

Published in final edited form as:

Ann Hum Genet. 2012 September ; 76(5): 342–351. doi:10.1111/j.1469-1809.2012.00721.x.

Genome-wide association and linkage study in the Amish detects a novel candidate late-onset Alzheimer disease gene

Anna C. Cummings¹, Lan Jiang¹, Digna R. Velez Edwards^{1,2}, Jacob L. McCauley³, Renee Laux¹, Lynne L. McFarland¹, Denise Fuzzell¹, Clare Knebusch¹, Laura Caywood³, Lori Reinhart-Mercer³, Laura Nations³, John R. Gilbert³, Ioanna Konidari³, Michael Tramontana⁴, Michael L. Cuccaro³, William K. Scott³, Margaret A. Pericak-Vance³, and Jonathan L. Haines¹

¹Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, TN, USA

²Vanderbilt Epidemiology Center, Vanderbilt University Medical Center, Nashville, TN, USA

³Dr. John T. Macdonald Foundation Department of Human Genetics and John P. Hussman Institute for Human Genomics, University of Miami, Miller School of Medicine, Miami, FL, USA

⁴School of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

Summary

To identify novel late-onset Alzheimer disease (LOAD) risk genes, we have analyzed Amish populations of Ohio and Indiana. We performed genome-wide SNP linkage and association studies on 798 individuals (109 with LOAD). We tested association using the Modified Quasi-Likelihood Score (MQLS) test and also performed two-point and multipoint linkage analyses. We found that LOAD was significantly associated with *APOE* ($P=9.0\times 10^{-6}$) in all our ascertainment regions except for the Adams County, Indiana, community ($P=0.55$). Genome-wide, the most strongly associated SNP was rs12361953 ($P=7.92\times 10^{-7}$). A very strong, genome-wide significant multipoint peak (recessive HLOD=6.14, dominant HLOD=6.05) was detected on 2p12. Three additional loci with multipoint HLOD scores >3 were detected on 3q26, 9q31, and 18p11. Converging linkage and association results, the most significantly associated SNP under the 2p12 peak was at rs2974151 ($P=1.29\times 10^{-4}$). This SNP is located in *CTNNA2*, which encodes catenin alpha 2, a neuronal-specific catenin known to have function in the developing brain. These results identify *CTNNA2* as a novel candidate LOAD gene, and implicate three other regions of the genome as novel LOAD loci. These results underscore the utility of using family-based linkage and association analysis in isolated populations to identify novel loci for traits with complex genetic architecture.

Keywords

GWAS; Linkage; founder population; Amish; Alzheimer

Corresponding author: Jonathan L. Haines, Center for Human Genetics Research, Vanderbilt University Medical Center, 519 Light Hall, Nashville, TN, 37232-0700, United States, jonathan@chgr.mc.vanderbilt.edu, Tel: 615-343-5851, Fax: 615-343-8619.

The authors have no conflict of interest to declare.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Table S1 Most significant genome-wide association results, stratified.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organised for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Introduction

Late-onset Alzheimer disease (LOAD) is a neurodegenerative disorder causing the majority of dementia cases in the elderly. A complex combination of genetic and environmental components likely determine susceptibility to LOAD (Bertram et al. 2010). The *APOE E4* allele is a well-established genetic risk factor for LOAD. Additional risk genes have been difficult to detect and replicate until recent successes using large consortia-derived genome-wide association study (GWAS) datasets, which have added *CRI*, *CLU*, *PICALM*, *BINI*, *EPHA1*, *MS4A*, *CD33*, *CD2AP*, and *ABCA7* to the list of confirmed LOAD susceptibility genes, each with modest effect (Harold et al. 2009; Hollingworth et al. 2011; Lambert et al. 2009; Naj et al. 2011; Seshadri et al. 2010).

Despite these recent successes the majority of the genetic risk for LOAD remains unknown. The remaining genetic risk may in part lie in additional loci with small effects at the population level, making most datasets underpowered. The use of a genetically isolated founder population, such as the Amish, represents an alternative to the use of large population based consortia-derived datasets in the search for genetic risk factors. In the case of a founder population, the number of disease variants is hypothesized to be fewer, thereby decreasing heterogeneity and increasing power.

We have taken this approach to discover at least one novel LOAD risk gene by studying the Amish communities of Holmes County, Ohio, and Adams, Elkhart and LaGrange Counties, Indiana (Hahs et al. 2006; McCauley et al. 2006). These communities are collectively part of a genetically isolated founder population originating from two waves of immigration of Swiss Anabaptists into the U.S in the 1700's and 1800's. The first wave of immigration brought the Anabaptists to Pennsylvania. In the early 1800's some of these immigrants moved to Holmes County, OH (Beachy 2011), while a second wave of immigration from Europe established more Amish communities in Ohio (including Wayne County but not Holmes County) and Indiana (including Adams County) (Hostetler 1993). Starting in 1841, the Elkhart and LaGrange Counties Amish community was founded by Amish families primarily from Somerset County, PA, and from Holmes and Wayne Counties, OH, who were seeking new farmland to settle (Amish Heritage Committee 2009). The Amish marry within their faith, limiting the amount of genetic variation introduced to the population. Not only are the Amish more genetically homogeneous, but because of their strict lifestyle, environmental exposures are also more homogeneous. The Amish have large families and a well-preserved comprehensive family history that can be queried via the Anabaptist Genealogy Database (AGDB) (Agarwala et al. 1999; Agarwala et al. 2003), making the Amish a valuable resource for genetic studies.

Our current study undertook a genome-wide approach, in a population isolate, using complementary linkage and association analyses to further elucidate the complex genetic architecture of LOAD. We utilized linkage analysis to look for sharing of genomic regions among affected individuals, while also using association analysis to look for differences in allele frequencies between affecteds and unaffecteds. We previously performed a genome-wide linkage study using microsatellites genotyped in only a small subset of the individuals included in this study (Hahs et al 2006). Here we use a much larger dataset with a much denser panel of markers using a genome-wide SNP chip. The results indicate that several novel regions likely harbor LOAD genes in the Amish, underscoring the genetic heterogeneity of this phenotype.

