

# UC Irvine

## UC Irvine Previously Published Works

### Title

Genomic Copy Number Analysis in Alzheimer's Disease and Mild Cognitive Impairment:  
An ADNI Study

### Permalink

<https://escholarship.org/uc/item/7vd866r8>

### Journal

International Journal of Alzheimer's Disease, 2011(1)

### ISSN

2090-8024

### Authors

Swaminathan, Shanker

Kim, Sungeun

Shen, Li

et al.

### Publication Date

2011

### DOI

10.4061/2011/729478

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License,  
available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

## Research Article

# Genomic Copy Number Analysis in Alzheimer's Disease and Mild Cognitive Impairment: An ADNI Study

**Shanker Swaminathan,<sup>1,2</sup> Sungeun Kim,<sup>1,3</sup> Li Shen,<sup>1,3</sup> Shannon L. Risacher,<sup>1</sup> Tatiana Foroud,<sup>1,2,3</sup> Nathan Pankratz,<sup>2</sup> Steven G. Potkin,<sup>4</sup> Matthew J. Huentelman,<sup>5</sup> David W. Craig,<sup>5</sup> Michael W. Weiner,<sup>6,7</sup> Andrew J. Saykin,<sup>1,2,3</sup> and The Alzheimer's Disease Neuroimaging Initiative (ADNI)**

<sup>1</sup> Center for Neuroimaging, Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN 46202, USA

<sup>2</sup> Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN 46202, USA

<sup>3</sup> Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN 46202, USA

<sup>4</sup> Department of Psychiatry and Human Behavior, University of California, Irvine, CA 92697, USA

<sup>5</sup> Neurogenomics Division, The Translational Genomics Research Institute, 445 N. Fifth Street, Phoenix, AZ 85004, USA

<sup>6</sup> Departments of Radiology, Medicine and Psychiatry, University of California, San Francisco, CA 94143, USA

<sup>7</sup> Department of Veterans Affairs Medical Center, San Francisco, CA 94121, USA

Correspondence should be addressed to Andrew J. Saykin, [asaykin@iupui.edu](mailto:asaykin@iupui.edu)

Received 5 October 2010; Revised 22 December 2010; Accepted 27 January 2011

Academic Editor: Benedetta Nacmias

Copyright © 2011 Shanker Swaminathan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copy number variants (CNVs) are DNA sequence alterations, resulting in gains (duplications) and losses (deletions) of genomic segments. They often overlap genes and may play important roles in disease. Only one published study has examined CNVs in late-onset Alzheimer's disease (AD), and none have examined mild cognitive impairment (MCI). CNV calls were generated in 288 AD, 183 MCI, and 184 healthy control (HC) non-Hispanic Caucasian Alzheimer's Disease Neuroimaging Initiative participants. After quality control, 222 AD, 136 MCI, and 143 HC participants were entered into case/control association analyses, including candidate gene and whole genome approaches. Although no excess CNV burden was observed in cases (AD and/or MCI) relative to controls (HC), gene-based analyses revealed CNVs overlapping the candidate gene *CHRFAM7A*, as well as *CSMD1*, *SLC35F2*, *HNRNPCL1*, *NRXN1*, and *ERBB4* regions, only in cases. Replication in larger samples is important, after which regions detected here may be promising targets for resequencing.

## 1. Introduction

Alzheimer's disease (AD) is the most common cause of dementia and accounts for 50–80% of dementia cases. Currently, an estimated 5.3 million Americans have AD, the seventh leading cause of death in the United States. The hallmark abnormalities of AD are deposits of the protein fragment amyloid  $\beta$  (plaques) and twisted strands of the protein tau (tangles). Amnesic mild cognitive impairment (MCI) is a clinical condition in which a person has problems with memory, with or without other cognitive deficits, that are noticeable to others and show up on psychometric testing

but are not severe enough to interfere significantly with daily functioning. About 14–18% of individuals aged 70 years and older have MCI, and these individuals are likely to progress to dementia, particularly AD, with an annual conversion rate of 10–15% [1].

Genetic factors play a key role in the development and progression of AD. AD has a high heritability, with 58–79% of phenotypic variation estimated to be caused by genetic factors [2]. Early-onset AD (onset < 65 years) accounts for a small percentage (<5%) of cases and is primarily caused by mutations in three genes that affect the cerebral levels

of amyloid  $\beta$  peptide: *A $\beta$ PP* (amyloid- $\beta$  protein precursor) on chromosome 21, *PSEN1* (presenilin 1) on chromosome 14, and *PSEN2* (presenilin 2) on chromosome 1 [3]. Late-onset AD (LOAD) accounts for the majority of AD cases, but only the  $\epsilon 4$  allele of the *APOE* (apolipoprotein E) gene on chromosome 19 has been consistently replicated across studies. Recently, three large genome-wide association studies (GWASs) identified five additional loci: *CLU* (clusterin), *CRI* (complement component (3b/4b) receptor 1), *PICALM* (phosphatidylinositol binding clathrin assembly protein), *BINI* (bridging integrator 1), and *EXOC3L2* (exocyst complex component 3-like 2) to be strongly associated with AD [4–6]. These loci also showed strong association in replication studies [7], further supporting a role in AD susceptibility.

Copy number variants (CNVs) are segments of DNA, ranging from 1 kilobase (kb) to several megabases (Mb), for which differences in the number of copies have been revealed by comparison of two or more genomes. These differences can be copy number gains (duplications or insertional transpositions), losses (deletions), gains or losses of the same locus, or multiallelic or complex rearrangements. CNVs have been implicated in various neuropsychiatric disorders such as autism and schizophrenia [8]. To date, the role of CNVs in LOAD has only been examined in one study [9]. These authors performed a genome-wide scan of AD in 331 dementia cases (in which >80% of patients had a clinical diagnosis of AD) and 368 controls. Although no CNVs, which are typically rare occurrences, were significant at genome-wide threshold, Heinzen et al. were able to identify a duplication in the *CHRNA7* gene warranting further investigation. To date, no study has looked at the role of CNVs in MCI.

In the present report, we conducted a preliminary CNV analysis using genotype data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort to examine the role of CNVs in susceptibility to MCI and LOAD. ADNI is an ongoing multiyear public-private partnership to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), genetic factors such as single nucleotide polymorphisms (SNPs) and CNVs, other biological markers, and clinical and neuropsychological assessments can be combined to improve early diagnosis and predict progression of MCI and early AD. Here, we used the genome-wide array data acquired on the ADNI cohort to determine whether AD and MCI participants (cases) showed an excess burden of CNVs relative to controls and to characterize any genomic regions where CNVs were detected in cases but not controls.

