

Analysis of Copy Number Variation in Alzheimer's Disease: the NIA-LOAD/NCRAD Family Study (89/120 characters)

Shanker Swaminathan^{1,2}, Li Shen^{1,3}, Sungeun Kim^{1,3}, Mark Inlow^{1,4}, John D. West¹, Kelley M. Faber², Tatiana Foroud^{1,2,3}, Richard Mayeux⁵, Andrew J. Saykin^{1,2,3,#}, for the Alzheimer's Disease Neuroimaging Initiative (ADNI)* and the NIA-LOAD/NCRAD Family Study Group**

¹*Department of Radiology and Imaging Sciences, Center for Neuroimaging, Indiana University School of Medicine, Indianapolis, IN, USA;*

²*Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA;*

³*Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN, USA;*

⁴*Department of Mathematics, Rose-Hulman Institute of Technology, Terre Haute, IN, USA;*

⁵*The Gertrude H. Sergievsky Center, The Taub Institute for Research on Alzheimer's Disease and the Aging Brain, and the Department of Neurology, Columbia University College of Physicians and Surgeons, New York, NY, USA*

#Please address correspondence to:

Andrew J. Saykin, PsyD

Department of Radiology and Imaging Sciences, Center for Neuroimaging, Indiana University School of Medicine, Indianapolis, IN, USA

Phone: 317-278-6947; Fax: 317-274-1067; E-mail: asaykin@iupui.edu

Running title: Analysis of CNVs in the NIA-LOAD/NCRAD Family Study

*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

**Investigators who contributed to the NIA-LOAD/NCRAD Family Study can be found on the dbGaP website

(http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000168.v1.p1).

ABSTRACT (247/250 words)

Copy number variants (CNVs) are DNA regions that have gains (duplications) or losses (deletions) of genetic material. CNVs may encompass a single gene or multiple genes and can affect their function. They are hypothesized to play an important role in certain diseases. We previously examined the role of CNVs in late-onset Alzheimer's disease (AD) and mild cognitive impairment (MCI) using participants from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study and identified gene regions overlapped by CNVs only in cases (AD and/or MCI) but not in controls. Using a similar approach as ADNI, we investigated the role of CNVs using 794 AD and 196 neurologically evaluated control non-Hispanic Caucasian NIA-LOAD/NCRAD Family Study participants with DNA derived from blood/brain tissue. The controls had no family history of AD and were unrelated to AD participants. CNV calls were generated and analyzed after detailed quality review. 711 AD cases and 171 controls who passed all quality thresholds were included in case/control association analyses, focusing on candidate gene and genome-wide approaches. We identified genes overlapped by CNV calls only in AD cases but not controls. A trend for lower CNV call rate was observed for deletions as well as duplications in cases compared to controls. Gene-based association analyses confirmed previous findings in the ADNI study (*ATXN1*, *HLA-DPB1*, *RELN*, *DOPEY2*, *GSTT1*, *CHRFAM7A*, *ERBB4*, *NRXN1*) and identified a new gene (*IMMP2L*) that may play a role in AD susceptibility. Replication in independent samples as well as further analyses of these gene regions is warranted.

Keywords (6 to 8 keywords): Alzheimer's disease, association study, *CHRFAM7A*, copy number variation, dementia, *IMMP2L*, NIA-LOAD/NCRAD, replication **(8 keywords).**

INTRODUCTION

Alzheimer's disease (AD) is the most common form of dementia characterized by loss of memory and other intellectual abilities, which eventually disrupts daily life activities. An estimated 5.4 million Americans have AD, the sixth leading cause of death across all ages in the United States [1]. The hallmark abnormalities of AD are deposits of the beta-amyloid protein fragments (amyloid plaques) and twisted strands of the tau protein (neurofibrillary tangles). Although no current treatments can slow or halt its progression, a large research effort is being undertaken to identify causes that can lead to earlier diagnosis and treatment. Amnesic mild cognitive impairment (MCI) is a clinical condition in which a person has memory problems that are not normal for the individual's age, but not severe enough to interfere significantly with daily functioning. Approximately 14-18% of individuals aged 70 years and older have MCI, and 10-15% of these individuals with MCI will likely progress to AD or another dementia every year [2].

Genetic variation is a key factor in the development and progression of AD, with approximately 58-79% of phenotypic variation estimated to be caused by genetic factors [3]. Early-onset AD (EOAD; onset < 60 years), which accounts for < 5% of AD cases is caused by mutations in the *APP*, *PSEN1* and *PSEN2* genes. The leading genetic risk factor for the more common late-onset form of AD (LOAD; onset \geq 60 years) is the *APOE* ϵ 4 allele [4]. A member of a three allele haplotype (composed of ϵ 2, ϵ 3 and ϵ 4 alleles), the ϵ 4 allele shows a dose-dependent increase in AD risk of approximately four-fold in carriers as compared to noncarriers [5-7]. Recently, other AD risk loci have been identified and replicated including: *CLU*, *CR1*, *PICALM*, *BIN1*, *EXOC3L2*, *MTHFD1L*, *MS4A4A/MS4A6E*, *CD2AP*, *CD33*, *ABCA7* and *CUGBP2* [8-15]. However, these loci do not account for all of the genetic variation associated with AD, and it is likely that other forms of genetic variation such as copy number variations (CNVs) play a role.

CNVs are DNA regions ranging in size from 1 kilobase (kb) to several megabases (Mb) present in variable number of copies in the genome. The regions can have addition of genetic material (copy number gains or duplications) or loss of genetic material (copy number losses or deletions). They often overlap one or more genes, and may affect gene function [16]. To our knowledge, only four studies have investigated the role of CNVs in LOAD [17-20]. Heinzen *et al.* performed a case-control genome-wide scan of AD and identified a duplication in the

CHRNA7 gene that they thought warranted further investigation [17]. In an in-depth analysis of the *CRI* region, Brouwers *et al.* observed a low-copy repeat associated CNV in *CRI*, that produced different CR1 isoforms, CR1-F and CR1-S [18]. They were able to obtain a significant association in carriers of CR1-S with AD and were able to replicate this finding in an independent cohort. In a case-only genome-wide CNV association study, Shaw *et al.* demonstrated that a chromosomal region on 14q11.2, encompassing a cluster of olfactory receptors, is associated with age at onset of AD [19].

In a previous study, we performed a case-control CNV analysis in 288 AD, 183 MCI, and 184 control non-Hispanic Caucasian participants in the Alzheimer's Disease Neuroimaging Initiative (ADNI) study who had DNA samples derived from peripheral blood [20]. The analyses included candidate gene and genome-wide approaches to identify genes overlapped by CNVs only in cases (AD and/or MCI) but not in controls. Although no excess CNV burden was observed in cases compared to controls, CNVs overlapping the candidate gene *CHRFAM7A*, as well as *CSMD1*, *SLC35F2*, *HNRNPCL1*, *NRXN1*, and *ERBB4* regions were identified only in cases. Using a similar approach, we analyzed the role of CNVs in AD using unrelated non-Hispanic Caucasian participants in the National Institute of Aging-LOAD/National Cell Repository for AD (NIA-LOAD/NCRAD) Family Study [15] who had DNA samples derived from blood or brain tissue. Case/control association analyses were performed to compare the CNV burden between AD participants (cases) and controls, and to characterize genomic regions where CNVs were detected in cases but not in controls.

MATERIAL AND METHODS

Data used in this study were obtained from the "NIA-Late Onset Alzheimer's Disease and National Cell Repository for Alzheimer's Disease Family Study: Genome-Wide Association Study for Susceptibility Loci" dataset (dbGaP Study Accession: phs000168.v1.p1, Project #2026) on the database of Genotypes and Phenotypes (dbGaP; http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000168.v1.p1) website.

Participants

Recruitment information for participants in the NIA-LOAD Family Study and NCRAD has been previously described [15]. Briefly, the AD sample contained individuals from families as well as unrelated individuals who had a family history of AD. All individuals were recruited after providing informed consent and with approval by the

relevant institutional review boards. The study was conducted according to the principles in the Declaration of Helsinki. The dataset contained information for 607 families (1516 affected, 1306 unaffected) from the NIA-LOAD Family Study, 138 families (337 affected, 166 unknown intermediate phenotypes) from NCRAD, and 471 unrelated patients from the NIA-LOAD Family Study and NCRAD. Three sources were used to ascertain unrelated controls: the NIA-LOAD Family Study (n=794), and NCRAD (n=144), with the NCRAD controls including 141 participants from the University of Kentucky. The NIA-LOAD and NCRAD recruited controls did not have a family history of LOAD in a first degree relative, whereas the University of Kentucky controls were not excluded if they had a family history of LOAD. Genome-wide genotyping for all samples was performed at the Center for Inherited Disease Research (CIDR; <http://www.cidr.jhmi.edu/>) using the Illumina Human610-Quad BeadChip. The *APOE* polymorphisms (based on rs7412 and rs429358) for all samples were genotyped at PreventionGenetics (<http://www.preventiongenetics.com/>). The phenotype and genotype information for all participants were available as part of the dataset.