Materials & Methods

Subjects

Methods for ascertainment were reviewed and approved by the individual Institutional Review Boards of the respective institutions. Participants were identified from published community directories, referral from other community members or due to close relationship with other participants, as previously described (Edwards et al. 2011). Informed consent was obtained from participants recruited from the Amish communities in Elkhart, LaGrange, and surrounding Indiana counties, and Holmes and surrounding Ohio counties with which we have had established working relationships for over 10 years.

Clinical Data

For individuals who agreed to participate, demographic, family, and environmental information was collected, informed consent was obtained, and both a functional assessment and the Modified-Mini-Mental State Exam (3MS) were administered (Teng & Chui 1987; Tschanz et al. 2002). Those scoring ≥ 87 on the 3MS were considered cognitively normal and were considered unaffected in our study. Those scoring < 87 were re-examined with further tests from the CERAD neuropsychological battery (Morris et al. 1989). Depression was also evaluated using the geriatric depression scale (GDS). Diagnoses for possible and probable AD were made according to the NINCDS-ADRDA criteria (McKhann et al. 1984). A yearly consensus case conference was held to confirm all diagnoses.

Genotyping

SNPs for *APOE* were genotyped for 823 individuals (127 with LOAD). To identify the six *APOE* genotypes determined by the *APOE* *E2, *E3 and *E4 alleles, two single nucleotide polymorphisms (SNPs) were assayed using the TaqMan method [Applied Biosystems Inc. (ABI), Foster City, CA, USA]. SNP-specific primers and probes were designed by ABI (TaqMan genotyping assays) and assays were performed according to the manufacturer's instructions in 5 μ l total volumes in 384-well plates. The polymorphisms distinguish the *E2 allele from the *E3 and *E4 alleles at amino acid position 158 (NCBI *rs7412*) and the *E4 allele from the *E2 and *E3 alleles at amino acid position 112 (NCBI *rs429358*).

Genome-wide genotyping was performed on 830 DNA samples using the Affymetrix 6.0 GeneChip® Human Mapping 1 million array set (Affymetrix®, Inc Santa Clara, CA). DNA for this project was allocated by the respective DNA banks at both the Hussman Institute of Human Genomics (HIHG) at the University of Miami and the Center for Human Genetics Research (CHGR) at Vanderbilt University. Genomic DNA was quantitated via the ND-8000 spectrophotometer and DNA quality was evaluated via gel electrophoresis. The genomic DNA (250 ng/5ul) samples were processed according to standard Affymetrix procedures for processing of the Affymetrix 6.0 GeneChip assay. The arrays were then scanned using the GeneChip Scanner 3000 7G operated by the Affymetrix® GeneChip® Command Console® (AGCC) software. The data were processed for genotype calling using the Affymetrix® Power Tools (APT) software using the birdseed calling algorithm version 2.0 Affymetrix®, Inc Santa Clara, CA (Korn et al. 2008).

We applied a number of quality control (QC) procedures to both samples and SNPs to ensure the accuracy of our genotype data prior to linkage and association analyses. Specific sample QC included: 1) Each individual DNA sample was examined via agarose to ensure that the sample was of high quality prior to inclusion on the array; 2) CEPH samples were placed across multiple arrays to ensure reproducibility of results across the arrays; 3) Samples with call rates $< 95\%$ were re-examined individually to ensure quality of genotypes. 4) Ultimately if the sample call rate remained below 95% after further

evaluation, attempts were made to rerun the array with a new DNA sample. If the sample still failed, it was dropped. Nine samples were dropped due to low genotyping efficiency. Three samples were excluded because they did not connect into a pedigree with the rest of the samples, and therefore, relationships of those individuals could not be accounted for. Sixteen samples with questionable gender based on X chromosome heterozygosity rates were eliminated. Three samples appearing to be aberrantly connected in the pedigree based on the genotype data were also excluded.

Specific SNP QC included: 1) Dropping 76,816 SNPs with call rates <98%. 2) Dropping 206,970 SNPs with minor allele frequencies (MAF) < 0.05. We additionally excluded 7,849 SNPs with a MAF less than 0.05 after adjusting for pedigree relationships using MQLS (see below). Due to the relatedness in this dataset we did not check SNPs for Hardy-Weinberg equilibrium. Following this extensive quality control, 798 samples (109 with LOAD, see Table 1) and 614,963 SNPs were analyzed. Because *APOE* genotyping and QC were done separately from genome-wide genotyping and QC, the sample sizes are different and the datasets are mostly, but not completely, overlapping. All 798 samples belong to one 4998-member pedigree with many consanguineous loops. The AGDB provided the pedigree information using an “all common paths” database query with all genotyped individuals (Agarwala et al 2003).

Statistical Analysis

Association analysis—We used the Modified Quasi-Likelihood Score (MQLS) test (software version 1.2) to correct for pedigree relationships (Thornton & McPeck 2007). MQLS is analogous to a χ^2 test, the most common approach for case-control data analysis with a binary trait, but MQLS incorporates kinship coefficients to correct for correlated genotypes of all the pedigree relationships. This test allows all samples to be included without dividing the pedigree. The MQLS test cannot be applied to X chromosome data, which were, therefore, eliminated from analysis. Because we previously found that Adams County has a lower *APOE*-4 allele frequency than the general population (Pericak-Vance et al. 1996), we did a stratified association analysis for *APOE* analyzing Adams County separately from the combined Elkhart, LaGrange, and Holmes Counties. Using the same stratification, we also re-analyzed our most significant SNPs from the GWAS analysis. To test the validity of the MQLS test in our pedigree, we performed simulation studies using this same pedigree structure to assess the type 1 error rate using MQLS for association. Type 1 error rates were not inflated (unpublished data).

