on primary rat cortical neurones. Ann N Y Acad Sci. 2001; 939:162–178. [PubMed: 11462768]

Wu H, Fan J, Zhu L, Liu S, Wu Y, Zhao T, Wu Y, Ding X, Fan W, Fan M. Sema4C expression in neural stem/progenitor cells and in adult neurogenesis induced by cerebral ischemia. J Mol Neurosci. 2009; 39:27–39. [PubMed: 19189244]

Yang Y, Cochran DA, Gargano MD, King I, Samhat NK, Burger BP, Sabourin KR, Hou Y, Awata J, Parry DA, Marshall WF, Witman GB, et al. Regulation of flagellar motility by the conserved flagellar protein CG34110/Ccdc135/FAP50. Mol Biol Cell. 2011; 22:976–987. [PubMed: 21289096]

Zhou T, Wang S, Ren H, Qi XR, Luchetti S, Kamphuis W, Zhou JN, Wang G, Swaab DF. Dendritic cell nuclear protein-1, a novel depression-related protein, upregulates corticotropin-releasing hormone expression. Brain. 2010; 133:3069–3079. [PubMed: 20693543]

Zubenko GS, Hughes HB 3rd. Predicted gene sequence C10orf112 is transcribed, exhibits tissue-specific expression, and may correspond to AD7. Genomics. 2009; 93:376–382. [PubMed: 19103277]

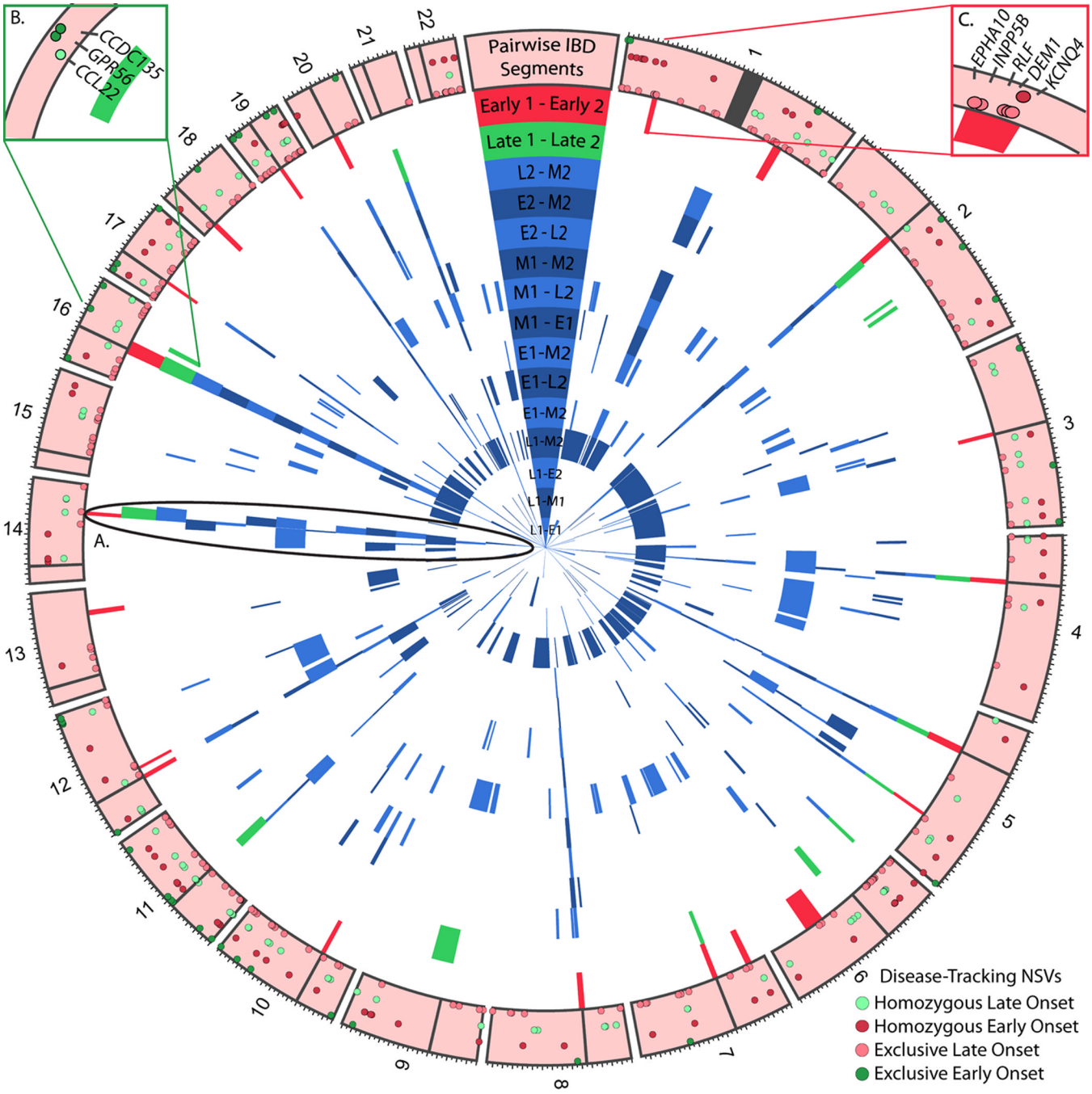

Figure 1.

