



Published in final edited form as:

*Neurogenetics*. 2009 July ; 10(3): 183–190. doi:10.1007/s10048-009-0182-4.

## Extended tracts of homozygosity identify novel candidate genes associated with late onset Alzheimer's Disease

M. A. Nalls<sup>\*,1</sup>, R. J. Guerreiro<sup>\*,1,2</sup>, J. Simon-Sanchez<sup>1</sup>, J. T. Bras<sup>1,2</sup>, B. J. Traynor<sup>3</sup>, J. R. Gibbs<sup>1</sup>, L. Launer<sup>4</sup>, J. Hardy<sup>5</sup>, and A. B. Singleton<sup>1,+</sup>

<sup>1</sup>Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland 20892, USA

<sup>2</sup>Center for Neurosciences and Cell Biology, Faculty of Medicine, University of Coimbra, 3004-504 Coimbra, Portugal

<sup>3</sup>Neurogenetics Branch, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20892, USA

<sup>4</sup>Laboratory of Epidemiology, Demography and Biometry, Intramural Research Program, National Institute on Aging, Bethesda, Maryland 20892, USA

<sup>5</sup>Department of Molecular Neuroscience and Reta Lila Weston Institute of Neurological Studies, Institute of Neurology, University College London, Queen Square, London WC1N 3BG, UK

### Abstract

Large tracts of extended homozygosity are more prevalent in outbred populations than previously thought. With the advent of high-density genotyping platforms, regions of extended homozygosity can be accurately located allowing for the identification of rare recessive risk variants contributing to disease. We compared measures of extended homozygosity (greater than 1 megabase in length) in a population of 837 late onset Alzheimer's disease (LOAD) cases and 550 controls. In our analyses, we identify one homozygous region on chromosome 8 that is significantly associated with LOAD after adjusting for multiple testing. This region contains seven genes from which the most biologically plausible candidates are *STAR*, *EIF4EBP1* and *ADRB3*. We also compared the total numbers of homozygous runs and the total length of these runs between cases and controls, showing a suggestive difference in these measures (p-values 0.052-0.062). This research suggests a recessive component to the etiology of LOAD.

### Keywords

Alzheimer's disease; homozygous regions; *APOE*; genetics

### INTRODUCTION

Recent research has noted that the human genome contains a higher frequency of extended tracts of contiguous homozygous single nucleotide polymorphisms (SNPs) than previously expected [1][2][3]. The occurrence of these long homozygous tracts in uninterrupted sequences may represent deletion polymorphisms, loss of heterozygosity, segmental

<sup>†</sup>Corresponding author: Andrew Singleton, Molecular Genetics Section and Laboratory of Neurogenetics, NIA, NIH, Building 35, Room 1A1014, 35 Convent Drive, Bethesda, MD 20892, USA, Tel 001 301 451 6079; Fax 001 301 451 5466,

§singleta@mail.nih.gov .

\*These authors contributed equally to this work

uniparental disomy, low minor allele frequency, or autozygosity. Recent data has suggested that these tracts most likely represent extended homozygosity by virtue of parental descent from a common ancestor [2][4][5][6]. The longest tracts are expected to occur in more recently inbred populations with subsequent recombination events interrupting long chromosomal segments [1]. Our group has reported the unexpected high degree of apparent parental consanguinity in control individuals from North America (9.5% of studied individuals harboring homozygous tracts larger than 5Mb) [7]. Similar numbers (6.6%) were presented by Li L et al when studying an outbred population of unrelated Han Chinese [2] and by Gibson et al who reported that 1393 tracts exceeding 1Mb in length were observed in the 209 unrelated HapMap individuals studied [1]. The same conclusion was obtained by McQuillan and colleagues in a recent manuscript [3]. Using pedigrees from two isolated and two more cosmopolitan populations of European origin, they demonstrated that runs of homozygosity up to 4 Mb are common in out bred individuals [3].

The development of platforms able to genotype over one million SNPs has provided an unparalleled opportunity to study tracts of extended homozygosity. One obvious application of this technology is the study of recessive families. Recent research our group has reported a whole genome analysis of a consanguineous family with early onset Alzheimer's disease (EOAD). This study presented the first catalog of extended homozygosity in EOAD and identified several regions as candidate loci for a recessive genetic lesion in AD [8]. This same technology may also be used to perform homozygosity mapping for disease in populations with unknown pedigree structures from ostensibly out bred populations [9].

Alzheimer's disease (AD) is a complex, multifactorial disorder in which a genetic component has been well established. A very small percentage of AD cases have an onset before 60-65 years of age, being classified as early onset AD. Most of these cases present a familial aggregation of the disease and some of them, an autosomal dominant pattern of inheritance [10]. In these latter cases the disease may be associated with the presence of mutations in three genes: *APP* (amyloid precursor protein OMIM, 104760), *PSEN1* (presenilin 1, OMIM 104311) and *PSEN2* (presenilin 2, OMIM 600759) [11].