## 2. Materials and Methods

**2.1. Alzheimer's Disease Neuroimaging Initiative.** The ADNI was launched in 2004 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies, and nonprofit organizations, as a \$60 million, multiyear public-private partnership. The Principal Investigator of this initiative is

Michael W. Weiner, M.D., VA Medical Center and University of California—San Francisco. ADNI is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations. Presently, more than 800 participants, aged 55 to 90, have been recruited from over 50 sites across the US and Canada, including approximately 200 cognitively normal older individuals (i.e., healthy controls or HCs) to be followed for 3 years, 400 patients diagnosed with MCI to be followed for 3 years, and 200 patients diagnosed with early AD to be followed for 2 years [10]. Longitudinal imaging, including structural 1.5T MRI scans collected on the full sample and [ $^{11}\text{C}$ ]PiB- and [ $^{18}\text{F}$ ]FDG-PET imaging on a subset, and performance on neuropsychological and clinical assessments were collected at baseline and at followup visits in 6–12-month intervals. Other biomarkers are also available including *APOE* and whole genome genotyping on the full ADNI sample, and longitudinal cerebrospinal fluid (CSF) markers on a subset of the sample. Written informed consent was obtained from all participants, and the study was conducted with prior institutional review boards approval. Further information about ADNI can be found in [11] and at <http://www.adni-info.org>.

**2.2. Participants.** Participants in the present report included 655 non-Hispanic Caucasian individuals from the ADNI cohort who had DNA samples extracted from peripheral blood. Those with DNA samples derived from cell lines were excluded from the present analysis because cell line transformation might influence CNV results [12, 13]. Current diagnoses were downloaded from the ADNI database as of 04/29/2010 (AD = 288, MCI = 183, HC = 184). In addition to AD participants who had a baseline and current diagnosis of AD, we included MCI participants who had converted from a baseline diagnosis of MCI to a current diagnosis of AD (MCI Converters) as well as one participant who had converted from a baseline diagnosis of HC to a current diagnosis of AD in the AD group. Similarly, in addition to MCI participants who had a baseline and current diagnosis of MCI, we included seven HC participants who had converted from a baseline diagnosis of HC to a current diagnosis of MCI in the MCI group. Data used in preparation of this report is publicly available on the ADNI web site (<http://loni.ucla.edu/ADNI/>). The focus of ADNI is on incident LOAD. To our knowledge, no participants in the present study carry a known causal mutation [14].

**2.3. Genotyping.** Blood samples from each participant were obtained and sent to Pfizer for DNA extraction and were also banked at The National Cell Repository for Alzheimer's Disease (NCRAD; <http://ncrad.iu.edu/>). Genotyping was performed by the Translational Genomics Research Institute (TGen) (Phoenix, AZ) using the Illumina Human610-Quad BeadChip as previously described [15, 16]. As indicated by the manufacturer's documentation, the Human610-Quad BeadChip contains 620,901 markers. This array provides

dense genomic coverage (89%) in the CEU (Utah residents with Northern and Western European ancestry from the Centre d'Etude du Polymorphisme Humain (CEPH) collection) population analyzed here with a median marker spacing of 2.7 kb. In addition, 27,635 markers are included in "unSNPable" regions likely to contain CNVs that are not easily assessed by SNPs. Coverage is provided for 3,938 CNV regions (184,064 markers) reported in the Toronto Database of Genomic Variants (<http://projects.tcag.ca/variation/>) at an average of 37.7 markers per region. Markers have an average of 15–18-fold redundancy to improve signal quality for detection of CNVs (mean Log R Ratio SD < 0.2, see below).

Normalized bead intensity data for each sample was loaded into GenomeStudioV2009.1 software (Illumina, Inc., CA) along with the manufacturer's cluster file to generate SNP genotypes. The Log R Ratio (LRR) and B Allele Frequency (BAF) values computed from the signal intensity files by GenomeStudio for each sample were exported and used for the generation of CNV calls. Initial genotyping was performed by TGen using BeadStudio software (Illumina, Inc., CA). In January 2010, we reprocessed the array data using GenomeStudioV2009.1, and this data set will be made available on the ADNI website in a followup data release.

**2.4. Inference of LRR and BAF.** The two alleles of an SNP are designated as allele A and allele B. GenomeStudio software uses a five-step six-degree of freedom affine transformation to normalize signal intensity values of the A and B alleles (referred to as X and Y). The normalized values are then transformed to a polar coordinate plot of normalized intensity  $R = X_{\text{norm}} + Y_{\text{norm}}$  and composition (copy angle)  $\theta = (2/\pi) * \arctan(Y_{\text{norm}}/X_{\text{norm}})$ , where  $X_{\text{norm}}$  and  $Y_{\text{norm}}$  represent transformed normalized signals from alleles A and B for a particular locus (Illumina's genotyping data normalization methods white paper). The LRR value for a sample is calculated as follows:  
 $LRR = \log_2(\text{normalized } R \text{ value/expected } R \text{ value})$  for the SNP.

Linear interpolation of the R value at the SNP's  $\theta$  value for a sample, relative to the R values of the surrounding clusters, is used to compute the expected R value.

The BAF for a sample shows the  $\theta$  value for an SNP, corrected for cluster positions, which were generated from a large set of previously studied normal individuals. BAF is described by the following equation:

$$\begin{aligned} \text{BAF} &= 0 \quad \text{if } \theta < \theta_{AA} \\ &= 0.5 * \frac{(\theta - \theta_{AA})}{(\theta_{AB} - \theta_{AA})} \quad \text{if } \theta < \theta_{AB} \\ &= 0.5 + 0.5 * \frac{(\theta - \theta_{AB})}{(\theta_{BB} - \theta_{AB})} \quad \text{if } \theta < \theta_{BB} \\ &= 1 \quad \text{if } \theta \geq \theta_{BB}, \end{aligned} \quad (1)$$

where  $\theta_{AA}$  = mean  $\theta$  value of all genotypes in AA cluster plotted in polar normalized coordinates,  $\theta_{AB}$  = mean  $\theta$  value of all genotypes in AB cluster plotted in polar normalized

coordinates, and  $\theta_{BB}$  = mean  $\theta$  value of all genotypes in BB cluster plotted in polar normalized coordinates (GenomeStudio Genotyping Module v1.0 User Guide).

**2.5. Generation of CNV Calls and Quality Control.** CNV calls were generated for the 655 non-Hispanic Caucasian participants whose DNA was derived from peripheral blood. PennCNV software (2009Aug27 version) (<http://www.openbioinformatics.org/penncnv/>), which implements a hidden Markov model (HMM) [17], was used to generate the CNV calls. The hg18 "all" PennCNV hidden Markov model (hmm), population frequency of B allele (pfb), and gcmodel files were used to ensure that CNV-specific markers were included. All samples were subjected to extensive quality control (QC). Since samples that have below optimal genomic wave QC values can be considered unreliable [18], we applied the GC-model wave adjustment procedure, using PennCNV's gcmodel file. A frequency distribution plot of the number of CNV calls for all samples was made, and samples were excluded if the number of CNV calls made for that individual was greater than the 90th percentile of the frequency distribution. One sample was observed to have multiple deletions and duplications on chromosome 18 (Supplementary Figure 1 in Supplementary Materials available Online at doi:10.4061/2011/729478) and was excluded from further analysis as it may be a mosaic sample [19]. Samples were also excluded if they met the following criteria: LRR SD > 0.35, BAF Drift > 0.002, or Waviness Factor (WF) > 0.04. The LRR SD is a measure of signal-to-noise ratio. Sometimes, when a sample has genotyping failure, many SNP markers will have abnormal BAF patterns (i.e., they do not cluster to 0, 0.5, or 1), yet their LRR looks normal. The BAF Drift takes into account these abnormal BAF patterns. The WF measures the waviness of the signal curves, as artificial gains and losses in the genome can be created by peaks and troughs of the wave.