Linkage analysis—Because of the large size and substantial consanguinity of the pedigree, we used PedCut (Liu et al. 2008) to find an optimal set of sub-pedigrees including the maximal number of subjects of interest within a bit-size limit (24 in this study) conducive to linkage analysis. This procedure resulted in 34 sub-pedigrees for analysis with an average of seven genotyped individuals (three genotyped affected) per sub-pedigree. Parametric heterogeneity two-point LOD (HLOD) scores were computed assuming affecteds-only autosomal dominant and recessive models using Merlin (Abecasis et al. 2002). A disease allele frequency of 10% was used to approximate Alzheimer disease prevalence. For the dominant model penetrances of 0 for no disease allele and 0.0001 for one or two copies of the disease allele, and under the recessive model penetrances of 0 for zero or one disease allele and 0.0001 for two disease alleles were used. Because the underlying genetic model is unknown, we tested both dominant and recessive models to maximize our ability to find a disease locus. SNPs on the X chromosome were analyzed using MINX (Merlin in X). Regions showing evidence for linkage, i.e. containing at least one two-point HLOD > 3.0, were followed up with parametric multipoint linkage analysis (also using Merlin). For the multipoint analyses, SNPs were pruned for linkage

disequilibrium (LD) in each region so that all pair-wise r^2 values were < 0.16 between all SNPs (Boyles et al. 2005). The LD from the HapMap CEPH samples (parents only) were used for pruning. Because the HapMap CEPH samples may not be an exact representation of LD in our Amish population, we also tested pruning using the data from this Amish dataset, but linkage results did not change using this approach (data not shown). Because linkage analyses can be biased when breaking larger pedigrees into a series of smaller ones (Liu et al. 2007; Liu et al. 2006), we performed simulation studies assuming no linkage (e.g. null distribution) and using the same large pedigree structure and the same pedigree splitting method. We determined empirical cut-offs for significance in our linkage studies to maintain a nominal type I error rate. We found that after 1000 replications, only 2.5% of the multipoint linkage scans generated a maximum HLOD > 3.0 (unpublished data).

All computations were done using either the Center for Human Genetics Research computational cluster or the Advanced Computing Center for Research and Education (ACCRE) cluster at Vanderbilt University.

Results

APOE

We found that LOAD was significantly associated with *APOE* (MQLS $P=9.0 \times 10^{-6}$) in our Amish population except for the Adams County, Indiana, community (MQLS $P=0.55$). The *E4* frequency, adjusted for pedigree relationships, in LOAD individuals in Elkhart, LaGrange, and Holmes Counties was 0.18 for affected individuals compared to 0.11 for unaffected individuals (Table 2). This compares to an *E4* allele frequency of 0.38 in Caucasian AD individuals (0.14 for controls) (alzgene.org). We also saw a progressively younger average age of onset with each additional copy of the *E4* allele (Table 3), consistent with other populations. We did not see evidence for linkage with *APOE* in our sub-pedigrees (dominant HLOD=0.50, recessive HLOD=0.29).

Genome-wide association: In the GWAS, the most significant MQLS P-value (7.92×10^{-7}), which did not surpass a Bonferroni-corrected genome-wide significance threshold of 8.13×10^{-8} , was at rs12361953 on chromosome 11 in *LUZP2* (leucine zipper protein 2) (Table 4, Fig 1). The pedigree-adjusted minor allele frequency was 0.26 for affected individuals versus 0.15 for unaffected individuals. Fourteen additional SNPs had p-values $< 1.0 \times 10^{-5}$ (Table 4). According to our simulation analyses, we have $> 80\%$ power to detect a p-value 0.005 under an additive model with an odds ratio of 2.0 (data not shown). After stratifying, each of the fifteen top SNPs had a more significant p-value in the non-Adams County dataset. Although some of the SNPs have very different minor allele frequencies in the two strata, the less significant p-values for the Adams County dataset can be explained mostly by the lack of power in that stratum (9 LOAD affected). All SNPs showed the same direction of effect in the two strata except for rs472926, rs12361953, and rs472926 (supplemental table 1). These association results did not fall within a megabase of any of the other 9 previously verified LOAD genes (*CRI1*, *CLU*, *PICALM*, *BINI*, *EPHA1*, *MS4A*, *CD33*, *CD2AP*, and *ABCA7*). However, four SNPs (rs10792820, rs11234505, rs10501608, and rs7131120) in *PICALM* generated nominally significant p-values ($P < 0.05$). Rs11234505 is only ~3.0 kb from rs561655, the most significant SNP published by Naj et al (2011), and rs10501608 is only ~10.5 kb from rs541458 the most significant SNP published by Harold et al (2009) and Lambert et al (2009). We also have a nominally significant SNP, rs6591625, in the *MS4A10* gene. The SNP is ~0.5 Mb from rs4938933, the most significant SNP published by Naj et al (2011).

Genome-wide linkage

In the genome-wide analysis, Forty five regions, among all chromosomes except 17, 21, and X, had at least one two-point HLOD ≥ 3.0 (data not shown). Multipoint linkage analysis for these regions resulted in four regions, one each on chromosomes 2, 3, 9, and 18 with a multipoint peak HLOD > 3 (Table 5, Fig 2). The highest peak occurred on chromosome 2 with a recessive peak HLOD of 6.14 (90.91 Mbp) and a dominant peak HLOD of 6.05 (81.03 Mbp). The most significant association results within the recessive and dominant ± 1 -LOD-unit support interval were rs1258411 ($P=5.29 \times 10^{-2}$) and rs2974151 ($P=1.29 \times 10^{-4}$), respectively. Rs1258411 is not located in a gene, but rs2974151 is located in an intron of *CTNNA2* (catenin, alpha 2). In addition to rs2974151, 10 other SNPs in this gene had P -values < 0.05 (data not shown). While this is less than 5% of the analyzed SNPs in *CTNNA2*, it still warrants attention.

The next highest multipoint result was on chromosome 3 with a dominant HLOD of 5.27 and a recessive HLOD of 3.53. The peak for both models is at 168.43 Mbp, and the most significant association result in the ± 1 -LOD-unit support interval was at rs9812366 ($P=4.00 \times 10^{-2}$), which is intergenic. The linkage peak on chromosome 9 reached an HLOD of 4.44 (107.76 Mbp) under the dominant model and 3.77 (101.7 Mbp) under the recessive model. This peak overlaps with the suggestive linkage peak found in the joint linkage analysis published by Hamshere et al (2007), however this region has not been consistently replicated in other studies. For both models the most significant association result in the ± 1 -LOD-unit support interval was at rs9969729 ($P=1.94 \times 10^{-6}$), which is intergenic. On chromosome 18 the dominant and recessive results both peaked at 8.77 Mbp with HLOD=3.97 for the dominant model and HLOD=4.43 for the recessive model. The most significant association result in this ± 1 -LOD-unit support interval was at rs632912 ($P=8.80 \times 10^{-4}$), which is intergenic. None of these regions overlap the linkage peaks found in our previous genome-wide microsatellite linkage study, which used only a subset of the individuals in the current dataset (Hahs et al 2006). As with our association results, these multipoint peaks did not encompass the previously known LOAD genes.