The great majority of Alzheimer cases are, however, late onset cases (LOAD) with no apparent familial aggregation. The E4 allele of the Apolipoprotein E gene (*APOE*, OMIM 107741) has been the only genetic factor consistently associated with both familial and sporadic forms of AD by different studies in several varied populations [12][13]. Although *APOE* is neither necessary nor sufficient for the development of AD, its risk has been shown to be dose dependent and correlates with the age at onset of the disease [14]. Thus, a large part of the genetic causes of AD are still unknown and the role of homozygosity has been considerably overlooked: only two studies of isolated populations with a high incidence of the disease have been reported and the primary analyses were performed using dominant or additive modes of inheritance [15][16].

In order to evaluate the role of runs of extended homozygosity in LOAD, we analyzed the association between Alzheimer's disease and quantitative measures of extended homozygosity. The underlying hypothesis of this work is that an excess of large homozygous tracts in AD patients versus controls would support the notion that a recessive component exists for the disease. This led us to perform a statistical comparison of the extended tracts of homozygosity in a series of 837 LOAD patients and 550 neurological normal controls [17].

## METHODS

### Data characterization

The data used for this analysis is publicly available at [http://www.tgen.org/research/neuro\\_gab2.cfm](http://www.tgen.org/research/neuro_gab2.cfm) and was generated by a genome wide analysis of unrelated LOAD cases and healthy controls [18]. Detailed quality control data for this publicly available dataset is described in Reiman et al., 2007 [17]. Due to the necessity of relatively complete genotyping coverage in analyses of extended homozygosity, we increased the genotyping success threshold to 97% successful genotypic calls per individual (allowing for an average call rate per individual of ~98.5%), but causing the exclusion of 24 participants from the initial dataset. The total population in our analyses after exclusions were made (total n=1387) was taken from three cohorts: the neuropathological discovery cohort (n= 722); the neuropathological replication cohort (n= 305), and the clinical replication cohort (n= 360). All participants included in the initial study were of self-reported European ancestry and population structure was evaluated using STRUCTURE [19]. Only participants identified as genetically most similar to European HapMap samples were included in the analytic dataset, with 14 outliers removed prior to data analysis [17] [20]. A summary of the information available from each cohort used in the analyses is presented in Table 1. Neither individual ages nor sex were available for public-access.

### Identification of runs of extended homozygosity

Runs of extended homozygosity were identified using the PLINKv1.02 software package [<http://pngu.mgh.harvard.edu/~purcell/plink/contact.shtml> - cite]. We utilized robust criteria for inclusion of genomic regions into runs of homozygosity as a means of reducing confounding by copy number. We identified only large runs of homozygosity (ROH), the primary criteria for genomic inclusion in a ROH being at least 1 Mb of consecutive homozygous genotypic calls at adjacent SNP loci. The minimum SNP density coverage was at least 50 SNPs per Mb to be included in a homozygous run, allowing for centromeric and SNP-poor regions to be algorithmically excluded from any analyses. The genome was scanned for ROHs using a sliding window of 50 SNPs, and allowed at most 2 missing genotypes and 1 heterozygote call per ROH. These large ROHs identified are suggestive of a recent origin, as opposed to shorter runs suggestive of an ancient ROH origin.

Total length of the genome comprised in homozygous runs (expressed in Mb), average length of each homozygous run (in Mb per ROH) and total number of homozygous runs per participant were calculated. The total length of the genome comprised of homozygous runs is the sum of the distance of each individual run per participant. The average length of homozygous runs was generated by dividing the total genomic length of the homozygous runs by the total number of homozygous runs per participant.

### Homozygosity mapping

From the initial structural analysis, 1090 consensus regions from overlapping ROHs were defined. These consensus regions were comprised of the loci shared by all overlapping segments of extended homozygosity in a particular genomic region and were used as positional loci for identification of common ROHs in our mapping efforts. Each consensus region was comprised of at least 3 SNPs and 100,00 base pairs and was found in no less than 10 participants. The largest consensus regions were 1,782,642 bp and contained 207 SNPs. The average consensus region defined was 351,744 bp in length and contained ~36 snps.

These consensus ROHs were analyzed using the maxT permutation test algorithm for case/control studies in PLINKv.1.02. This algorithm, which tests the frequency of segmental occurrence in cases compared to that in controls, is identical to that used in copy number

variant analyses of disease association although it has been adapted to incorporate the larger ROH segments. 50,000 permutations of the maxT permutation testing algorithm were used to generate empirical p-values and association coefficients to evaluate risk of LOAD attributable to individual runs of homozygosity. These empirical p-values were then corrected for multiple testing incorporating the possibility of false positive results in each of the possible permutations per ROH tested during the label-swapping function.