Analyses were also restricted to autosomes due to the complications of hemizyosity in males and X-chromosome inactivation in females. Finally, to ensure only high-confidence CNVs were included in the analysis, CNVs for which the difference of the log likelihood of the most likely copy number state and less likely copy number state was less than 10 (generated using the confidence function in PennCNV), CNVs that were called based on data from fewer than 10 SNPs, and CNVs that had >50% overlap with centromeric, telomeric, and immunoglobulin regions as defined in Need et al. [20] were excluded. 501 participants (AD = 222, MCI = 136, HC = 143) passed all QC checks and were included in further CNV analyses.

**2.6. Case/Control Association Analyses.** Case/control association analyses using CNV calls generated for the AD, MCI, and HC participants were performed using PLINK v1.07 [21] (<http://pngu.mgh.harvard.edu/purcell/plink/>) to investigate any differences in CNV calls between cases and controls (AD versus HC; MCI versus HC). Two approaches were used: (1) a candidate gene approach using AD genes, identified from the AlzGene database [22] (<http://www.alzgene.org/>)

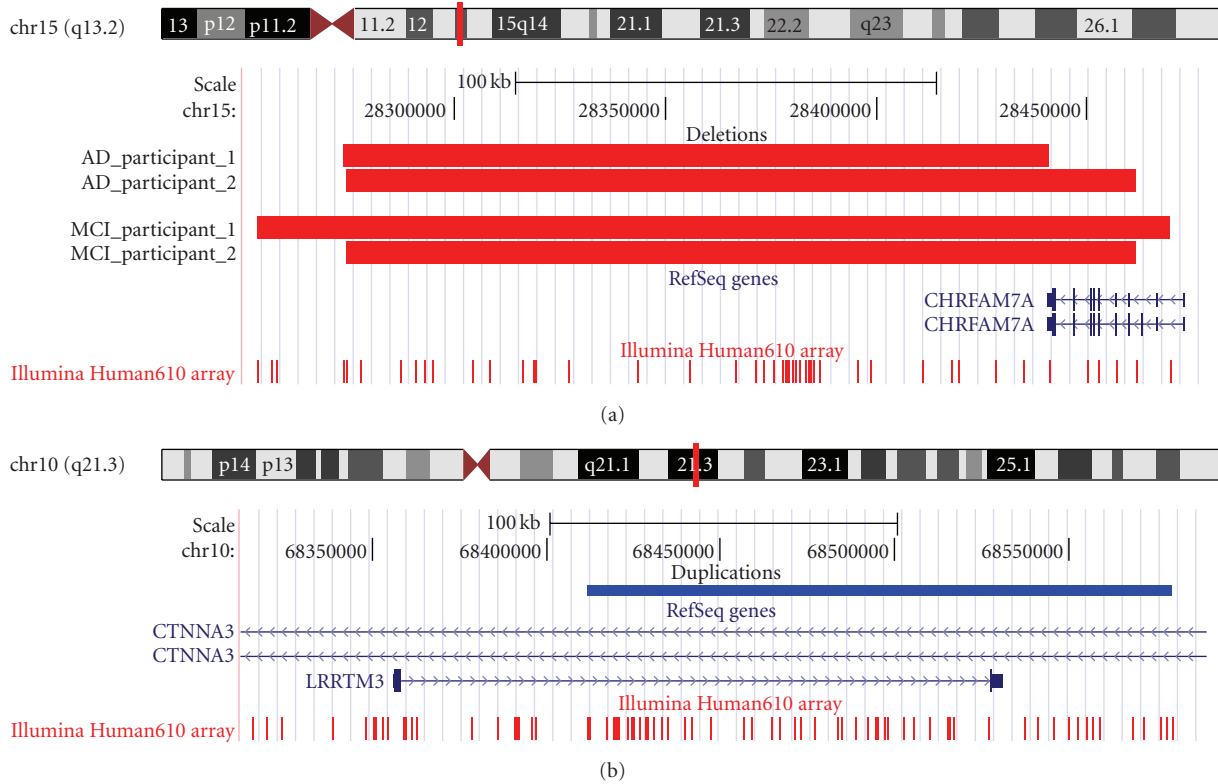


FIGURE 1: Examples of candidate genes (a) *CHRFAM7A* and (b) *LRRTM3*, overlapped by CNV calls from at least one Alzheimer's disease and/or one mild cognitive impairment participant, but no healthy controls. The red rectangles represent deletions, and the blue rectangles represent duplications (March 2006 (NCBI36/hg18) assembly).