Discussion

APOE was clearly associated with dementia in our population; however, it did not explain the majority of affected individuals. In the Adams County communities, there were only 8/74 individuals who carried at least one *APOE-E4* allele. In the remaining Amish communities, the *APOE-E4* allele was more common, but still less common than in the general population. In addition, the majority of affected individuals (81/127, 64% for all counties; 45/115, 39% for non-Adams counties) did not carry an *APOE-E4* allele. The specific deficit of the *APOE-E4* allele in Adams County as well as differences in allele frequencies for some of the top GWAS SNPs indicates at least some level of locus heterogeneity underlying LOAD in the Amish population.

Additional support for locus heterogeneity arises from the linkage results. Examination of the subpedigree-specific lod scores for the four significant loci indicates that 13 of the 34 subpedigrees generate no lod scores > 0.50 for any of the loci, and 14/21 (67%) of the remaining subpedigrees generate lod scores > 0.50 for only one of the four loci. In addition, the vast majority of the remaining SNPs across the genome generated HLOD scores with alpha values (proportion of linked pedigrees) < 1.0 . Finally, the suggestion of locus heterogeneity is consistent with the societal differences across church districts, which can further restrict marriages even within the Amish.

Because of the relatedness of individuals in our dataset we could take advantage of both linkage and association approaches to identify potential LOAD loci. In our examination, we

found that our most significant association results did not fall under any of the four linkage peaks. However, under the linkage peaks we did see some evidence of association. Within our most significant region of linkage lies *CTNNA2*, which also had suggestive evidence for association. In addition to the result at rs2974151 ($P=1.29\times 10^{-4}$), multiple SNPs in *CTNNA2* had P-values < 0.05 , decreasing the likelihood of a false positive association for this gene. However, because of the relatedness in our dataset it was difficult to get an accurate measurement of LD structure to determine if the SNPs in this region were more highly correlated due to a founder effect.

CTNNA2 encodes the catenin alpha 2 protein, which is a neuronal-specific catenin. Catenins are cadherin-associated proteins and are thought to link cadherins to the cytoskeleton to regulate cell-cell adhesion. Catenin alpha 2 can form complexes with other catenins such as beta-catenin, which interacts with presenilin. Mutations in presenilin lead to destabilization of beta-catenin which potentiates neuronal apoptosis (Zhang et al. 1998). Catenin alpha 2 is also thought to regulate morphological plasticity of synapses and cerebellar and hippocampal lamination during development in mice (Park et al. 2002). It also functions in the control of startle modulation in mice (Park et al 2002).

It was not completely unexpected to see some discordance between the linkage and association results, as was demonstrated in our *APOE* results where we saw evidence for association but not for linkage. Because we needed to divide the pedigree to facilitate linkage analysis and because we used an affecteds-only analysis, only a subset of the individuals analyzed in association analysis were analyzed in linkage analysis. The breaking of the pedigree likely reduces the observed genomic sharing between relatives as the tracking of the natural flow of alleles was somewhat disrupted, as we saw when we tested *APOE* for linkage. Also, the very nature of association analysis versus linkage analysis will provide some different results. Linkage analysis locates shared genomic regions between affected individuals in the same pedigree by testing for co-segregation of a chromosomal segment from a common ancestor. Association using MQS tests for differences in allele frequencies between affected and unaffected individuals while correcting for the pedigree relationships. Association analysis is more powerful in detecting protective effects as well as smaller effects in the population compared to affecteds-only linkage analysis but is underpowered when sample sizes are small and genetic heterogeneity is present. Conversely, linkage analysis is more suitable for finding large effects in a small number of related individuals and is more robust to genetic heterogeneity.

Our results confirmed the complex genetic architecture of LOAD even in this more homogeneous set of individuals. Multiple genes appeared to be significantly contributing to LOAD risk in the Amish. We replicated the effect of *APOE*, replicated the evidence for linkage on 9q22, and also found modest evidence for association of both *PICALM* and *MS4A* in this population. Most importantly, this unique population allowed us to find additional candidate loci, particularly in the *CTNNA2* region in which we saw strong evidence for both linkage and association. The role of *CTNNA2* in the brain also makes this gene a promising candidate. The *CTNNA2* region, in addition to other potential risk regions, needs to be more closely examined to identify the underlying responsible variants and their functional consequences.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the family participants and community members for graciously agreeing to participate, making this research possible. This study is supported by the National Institutes of Health grants AG019085 (to JLH and MAP-V) and AG019726 (to WKS). Some of the samples used in this study were collected while WKS, JRG, and MAP-V were faculty members at Duke University. The authors would like to thank Gene Jackson of Scott & White for his effort and support on this project. Additional work was performed using the Vanderbilt Center for Human Genetics Research Core facilities: the Genetic Studies Ascertainment Core, the DNA Resources Core, and the Computational Genomics Core.