### Additional statistical analyses

We tested hypotheses of increased summary measures of extended homozygosity (i.e., number of ROHs, total and average ROH lengths) being associated with LOAD using simple two tailed T-tests to compare means between cases and controls. These measures were relatively normally distributed.

## RESULTS

### Identification of risk homozygous regions

In order to localize specific homozygous regions significantly associated with LOAD, empirical p-values for 50,000 permutations of homozygous consensus overlapping regions were determined using a modified CNV algorithm from PLINK toolset, adapted to larger segments (Figure 1). One homozygous consensus region in chromosome 8 was found to be significantly overrepresented in cases when compared to controls (Table 2). This region is not concurrent with homozygous regions previously reported to have a frequency above 25% in healthy individuals [21] nor was it found as an overlapping long contiguous stretch of homozygosity in three different populations (Han Chinese, Taiwan aborigines and Caucasians) [2]. The other top regions most significantly overrepresented in LOAD cases include additional loci on chromosomes 8, 9 and 10 and are associated with empirical p values between 0.001 and 0.01 (0.02 and 0.1 after adjustment for multiple testing corrections).

### Genes included in the risk homozygous regions

The genes present within the top 3% homozygous consensus regions overrepresented in cases when compared with controls are shown in Table 2. From the seven genes located in the homozygous region in chromosome 8 found to be most significantly different between cases and controls (*RAB11FIP1*; *MGC33309*; *ADRB3*; *EIF4EBP1*; *ASH2L*; *STAR* and *LSMI*), only the *ADRB3* gene (GeneID: 155), coding for the beta-3-adrenergic receptor is represented in the AlzGene database [22]. One SNP in this gene (rs4998) was found to be associated with AD in ApoE4 negative samples by Hamilton G. et al, in a case-control association study performed for candidate insulin signaling genes [23]. Additionally, the steroidogenic acute regulatory protein (StAR), coded by the *STAR* gene (GeneID: 6770), was found to be markedly increased in the cytoplasm of hippocampal pyramidal neurons and in the cytoplasm of other non-neuronal cell types from AD brains, when compared with age-matched controls [24]. Also, the levels of phosphorylated eukaryotic translation initiation factor 4E binding protein 1, encoded by the gene *EIF4EBP1* (GeneID: 1978), have been previously found to be dramatically increased in AD and positively significantly correlated with total tau and p-tau [25].

### Analysis of extended homozygosity

The AD cases studied presented a slightly higher degree of extended homozygosity when compared with the control group. The most variability observed was associated with the total number of homozygous runs and total run length with suggestive p-values of 0.062 and

0.052 respectively, while the average run length difference between these two groups was not statistically significant ( $p > 0.10$ ).

## DISCUSSION

In order to identify risk loci significantly different between cases and controls that might contain genetic lesions responsible for AD, empirical p-values for consensus homozygous regions were determined in both cases and controls groups. One consensus ROH in chromosome 8 was found to be significantly overrepresented in cases when compared to controls after correction for multiple testing. This region contains seven genes, three of which have previously been studied in relation to AD (Table 2).

The *ADRB3* gene product, beta-3-adrenergic receptor, is one of three beta-adrenergic receptor subtypes. The product is located mainly in adipose tissue and is involved in the regulation of adipocyte lipolysis, thermogenesis and oxygen consumption [26]. There is evidence that a disturbance in the insulin signal transduction pathway may be a central and early pathophysiological event in sporadic AD [27]. Patients with diabetes and insulin resistance have an increased risk of impaired cognition or dementia [28]. Hamilton et al performed a two stage association study to investigate the role of insulin signaling genes in the risk of developing AD showing one SNP in *ADRB3* (rs4998) was studied and found to be associated with AD in ApoE4 negative samples [23].

The steroidogenic acute regulatory protein (StAR), coded by the *STAR* gene, plays a key role in the acute regulation of steroid hormone synthesis by enhancing the conversion of cholesterol into pregnenolone [29]. To test the hypothesis that hormones of the hypothalamic-pituitary-gonadal axis have a role in AD pathogenesis, K.M. Webber et al studied the expression levels of StAR in AD and control brains. The authors found an increase of this protein in AD hippocampal neurons as well as other non-neuronal cells compared to aged matched controls. These results, together with the finding that StAR is present in the same brain regions as the luteinizing hormone (LH) receptors, suggest that steroidogenic pathways regulated by LH may play a role in AD [24].