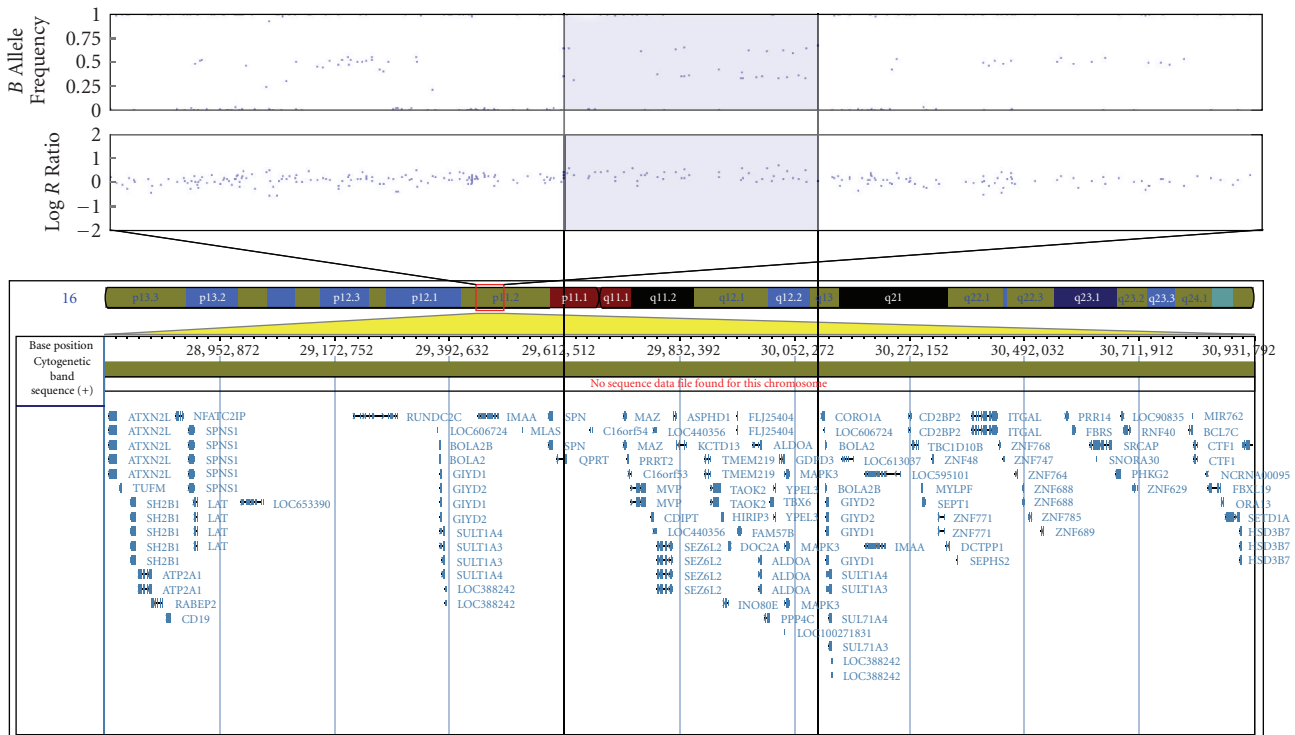


FIGURE 2: Representative image of B Allele Frequency and Log R Ratio of the participant who had a duplication at 16p11.2. The purple shaded portion indicates the duplicated region (Human Genome Build 36.1).

as having a positive association with AD in at least one study, consisting of 294 genes as of 04/22/2010, and (2) a whole genome approach using PLINK's entire gene list (hg18 coordinates), consisting of 17,938 genes. The AlzGene database provides a comprehensive and regularly updated synopsis of genetic studies in AD. In both approaches, CNV segments either partially or completely overlapping gene regions were analyzed. Both deletions and duplications were included in the analyses.

Representative plots of CNV calls (Figure 1) were generated in UCSC Genome Browser (<http://genome.ucsc.edu/>) [23] (March 2006 (NCBI36/hg18) assembly). Plots were produced using the genome browser track for the Illumina Human-610 array obtained from the PennCNV website. Representative plots of LRR and BAF values for samples (Figure 2, Supplementary Figures 1 to 3) were generated using the Illumina Genome Viewer plugin within GenomeStudio (Human Genome Build 36.1).

### 3. Results

**3.1. Description of CNV Calls by Current Diagnostic Group.** The sample demographics and CNV call characteristics of the 501 participants who passed all QC checks are shown in Tables 1 and 2. A total of 6,737 CNV calls (4,746 deletions and 1,991 duplications) were observed in these participants. The average number of SNPs per CNV call was 25 and the average length of a CNV call was 105.93 kb. A higher CNV call rate and a lower average CNV call size were observed in deletions compared to duplications. On comparing the three diagnostic groups, AD and MCI participants appeared to have a higher CNV call rate for deletions and a lower CNV call rate for duplications, but these were not statistically significant ( $P < .05$ ) when evaluated by permutation. We also evaluated whether CNV burden was higher in cases than controls in the *APOE* e4 negative participants. There was a similar trend toward a higher CNV call rate for deletions and lower CNV call rate for duplications in AD and MCI participants, but these were not statistically significant ( $P < .05$ ; data not shown). A large proportion of deletions and duplications were found in the 0.1–0.5 Mb size range (Table 3). Two AD participants were found to have very large CNV calls (>2 Mb) (Supplementary Figures 2 and 3). One AD participant had a deletion on chromosome 4 (Supplementary Figure 2), which includes the following genes: *NDST4* (N-deacetylase/N-sulfotransferase 4), *TRAMIL1* (translocation-associated membrane protein 1-like 1), and *MIR1973* (microRNA 1973). The other AD participant had a duplication on chromosome 11 (Supplementary Figure 3), which includes the gene *LUZP2* (leucine zipper protein 2).

#### 3.2. Case/Control Association Analyses

**3.2.1. Candidate Gene Approach.** We identified regions overlapping 294 AD candidate genes with CNV calls from at least one case (AD and/or MCI) but no controls (HC). As expected, cell sizes were very small in each group

leading to low power. Resulting CNV calls along with *APOE* genotype and age at onset (for the AD at baseline group) are presented in Table 4 for reference although these did not meet conventional significance ( $P < .05$ ). A number of genes, such as *CHRFAM7A* (*CHRNA7* (cholinergic receptor, nicotinic, alpha 7, exons 5–10) and *FAM7A* (family with sequence similarity 7A, exons A–E) fusion), had CNV calls from only AD or MCI participants partially overlapping them. Figure 1 shows representative plots of two of these genes (*CHRFAM7A* and *LRRTM3*).

**3.2.2. Whole Genome Approach.** We also identified CNV calls present in cases (AD and/or MCI) but not controls (HC) within regions overlapping 17,938 genes. There was no significant ( $P < .05$ ) gene after correction for multiple testing. We, therefore, focused on genes that had an uncorrected  $P < .05$ . The genes identified included *CSMD1* (CUB and Sushi multiple domains 1), *HNRNPCL1* (heterogeneous nuclear ribonucleoprotein C-like 1), and *SLC35F2* (solute carrier family 35, member F2) (Table 5). We also observed CNVs overlapping two genes associated with neuropsychiatric disorders: *NRXN1* (neurexin 1) [24, 25] and *ERBB4* (v-erb-a erythroblastic leukemia viral oncogene homolog 4 (avian)) [26], but these did not reach significance ( $P < .05$ ). An MCI participant, who subsequently converted to clinical AD, was also observed to have a duplication comprising 23 genes in the 16p11.2 region (Figure 2).

### 4. Discussion

The present report represents an initial analysis of CNVs in the ADNI dataset and is the first CNV analysis of patients with MCI. After extensive QC, we analyzed CNV calls generated in cases (AD and MCI) compared to controls (HC), using whole genome and candidate gene association approaches.

Comparison of the CNV calls between the three diagnostic groups showed no excess CNV burden (rate of calls) in AD and MCI participants compared to controls. This is consistent with previously published results [9]. Two AD participants were found to have CNV calls >2 Mb. One AD participant had a duplication on chromosome 11 (Supplementary Figure 3) which includes the gene *LUZP2* (leucine zipper protein 2). This gene has been shown to be expressed only in the brain and spinal cord in adult mouse tissues [27]. The authors of the study also found this gene to be deleted in some patients with Wilms tumor-aniridia-genitourinary anomalies-mental retardation (WAGR) syndrome. Another AD participant had a deletion on chromosome 4 (Supplementary Figure 2), which includes the following genes: *NDST4* (N-deacetylase/N-sulfotransferase 4), *TRAMIL1* (translocation-associated membrane protein 1-like 1), and *MIR1973* (microRNA 1973). None of these genes have been previously associated with AD susceptibility. Further investigation by either cytogenetic techniques such as fluorescence in situ hybridization (FISH) or molecular biology techniques such as real-time quantitative polymerase

TABLE 1: Sample demographics.