References

- Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet.* 2002; 30:97–101. [PubMed: 11731797]
- Agarwala R, Biesecker LG, Schäffer AA. Anabaptist genealogy database. *Am J Med Genet C Semin Med Genet.* 2003; 121:32–37. [PubMed: 12888984]
- Agarwala R, Biesecker LG, Tomlin JF, Schäffer AA. Towards a complete North American Anabaptist genealogy: A systematic approach to merging partially overlapping genealogy resources. *Am J Med Genet.* 1999; 86:156–161. [PubMed: 10449652]
- Amish Heritage Committee. Amish and Mennonites in Eastern Elkhart & LaGrange Counties, Indiana 1841–1991. Goshen, Indiana: Amish Heritage Committee; 2009. 2nd printing
- Beachy, L. Unser Leit: The Story of the Amish. Millersburg, OH: Goodly Heritage Books; 2011.
- Bertram L, Lill CM, Tanzi RE. The genetics of Alzheimer disease: back to the future. *Neuron.* 2010; 68:270–281. [PubMed: 20955934]
- Boyles AL, Scott WK, Martin ER, Schmidt S, Li YJ, Ashley-Koch A, Bass MP, Schmidt M, Pericak-Vance MA, Speer MC, Hauser ER. Linkage disequilibrium inflates type I error rates in multipoint linkage analysis when parental genotypes are missing. *Hum Hered.* 2005; 59:220–227. [PubMed: 16093727]
- Edwards DR, Gilbert JR, Jiang L, Gallins PJ, Caywood L, Creason M, Fuzzell D, Knebusch C, Jackson CE, Pericak-Vance MA, Haines JL, Scott WK. Successful aging shows linkage to chromosomes 6, 7, and 14 in the amish. *Ann Hum Genet.* 2011; 75:516–528. [PubMed: 21668908]
- Hahs DW, McCauley JL, Crunk AE, McFarland LL, Gaskell PC, Jiang L, Slifer SH, Vance JM, Scott WK, Welsh-Bohmer KA, Johnson SR, Jackson CE, Pericak-Vance MA, Haines JL. A genome-wide linkage analysis of dementia in the Amish. *Am J Med Genet B Neuropsychiatr Genet.* 2006; 141:160–166. [PubMed: 16389594]
- Hamshere ML, Holmans PA, Avramopoulos D, Bassett SS, Blacker D, Bertram L, Wiener H, Rochberg N, Tanzi RE, Myers A, Wavrant-De Vrièze F, Go R, Fallin D, Lovestone S, Hardy J, Goate A, O'Donovan M, Williams J, Owen MJ. Genome-wide linkage analysis of 723 affected relative pairs with late-onset Alzheimer's disease. *Hum Mol Genet.* 2007; 16:2703–2712. [PubMed: 17725986]
- Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, Pahwa JS, Moskva V, Dowzell K, Williams A, Jones N, Thomas C, Stretton A, Morgan AR, Lovestone S, Powell J, Proitsi P, Lupton MK, Brayne C, Rubinsztein DC, Gill M, Lawlor B, Lynch A, Morgan K, Brown KS, Passmore PA, Craig D, McGuinness B, Todd S, Holmes C, Mann D, Smith AD, Love S, Kehoe PG, Hardy J, Mead S, Fox N, Rossor M, Collinge J, Maier W, Jessen F, Schürmann B, van den Bussche H, Heuser I, Kornhuber J, Wiltfang J, Dichgans M, Frölich L, Hampel H, Hüll M, Rujescu D, Goate AM, Kauwe JS, Cruchaga C, Nowotny P, Morris JC, Mayo K, Sleegers K, Bettens K, Engelborghs S, De Deyn PP, Van Broeckhoven C, Livingston G, Bass NJ, Gurling H, McQuillin A, Gwilliam R, Deloukas P, Al-Chalabi A, Shaw CE, Tzolaki M, Singleton AB, Guerreiro R, Mühleisen TW, Nöthen MM, Moebus S, Jöckel KH, Klopp N, Wichmann HE, Carrasquillo MM, Pankratz VS, Younkin SG, Holmans PA, O'Donovan M, Owen MJ, Williams J. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet.* 2009; 41:1088–1093. [PubMed: 19734902]
- Hollingworth P, Harold D, Sims R, Gerrish A, Lambert JC, Carrasquillo MM, Abraham R, Hamshere ML, Pahwa JS, Moskva V, Dowzell K, Jones N, Stretton A, Thomas C, Richards A, Ivanov D, Widdowson C, Chapman J, Lovestone S, Powell J, Proitsi P, Lupton MK, Brayne C, Rubinsztein

DC, Gill M, Lawlor B, Lynch A, Brown KS, Passmore PA, Craig D, McGuinness B, Todd S, Holmes C, Mann D, Smith AD, Beaumont H, Warden D, Wilcock G, Love S, Kehoe PG, Hooper NM, Vardy ER, Hardy J, Mead S, Fox NC, Rossor M, Collinge J, Maier W, Jessen F, R  ther E, Sch  rman B, Heun R, K  lsch H, van den Bussche H, Heuser I, Kornhuber J, Wiltfang J, Dichgans M, Fr  lich L, Hampel H, Gallacher J, H  ll M, Rujescu D, Giegling I, Goate AM, Kauwe JS, Cruchaga C, Nowotny P, Morris JC, Mayo K, Sleegers K, Bettens K, Engelborghs S, De Deyn PP, Van Broeckhoven C, Livingston G, Bass NJ, Gurling H, McQuillin A, Gwilliam R, Deloukas P, Al-Chalabi A, Shaw CE, Tsolaki M, Singleton AB, Guerreiro R, M  hleisen TW, N  then MM, Moebus S, J  ckel KH, Klopp N, Wichmann HE, Pankratz VS, Sando SB, Aasly JO, Barcikowska M, Wszolek ZK, Dickson DW, Graff-Radford NR, Petersen RC, van Duijn CM, Breteler MM, Ikram MA, DeStefano AL, Fitzpatrick AL, Lopez O, Launer LJ, Seshadri S, Berr C, Campion D, Epelbaum J, Dartigues JF, Tzourio C, Alperovitch A, Lathrop M, Feulner TM, Friedrich P, Riehle C, Krawczak M, Schreiber S, Mayhaus M, Nicolhaus S, Wagenpfeil S, Steinberg S, Stefansson H, Stefansson K, Snaedal J, Bj  rnsson S, Jonsson PV, Chouraki V, Genier-Boley B, Hiltunen M, Soininen H, Combarros O, Zelenika D, Delepine M, Bullido MJ, Pasquier F, Mateo I, Frank-Garcia A, Porcellini E, Hanon O, Coto E, Alvarez V, Bosco P, Siciliano G, Mancuso M, Panza F, Solfrizzi V, Nacmias B, Sorbi S, Boss   P, Piccardi P, Arosio B, Annoni G, Seripa D, Pilotto A, Scarpini E, Galimberti D, Brice A, Hannequin D, Licastr   F, Jones L, Holmans PA, Jonsson T, Riemenschneider M, Morgan K, Younkin SG, Owen MJ, O'Donovan M, Amouyel P, Williams J. Alzheimer's Disease Neuroimaging Initiative; CHARGE consortium; EADI1 consortium. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet.* 2011; 43:429–435. [PubMed: 21460840]

Hostetler, J. Amish Society. 4. Baltimore, MD: Johns Hopkins University Press; 1993.