The *EIF4EBP1* gene encodes one member of a family of translation repressor proteins, which interacts with eukaryotic translation initiation factor 4E (eIF4E). This is a limiting component of the multi-subunit complex that recruits 40S ribosomal subunits to the 5' end of mRNAs. Interaction of this protein with eIF4E inhibits the complex assembly and represses translation. The 4E-BP1 protein is phosphorylated in response to several signals, including insulin signaling and together with several other components of the translational machinery is regulated through signaling events that require the mammalian target of rapamycin (mTOR) [30]. Neurofibrillary tangles, composed mainly of hyperphosphorylated tau protein, is one of the neuropathological hallmarks of AD. Tau mRNA levels have been shown not to be altered in sporadic AD brains. Nevertheless, Li X, et al studied the possibility that tau mRNAs in AD brains could be abnormally regulated by investigating the levels of various translation control elements including 4E-BP1, in the brains of AD and controls subjects. Together with increased levels of p-mTOR, they found increased levels of phosphorylated 4E-BP1 in AD and a positive significant correlation with total tau and phosphorylated tau [25].

Genetic variation in one or more of these genes may account for the development of different phenotypes related to LOAD and it is possible that this form of genetic heterogeneity coexists with the multifactorial, common-disease/common-variant mode of inheritance that is generally studied in whole-genome association [21].

From a methodological perspective, using ROHs to identify possible candidate loci for small effects size or rare recessive variants in unrelated individuals may be useful. This method uses relatively few statistical tests compared to conventional genome-wide association studies to probabilistically model recessive genome-wide associations with disease, increasing power to some degree. The method itself likely trades specificity of results for sensitivity, by seeking to identify large regions of homozygosity harboring as-of-yet unknown recessive disease components. In particular, homozygosity mapping is limited by fine mapping within the risk ROH not being feasible in the discovery population. To identify a more focused risk region in replication of a homozygosity mapping finding, a population of a different ethnic background or an admixed population would have to be studied, as boundaries of the ROHs would likely be different from the discovery cohort. The summary measures of genome-wide extended homozygosity may also prove useful in quantifying distant consanguinity in a single unrelated population, for which  $F_{st}$  calculations may not be applicable to gauge genetic distance from a founder group.

Using public whole genome SNP analysis data on late onset Alzheimer's disease and neurologically normal controls we identified an average of 52.1 homozygous runs larger than 1 Mb containing more than 50 consecutive SNPs in 1387 samples. In order to determine if Alzheimer's disease populations are more consanguineous than healthy controls we statistically compared the total number of homozygous runs, average length of homozygous runs and total length of genome contained in runs of homozygosity between cases and controls. For this analysis we used a stringent method, assuming that the presence of large tracts of contiguous homozygous SNPs represent a direct association with increased levels of relatively recent consanguinity [17][7]. Only homozygous runs with sizes over 1 Mb were included in the analysis in order to rule out the effect of copy number variation, which could confound results if smaller minimum run sizes were used. Our borderline significant results suggest that the AD cases in this study may be characterized by a higher degree of extended homozygosity than the control group, although this generalization would need to be tested in a larger sample size.

As the individual ages were not available to the public, it was not possible to establish a relation between age and homozygosity. Still, our group has recently reported the first genetic proof that younger individuals present less homozygosity than older ones, presumably due to populations increased mobility in the last generations. Linear predictive models of homozygosity and current age, showed that younger individuals have smaller percentage of their genome contained in homozygous runs; which are significantly shorter in length (Nalls et al unpublished data). Similar results would be expected for the present populations.

The E4 allele of *APOE* has been the only genetic risk factor consistently associated with familial and sporadic forms of AD [12][13]. In the present study no association between homozygosity at the *APOE* gene and AD was found. This is explained by the size of the statistical window used (1Mb), as the methodology used was designed to pick up variants with recently created ROHs that have not been broken up by recombination. High recombination rates in this genomic region in the European American population, prevent detection of any signal at the *APOE* gene using homozygosity mapping [31].

In summary, we conclude that extended runs of homozygosity are common in out bred populations and present statistical data that show specific large tracts of homozygosity are a risk factor for Alzheimer's disease. The present results, together with the previously reported cases of consanguineous families with AD [8] and the speculation that recessive genes for AD are responsible for the high AD prevalence in the Wadi Ara [33], demonstrate that particular regions of homozygosity have a significant role in AD genetics. It is possible

that future sequencing and additional follow-up analyses will allow the identification of small-effect size recessive risk variant(s) within the described ROHs.