Alzheimer's Disease Neuroimaging Initiative (ADNI)

The Alzheimer's Disease Neuroimaging Initiative (ADNI) data used in the present study was obtained from the ADNI database (<http://adni.loni.ucla.edu>). Launched in 2003 as a \$60 million, multiyear public-private partnership, the primary goal of ADNI has been to test whether serial magnetic resonance imaging, positron emission tomography, other biological markers, and clinical and neuropsychological assessments can be combined to measure the progression of MCI and early AD. Michael W. Weiner, MD, VA Medical Center and University of California-San Francisco is the Principal Investigator of this initiative. Further information about ADNI can be found at <http://www.adni-info.org>. Genome-wide genotyping of the ADNI sample was performed using the Illumina Human610-Quad BeadChip as previously described [20, 21]. The *APOE* polymorphisms (rs429358 and rs7412) were genotyped separately. Clinical, imaging, biomarker and genetic information for all ADNI participants is available in the ADNI database.

Sample Selection Criteria

For the present analysis, we selected unrelated AD participants (n=794) and controls (n=196) of non-Hispanic Caucasian descent who had DNA samples extracted from blood or brain tissue as described below.

Genotype Data

Genotype data for 5573 participants were available as part of this dataset. Participants who had DNA samples extracted from "Blood" or "Brain Tissue" were selected for the present analyses. Those with DNA samples extracted from lymphoblastoid cell lines (LCLs) were excluded because cell line transformation may bias CNV results [22, 23]. Also, only samples that were determined to have "High Quality Genotyping Data; High Quality Intensity Data for CNV detection" or "High Quality Genotyping Data; Low Quality Intensity Data for CNV detection" in order to have maximum power and representativeness and no known chromosomal abnormality were included in the analyses. 2843 samples with genotype data were available after these filtering steps.

Phenotype Data

Phenotype data for 5220 participants were available as part of this dataset. Participants with a diagnosis of "Alzheimer disease" (AD) or "Neurologically evaluated control" (controls) were selected for the analyses. An AD participant was defined as an individual meeting NINCDS-ADRDA criteria for probable or possible AD [24], or meeting NINCDS-ADRDA criteria for definite AD when clinical and pathological information was available, or CERAD pathological criteria [25] for AD based on postmortem information alone. A control was defined as an individual who demonstrated or had a documented history of normal cognitive function for age. Controls were determined by clinical or neuropathological examination to not meet criteria for AD. Only participants who were "recruited as controls" and not part of an AD family were selected as controls. We also restricted our analyses to non-Hispanic Caucasian participants identified from principal component analyses using the smartpca program from the EIGENSOFT package [26]. 2761 participants (1852 AD and 909 controls) with phenotype data remained after these filtering steps. 1592 participants (1312 AD and 280 controls) had both genotype and phenotype data.

Selection of One Participant per Family

One AD participant per family was selected based on the following criteria: 1) diagnosis level (definite AD>probable AD>possible AD); 2) lower age of dementia symptoms; 3) lower age at which participant was diagnosed with AD; and 4) higher genotype call rate. One control per family was selected based on a higher genotype call rate. Controls were not biologically related to the portions of the family affected by AD. 795 cases and 249 controls remained after this filtering step. Although controls with no family history of AD were recruited, over

time they may have family members develop AD. 54 such recruited controls (53 controls and one control who converted to AD) had first degree relatives with AD, and were excluded from the analyses. After all filtering steps, 990 participants (794 AD and 196 controls) were available for performing the case/control association analyses.

Generation of CNV calls and Quality Control

CNV call generation and quality control (QC) measures in the present study were performed as described in a previous similar study, Swaminathan *et al.* [20]. Briefly, the Log R Ratio (LRR) and B Allele Frequency (BAF) values for each sample were used for the generation of CNV calls for the sample. The LRR of a sample is the log (base 2) ratio of the normalized intensity value for the SNP divided by the expected normalized intensity value. The BAF for a sample is the allelic composition (copy angle) for a SNP that is corrected for cluster position. A large set of normal individuals were used to generate the cluster positions (Illumina GenomeStudio Genotyping Module v1.0 User Guide). PennCNV software [27] (2010Jun22 version) (<http://www.openbioinformatics.org/penncnv/>), which implements a hidden Markov model (HMM) was used for the generation of CNV calls. A number of detection algorithms are available for analyzing CNVs from genome-wide SNP array data including HMM, segmentation algorithms, t-tests and standard deviations of the LRR. A comparison of these methods has been performed by Dellinger *et al.* [28]. Although the PennCNV program was observed to have moderate power in detecting CNVs, it also had a low false positive call rate and hence was used for the generation of CNV calls. A genomic wave adjustment procedure [29] as implemented in PennCNV was carried out. A frequency distribution plot of the number of CNV calls for all samples was made, and a sample was excluded if the number of CNV calls for the sample was greater than the 90th percentile. Samples were also excluded if: LRR SD>0.35, BAF Drift>0.002 or Waviness Factor>0.04. An AD sample was observed to have a very large (~11.5 Mb) deletion on chromosome 14 and was excluded from the analyses as it may be a possible outlier. Due to complications of hemizyosity in males and X-chromosome inactivation in females, we restricted our analyses to autosomes. To ensure only high quality samples were included in the analyses, CNVs for which the difference of the most likely copy number state and less likely copy number state was less than 10, CNV calls generated based on data from less than 10 SNPs, and CNVs that had >50% overlap with centromeric, telomeric, and immunoglobulin regions as defined in Need *et al.* [30] were excluded. 882 participants (711 AD, 171 controls) with 8211 CNV calls remained after all QC measures and were entered into the case/control association analyses.

Case/Control Association Analyses

Case/control association analyses were performed using CNV calls generated for the AD participants and controls using PLINK v1.07 [31] (<http://pngu.mgh.harvard.edu/~purcell/plink/>) to determine CNV call differences between cases (AD) and controls. As in the ADNI study, two approaches were used: (1) a candidate gene approach consisting of AD genes identified from the AlzGene database [32] (Updated 5 January 2011) (<http://www.alzgene.org/>) as having a positive association with AD in at least one study, consisting of 317 genes tested, and (2) a genome-wide approach using PLINK's gene list (hg18 coordinates), consisting of 17938 genes tested. We also performed the candidate gene analysis in the ADNI study using the same AD candidate gene list as used in the present study to compare the results obtained in the two studies. The AlzGene database is a publicly available and regularly updated online resource that comprehensively catalogs genetic case/control and family association studies in AD. In both approaches, CNV segments either partially or completely overlapping gene regions were analyzed, and both deletions and duplications were included in the analyses.

Representative plots of the CNV calls Fig. (2) were created in UCSC Genome Browser [33] (<http://genome.ucsc.edu/>) (March 2006 (NCBI36/hg18) assembly). The Genome Browser track for the Illumina Human-610 array was obtained from the PennCNV website (http://www.openbioinformatics.org/penncnv/penncnv_download.html). The Illumina Genome Viewer plug-in (Human Build 36.1) within GenomeStudio was used to generate representative plots of LRR and BAF values for the participants Fig. (1) and Fig. (3).

RESULTS

Sample Demographics and CNV Call Characteristics

The sample demographics of the 882 participants (711 AD and 171 controls) are shown in Table 1. Of the 711 AD participants, 263 participants had a diagnosis of definite AD, 425 participants had a diagnosis of probable AD, and 23 participants had a diagnosis of possible AD. Significant ($p < 0.05$; two-sided) differences in the number of years of education, the absence or presence of the *APOE* $\epsilon 4$ allele and the DNA source were observed between the AD participants and controls. The CNV call characteristics of the 882 participants are shown in Table 2. 8211 CNV calls (5586 deletions and 2625 duplications) were observed with an average CNV call length of 87.07 kb and

an average of 26 SNPs per CNV call. A higher CNV call rate and a lower average CNV call size were observed in deletions compared to duplications. A trend towards a lower CNV call rate was observed for deletions as well as for duplications in AD participants compared to controls, but this was not significant ($p>0.05$; two-sided) when evaluated by permutation. The largest proportion of deletions and duplications were found in the 0.1-0.5 Mb range (Table 3). Two AD participants were observed to have very large CNV calls (>2 Mb) Fig. (1). The first AD participant had a 2.4 Mb deletion on chromosome 11 Fig. (1A), which includes many genes. The second AD participant had a 3.2 Mb duplication on chromosome 3 Fig. (1B), which includes the *GBE1* gene.

Case/Control Association Analyses

Candidate Gene Approach

CNV calls overlapping 317 AD candidate genes from at least one case (AD) but no controls were identified. The 30 genes identified are present in Table 4 for reference although these do not meet conventional significance ($p<0.05$; one-sided) due to low power. On performing a similar analysis using the same AD gene list in the ADNI study, 15 genes were also identified as being overlapped by CNV calls from at least one case (AD and/or MCI) but no controls. Five genes were identified by both studies: *ATXN1*, *HLA-DPB1*, *RELN*, *DOPEY2* and *GSTT1*. The conditional probability of five or more genes being simultaneously identified by the two studies was evaluated by combinatorial calculation ($p=0.0083$). The *CHRFAM7A* gene reported in the ADNI study [20] was also identified in 12 AD participants and one control in the present study Fig. (2A).