Current diagnosis	Alzheimer's disease	Mild cognitive impairment	Healthy controls	P value
Number of participants	222	136	143	—
Gender (Males/Females)	133/89	87/49	82/61	not significant
Baseline age (Mean $\pm$ SD)	75.10 $\pm$ 7.27	75.88 $\pm$ 7.17	75.83 $\pm$ 5.32	not significant
Years of education (Mean $\pm$ SD)	15.30 $\pm$ 3.05	15.85 $\pm$ 3.01	16.24 $\pm$ 2.62	0.009
APOE group (e4 negative/e4 positive)	73/149	70/66	108/35	<0.001
Age of onset (Mean $\pm$ SD)	74.08 $\pm$ 7.73	—	—	—

TABLE 2: Characteristics of CNV calls in the three diagnostic groups.

	Alzheimer's disease ( <i>n</i> = 222)	Mild cognitive impairment ( <i>n</i> = 136)	Healthy controls ( <i>n</i> = 143)
Deletions:			
Number of CNVs	2128	1340	1278
Rate per participant	9.59	9.85	8.94
Average size (kb)	73.24	76.32	79.38
Duplications:			
Number of CNVs	886	498	607
Rate per participant	3.99	3.66	4.24
Average size (kb)	157.24	154.06	170.30

chain reaction (PCR) and deep resequencing is required to determine the clinical relevance of these regions.

A case/control association analysis was then performed using a candidate gene approach and a whole genome approach to determine if there was an excess of CNV calls partially overlapping genes in AD or MCI participants relative to controls, suggesting potential involvement of these genes in AD or MCI susceptibility.

The candidate gene approach revealed several interesting genes (Table 4 and Figure 1). The *CHRFAM7A* gene had CNV calls in cases (two AD and two MCI) but not in controls. *CHRFAM7A*, located on chromosome 15, consists of a partial duplication of the *CHRNA7* (cholinergic receptor, nicotinic, alpha 7) gene (exons 5-10) fused to a copy of the *FAM7A* (family with sequence similarity 7A) gene (exons A-E) [28]. The *CHRFAM7A* gene contains a polymorphism consisting of a 2-base pair deletion (−2 bp) at position 497-498 of exon 6. This −2 bp polymorphism has been associated with schizophrenia [29]. The *CHRFAM7A* genotype without the −2 bp allele has also been shown to be significantly overrepresented in AD ( $P = .011$ ), dementia with Lewy bodies ( $P = .001$ ), and Pick's disease ( $P < .0001$ ) participants [30]. Heinzen et al. found a duplication in six out of 276 dementia cases (2%) and one out of 322 controls (0.3%) within the schizophrenia and epilepsy-associated risk region at 15q13.3, affecting the *CHRNA7* gene [9]. In the present study, we found a deletion in one out of 222 AD participants (0.45%) and one out of 136 MCI participants (0.74%), as well as a duplication in two out of 143 HC participants (1.40%). This gene codes for one of several neuronal cholinergic nicotinic receptors. Genetic variants in *CHRNA7* and other cholinergic receptor genes have been implicated in AD susceptibility [9], and further investigation

of this gene family is warranted. The number of CNV calls overlapping the identified genes is small, as we had a small sample size ( $n = 501$ ) after QC for analysis limiting power. Nevertheless, all identified genes have been previously investigated in AD studies and thus represent potential candidate genes. Replication studies with larger sample sizes as well as laboratory validation are required to confirm the role of these genes in AD susceptibility.

The whole genome approach revealed three genes at uncorrected  $P < .05$ , as shown in Table 5. *CSMD1* (CUB and Sushi multiple domains 1) has been shown to be primarily synthesized in the developing central nervous system (CNS) and epithelial tissues [31]. It is enriched in the nerve growth cone, suggesting that it may be an important regulator of complement activation and inflammation in the developing CNS. *HNRNPCL1* (heterogeneous nuclear ribonucleoprotein C-like 1) is predicted to play a role in nucleosome assembly by neutralizing basic proteins such as A and B core hnRNPs (Uniprot: <http://www.uniprot.org/>). *SLC35F2* (solute carrier family 35, member F2), also known as lung squamous cell cancer-related protein LSCC-3, is integral to membrane and transport (Gene Ontology: <http://www.geneontology.org/>).

We also identified CNVs overlapping two candidate genes associated with neuropsychiatric disorders: *NRXN1* and *ERBB4*, from the whole genome approach in cases, but not in controls. Deletions in the *NRXN1* (neurexin 1) gene were observed in four AD participants and three MCI participants; deletions in the *ERBB4* (v-erb-a erythroblastic leukemia viral oncogene homolog 4) gene were observed in four AD participants and one MCI participant, respectively. *NRXN1*, a member of the neurexin family on chromosome 2, is a cell surface receptor that binds neuroligins. The  $Ca^{2+}$ -dependent neurexin-neuroligin complex

TABLE 3: Participants grouped by CNV call size.

Call size	Alzheimer's disease ( <i>n</i> = 222)		Mild cognitive impairment ( <i>n</i> = 136)		Healthy controls ( <i>n</i> = 143)	
	Deletions <i>n</i> (%)	Duplications <i>n</i> (%)	Deletions <i>n</i> (%)	Duplications <i>n</i> (%)	Deletions <i>n</i> (%)	Duplications <i>n</i> (%)
0.1–0.5 Mb	174 (78.38)	183 (82.43)	104 (76.47)	100 (73.53)	114 (79.72)	120 (83.92)
0.5–1.0 Mb	6 (2.70)	27 (12.16)	8 (5.88)	18 (13.24)	8 (5.94)	27 (18.88)
1.0–1.5 Mb	0 (0.00)	8 (3.60)	0 (0.00)	4 (2.94)	2 (1.40)	8 (5.59)
1.5–2.0 Mb	0 (0.00)	2 (0.90)	0 (0.00)	1 (0.74)	1 (0.70)	0 (0.00)
>2.0 Mb	1 (0.45)	1 (0.45)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)

TABLE 4: Genes that have CNV calls from at least one Alzheimer's disease (AD) and/or one mild cognitive impairment (MCI) participant and no healthy controls using the candidate gene approach.

Chromosome	Region	Start (bp)	End (bp)	Number of AD participants	<i>APOE</i> genotype	Age at onset <sup>a</sup>	Number of MCI participants	<i>APOE</i> genotype
5	<i>PPP2R2B</i>	145949260	146441226	1	e3/e3	N/A	0	—
6	<i>ATXN1</i>	16407321	16869700	1	e3/e4	83 years	0	—
7	<i>MAGI2</i>	77484309	78920826	1	e2/e4	N/A	0	—
7	<i>RELN</i>	102899472	103417198	0	—	—	1	e4/e4
9	<i>GRIN3A</i>	103371455	103540683	1	e3/e3	74 years	0	—
10	<i>LRRTM3</i>	68355797	68530873	1	e3/e3	55 years	0	—
10	<i>LIPA</i>	90963305	91001640	0	—	—	1 <sup>b</sup>	e3/e3
12	<i>PPM1H</i>	61324030	61614932	1	e2/e3	N/A	1	e2/e3
15	<i>CHRFAM7A</i>	28440734	28473156	2	e3/e3	N/A	2	e3/e3
					e3/e4	N/A		e3/e4
15	<i>ADAM10</i>	56675801	56829469	1	e3/e3	N/A	1 <sup>b</sup>	e3/e3
21	<i>DNAJC28</i>	33782107	33785893	1	e3/e4	74 years	0	—
21	<i>DOPEY2</i>	36458708	36588442	0	—	—	1	e3/e4
22	<i>GSTT1</i>	22706138	22714284	1	e3/e3	59 years	0	—

<sup>a</sup>Age at onset of AD symptoms, available only for participants with a baseline diagnosis of AD; N/A: Not available.