## Acknowledgments

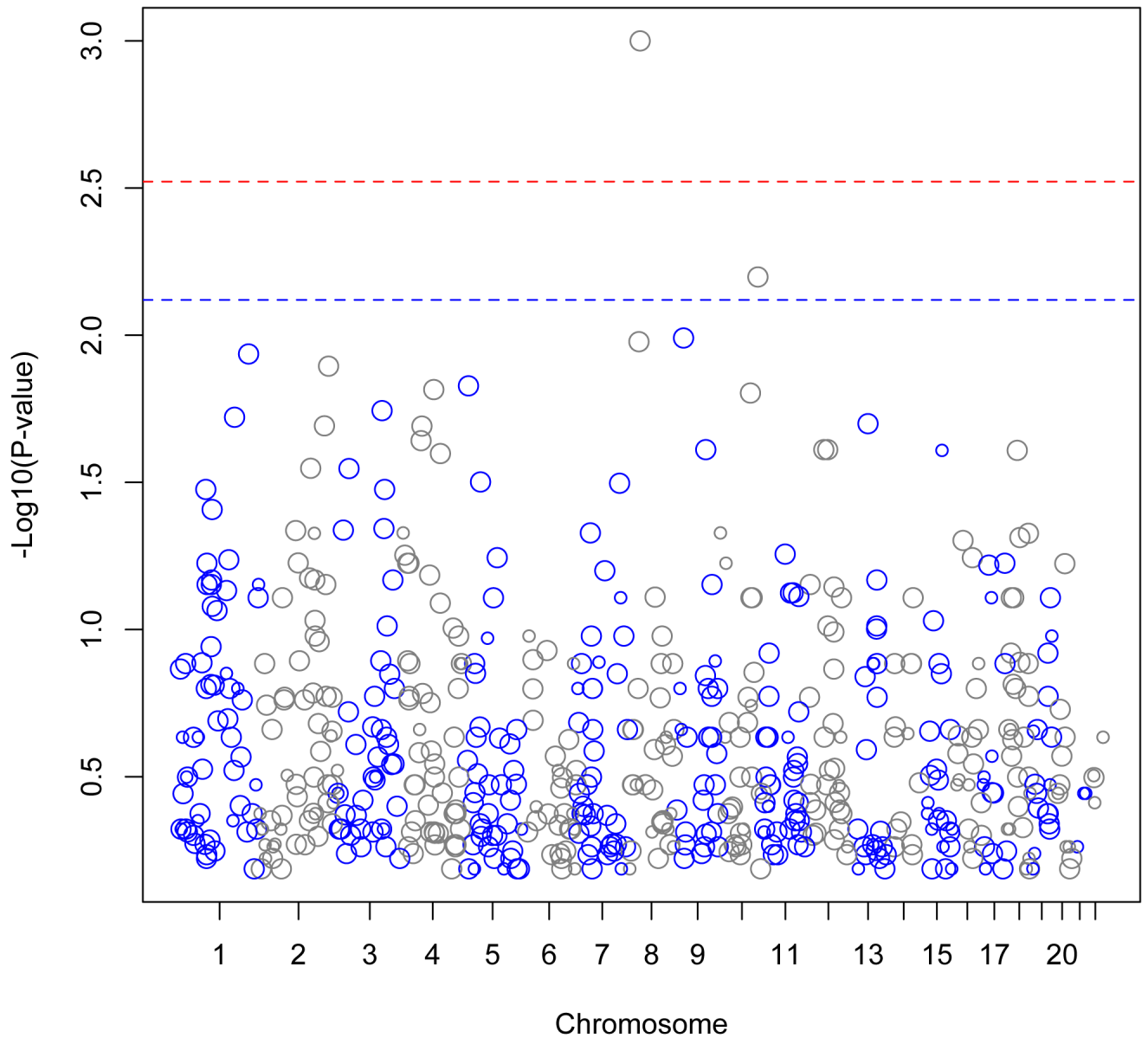
This work was supported in part by the Intramural Research Program of the National Institute on Aging, National Institutes of Health, Department of Health and Human Services (Z01 AG000950-06) and the Portuguese Fundacao para a Ciencia e Tecnologia grants (SFRH/BD/29647/2006) and (SFRH/BD/27442/2006). The experiments here presented comply with the current laws of United States of America.