Genome-wide Approach

We also identified CNV calls present in cases (AD) but not controls overlapping 17938 genes in the human genome. The *IMMP2L* gene was identified in 13 AD participants although it did not reach conventional significance (Uncorrected $p=0.059$; one-sided) Fig. (2B). This gene was also found to be overlapped by CNV calls from four AD, one MCI and one control in the ADNI study. The *CSMD1*, *HNRNPCL1* and *SLC35F2* genes reported in the ADNI study were found to be overlapped by CNV calls from 11 AD participants and six controls, 23 AD participants and 11 controls, and 20 AD participants and three controls respectively from the NIA-LOAD/NCRAD sample. Two genes associated with neuropsychiatric disorders reported in the ADNI study: *NRXN1* and *ERBB4*, were also identified in the present study in five and four AD participants, but not in controls. An AD participant was

observed to have a duplication in the 16p13.11 region Fig. (3A) and another AD participant was observed to have a deletion in the 17p12 region Fig. (3B).

DISCUSSION

The present report represents an initial analysis of CNVs in the NIA-LOAD/NCRAD Family Study and a follow-up report and partial replication of the CNV analyses in the ADNI study. After extensive QC, we performed case (AD)/control association analyses using candidate gene and genome-wide approaches similar to those used in the ADNI study.

Comparison of the CNV calls in the two diagnosis groups showed a trend towards a lower CNV call rate for deletions as well as duplications in AD participants compared to controls. In the ADNI study, a trend towards a higher CNV call rate for deletions and a lower CNV call rate for duplications in AD and MCI participants compared to controls was observed. The differences observed between the two studies may be due to random sampling variation, different participant selection criteria, and that the ADNI study analyses also included MCI participants in addition to AD participants and controls, whereas the NIA-LOAD/NCRAD Family Study analyses included only AD participants and controls. Two AD participants in the present study were identified as having very large CNV calls (>2 Mb) Fig. (1). The first AD participant had a 2.4 Mb deletion on chromosome 11 Fig. (1A), overlapping a number of genes including olfactory receptor genes. Olfactory receptor genes are members of a large multigene family encoding signaling transduction pathway components involved in odorant discrimination [34]. Many of the olfactory receptor genes are located on chromosomes 6, 11 and 17, as well as distributed on other chromosomes. Odor identification has been shown to be impaired early in AD [35] and an increase in odor identification deficits has been shown in MCI participants compared to participants without MCI [36]. A high copy number in a region on chromosome 14 encompassing a cluster of olfactory receptors was recently shown to be associated with younger age at onset of AD [19]. The second AD participant had a 3.2 Mb duplication on chromosome 3 Fig. (1B), which includes the *GBE1* (glucan (1,4-alpha-), branching enzyme 1) gene. The protein encoded by this gene is a glycogen branching enzyme involved in glycogen biosynthesis and mutations in the gene have been associated with glycogen storage disease IV [37]. This gene has not been previously associated with AD susceptibility. Further

characterization of these regions by cytogenetic or molecular techniques is required to determine their clinical relevance.

Case/control association analyses using a candidate gene approach and a genome-wide approach similar to those used in the ADNI study were performed. CNV calls partially overlapping genes in AD participants relative to controls were determined, suggesting a possible role for these genes in AD susceptibility.

Several interesting genes were identified using a candidate gene approach (Table 4). The *CHRFAM7A* (*CHRNA7* (cholinergic receptor, nicotinic, alpha 7, exons 5-10) and *FAM7A* (family with sequence similarity 7A, exons A-E) fusion) gene reported in the ADNI study was also identified in 12 AD participants and one control in the present study Fig. (2A). Located on chromosome 15, the gene is a hybrid consisting of a partial duplication of the *CHRNA7* gene fused to a copy of the *FAM7A* gene [38, 39]. It is highly polymorphic and individuals with and without the gene have been identified. A 2-bp deletion polymorphism at position 497-498 in exon 6 of the gene has been associated with schizophrenia [40], and a significant over-representation of the *CHRFAM7A* genotype without the 2-bp allele has been observed in AD participants, dementia with Lewy bodies participants and Pick's disease participants compared to controls [41]. Although *CHRFAM7A* is transcribed, its translation and the possible role of the resulting protein is uncertain. Recently, de Lucas-Cerrillo *et al.* cloned and expressed the full-length coding sequence of the *CHRFAM7A* transcript in pituitary-derived GH4C1 cells and oocytes [42]. On performing a functional study of the protein in oocytes, they observed a dominant negative regulatory function of the protein on $\alpha 7$ nicotinic acetylcholine receptors ($\alpha 7$ nAChRs) activity through reduction in the number of functional $\alpha 7$ nAChRs incorporated into the oocyte surface. Based on these and other results, they suggested that the *CHRFAM7A* gene product could possibly modulate $\alpha 7$ receptor-mediated synaptic transmission and cholinergic anti-inflammatory response. Another recent study by Araud *et al.* suggests *CHRFAM7A* to be a dominant negative modulator of *CHRNA7* function and important for receptor regulation in humans [43]. In order to determine how $\alpha 7$ nAChR activation would affect *APP* processing, Nie *et al.* constructed a SH-EP1- $\alpha 7$ nAChR-hAPP695 cell line model co-expressing $\alpha 7$ nAChR gene and human amyloid precursor protein 695 (hAPP695) gene [44]. Their results demonstrated that, by regulating γ -secretase activity, activation of $\alpha 7$ nAChR reduced *APP* processing in the amyloidogenic pathway. At the same time, the activation was found to enhance *APP* processing in the non-amyloidogenic pathway, thus suggesting a role of $\alpha 7$ nAChR in *APP* processing. Although the identified genes were

overlapped by a small number of CNV calls, these genes have been previously investigated in AD studies and thus represent potential candidate genes. The role of these genes in AD susceptibility can be confirmed by performing replication studies in other samples and laboratory validation.

In order to compare NIA-LOAD/NCRAD results with those obtained in the previous ADNI study, we performed candidate gene analyses using the same AD gene list as the present study. Five genes (*ATXN1*, *HLA-DPBI*, *RELN*, *DOPEY2* and *GSTT1*) identified in the NIA-LOAD/NCRAD Family Study were also identified in the ADNI study as being overlapped by CNV calls from cases (AD and/or MCI participants), but not controls. The present study identified 30 genes and the ADNI study identified 15 genes. There was a statistically significant agreement (conditional probability, $p=0.0083$) for these five genes showing a signal across the two studies. The *ATXN1* (ataxin 1) gene on chromosome 6 encodes a protein ataxin-1, the mutated form of which is associated with spinocerebellar ataxia type 1. The loss of function of this gene has been shown to increase amyloid-beta 40 (A β 40) and amyloid-beta 42 (A β 42) levels by potentiating β -secretase processing of amyloid precursor protein [45]. *ATXN1* and the related Brother of *ATXN1* (*BOATI*) have been recently shown to be important components of the Notch signaling pathway, and may play a role in several Notch-controlled development and disease processes [46]. An evidence of association has been suggested between an intronic SNP (rs179943) in this gene and AD [47]. The *HLA-DPBI* (major histocompatibility complex, class II, DP beta 1) also on chromosome 6 is a member of the HLA class II beta chain paralogues and plays a major role in the immune system by presenting peptides derived from extracellular proteins. The *HLA* alleles have been previously investigated for association to AD [48, 49]. The *RELN* (reelin) gene on chromosome 7 encodes for a large secreted extracellular matrix protein thought to control cell-cell interactions important during brain development for cell positioning and neuronal migration. Variants in the gene have been associated with AD [50]. Increased expression of reelin has been observed in the pyramidal neurons of the hippocampus in AD individuals and in cognitively intact controls with AD-associated pathology [51]. The authors of the study suggest that the reelin up-regulation may be a compensatory response to amyloid-beta or tau-related stress associated with AD even prior to the onset of dementia. The *DOPEY2* (dopey family member 2) or *C21orf5* gene located in the Down syndrome critical region on chromosome 21 has been considered a potential Down syndrome (DS) candidate gene [52]. The gene has been shown to be differentially expressed and overexpressed in DS brains and it is thought that its overexpression could play an important role in the neurological

phenotypes and mental retardation in DS patients. The *GSTT1* (glutathione S-transferase theta 1) gene on chromosome 22 is a member of a protein superfamily that catalyses the conjugation of reduced glutathione to a number of electrophilic and hydrophobic compounds. Glutathione S-transferase variants have been previously investigated for association to AD [53, 54].