Korn JM, Kuruvilla FG, McCarroll SA, Wysoker A, Nemesh J, Cawley S, Hubbell E, Veitch J, Collins PJ, Darvishi K, Lee C, Nizzari MM, Gabriel SB, Purcell S, Daly MJ, Altshuler D. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet.* 2008; 40:1253–1260. [PubMed: 18776909]

Lambert JC, Heath S, Even G, Campion D, Sleegers K, Hiltunen M, Combarros O, Zelenika D, Bullido MJ, Tavernier B, Letenneur L, Bettens K, Berr C, Pasquier F, Fi  vet N, Barberger-Gateau P, Engelborghs S, De Deyn P, Mateo I, Franck A, Helisalmi S, Porcellini E, Hanon O, de Pancorbo MM, Lendon C, Dufouil C, Jaillard C, Leveillard T, Alvarez V, Bosco P, Mancuso M, Panza F, Nacmias B, Boss   P, Piccardi P, Annoni G, Seripa D, Galimberti D, Hannequin D, Licastr   F, Soininen H, Ritchie K, Blanch   H, Dartigues JF, Tzourio C, Gut I, Van Broeckhoven C, Alperovitch A, Lathrop M, Amouyel P. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet.* 2009; 41:1094–1099. [PubMed: 19734903]

Liu F, Arias-V  squez A, Sleegers K, Aulchenko YS, Kayser M, Sanchez-Juan P, Feng BJ, Bertoli-Avella AM, van Swieten J, Axenovich TI, Heutink P, van Broeckhoven C, Oostra BA, van Duijn CM. A genomewide screen for late-onset Alzheimer disease in a genetically isolated Dutch population. *Am J Hum Genet.* 2007; 81:17–31. [PubMed: 17564960]

Liu F, Elefante S, van Duijn CM, Aulchenko YS. Ignoring distant genealogic loops leads to false-positives in homozygosity mapping. *Ann Hum Genet.* 2006; 70:965–970. [PubMed: 17044871]

Liu F, Kirichenko A, Axenovich TI, van Duijn CM, Aulchenko YS. An approach for cutting large and complex pedigrees for linkage analysis. *Eur J Hum Genet.* 2008; 16:854–860. [PubMed: 18301450]

McCauley JL, Hahs DW, Jiang L, Scott WK, Welsh-Bohmer KA, Jackson CE, Vance JM, Pericak-Vance MA, Haines JL. Combinatorial Mismatch Scan (CMS) for loci associated with dementia in the Amish. *BMC Med Genet.* 2006; 7:19. [PubMed: 16515697]

McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology.* 1984; 34:939–944. [PubMed: 6610841]

Morris JC, Heyman A, Mohs RC, Hughes JP, van Belle G, Fillenbaum G, Mellits ED, Clark C. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology.* 1989; 39:1159–1165. [PubMed: 2771064]

- Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buross J, Gallins PJ, Buxbaum JD, Jarvik GP, Crane PK, Larson EB, Bird TD, Boeve BF, Graff-Radford NR, De Jager PL, Evans D, Schneider JA, Carrasquillo MM, Ertekin-Taner N, Younkin SG, Cruchaga C, Kauwe JS, Nowotny P, Kramer P, Hardy J, Huentelman MJ, Myers AJ, Barmada MM, Demirci FY, Baldwin CT, Green RC, Rogava E, St George-Hyslop P, Arnold SE, Barber R, Beach T, Bigio EH, Bowen JD, Boxer A, Burke JR, Cairns NJ, Carlson CS, Carney RM, Carroll SL, Chui HC, Clark DG, Corneveaux J, Cotman CW, Cummings JL, DeCarli C, DeKosky ST, Diaz-Arrastia R, Dick M, Dickson DW, Ellis WG, Faber KM, Fallon KB, Farlow MR, Ferris S, Frosch MP, Galasko DR, Ganguli M, Gearing M, Geschwind DH, Ghetti B, Gilbert JR, Gilman S, Giordani B, Glass JD, Growdon JH, Hamilton RL, Harrell LE, Head E, Honig LS, Hulette CM, Hyman BT, Jicha GA, Jin LW, Johnson N, Karlawish J, Karydas A, Kaye JA, Kim R, Koo EH, Kowall NW, Lah JJ, Levey AI, Lieberman AP, Lopez OL, Mack WJ, Marson DC, Martiniuk F, Mash DC, Masliah E, McCormick WC, McCurry SM, McDavid AN, McKee AC, Mesulam M, Miller BL, Miller CA, Miller JW, Parisi JE, Perl DP, Peskind E, Petersen RC, Poon WW, Quinn JF, Rajbhandary RA, Raskind M, Reisberg B, Ringman JM, Roberson ED, Rosenberg RN, Sano M, Schneider LS, Seeley W, Shelanski ML, Slifer MA, Smith CD, Sonnen JA, Spina S, Stern RA, Tanzi RE, Trojanowski JQ, Troncoso JC, Van Deerlin VM, Vinters HV, Vonsattel JP, Weintraub S, Welsh-Bohmer KA, Williamson J, Woltjer RL, Cantwell LB, Dombroski BA, Beekly D, Lunetta KL, Martin ER, Kamboh MI, Saykin AJ, Reiman EM, Bennett DA, Morris JC, Montine TJ, Goate AM, Blacker D, Tsuang DW, Hakonarson H, Kukull WA, Foroud TM, Haines JL, Mayeux R, Pericak-Vance MA, Farrer LA, Schellenberg GD. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet.* 2011; 43:436–441. [PubMed: 21460841]
- Park C, Falls W, Finger JH, Longo-Guess CM, Ackerman SL. Deletion in *Catn2*, encoding alpha-N-catenin, causes cerebellar and hippocampal lamination defects and impaired startle modulation. *Nat Genet.* 2002; 31:279–284. [PubMed: 12089526]
- Pericak-Vance MA, Johnson CC, Rimmler JB, Saunders AM, Robinson LC, D'Hondt EG, Jackson CE, Haines JL. Alzheimer's disease and apolipoprotein E-4 allele in an Amish population. *Ann Neurol.* 1996; 39:700–704. [PubMed: 8651641]
- Seshadri S, Fitzpatrick AL, Ikram MA, DeStefano AL, Gudnason V, Boada M, Bis JC, Smith AV, Carrasquillo MM, Lambert JC, Harold D, Schrijvers EM, Ramirez-Lorca R, Debette S, Longstreth WT Jr, Janssens AC, Pankratz VS, Dartigues JF, Hollingworth P, Aspelund T, Hernandez I, Beiser A, Kuller LH, Koudstaal PJ, Dickson DW, Tzourio C, Abraham R, Antunez C, Du Y, Rotter JJ, Aulchenko YS, Harris TB, Petersen RC, Berr C, Owen MJ, Lopez-Arrieta J, Vardarajan BN, Becker JT, Rivadeneira F, Nalls MA, Graff-Radford NR, Campion D, Auerbach S, Rice K, Hofman A, Jonsson PV, Schmidt H, Lathrop M, Mosley TH, Au R, Psaty BM, Uitterlinden AG, Farrer LA, Lumley T, Ruiz A, Williams J, Amouyel P, Younkin SG, Wolf PA, Launer LJ, Lopez OL, van Duijn CM, Breteler MM. European Alzheimer's Disease Initiative Investigators; CHARGE Consortium; GERAD1 Consortium; EADI1 Consortium. Genome-wide analysis of genetic loci associated with Alzheimer disease. *Journal of the American Medical Association.* 2010; 303:1832–1840. [PubMed: 20460622]
- Teng EL, Chui HC. The Modified Mini-Mental State (3MS) examination. *J Clin Psychiatry.* 1987; 48:314–318. [PubMed: 3611032]
- Thornton T, McPeck MS. Case-control association testing with related individuals: a more powerful quasi-likelihood score test. *Am J Hum Genet.* 2007; 81:321–337. [PubMed: 17668381]
- Tschanz JT, Welsh-Bohmer KA, Plassman BL, Norton MC, Wyse BW, Breitner JC. Cache County Study Group. An adaptation of the modified mini-mental state examination: analysis of demographic influences and normative data: the cache county study. *Neuropsychiatry Neuropsychol Behav Neurol.* 2002; 15:28–38. [PubMed: 11877549]
- Zhang Z, Hartmann H, Do VM, Abramowski D, Sturchler-Pierrat C, Staufenbiel M, Sommer B, van de Wetering M, Clevers H, Saftig P, De Strooper B, He X, Yankner BA. Destabilization of beta-catenin by mutations in presenilin-1 potentiates neuronal apoptosis. *Nature.* 1998; 395:698–702. [PubMed: 9790190]