## REFERENCES

- Gibson J, Morton NE, Collins A. Extended tracts of homozygosity in outbred human populations. *Hum. Mol. Genet* 2006;15:789–95. doi:ddi493. [PubMed: 16436455]
- Li L, Ho S, Chen C, Wei C, Wong W, et al. Long contiguous stretches of homozygosity in the human genome. *Hum. Mutat* 2006;27:1115–21. doi:10.1002/humu.20399. [PubMed: 16955415]
- McQuillan R, Leutenegger A, Abdel-Rahman R, Franklin CS, Pericic M, et al. Runs of homozygosity in European populations. *Am J Hum Genet* 2008;83:359–72. doi:10.1016/j.ajhg.2008.08.007. [PubMed: 18760389]
- Devilee P, Cleton-Jansen AM, Cornelisse CJ. Ever since Knudson. *Trends Genet* 2001;17:569–73. doi:11585662. [PubMed: 11585662]
- Raghavan M, Lillington DM, Skoulakis S, Debernardi S, Chaplin T, et al. Genome-wide single nucleotide polymorphism analysis reveals frequent partial uniparental disomy due to somatic recombination in acute myeloid leukemias. *Cancer Res* 2005;65:375–8. doi:65/2/375. [PubMed: 15695375]
- Woods CG, Valente EM, Bond J, Roberts E. A new method for autozygosity mapping using single nucleotide polymorphisms (SNPs) and EXCLUDEAR. *J. Med. Genet* 2004;41:e101. doi:15286161. [PubMed: 15286161]
- Simon-Sanchez J, Scholz S, Fung H, Matarin M, Hernandez D, et al. Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals. *Hum. Mol. Genet* 2007;16:1–14. doi:ddl436. [PubMed: 17116639]
- Clarimón J, Djaldetti R, Lleó A, Guerreiro RJ, Molinuevo JL, et al. Whole genome analysis in a consanguineous family with early onset alzheimer's disease. *Neurobiol Aging*. 2008 doi:S0197-4580(08)00055-9.
- Leutenegger A, Prum B, Génin E, Verny C, Lemainque A, et al. Estimation of the inbreeding coefficient through use of genomic data. *Am. J. Hum. Genet* 2003;73:516–23. doi:12900793. [PubMed: 12900793]
- Rademakers R, Cruts M, Van Broeckhoven C. Genetics of early-onset alzheimer dementia. *ScientificWorldJournal* 2003;3:497–519. doi:12847300. [PubMed: 12847300]
- Bertram L, Tanzi RE. Alzheimer's disease: one disorder, too many genes? *Hum Mol Genet* 2004;13(1):R135–41. doi:14764623. [PubMed: 14764623]
- Saunders AM, Strittmatter WJ, Schmechel D, George-Hyslop PH, Pericak-Vance MA, et al. Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology* 1993;43:1467–72. doi:8350998. [PubMed: 8350998]
- van Duijn CM, de Knijff P, Cruts M, Wehnert A, Havekes LM, et al. Apolipoprotein E4 allele in a population-based study of early-onset alzheimer's disease. *Nat Genet* 1994;7:74–8. doi:8075646. [PubMed: 8075646]
- Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 1993;261:921–3. doi:8346443. [PubMed: 8346443]
- Farrer LA, Bowirrat A, Friedland RP, Waraska K, Korczyn AD, et al. Identification of multiple loci for Alzheimer disease in a consanguineous Israeli-Arab community. *Hum Mol Genet* 2003;12:415–22. doi:12566388. [PubMed: 12566388]