The genome-wide approach revealed CNV calls overlapping the *IMMP2L* (inner mitochondrial membrane peptidase-like (*S. cerevisiae*)) gene in 13 AD participants, but no controls. This gene was also identified in four AD, three MCI and one control in the ADNI study Fig. (2B). Located on chromosome 7, *IMMP2L* is a catalytic subunit of the mitochondrial inner membrane peptidase (IMP) complex [55]. The IMP complex proteolytically removes the mitochondrial targeting presequence of nuclear-encoded proteins to generate mature, active proteins in the mitochondrial intermembrane space. *IMMP2L* was first identified in a patient with Gilles de la Tourette syndrome (GTS) who carried a *de novo* inverted duplication of a segment of the long arm of chromosome 7 [56]. Recently, it has been identified as being disrupted in another patient with GTS who had a *de novo* translocation between chromosomes 2 and 7 [57]. Reverse-transcription polymerase chain reaction analyses showed ubiquitous expression of this gene except in the adult liver and lung. In mutant mice with a mutation in the *IMMP2L* gene, the mutation has been shown to affect the signal peptide sequence processing of mitochondrial proteins cytochrome c1 and glycerol phosphate dehydrogenase 2 [58]. This affected the mitochondrial function such that mitochondria from mutant mice generated higher than normal superoxide ion levels and there was impaired fertility in both sexes. Mutant mice have also been shown to manifest multiple aging-associated phenotypes including wasting, loss of subcutaneous fat, sarcopenia, kyphosis and ataxia, and the loss of subcutaneous fat was due to impaired adipose progenitor/stem cells self-renewal [59]. The authors suggest that accelerated aging is driven by mitochondrial reactive oxygen species and that reactive oxygen species damage to adult stem cells could be a possible mechanism for age-associated disorders. The *CSMD1* (CUB and Sushi multiple domains 1), *HNRNPCL1* (heterogeneous nuclear ribonucleoprotein C-like 1) and *SLC35F2* (solute carrier family 35, member F2) genes reported in the ADNI study were found to be overlapped by CNV calls from 11 AD participants and six controls, 23 AD participants and 11 controls, and 20 AD participants and three controls respectively in the present study. None of these genes have been previously associated with AD. These genes warrant further investigation in independent samples to determine their role in AD susceptibility.

Two neuropsychiatric disorder candidate genes: *NRXN1* and *ERBB4*, reported in the ADNI study, were also identified in the present study to be overlapped by CNV calls only in AD participants, but not in controls. Deletions in four AD participants were observed in the *NRXN1* (neurexin 1) gene; deletions in four AD participants and a duplication in one AD participant were observed in the *ERBB4* (v-erb-a erythroblastic leukemia viral oncogene homolog 4 (avian)) gene. *NRXN1* is a member of the neurexin family located on chromosome 2. Neurexins are molecules on the cell surface that form a heterophilic, Ca^{2+} -dependent complex at central nervous system synapses by binding to neuroligins [60]. This transsynaptic complex is essential for formation of synaptic contacts and efficient neurotransmission. Although *NRXN1* has not been previously associated with AD, the gene has been shown to have decreased expression with increasing AD severity [61]. The gene has also been associated with autism [62], autism spectrum disorders [63] and schizophrenia [64]. The *ERBB4* gene, also located on chromosome 2, belongs to the type I receptor tyrosine kinase subfamily, and encodes a receptor for neuregulin 1 (*NRG1*). It is widely expressed in many adult and fetal tissues, with high levels of expression observed in the developing brain and heart [65]. *NRG1* and ErbB4 immunoreactivity have been found to be associated with neuronal plaques in AD brains as well in a transgenic mouse model of AD [66]. A significant increase in ErbB4 immunoreactivity was also observed in AD human brains, and this was demonstrated to colocalize with the apoptotic signal Bax in apoptotic hippocampal pyramidal neurons, suggesting the possible role of *NRG1*/ErbB4 signaling as a survival signal in AD progression [67]. The authors in a more recent study showed that the immunoreactivities of ErbB4 and phospho-ErbB4 were of higher intensity in the neurons of the CA1-2 transitional field of AD brains compared to age-matched controls [68]. They also observed an increased ErbB4 expression in the neurons of the cortico medial nucleus amygdala, human basal forebrain and superior frontal gyrus of AD brains. ErbB4 immunoreactivity was also found to be significantly increased in the cerebral cortex and hippocampus of amyloid precursor protein/presenilin 1 double transgenic mice compared to age-matched wild type control. The authors thus suggested that ErbB4 immunoreactivity upregulation may be involved in the progression of AD pathology. An AD participant was observed to have a duplication in the 16p13.11 region Fig. (3A). Duplications at 16p13.11 have been observed to be significantly enriched in individuals with neurocognitive disease [69], attention-deficit hyperactivity disorder [70] and schizophrenia [71] compared to controls, but have not been previously associated with AD. A different AD participant was observed to have a deletion in the 17p12 region Fig. (3B). Although this region is a possible schizophrenia candidate loci [72, 73], it has not been previously associated with AD.

A number of gene regions overlapped by CNV calls were identified in the present study. Molecular characterization of these regions will be required to determine their specific mechanistic relevance in normal aging and AD. With further study, some of these regions may hold promise for biomarker or drug target development.

The present study has some limitations regarding the selection criteria chosen for the analysis as mentioned in the Material and Methods section. We excluded any controls that were biologically related to individuals with AD as it was possible that these participants could develop AD at a later time, and might have shared genetic regions with the affected family members. Thus, a higher number of AD participants (n=794) compared to controls (n=196) were included in the present study. Also, it has been shown that Epstein-Barr virus (EBV)-transformed peripheral B lymphocytes, in the postimmortal stage, develop strong telomerase activity and aneuploidy, as well as gene mutations and reprogramming [23]. These changes may lead to biological artifacts [22, 74], and hence we chose to include only participants whose DNA samples were derived from blood or brain tissue. There does not yet appear to be a well-defined consensus set of QC criteria to ensure that the most appropriate samples are included in CNV analyses. Thus, it is possible that the QC measures used in the present study may have been too stringent, and samples that may have had informative CNV data have been excluded.

CONCLUSION

In sum, we have conducted an initial CNV analysis in the NIA-LOAD/NCRAD Family Study dataset. A trend towards a lower CNV call rate for deletions as well as duplications was observed in AD participants compared to controls. A number of gene regions were identified from the gene-based association analyses including those reported in the ADNI study (*CHRFAM7A*, *NRXN1* and *ERBB4*) as well as a new gene (*IMMP2L*). The candidate gene analysis of this dataset and the ADNI study simultaneously identified a statistically significant set of five genes (*ATXN1*, *HLA-DPBI*, *RELN*, *DOPEY2* and *GSTT1*). Additional replication in other independent data sets will be important to confirm these findings. Future studies to elucidate the biological role of these variants appear warranted. Overall, some consistency of CNVs across AD cohorts is emerging and this variation holds promise for revealing novel risk factors and disease mechanisms.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank the NIH GWAS Data Repository for providing the "Genetic Consortium for Late Onset Alzheimer's Disease" dataset (dbGaP Study Accession: phs000168.v1.p1, Project #2026). Funding support for this study was provided through the following grants from the National Institute of Aging: U24AG026395 (NIA-LOAD Family Study); U24AG021886 (National Cell Repository for Alzheimer's Disease); R01AG027224 and P30AG005133 University of Pittsburgh; P30AG10161 Rush University Medical Center; P30AG013846 Boston University; P50AG08702, P01AG07232, and R37AG015473 Columbia University; P30AG028377 and P50AG05128 Duke University; P30AG010133 Indiana University; P50AG05134 Massachusetts General Hospital; P50AG165574 Mayo Clinic, Rochester, and Mayo Clinic, Jacksonville; P01AG05138, P01AG02219, and P50AG05138 Mount Sinai School of Medicine; P30AG13854 Northwestern University Medical School; P30AG008017 Oregon Health and Science University; P50AG016582 University of Alabama at Birmingham; P50AG016579 David Geffen School of Medicine, University of California Los Angeles; P30AG028383 University of Kentucky, Lexington; P30AG10124 University of Pennsylvania; P50AG05142 University of Southern California; P30AG12300 The University of Texas Southwestern Medical Center; P50AG05136 University of Washington; and P50AG05681 and P01AG03991 Washington University School of Medicine. Genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The John Hopkins University, contract number HHSN268200782096C.

We thank the contributors who collected samples used in the present study. We would also like to particularly thank the patients and their families, whose help and participation made this work possible. Susan LaRusse Eckert and Stephanie Doan (Columbia University), and Michele Goodman (Indiana University), helped coordinate the project across the United States. Creighton H. Phelps, PhD, Marcelle Morrison-Bogorod, PhD, and Marilyn Miller, PhD, at the National Institute of Aging provided guidance. We would also like to thank Jennifer Williamson-Catania at Columbia University and Stephen Snyder, PhD, at the National Institute of Aging. Samples used in this study were obtained from the National Cell Repository for Alzheimer's Disease.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Amorfix Life Sciences Ltd.; AstraZeneca; Bayer HealthCare; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129, K01 AG030514, the Dana Foundation, and U01AG032984 Alzheimer's Disease Genetics Consortium (ADGC) grant (PI: Schellenberg). Samples from the National Cell Repository for Alzheimer's Disease (NCRAD), which receives government support under a cooperative agreement (U24AG021886) awarded by the National Institute on Aging, were used in this study. Additional support for data analysis was provided by NIA R01AG19771, P30AG010133, the Indiana Economic Development Corporation (IEDC #87884), and the Foundation for the NIH. The authors thank contributors, including the ADNI sites that collected samples used in this study, as well as patients and their families, whose help and participation made this work possible. We thank the following people for their contributions to the ADNI genotyping project: (1) genotyping at the Translational Genomics Research Institute, Phoenix, AZ: Drs. Matthew Huentelman and David Craig and (2) sample processing, storage and distribution at the NCRAD: Kelley Faber and Colleen Mitchell.