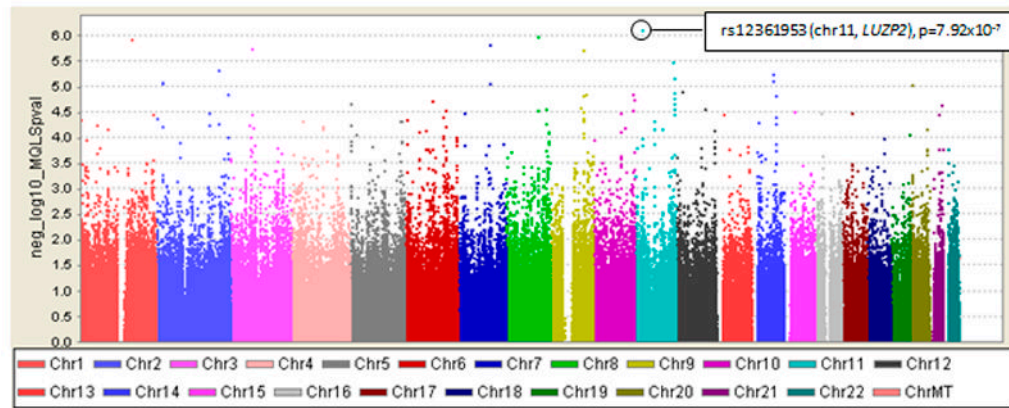


Figure 1. MQLS Manhattan plot

Genome-wide association results were calculated using MQLS for 798 individuals (109 Late-onset Alzheimer disease affected). The lowest P-value (7.92×10^{-7}) was calculated on chromosome 11 at rs12361953 which is located in the Leucine zipper protein 2 (*LUZP2*) gene.