16. Liu F, Arias-Vásquez A, Sleegers K, Aulchenko YS, Kayser M, et al. A genomewide screen for late-onset alzheimer disease in a genetically isolated Dutch population. *Am J Hum Genet* 2007;81:17–31. doi:S0002-9297(07)62813-4. [PubMed: 17564960]
17. Gibbs JR, Singleton A. Application of genome-wide single nucleotide polymorphism typing: simple association and beyond. *PLoS Genet* 2006;2:e150. doi:06-PLGE-RV-0291R2. [PubMed: 17029559]
18. Reiman EM, Webster JA, Myers AJ, Hardy J, Dunckley T, et al. GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers. *Neuron* 2007;54:713–20. doi:S0896-6273(07)00379-0. [PubMed: 17553421]
19. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics* 2000;155:945–59. [PubMed: 10835412]
20. A haplotype map of the human genome. *Nature* 2005;437:1299–320. doi:10.1038/nature04226. [PubMed: 16255080]
21. Lencz T, Lambert C, DeRosse P, Burdick KE, Morgan TV, et al. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proc. Natl. Acad. Sci. U.S.A* 2007;104:19942–7. doi:0710021104. [PubMed: 18077426]
22. Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nat. Genet* 2007;39:17–23. doi:ng1934. [PubMed: 17192785]
23. Hamilton G, Proitsis P, Jehu L, Morgan A, Williams J, et al. Candidate gene association study of insulin signaling genes and Alzheimer's disease: evidence for SOS2, PCK1, and PPARgamma as susceptibility loci. *Am J Med Genet B Neuropsychiatr Genet* 2007;144B:508–16. doi:10.1002/ajmg.b.30503. [PubMed: 17440948]
24. Webber KM, Stocco DM, Casadesus G, Bowen RL, Atwood CS, et al. Steroidogenic acute regulatory protein (StAR): evidence of gonadotropin-induced steroidogenesis in Alzheimer disease. *Mol Neurodegener* 2006;1:14. doi:1750-1326-1-14. [PubMed: 17018137]
25. Li X, Alafuzoff I, Soininen H, Winblad B, Pei J. Levels of mTOR and its downstream targets 4E-BP1, eEF2, and eEF2 kinase in relationships with tau in Alzheimer's disease brain. *FEBS J* 2005;272:4211–20. doi:EJB4833. [PubMed: 16098202]
26. Meyers DS, Skwish S, Dickinson KE, Kienzle B, Arbeeny CM. Beta 3-adrenergic receptor-mediated lipolysis and oxygen consumption in brown adipocytes from cynomolgus monkeys. *J Clin Endocrinol Metab* 1997;82:395–401. doi:9024225. [PubMed: 9024225]
27. Hoyer S. Causes and consequences of disturbances of cerebral glucose metabolism in sporadic Alzheimer disease: therapeutic implications. *Adv Exp Med Biol* 2004;541:135–52. doi:14977212. [PubMed: 14977212]
28. Craft S, Watson GS. Insulin and neurodegenerative disease: shared and specific mechanisms. *Lancet Neurol* 2004;3:169–78. doi:14980532. [PubMed: 14980532]
29. Sugawara T, Holt JA, Kiriakidou M, Strauss JF. Steroidogenic factor 1-dependent promoter activity of the human steroidogenic acute regulatory protein (StAR) gene. *Biochemistry* 1996;35:9052–9. doi:8703908. [PubMed: 8703908]
30. Proud CG. Regulation of mammalian translation factors by nutrients. *Eur J Biochem* 2002;269:5338–49. doi:12423332. [PubMed: 12423332]
31. Yu C, Seltman H, Peskind ER, Galloway N, Zhou PX, et al. Comprehensive analysis of APOE and selected proximate markers for late-onset alzheimer's disease: patterns of linkage disequilibrium and disease/marker association. *Genomics* 2007;89:655–65. doi:S0888-7543(07)00050-X. [PubMed: 17434289]
32. Smith JD. Apolipoprotein E4: an allele associated with many diseases. *Ann. Med* 2000;32:118–27. doi:10766403. [PubMed: 10766403]
33. Bowirrat A, Friedland RP, Chapman J, Korczyn AD. The very high prevalence of AD in an Arab population is not explained by APOE epsilon4 allele frequency. *Neurology* 2000;55:731. doi:10980749. [PubMed: 10980749]





**Figure 1.** Empirical p-values for runs of homozygosity between cases and controls. The red line denotes significance based on permutation-based adjustments for multiple testing phenomena.

**Table 1**

Summary of the information available online for the data used in this analysis

	Neuropathological discovery cohort		Neuropathological replication cohort		Clinical replication cohort	
N	722		305		360	
Cases/Controls	LOAD cases	Controls	LOAD cases	Controls	LOAD cases	Controls
	432 <sup>a</sup>	290 <sup>b</sup>	191 <sup>a</sup>	114 <sup>b</sup>	214 <sup>c</sup>	146 <sup>d</sup>
ApoE4 carriers	291 (67%)	61 (21%)	109 (57%)	27 (24%)	113 (53%)	29 (20%)
	<b>Neuropathological cohorts</b>		<b>Clinical replication cohort</b>			
	LOAD cases	Controls	LOAD cases	Controls		
Mean age <sup>±</sup> standard deviation	73.5 <sup>±</sup> 6.2 at death	75.8 <sup>±</sup> 7.5 at death	78.9 <sup>±</sup> 7.8 at last clinical visit	81.7 <sup>±</sup> 6.6 at last clinical visit		

<sup>a</sup> Brain donor cases satisfied clinical and neuropathological criteria for LOAD.

<sup>b</sup> Brain donor controls did not have significant cognitive impairment or significant neuropathological features of AD.

<sup>c</sup> Clinical cases satisfied criteria for probable AD (living subjects who were at least 65 years old at the time of their death or last clinical assessment and who were independently assessed for their APOE genotype).

<sup>d</sup> Clinical controls did not have clinically significant cognitive impairment or significant neuropathological features of AD.