REFERENCES

1. Thies W, Bleiler L. 2011 Alzheimer's disease facts and figures. *Alzheimers Dement* 7: 208-244 (2011).
2. Petersen RC, Roberts RO, Knopman DS, *et al.* Mild cognitive impairment: ten years later. *Arch Neurol* 66: 1447-1455 (2009).
3. Gatz M, Reynolds CA, Fratiglioni L, *et al.* Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry* 63: 168-174 (2006).
4. Bertram L, Lill CM, Tanzi RE. The genetics of Alzheimer disease: back to the future. *Neuron* 68: 270-281 (2010).
5. Corder EH, Saunders AM, Strittmatter WJ, *et al.* Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 261: 921-923 (1993).
6. Farrer LA, Cupples LA, Haines JL, *et al.* Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *JAMA* 278: 1349-1356 (1997).
7. Saunders AM, Strittmatter WJ, Schmechel D, *et al.* Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology* 43: 1467-1472 (1993).
8. Harold D, Abraham R, Hollingworth P, *et al.* Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* 41: 1088-1093 (2009).
9. Lambert JC, Heath S, Even G, *et al.* Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet* 41: 1094-1099 (2009).
10. Jun G, Naj AC, Beecham GW, *et al.* Meta-analysis confirms CR1, CLU, and PICALM as Alzheimer disease risk loci and reveals interactions with APOE genotypes. *Arch Neurol* 67: 1473-1484 (2010).
11. Seshadri S, Fitzpatrick AL, Ikram MA, *et al.* Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA* 303: 1832-1840 (2010).
12. Naj AC, Beecham GW, Martin ER, *et al.* Dementia revealed: novel chromosome 6 locus for late-onset Alzheimer disease provides genetic evidence for folate-pathway abnormalities. *PLoS Genet* 6: (2010).
13. Naj AC, Jun G, Beecham GW, *et al.* Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet* 43: 436-441 (2011).
14. Hollingworth P, Harold D, Sims R, *et al.* Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet* 43: 429-435 (2011).
15. Wijsman EM, Pankratz ND, Choi Y, *et al.* Genome-wide association of familial late-onset Alzheimer's disease replicates BIN1 and CLU and nominates CUGBP2 in interaction with APOE. *PLoS Genet* 7: e1001308 (2011).
16. Cook EH, Jr., Scherer SW. Copy-number variations associated with neuropsychiatric conditions. *Nature* 455: 919-923 (2008).
17. Heinzen EL, Need AC, Hayden KM, *et al.* Genome-wide scan of copy number variation in late-onset Alzheimer's disease. *J Alzheimers Dis* 19: 69-77 (2010).
18. Brouwers N, Van Cauwenberghe C, Engelborghs S, *et al.* Alzheimer risk associated with a copy number variation in the complement receptor 1 increasing C3b/C4b binding sites. *Mol Psychiatry* (2011).
19. Shaw CA, Li Y, Wiszniewska J, *et al.* Olfactory copy number association with age at onset of Alzheimer disease. *Neurology* 76: 1302-1309 (2011).
20. Swaminathan S, Kim S, Shen L, *et al.* Genomic Copy Number Analysis in Alzheimer's Disease and Mild Cognitive Impairment: An ADNI Study. *Int J Alzheimers Dis* 2011: 729478 (2011).
21. Saykin AJ, Shen L, Foroud TM, *et al.* Alzheimer's Disease Neuroimaging Initiative biomarkers as quantitative phenotypes: Genetics core aims, progress, and plans. *Alzheimers Dement* 6: 265-273 (2010).
22. Craddock N, Hurles ME, Cardin N, *et al.* Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* 464: 713-720 (2010).
23. Sie L, Loong S, Tan EK. Utility of lymphoblastoid cell lines. *J Neurosci Res* 87: 1953-1959 (2009).
24. 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* 34: 939-944 (1984).
25. Mirra SS, Heyman A, McKeel D, *et al.* The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology* 41: 479-486 (1991).

26. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38: 904-909 (2006).
27. Wang K, Li M, Hadley D, *et al.* PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* 17: 1665-1674 (2007).
28. Dellinger AE, Saw SM, Goh LK, Seielstad M, Young TL, Li YJ. Comparative analyses of seven algorithms for copy number variant identification from single nucleotide polymorphism arrays. *Nucleic Acids Res* 38: e105 (2010).
29. Diskin SJ, Li M, Hou C, *et al.* Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. *Nucleic Acids Res* 36: e126 (2008).
30. Need AC, Ge D, Weale ME, *et al.* A genome-wide investigation of SNPs and CNVs in schizophrenia. *PLoS Genet* 5: e1000373 (2009).
31. Purcell S, Neale B, Todd-Brown K, *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81: 559-575 (2007).
32. Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nat Genet* 39: 17-23 (2007).
33. Kent WJ, Sugnet CW, Furey TS, *et al.* The human genome browser at UCSC. *Genome Res* 12: 996-1006 (2002).
34. Buettner JA, Glusman G, Ben-Arie N, Ramos P, Lancet D, Evans GA. Organization and evolution of olfactory receptor genes on human chromosome 11. *Genomics* 53: 56-68 (1998).
35. Serby M, Larson P, Kalkstein D. The nature and course of olfactory deficits in Alzheimer's disease. *Am J Psychiatry* 148: 357-360 (1991).
36. Devanand DP, Tabert MH, Cuasay K, *et al.* Olfactory identification deficits and MCI in a multi-ethnic elderly community sample. *Neurobiol Aging* 31: 1593-1600 (2010).
37. Bruno C, Cassandrini D, Assereto S, Akman HO, Minetti C, Di Mauro S. Neuromuscular forms of glycogen branching enzyme deficiency. *Acta Myol* 26: 75-78 (2007).
38. Riley B, Williamson M, Collier D, Wilkie H, Makoff A. A 3-Mb map of a large Segmental duplication overlapping the alpha7-nicotinic acetylcholine receptor gene (CHRNA7) at human 15q13-q14. *Genomics* 79: 197-209 (2002).
39. Gault J, Robinson M, Berger R, *et al.* Genomic organization and partial duplication of the human alpha7 neuronal nicotinic acetylcholine receptor gene (CHRNA7). *Genomics* 52: 173-185 (1998).
40. Sinkus ML, Lee MJ, Gault J, *et al.* A 2-base pair deletion polymorphism in the partial duplication of the alpha7 nicotinic acetylcholine gene (CHRFAM7A) on chromosome 15q14 is associated with schizophrenia. *Brain Res* 1291: 1-11 (2009).
41. Feher A, Juhasz A, Rimanoczy A, Csibri E, Kalman J, Janka Z. Association between a genetic variant of the alpha-7 nicotinic acetylcholine receptor subunit and four types of dementia. *Dement Geriatr Cogn Disord* 28: 56-62 (2009).
42. de Lucas-Cerrillo AM, Maldifassi MC, Arnalich F, *et al.* Function of partially duplicated human alpha77 nicotinic receptor subunit CHRFAM7A gene: potential implications for the cholinergic anti-inflammatory response. *J Biol Chem* 286: 594-606 (2011).
43. Araud T, Graw S, Berger R, *et al.* The chimeric gene CHRFAM7A, a partial duplication of the CHRNA7 gene, is a dominant negative regulator of alpha7*nAChR function. *Biochem Pharmacol* 82: 904-914 (2011).
44. Nie HZ, Shi S, Lukas RJ, Zhao WJ, Sun YN, Yin M. Activation of alpha7 nicotinic receptor affects APP processing by regulating secretase activity in SH-EP1-alpha7 nAChR-hAPP695 cells. *Brain Res* 1356: 112-120 (2010).
45. Zhang C, Browne A, Child D, Divito JR, Stevenson JA, Tanzi RE. Loss of function of ATXN1 increases amyloid beta-protein levels by potentiating beta-secretase processing of beta-amyloid precursor protein. *J Biol Chem* 285: 8515-8526 (2010).
46. Tong X, Gui H, Jin F, *et al.* Ataxin-1 and Brother of ataxin-1 are components of the Notch signalling pathway. *EMBO Rep* 12: 428-435 (2011).
47. Bettens K, Brouwers N, Van Miegroet H, *et al.* Follow-up study of susceptibility loci for Alzheimer's disease and onset age identified by genome-wide association. *J Alzheimers Dis* 19: 1169-1175 (2010).
48. Ma SL, Tang NL, Tam CW, *et al.* Association between HLA-A alleles and Alzheimer's disease in a southern Chinese community. *Dement Geriatr Cogn Disord* 26: 391-397 (2008).