<sup>b</sup>The same participant had CNV calls overlapping the two genes.

is present in the CNS at synapses and is required for efficient neurotransmission and formation of synaptic contacts [32]. This gene has been found to have reduced expression with AD severity [33], and its disruption has been shown to be associated with schizophrenia [20, 25, 34] and autism [24, 35]. Deletions in this gene have also been shown to predispose to a variety of developmental disorders including autism spectrum disorders, language delays, and mental retardation [36]. Interestingly, an SNP (rs6463843) flanking the *NXP1* (neurexophilin 1) gene was identified by our group in a GWAS of neuroimaging phenotypes in the ADNI cohort [37]. The *NXP1* gene, a member of the neurexophilin family, forms a tight complex with alpha neurexins, and the SNP was found to be associated with reduced global and regional grey matter density. The *ERBB4* gene, also on chromosome 2, is a member of the type I receptor kinase subfamily, that encodes a receptor for neuregulin 1 (*NRG1*). The neuregulin-ErbB receptor signaling pathway plays a role in development, synaptic function, and neural network activity and has been implicated in schizophrenia [38]. One AD participant had

a large duplication that included 23 genes in the 16p11.2 region (Figure 2). CNVs in this region have previously been associated with autism [39–41], schizophrenia [42], cognitive impairment and speech/language delay [43], and obesity [44, 45], but not AD or MCI. Because the ADNI employed a case/control design, DNA from family members was not available for linkage analysis. This limitation precluded determination as to whether CNVs were *de novo* or inherited.

The ADNI cohort provides a unique opportunity for discovery analyses such as this initial CNV analysis. With multiple types of potential biomarkers, including structural and molecular imaging, blood and CSF markers, genetic information, and behavioral data, analysis of the ADNI data has the potential to enhance knowledge of the underlying mechanisms leading to MCI and to AD.

The present study has several limitations related to participant inclusion and exclusion and the software and algorithms used in the analyses. CNV calls in the present report were generated from DNA samples derived only from peripheral blood-78 participants whose DNAs were derived



TABLE 5: Significant (uncorrected  $P < .05$ , relative to healthy controls) genes present in either Alzheimer's disease (AD) and/or mild cognitive impairment (MCI) participants, but not healthy controls using the whole genome approach.

Chromosome	Region	Start (bp)	End (bp)	Number of AD calls	$P$ value for AD calls	Number of MCI calls	$P$ value for MCI calls
8	<i>CSMD1</i>	2780281	4839736	9	0.0114	4	0.0556
1	<i>HNRNPCL1</i>	12829847	12831165	6	0.0493	4	0.0549
11	<i>SLC35F2</i>	107166926	107234864	5	0.0820	6	0.0120

from lymphoblastoid cell lines (LCLs) were excluded. LCLs are generated by transforming peripheral B lymphocytes by the Epstein-Barr virus (EBV). EBV-transformed cells are shown to have significant telomerase activity and develop aneuploidy, along with other cellular changes such as gene mutations and reprogramming in the postimmortal cellular stage of transformation [13]. Thus, to avoid CNV call discrepancies that may arise between the different DNA sources, we chose to include only those participants whose DNA was derived from peripheral blood. Additional QC was also performed, resulting in only 501 samples that passed all QC checks. To date, no definitive QC criterion has been established to ensure only high-quality samples are included in CNV analyses. Therefore, the QC criterion applied in the present study may have been too stringent leading to the exclusion of samples which otherwise may have had informative CNV data. In future studies, we propose to analyze multiple QC thresholds to determine the optimum QC criteria.

Another limitation is that the CNV calls analyzed in the current study were generated using only one software program (PennCNV). Several detection algorithms including HMMs, segmentation algorithms,  $t$ -tests, and standard deviations of the LRR are available for identifying CNVs from genome-wide SNP array data. A comparison of these methods has been performed by Dellinger et al. Even though the PennCNV program was found to have moderate power in detecting CNVs, it also had a low false positive call rate. The program was found to detect less CNV calls in comparison to other methods and did not accurately detect small CNVs (3-4 SNP CNVs) [46]. However, in our analyses, we have included CNV calls that had at least 10 SNPs. Obtaining the same CNV calls from another algorithm would help further reduce false positive CNV calls.

The heterogeneity of the MCI group of participants also represents a possible limitation of the present study. Although biomarkers such as CSF and PiB-PET can help differentiate MCI participants who have an AD-like profile from those who have a normal profile, this data was only available for a small number of ADNI-1 participants which would have limited power to detect differences in CNVs. In the next phases of the project (ADNI-GO and ADNI-2), all subjects will have CSF and amyloid PET data, enabling further examination of this issue.

## 5. Conclusion

In sum, we have conducted an initial CNV analysis in the ADNI cohort dataset. Although no excess CNV burden was found in cases relative to controls, a number of interesting

candidate genes and regions were identified. Replication in larger samples will be critical to confirm these findings. Additional region-based analyses may help elucidate the role of these CNVs, and deep resequencing studies may be warranted for some of these regions if they replicate in other cohorts.

## Disclosure

Dr. L. Shen receives support from an NIBIB R03 EB008674 Grant and an Indiana CTSI award (IUSM/CTR based on NCRR RR025761). Dr. T. Foroud receives support from an NIH/NIA 5U24AG021886 Grant and U01 AG032984 Alzheimer's Disease Genetics Consortium (ADGC) (PI: Schellenberg). Dr. S. G. Potkin receives Grant support from the Transdisciplinary Imaging Genetics Center (TIGC) P20 RR020837-01 and NIH/NCRR U24 RR021992 and serves on the Editorial Board of Brain Imaging and Behavior. Dr. M. J. Huentelman receives Grant support from NIH-NINDS R01 N5059873 (awarded to Dr. M. J. Huentelman). Dr. M. W. Weiner serves on the scientific advisory boards for Alzheimer's Study Group, Bayer Schering Pharma, Eli Lilly and Company, CoMentics Inc., Neurochem Inc., SIRA UCSD, Eisai Inc., Avid Radiopharmaceuticals Inc., Aegis therapies, Genentech Inc., Allergen Inc., Lippincott Williams and Wilkins, Bristol-Myers Squibb, Forest Laboratories, Pfizer Inc., McKinsey and Company, Mitsubishi Tanabe Pharma Corporation, and Novartis; has received funding for travel from Nestle and Kenes International and to attend conferences not funded by industry; serves on the Editorial Board of Alzheimer's and Dementia, and Brain Imaging and Behavior; has received honoraria from Rotman Research Institute and BOLT International; receives research support from NIH, DOD, VA, Merck and Avid; holds stock in Synarc and Elan Corporation. Dr. A. J. Saykin receives support from the NIH (R01 CA101318, R01 AG19771, RC2 AG036535-01, P30 AG10133-18S1, U01 AG032984, and U01 AG032984 Alzheimer's Disease Genetics Consortium (ADGC) (PI: Schellenberg)), and investigator-initiated research support from Siemens Medical Solutions and Welch Allyn, Inc. He serves as Editor-in-Chief of *Brain Imaging and Behavior*, a Springer Journal. He has received support from the Indiana Economic Development Corporation (IEDC #87884) and has served as a consultant to Baxter BioScience, Bristol-Myers Squibb and Pfizer, Inc. and received funding from Eli Lilly and Company to provide postdoctoral training in neuroimaging. He serves as a consultant to NIH-sponsored research programs at Dartmouth Medical School, University of Michigan and the University of Vermont. All other authors report no disclosures.