\* denotes taken from Reiman et al., 2007 published data

Table 2

Top 3% homozygous consensus regions overrepresented in cases when compared to controls and genes present within these regions

Chr	Start of region (bp)	End of region (bp)	Region length (bp)	Empirical p-value	Permutation p-value after adjustment for multiple testing	Candidate genes	Number of cases with homozygous runs	Number of controls with homozygous runs	OR
8	37835460	38143780	308320	0.00099998	0.0170397	RAB11FIP1; MGC33309; ADRB3; EIF4EBP1; ASH2L; STAR; LSMI	40	9	3.02
10	116826992	117301088	474096	0.00633987	0.0908982	ATRNL1	45	14	2.18
9	26314013	26648658	334645	0.0102198	0.129957	-	9	0	n.a.
8	34251617	34375482	123865	0.0105198	0.151257	LOC137107	109	49	1.53
1	213279861	213715720	435859	0.0115798	0.152477	ESRRG	32	9	2.39
2	214386585	214511938	125353	0.0127397	0.155837	PF20	34	10	2.29
5	15390663	15634707	244044	0.0148597	0.161017	LOC391741; FBXL7	27	7	2.59
4	98911102	99056376	145274	0.0152997	0.198896	MGC46496	162	81	1.39
10	93788860	94018536	229676	0.0157197	0.192296	CPEB3	63	25	1.71
3	137687127	137949289	262162	0.0180396	0.213336	STAG1	149	74	1.39
1	169919601	170135865	216264	0.0189996	0.190176	LOC343153	14	2	4.66
13	56558450	56726080	167630	0.0199796	0.241755	FLJ40296	58	23	1.71
2	201456660	201882057	425397	0.0203196	0.211676	AOX2; BZWI; LOC391472; CLK1; PPL3; NIF3L1; ORC2L; MGC39518; NDUFB3; CFLAR; CASPI0	28	8	2.34
4	62126310	62256398	130088	0.0203596	0.196436	LPHN3	11	1	7.31
4	59420217	59599269	179052	0.0228195	0.235695	-	21	5	2.81
9	95407641	95524041	116400	0.0244795	0.240655	PSMA7P	16	3	3.55

Chr	Start of region (bp)	End of region (bp)	Region length (bp)	Empirical p-value	Permutation p-value after adjustment for multiple testing	Candidate genes	Number of cases with homozygous runs	Number of controls with homozygous runs	OR
12	51809479	51949002	139523	0.0244795	0.240655	LOC400039; CSAD; LOC283337; ITGB7; RARG; MGCI1308; ESPLI	16	3	3.55
12	63691379	64048325	356946	0.0244795	0.240655	WIFI; MANI; LOC253827	16	3	3.55
15	65557360	66059106	501746	0.0246195	0.251595	FLJ12476; LOC145853; MAP2K5; LOC390598; LOC388129	23	6	2.56
18	31690766	31818000	127234	0.0246195	0.251595	C18orf21	23	6	2.56
4	119462492	119635887	173395	0.0252395	0.278194	NDST3; PRSS12	53	21	1.70
2	158898786	159035950	137164	0.0283194	0.272655	LOC130940	18	4	3.00
3	35117703	35239453	121750	0.0283794	0.257595	LOC442078	13	2	4.32
5	52730874	53125379	394505	0.0315194	0.268275	FST; NDUFS4	10	1	6.64
7	13283685	133208570	369885	0.0318394	0.343753	SEC8L1	63	27	1.58
1	80639887	80811557	171670	0.0334193	0.300274	LOC441889; COX6A1P	22	6	2.45
3	145678190	145958179	279989	0.0334193	0.300274	-	22	6	2.45
1	100201372	100553949	352577	0.0390992	0.376492	HIAT1; DKFZp761A078; FLJ10287; MGC14816; DBT; RTCD1; CDC14A	49	20	1.65
3	143407729	143828507	420778	0.0453791	0.422592	MGC40579; XRNI; ATR; PLS1	35	13	1.80
3	17477922	17773653	295731	0.0459391	0.448171	TBC1D5	84	40	1.42
2	112487241	112641378	154137	0.0460991	0.385732	MERTK; FLJ14681; FLJ37440	27	9	2.00
7	42294711	42463020	168309	0.0469591	0.329253	-	6	0	n.a.

Chr – Chromosome; bp – Base Pairs; OR – Odds Ratio; n.a. – not available.

**Table 3**

Descriptive statistics for measures of extended homozygosity

	<b>N</b>	<b>Measure</b>	<b>Mean</b>	<b>SD</b>
<b>Controls</b>	550	Number Runs	51.35	16.43
		Total Run Length	73.02	28.13
		Average Run Length	1.41	0.22
<b>Cases</b>	837	Number Runs	52.94	15.24
		Total Run Length	76.15	30.48
		Average Run Length	1.42	0.28
<b>Total</b>	1387	Number Runs	52.31	15.73
		Total Run Length	74.91	29.60
		Average Run Length	1.42	0.26