49. Lehmann DJ, Barnardo MC, Fuggle S, *et al.* Replication of the association of HLA-B7 with Alzheimer's disease: a role for homozygosity? *J Neuroinflammation* 3: 33 (2006).
50. Seripa D, Matera MG, Franceschi M, *et al.* The RELN locus in Alzheimer's disease. *J Alzheimers Dis* 14: 335-344 (2008).
51. Kramer PL, Xu H, Woltjer RL, *et al.* Alzheimer disease pathology in cognitively healthy elderly: A genome-wide study. *Neurobiol Aging* (2010).
52. Rachidi M, Delezoide AL, Delabar JM, Lopes C. A quantitative assessment of gene expression (QAGE) reveals differential overexpression of DOPEY2, a candidate gene for mental retardation, in Down syndrome brain regions. *Int J Dev Neurosci* 27: 393-398 (2009).
53. Pinhel MA, Nakazone MA, Cacao JC, *et al.* Glutathione S-transferase variants increase susceptibility for late-onset Alzheimer's disease: association study and relationship with apolipoprotein E epsilon4 allele. *Clin Chem Lab Med* 46: 439-445 (2008).
54. Spalletta G, Bernardini S, Bellincampi L, *et al.* Glutathione S-transferase P1 and T1 gene polymorphisms predict longitudinal course and age at onset of Alzheimer disease. *Am J Geriatr Psychiatry* 15: 879-887 (2007).
55. Burri L, Strahm Y, Hawkins CJ, *et al.* Mature DIABLO/Smac is produced by the IMP protease complex on the mitochondrial inner membrane. *Mol Biol Cell* 16: 2926-2933 (2005).
56. Petek E, Windpassinger C, Vincent JB, *et al.* Disruption of a novel gene (IMMP2L) by a breakpoint in 7q31 associated with Tourette syndrome. *Am J Hum Genet* 68: 848-858 (2001).
57. Patel C, Cooper-Charles L, McMullan DJ, Walker JM, Davison V, Morton J. Translocation breakpoint at 7q31 associated with tics: further evidence for IMMP2L as a candidate gene for Tourette syndrome. *Eur J Hum Genet* 19: 634-639 (2011).
58. Lu B, Poirier C, Gaspar T, *et al.* A mutation in the inner mitochondrial membrane peptidase 2-like gene (Immp2l) affects mitochondrial function and impairs fertility in mice. *Biol Reprod* 78: 601-610 (2008).
59. George SK, Jiao Y, Bishop CE, Lu B. Mitochondrial peptidase IMMP2L mutation causes early onset of age-associated disorders and impairs adult stem cell self-renewal. *Aging Cell* 10: 584-594 (2011).
60. Reissner C, Klose M, Fairless R, Missler M. Mutational analysis of the neurexin/neuroigin complex reveals essential and regulatory components. *Proc Natl Acad Sci U S A* 105: 15124-15129 (2008).
61. Gomez Ravetti M, Rosso OA, Berretta R, Moscato P. Uncovering molecular biomarkers that correlate cognitive decline with the changes of hippocampus' gene expression profiles in Alzheimer's disease. *PLoS One* 5: e10153 (2010).
62. Szatmari P, Paterson AD, Zwaigenbaum L, *et al.* Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet* 39: 319-328 (2007).
63. Glessner JT, Wang K, Cai G, *et al.* Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* 459: 569-573 (2009).
64. Rujescu D, Ingason A, Cichon S, *et al.* Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum Mol Genet* 18: 988-996 (2009).
65. Srinivasan R, Poulosom R, Hurst HC, Gullick WJ. Expression of the c-erbB-4/HER4 protein and mRNA in normal human fetal and adult tissues and in a survey of nine solid tumour types. *J Pathol* 185: 236-245 (1998).
66. Chaudhury AR, Gerecke KM, Wyss JM, Morgan DG, Gordon MN, Carroll SL. Neuregulin-1 and erbB4 immunoreactivity is associated with neuritic plaques in Alzheimer disease brain and in a transgenic model of Alzheimer disease. *J Neuropathol Exp Neurol* 62: 42-54 (2003).
67. Woo RS, Lee JH, Yu HN, Song DY, Baik TK. Expression of ErbB4 in the apoptotic neurons of Alzheimer's disease brain. *Anat Cell Biol* 43: 332-339 (2010).
68. Woo RS, Lee JH, Yu HN, Song DY, Baik TK. Expression of ErbB4 in the neurons of Alzheimer's disease brain and APP/PS1 mice, a model of Alzheimer's disease. *Anat Cell Biol* 44: 116-127 (2011).
69. Mefford HC, Cooper GM, Zerr T, *et al.* A method for rapid, targeted CNV genotyping identifies rare variants associated with neurocognitive disease. *Genome Res* 19: 1579-1585 (2009).
70. Williams NM, Zaharieva I, Martin A, *et al.* Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. *Lancet* 376: 1401-1408 (2010).
71. Ingason A, Rujescu D, Cichon S, *et al.* Copy number variations of chromosome 16p13.1 region associated with schizophrenia. *Mol Psychiatry* 16: 17-25 (2011).
72. Kirov G, Grozeva D, Norton N, *et al.* Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. *Hum Mol Genet* 18: 1497-1503 (2009).

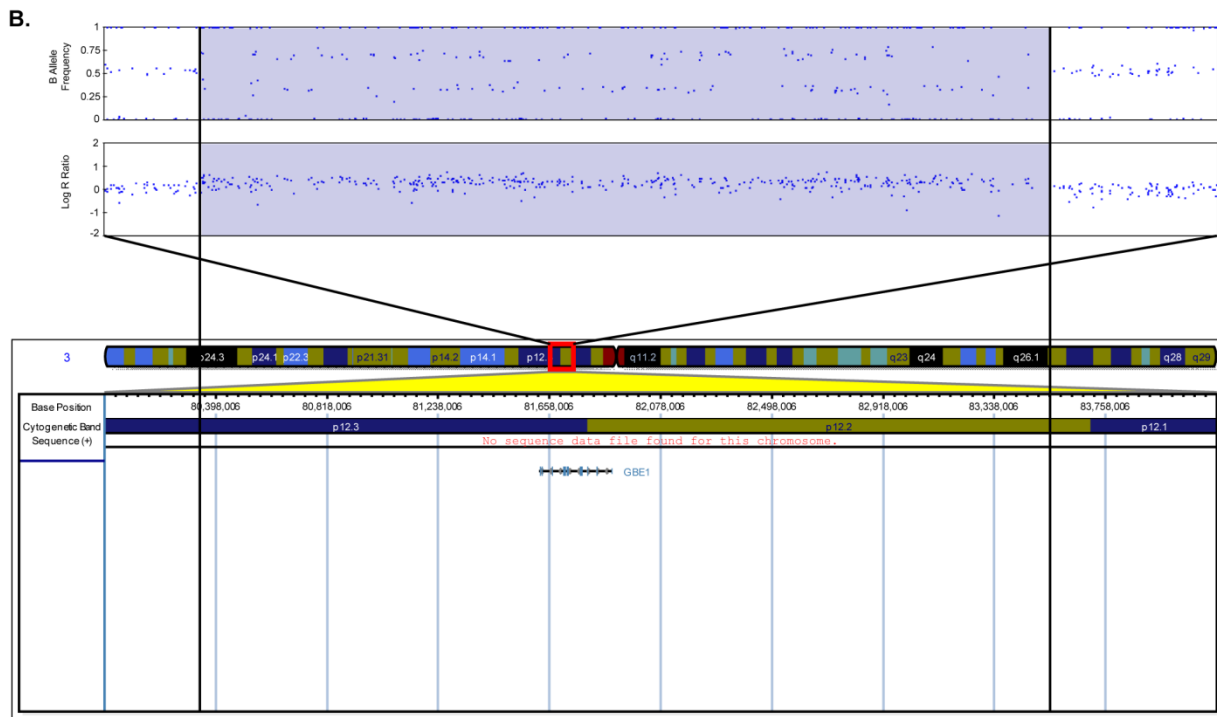
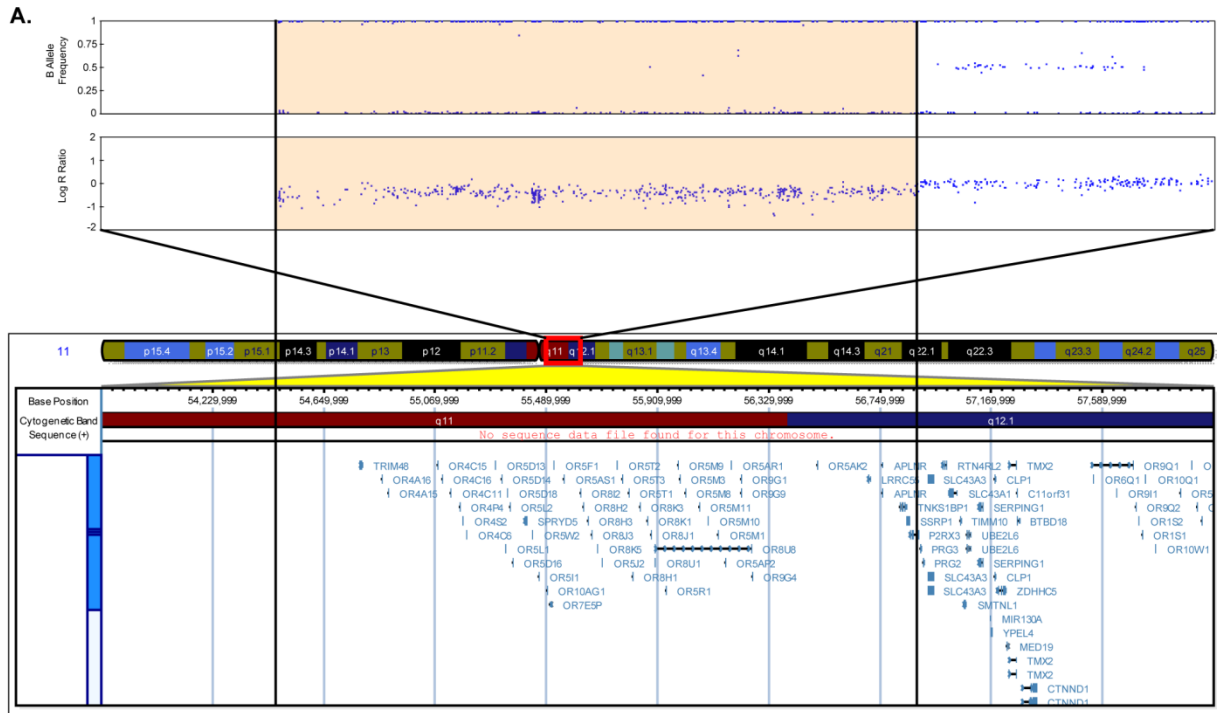
73. Magri C, Sacchetti E, Traversa M, *et al.* New copy number variations in schizophrenia. PLoS One 5: e13422 (2010).
74. Awadalla P, Gauthier J, Myers RA, *et al.* Direct measure of the de novo mutation rate in autism and schizophrenia cohorts. Am J Hum Genet 87: 316-324 (2010).