Identity By Descent (IBD) Mapping. IBD segments were determined for each pair of genomes in our study and plotted as tracks within the center of the circle. Labels indicate which two genomes are being compared in each track. The six individuals in the study are labeled Early (E) 1 and 2, Mean (M) 1 and 2, and Late (L) 1 and 2, depending on the age at onset. IBD segments between the two early outliers (Early 1 and Early 2) are shown red. IBD segments between the two late outliers (Late 1 and Late 2) are shown in green. Numbered chromosomes are shown in the outermost track, with centromeres drawn in black. Within the chromosomes, nonsynonymous variants (NSVs) tracking with AAO are plotted as colored circles. A: The only region of IBD shared across all six genomes contains

the causal PSEN1 p.Glu280Ala mutation. B: A stretch of IBD unique to the late-onset pair contains 3 rare variants. C: A stretch of IBD unique to the early-onset pair contains 6 nonsynonymous variants in 5 genes.

Table 1

Nonsynonymous variants potentially implicated in modifying age at onset

Onset	Gene	Variant	Filters	Putative Function*
Early	*ACE:* Angiotensin I converting enzyme MIM# 106180	NM_000789.3 c.731A>G, p.Tyr244Cys *chr17:61557773*	SIFT score = 0 MAF = 0.008	Several variants in this gene are associated with Alzheimer disease. ACE degrades Aβ [Hu et al., 2001]
Early	*EPHA10:* EPH receptor A10 MIM# 611123	NM_173641.2 c.205C>T, p.Arg69Cys *chr1:38227722*	100% conserved across orthologs IBD in early onset	Neuronal cell communication [Aasheim et al., 2005]
Early	*LIF:* Leukemia Inhibitory Factor MIM# 159540	NM_002309.3 c.280G>A, p.Ala94Thr *chr22:30639969*	Conserved across domains/species MAF = 0.010	Cytokine with roles in neuronal differentiation and signaling
Early	*MTFP1:* Mitochondrial fission process 1 MIM# 610235	NM_001003704.2 c.368A>T, p.Asp123Val *chr22:30824621*	MAF = 0.023	Loss of function results in caspase cascade activation, apoptosis
Early	*NEFM:* Neurofilament, medium polypeptide MIM# 162250	NM_005382.2 c.1423G>A, p.Ala475Thr *chr8:24774791*	MAF = 0.004	Neuron intracellular transport and axoskeletal component
Early	*SEC14L2:* MIM# 607558	NM_012429.3 c.410A>G, p.His137Arg *chr22:30803579*	MAF = 0.008	Cholesterol biosynthesis and phospholipid transport
Early	*SEC14L4:* MIM# 612825	NM_174977.3 c.907C>T, p.Leu303Phe *chr22:30887825* c.973G>C, p.Gly325Arg *chr22:30887668*	90% conserved across species MAF = 0.0082	Biogenesis of Golgi-derived transport vesicles
Late	*NPC1:* Niemann-Pick disease C1 MIM# 607623	NM_000271.3 c.709C>T, p.Pro237Ser *chr18:21140367*	MAF = 0.006	Intracellular trafficking of cholesterol, Niemann-Pick disease
Late	*RIC8A:* Resistance to inhibitors of cholinesterase MIM# 609146	NM_021932.4 c.624_626delCCC, p.Pro208_Pro209? *chr11:209898*	Amino Acid Deletion	C.elegans RIC8A mutants have defective synaptic transmission [Miller et al., 1996]
Late	*SEMA4C:* sema domain, Ig domain TM domain, and short cytoplasmic domain 4C MIM# 604462	NM_017789.4 c.2077G>T, c.2078A>T, p.Glu693Leu *chr2:97526787*	Substitution MAF = 0.018 MAF = 0.018	Axon guidance Cortical neurons expression Interacts with post-synaptic density protein 95 (*DLG4*, MIM# 602887) [Wu et al., 2009]

After all bioinformatics analyses, a promising list of candidate factors was compiled.

*Unless otherwise stated, functions are gathered from UniProtKB/Swiss-Prot Function Entries. Please refer to these for relevant citations.

Mutations are reported using standard nomenclature (http://www.hgvs.org/mutnomen/) with reference transcript listed, complementary DNA (cDNA) sequence position numbered such that +1 corresponds to the A of the ATG initiation codon, and a protein sequence derived from and numbered based on the same reference. Chromosome and starting positions are given in GRCH37 coordinates. MAF = minor allele frequency. IBD = identity by descent.