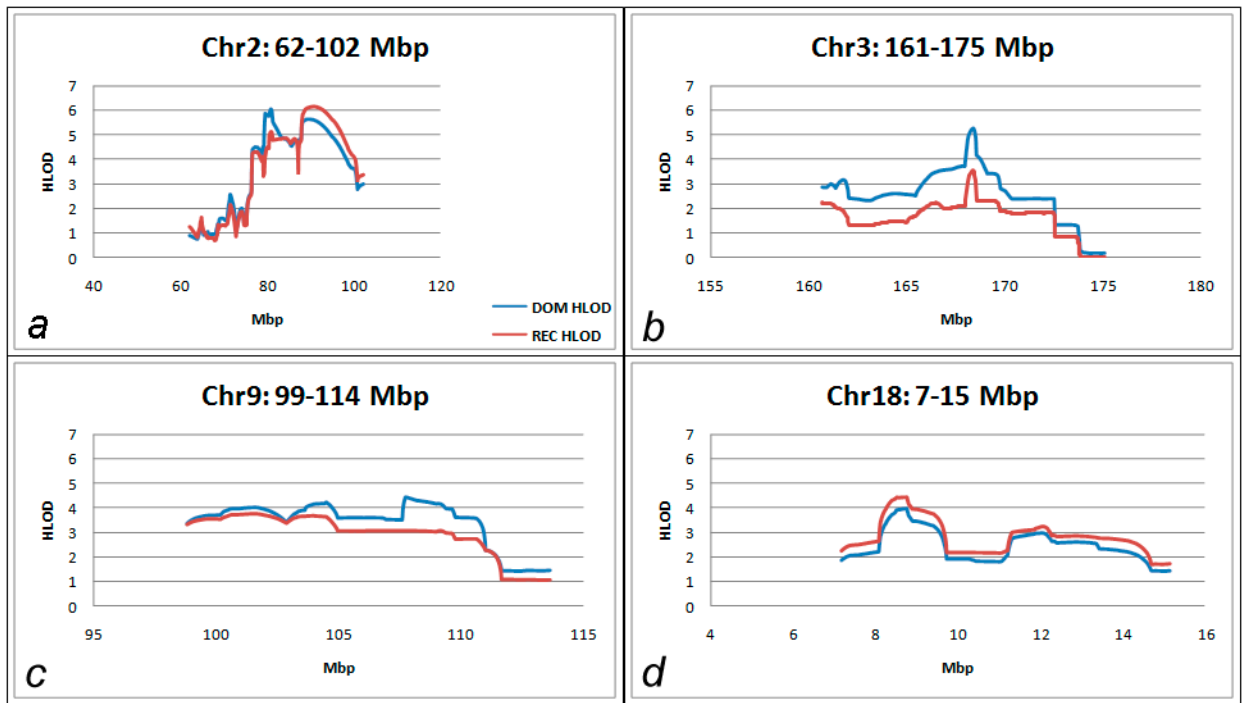


Figure 2. Strongest multipoint linkage peaks

Parametric dominant (blue) and recessive (red) multipoint linkage peaks with HLOD scores >3 were calculated on chromosomes 2 (a), 3 (b), 9 (c), and 18 (d). Red=recessive, Blue=dominant

Table 1

Genome-wide dataset

Ages of exam and onset averages and standard deviations were calculated for the 798 samples—Late-onset AD (LOAD) samples, cognitively normal (unaffected) samples, and unclear or unknown samples—which passed QC for genome-wide genotyping.

	Males	Females	Total	Average Age of Exam (Standard Deviation)	Average Age of Onset (Standard Deviation)
LOAD Affected	43	66	109	83 (7.57)	79 (6.68)
Cognitively Normal	192	258	450	78 (7.67)	-
Unclear or Unknown	117	122	239	74 (15.52)	-

Table 2
MQLS-corrected *APOE* allele frequencies

APOE allele frequencies of Late-onset AD (LOAD) affected individuals versus cognitively normal individuals (unaffecteds) were calculated using MQLS to correct for pedigree relationships. Frequencies were calculated in the Adams County individuals separately from Elkhart, LaGrange, and Holmes Counties.

<i>APOE</i> allele frequencies			
Elkhart, LaGrange, and Holmes Counties			
	<i>E2</i>	<i>E3</i>	<i>E4</i>
LOAD Affected	0.07	0.75	0.18
Cognitively Normal	0.08	0.82	0.11
Adams County			
LOAD Affected	0.00	0.94	0.06
Cognitively Normal	0.04	0.88	0.08

Table 3
Age of onset and number of affected versus unaffected individuals by APOE genotype

Average ages of onset and standard deviations by APOE genotype and number LOAD affected and unaffected by APOE genotype

	APOE Genotype				
	4/4	3/4	2/4	3/3	2/3
Average Age of Onset (stdev)	71 (7.59)	76 (7.94)	74 (3.54)	80 (6.70)	84 (7.42)
Number LOAD Affected	10	34	2	69	12
Number Cognitively normal	6	90	6	308	35

Table 4

Most significant genome-wide association results

Top genome-wide association results calculated using QLS. Minor allele frequencies (MAF) are MQLS-corrected for pedigree relationships. A gene is only listed if SNP falls within specified gene. Megabase pair (Mpb) positions are based on NCBI Build 36.

Chromosome	SNP	Position (Mpb)	Minor Allele	Affected MAF	Unaffected MAF	MQLS P-value	Gene
1	rs4145462	165.99	T	0.10	0.05	1.22×10-06	<i>MPZLI</i>
2	rs41458646	23.09	G	0.27	0.17	8.44×10-06	-
2	rs41476545	23.09	G	0.27	0.17	9.02×10-06	-
2	rs6738181	204.84	A	0.35	0.19	4.97×10-06	-
3	rs7638995	69.26	A	0.20	0.11	1.82×10-06	-
7	rs679974	105.06	C	0.18	0.08	8.67×10-06	<i>ATXN7LI</i>
7	rs11983798	105.07	T	0.17	0.08	1.49×10-06	<i>ATXN7LI</i>
8	rs6468852	104.05	G	0.24	0.13	1.06×10-06	-
9	rs9969729	107.67	A	0.14	0.08	1.94×10-06	-
11	rs12361953	24.57	C	0.26	0.15	7.92×10-07	<i>LUZP2</i>
11	rs472926	125.41	C	0.25	0.15	3.28×10-06	<i>CDON</i>
11	rs4937314	127.69	C	0.27	0.17	7.00×10-06	-
14	rs11848070	70.58	C	0.38	0.25	5.64×10-06	<i>PCNX</i>
14	rs17767225	70.74	T	0.32	0.2	7.88×10-06	-
20	rs6085820	6.93	A	0.17	0.09	9.31×10-06	-

Table 5

Most significant multipoint linkage results

Parametric dominant (Dom) and recessive (Rec) multipoint maximum heterogeneity (HLOD) scores were calculated using Merlin. Regions are determined by ± 1 -LOD-unit support intervals.

Chr	Dom Peak HLOD	Dom peak alpha	Position (Mbp)	Region	Lowest MQLS p-value in Region	Rec Peak HLOD	Rec peak alpha	Position (Mbp)	Region	Lowest MQLS p-value in Region
2	6.05	0.48	81.03	79.46–82.95	1.29E-4 (rs2974151)	6.14	0.39	90.81	87.97–97.46	5.29E-2 (rs1258411)
3	5.27	0.49	168.43	168.06–168.60	4.00E-2 (rs9812366)	3.53	0.26	168.43	168.06–168.60	4.00E-2 (rs9812366)
9	4.44	0.34	107.77	102.94–110.80	1.94E-6 (rs9969729)	3.77	0.23	101.7	98.86–109.79	1.94E-6 (rs9969729)
18	3.97	0.27	8.77	8.12–9.59	8.80E-4 (rs632912)	4.43	0.21	8.77	8.12–9.59	8.80E-4 (rs632912)