FIGURE LEGENDS

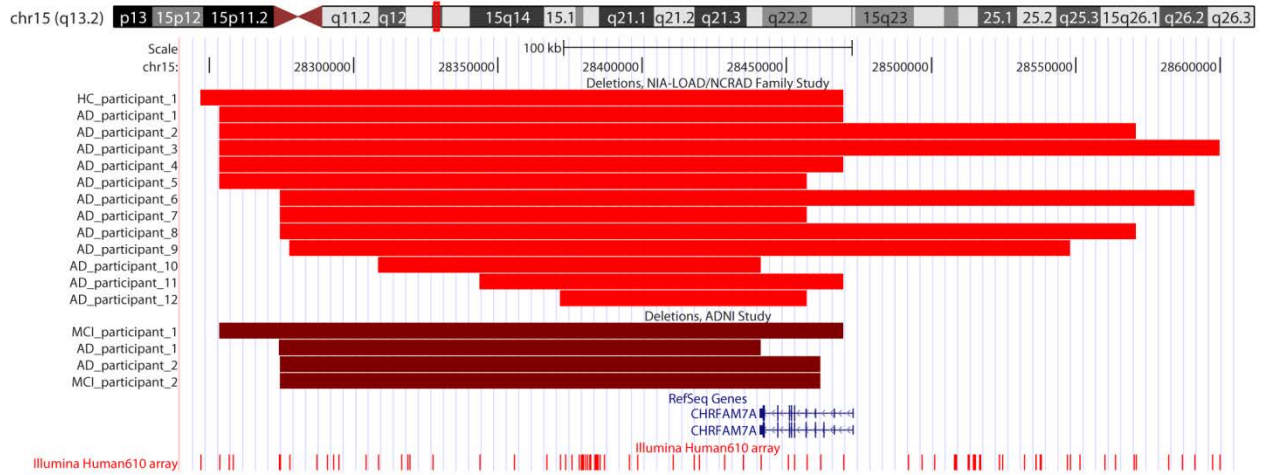
Fig. (1). Representative images of B Allele Frequency and Log R Ratio of the Alzheimer's disease participants who had a deletion >2 Mb on chromosome 11 (A) and a duplication >2 Mb on chromosome 3 (B) including chromosomal coordinates and the genes in the corresponding regions. Each blue dot in the B Allele Frequency and Log R Ratio plots represents a SNP. The orange shaded portion (A) indicates the region with the deletion and the blue shaded portion (B) indicates the region with duplication (Human Genome Build 36.1).

Fig. (2). Representative UCSC Genome Browser (March 2006 (NCBI36/hg18) assembly) plots of deletions overlapping the *CHRFAM7A* (A) and *IMMP2L* (B) genes in participants from the NIA-LOAD/NCRAD Family and ADNI studies. The chromosomal location of the gene and the probes on the Illumina Human610 array are shown. The region with the deletion for each participant relative to the gene is highlighted by a lighter red shade for the NIA-LOAD/NCRAD Family Study participants and a darker red shade for the ADNI participants.

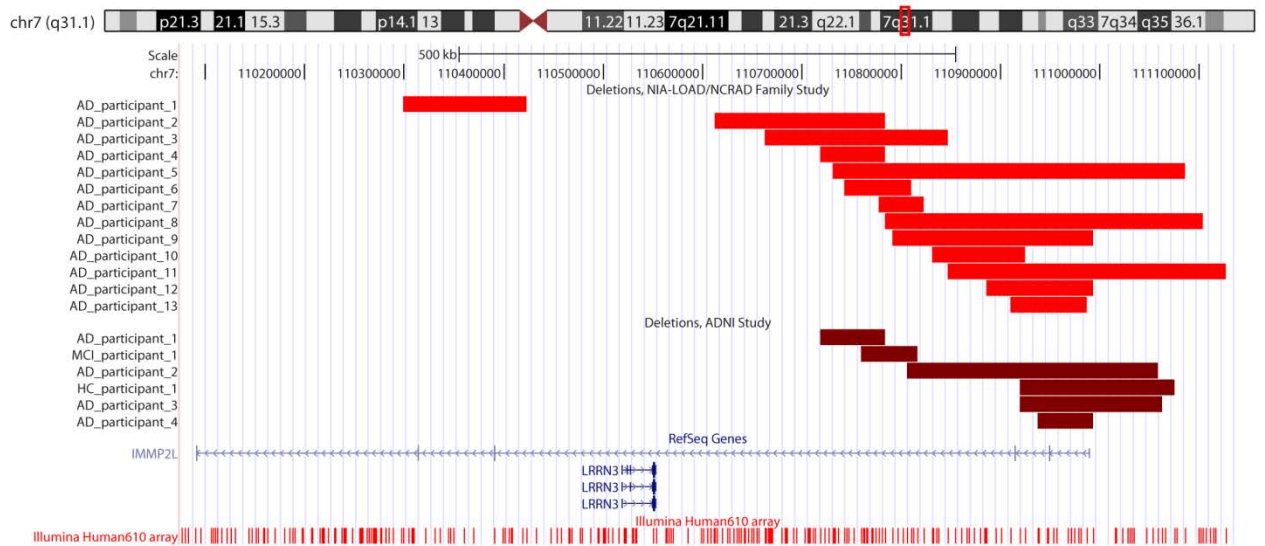
Fig. (3). Representative images of B Allele Frequency and Log R Ratio of the Alzheimer's disease participants who had a duplication at 16p13.11 (A) and a deletion at 17p12 (B) along with chromosomal coordinates and genes in corresponding regions are shown. Blue dots in the B Allele Frequency and Log R Ratio plots represent SNPs. The blue shaded portion (A) indicates the duplicated region and the orange shaded portion (B) the deleted region (Human Genome Build 36.1).



A.



B.