## The Alzheimer's Disease Neuroimaging Initiative (ADNI)

Data used in the preparation of this article were obtained from the ADNI database (<http://loni.ucla.edu/ADNI>). As such, the investigators within 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 the ADNI investigators is available at [http://loni.ucla.edu/ADNI/Collaboration/ADNI\\_Authorship\\_list.pdf](http://loni.ucla.edu/ADNI/Collaboration/ADNI_Authorship_list.pdf).

## Acknowledgments

Data collection and sharing for this project was funded by the The Alzheimer's Disease Neuroimaging Initiative (ADNI) (NIH Grant U01 AG024904). ADNI is funded by the NIA, NIBIB, and through generous contributions from the following: Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson, Eli Lilly and Co., Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer Inc, F. Hoffman-La Roche, Schering-Plough, Synarc, Inc., and Wyeth, as well as nonprofit partners like the Alzheimer's Association and Alzheimer's Drug Discovery Foundation, with participation from the US Food and Drug Administration. Private sector contributions to ADNI are facilitated by the Foundation for the National Institutes of Health (<http://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 nos. P30 AG010129 and K01 AG030514, the Dana Foundation and U01 AG032984 Alzheimer's Disease Genetics Consortium (ADGC) Grant (PI: Schellenberg). Samples from NCRAD, which receives government support under a cooperative agreement Grant (U24 AG21886) awarded by the NIA, were used in this study. The authors thank contributors, including the Alzheimer's Disease Centers, who collected samples used in this study, as well as patients and their families, whose help and participation made this work possible. They also thank the following people for their contributions to the ADNI genotyping project: (1) genotyping at the Translational Genomics Institute, Phoenix AZ: Jennifer Webster, Jill D. Gerber, April N. Allen, and Jason J. Corneveaux; (2) sample processing, storage, and distribution at the NIA-sponsored NCRAD: Kelley Faber. They also thank John West for providing them with technical support.

## References

- [1] R. C. Petersen, R. O. Roberts, D. S. Knopman et al., "Mild cognitive impairment: ten years later," *Archives of Neurology*, vol. 66, no. 12, pp. 1447–1455, 2009.
- [2] M. Gatz, C. A. Reynolds, L. Fratiglioni et al., "Role of genes and environments for explaining Alzheimer disease," *Archives of General Psychiatry*, vol. 63, no. 2, pp. 168–174, 2006.
- [3] K. Sleegers, J. C. Lambert, L. Bertram, M. Cruts, P. Amouyel, and C. Van Broeckhoven, "The pursuit of susceptibility genes for Alzheimer's disease: progress and prospects," *Trends in Genetics*, vol. 26, no. 2, pp. 84–93, 2010.
- [4] D. Harold, R. Abraham, P. Hollingworth et al., "Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease," *Nature Genetics*, vol. 41, no. 10, pp. 1088–1093, 2009.
- [5] J. C. Lambert, S. Heath, G. Even et al., "Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease," *Nature Genetics*, vol. 41, no. 10, pp. 1094–1099, 2009.
- [6] S. Seshadri, A. L. Fitzpatrick, M. A. Ikram et al., "Genome-wide analysis of genetic loci associated with Alzheimer disease," *Journal of the American Medical Association*, vol. 303, no. 18, pp. 1832–1840, 2010.
- [7] G. Jun, A. C. Naj, G. W. Beecham et al., "Meta-analysis confirms CR1, CLU, and PICALM as Alzheimer disease risk loci and reveals interactions with APOE genotypes," *Archives of Neurology*, vol. 67, no. 12, pp. 1473–1484, 2010.
- [8] E. H. Cook Jr. and S. W. Scherer, "Copy-number variations associated with neuropsychiatric conditions," *Nature*, vol. 455, no. 7215, pp. 919–923, 2008.
- [9] E. L. Heinzen, A. C. Need, K. M. Hayden et al., "Genome-wide scan of copy number variation in late-onset Alzheimer's disease," *Journal of Alzheimer's Disease*, vol. 19, no. 1, pp. 69–77, 2010.
- [10] R. C. Petersen, P. S. Aisen, L. A. Beckett et al., "Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization," *Neurology*, vol. 74, no. 3, pp. 201–209, 2010.
- [11] M. W. Weiner, P. S. Aisen, C. R. Jack Jr. et al., "The Alzheimer's disease neuroimaging initiative: progress report and future plans," *Alzheimer's and Dementia*, vol. 6, no. 3, pp. 202–211.e3, 2010.
- [12] The Wellcome Trust Case Control Consortium, "Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls," *Nature*, vol. 464, no. 7289, pp. 713–720, 2010.
- [13] L. Sie, S. Loong, and E. K. Tan, "Utility of lymphoblastoid cell lines," *Journal of Neuroscience Research*, vol. 87, no. 9, pp. 1953–1959, 2009.
- [14] L. M. Bekris, C.-E. Yu, T. D. Bird, and D. W. Tsuang, "Review article: genetics of Alzheimer disease," *Journal of Geriatric Psychiatry and Neurology*, vol. 23, no. 4, pp. 213–227, 2010.
- [15] A. J. Saykin, L. Shen, T. M. Foroud et al., "Alzheimer's Disease Neuroimaging Initiative biomarkers as quantitative phenotypes: genetics core aims, progress, and plans," *Alzheimer's and Dementia*, vol. 6, no. 3, pp. 265–273, 2010.
- [16] S. G. Potkin, G. Guffanti, A. Lakatos et al., "Hippocampal atrophy as a quantitative trait in a genome-wide association study identifying novel susceptibility genes for Alzheimer's Disease," *PLoS One*, vol. 4, no. 8, Article ID e6501, 2009.
- [17] K. Wang, M. Li, D. Hadley et al., "PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data," *Genome Research*, vol. 17, no. 11, pp. 1665–1674, 2007.
- [18] S. J. Diskin, M. Li, C. Hou et al., "Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping

- platforms," *Nucleic Acids Research*, vol. 36, no. 19, article no. e126, 2008.
- [19] L. K. Conlin, B. D. Thiel, C. G. Bonnemann et al., "Mechanisms of mosaicism, chimerism and uniparental disomy identified by single nucleotide polymorphism array analysis," *Human Molecular Genetics*, vol. 19, no. 7, Article ID ddq003, pp. 1263–1275, 2010.
- [20] A. C. Need, D. Ge, M. E. Weale et al., "A genome-wide investigation of SNPs and CNVs in schizophrenia," *PLoS Genetics*, vol. 5, no. 2, Article ID e1000373, 2009.
- [21] S. Purcell, B. Neale, K. Todd-Brown et al., "PLINK: a tool set for whole-genome association and population-based linkage analyses," *American Journal of Human Genetics*, vol. 81, no. 3, pp. 559–575, 2007.
- [22] L. Bertram, M. B. McQueen, K. Mullin, D. Blacker, and R. E. Tanzi, "Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database," *Nature Genetics*, vol. 39, no. 1, pp. 17–23, 2007.
- [23] W. James Kent, C. W. Sugnet, T. S. Furey et al., "The human genome browser at UCSC," *Genome Research*, vol. 12, no. 6, pp. 996–1006, 2002.
- [24] P. Szatmari, A. D. Paterson, L. Zwaigenbaum et al., "Mapping autism risk loci using genetic linkage and chromosomal rearrangements," *Nature Genetics*, vol. 39, no. 3, pp. 319–328, 2007.
- [25] D. Rujescu, A. Ingason, S. Cichon et al., "Disruption of the neurexin 1 gene is associated with schizophrenia," *Human Molecular Genetics*, vol. 18, no. 5, pp. 988–996, 2009.
- [26] C. -L. Lu, Y. -C. Wang, J. -Y. Chen, I. -C. Lai, and Y. -J. Liou, "Support for the involvement of the ERBB4 gene in schizophrenia: a genetic association analysis," *Neuroscience Letters*, vol. 481, no. 2, pp. 120–125, 2010.
- [27] M. Wu, E. J. Michaud, and D. K. Johnson, "Cloning, functional study and comparative mapping of Luzp2 to mouse Chromosome 7 and human Chromosome 11p13-11p14," *Mammalian Genome*, vol. 14, no. 5, pp. 323–334, 2003.
- [28] B. Riley, M. Williamson, D. Collier, H. Wilkie, and A. Makoff, "A 3-Mb map of a large segmental duplication overlapping the  $\alpha 7$ -nicotinic acetylcholine receptor gene (CHRNA7) at human 15q13-q14," *Genomics*, vol. 79, no. 2, pp. 197–209, 2001.
- [29] M. L. Sinkus, M. J. Lee, J. Gault et al., "A 2-base pair deletion polymorphism in the partial duplication of the  $\alpha 7$  nicotinic acetylcholine gene (CHRFAM7A) on chromosome 15q14 is associated with schizophrenia," *Brain Research*, vol. 1291, pp. 1–11, 2009.
- [30] Á. Fehér, A. Juhász, Á. Rimanóczy, É. Csibri, J. Kálmán, and Z. Janka, "Association between a genetic variant of the  $\alpha 7$  nicotinic acetylcholine receptor subunit and four types of dementia," *Dementia and Geriatric Cognitive Disorders*, vol. 28, no. 1, pp. 56–62, 2009.
- [31] D. M. Kraus, G. S. Elliott, H. Chute et al., "CSMD1 is a novel multiple domain complement-regulatory protein highly expressed in the central nervous system and epithelial tissues," *Journal of Immunology*, vol. 176, no. 7, pp. 4419–4430, 2006.
- [32] C. Reissner, M. Klose, R. Fairless, and M. Missler, "Mutational analysis of the neurexin/neurologin complex reveals essential and regulatory components," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 39, pp. 15124–15129, 2008.
- [33] M. G. Ravetti, O. A. Rosso, R. Berretta, and P. Moscato, "Uncovering molecular biomarkers that correlate cognitive decline with the changes of hippocampus' gene expression profiles in Alzheimer's disease," *PLoS One*, vol. 5, no. 4, Article ID e10153, 2010.
- [34] M. Ikeda, B. Aleksic, G. Kirov et al., "Copy number variation in Schizophrenia in the Japanese population," *Biological Psychiatry*, vol. 67, no. 3, pp. 283–286, 2010.
- [35] J. T. Glessner, K. Wang, G. Cai et al., "Autism genome-wide copy number variation reveals ubiquitin and neuronal genes," *Nature*, vol. 459, no. 7246, pp. 569–572, 2009.
- [36] M. S. L. Ching, Y. Shen, W. H. Tan et al., "Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders," *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, vol. 153, no. 4, pp. 937–947, 2010.
- [37] L. I. Shen, S. Kim, S. L. Risacher et al., "Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: a study of the ADNI cohort," *NeuroImage*, vol. 53, no. 3, pp. 1051–1063, 2010.
- [38] A. Buonanno, "The neuregulin signaling pathway and schizophrenia: from genes to synapses and neural circuits," *Brain Research Bulletin*, vol. 83, no. 3–4, pp. 122–131, 2010.
- [39] L. A. Weiss, Y. Shen, J. M. Korn et al., "Association between microdeletion and microduplication at 16p11.2 and autism," *New England Journal of Medicine*, vol. 358, no. 7, pp. 667–675, 2008.
- [40] R. A. Kumar, S. Karamohamed, J. Sudi et al., "Recurrent 16p11.2 microdeletions in autism," *Human Molecular Genetics*, vol. 17, no. 4, pp. 628–638, 2008.
- [41] C. R. Marshall, A. Noor, J. B. Vincent et al., "Structural variation of chromosomes in autism spectrum disorder," *American Journal of Human Genetics*, vol. 82, no. 2, pp. 477–488, 2008.
- [42] S. E. McCarthy, V. Makarov, G. Kirov et al., "Microduplications of 16p11.2 are associated with schizophrenia," *Nature Genetics*, vol. 41, no. 11, pp. 1223–1227, 2009.
- [43] M. Shinawi, P. Liu, S. H. L. Kang et al., "Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size," *Journal of Medical Genetics*, vol. 47, no. 5, pp. 332–341, 2010.
- [44] R. G. Walters, S. Jacquemont, A. Valsesia et al., "A new highly penetrant form of obesity due to deletions on chromosome 16p11.2," *Nature*, vol. 463, no. 7281, pp. 671–675, 2010.
- [45] E. G. Bochukova, N. I. Huang, J. Keogh et al., "Large, rare chromosomal deletions associated with severe early-onset obesity," *Nature*, vol. 463, no. 7281, pp. 666–670, 2010.
- [46] A. E. Dellinger, S. M. Saw, L. K. Goh, M. Seielstad, T. L. Young, and Y. J. Li, "Comparative analyses of seven algorithms for copy number variant identification from single nucleotide polymorphism arrays," *Nucleic acids research*, vol. 38, no. 9, p. e105, 2010.