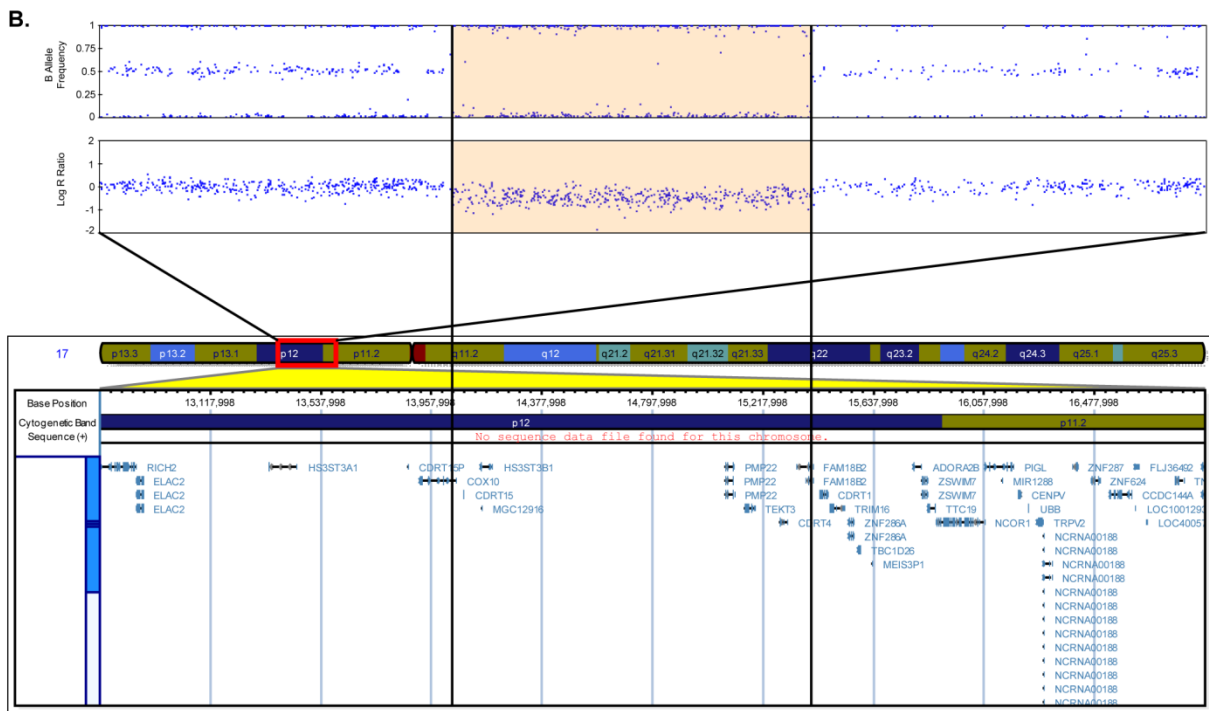
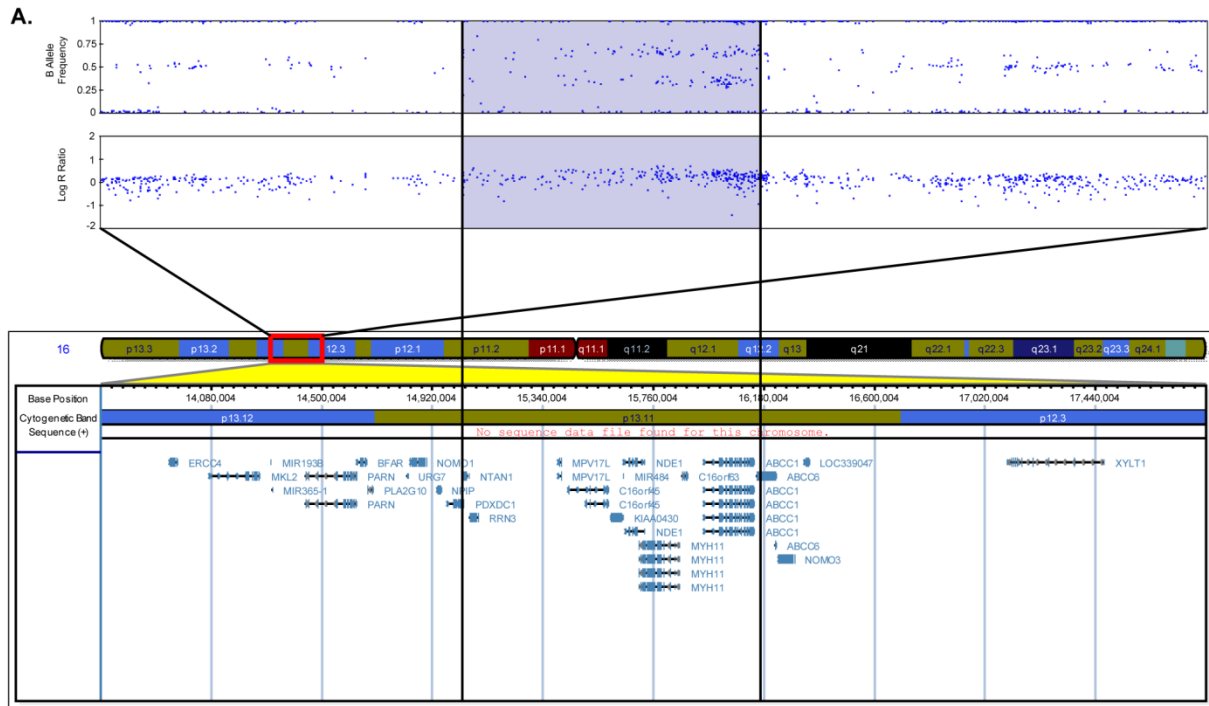


Table 1. Sample Demographics.

	Alzheimer's disease	Controls	<i>p</i> (two-sided)
Number of participants	711	171	-
Gender (Males/Females)	243/468 (n=711)	72/99 (n=171)	0.052
Years of education (Mean±SD)	13.20±3.02 (n=364)	15.16±3.07 (n=160)	<0.001
<i>APOE</i> group (ϵ 4 negative/ ϵ 4 positive)	174/537 (n=711)	136/35 (n=171)	<0.001
Age at last evaluation (Mean±SD)	-	79.33±11.20 (n=171)	-
Age participant developed dementia symptoms (Mean±SD)	72.02±6.77 (n=705)	-	-
Age participant diagnosed with Alzheimer's disease (Mean±SD)	75.85±6.88 (n=586)	-	-
DNA Source (Blood/Brain Tissue)	673/38 (n=711)	136/35 (n=171)	<0.001

Table 2. Characteristics of CNV Calls.

		Alzheimer's disease (n=711)	Controls (n=171)
Deletions	Number of CNVs	4453	1133
	Rate per participant	6.26	6.63
	Average size (kb)	65.94	61.92
Duplications	Number of CNVs	2074	551
	Rate per participant	2.92	3.22
	Average size (kb)	133.1	124.4

Table 3. Participants Grouped by CNV Call Size.

Call size	Alzheimer's disease (n=711)		Controls (n=171)	
	Deletions n (%)	Duplications n (%)	Deletions n (%)	Duplications n (%)
0.1-0.5 Mb	440 (61.88)	477 (67.09)	102 (59.65)	116 (67.84)
0.5-1.0 Mb	12 (1.69)	39 (5.49)	4 (2.34)	12 (7.02)
1.0-1.5 Mb	2 (0.28)	10 (1.41)	1 (0.58)	0 (0.00)
1.5-2.0 Mb	3 (0.42)	4 (0.56)	0 (0.00)	0 (0.00)
>2.0 Mb	1 (0.14)	1 (0.14)	0 (0.00)	0 (0.00)

Table 4. Genes Overlapped by CNV Calls from at Least One Alzheimer's Disease Participant and No Controls Using the Candidate Gene Approach.

Chromosome	Region	Start (bp)	End (bp)	Number of Alzheimer's disease participants
1	<i>CFH</i>	194887630	194983257	1
6	<i>ATXN1^a</i>	16407321	16869700	3
6	<i>HLA-A</i>	30018309	30021633	2
6	<i>MICA</i>	31479349	31491069	4
6	<i>HLA-DQA1</i>	32713160	32719407	5 ^b
6	<i>HLA-DOA</i>	33079937	33085367	1 ^c
6	<i>HLA-DPBI^a</i>	33151737	33162954	1 ^c
7	<i>CD36</i>	80069439	80146529	1
7	<i>RELN^a</i>	102899472	103417198	1
9	<i>APBA1</i>	71235021	71477042	4
9	<i>ABCA1</i>	106583104	106730257	3
9	<i>RXRA</i>	136358230	136472252	1
10	<i>ABCC2</i>	101532452	101601652	1
11	<i>PICALM</i>	85346132	85457756	1
15	<i>CYP19A1</i>	49287545	49418087	1
15	<i>CHRNA3</i>	76674705	76700377	1
15	<i>MEF2A</i>	97956184	98071524	1
17	<i>TP53</i>	7512444	7531588	1
17	<i>COX10</i>	13913443	14052721	1
17	<i>SREBF1</i>	17656110	17681050	3 ^b
17	<i>CCL3</i>	31439715	31441619	2

17	<i>KIF18B</i>	40358973	40380608	1
18	<i>DSCI</i>	26963211	26996817	1
21	<i>NCAM2</i>	21292503	21833085	1
21	<i>APP</i>	26174731	26465003	1
21	<i>DOPEY2^a</i>	36458708	36588442	2
21	<i>KCNJ6</i>	37918656	38210566	2
22	<i>COMT</i>	18309308	18336530	2
22	<i>BCR</i>	21852551	21990224	1
22	<i>GSTT1^a</i>	22706138	22714284	1

^aGenes also identified in the Alzheimer's Disease Neuroimaging Initiative Study as being overlapped by CNV calls in Alzheimer's disease and/or mild cognitive impairment participants, but not controls.

^bOne AD participant had CNV calls overlapping the two genes.

^cA different AD participant had CNV calls overlapping the two genes.