ROLE OF GENOMIC COPY NUMBER VARIATION IN ALZHEIMER'S DISEASE AND MILD COGNITIVE IMPAIRMENT

Shanker Swaminathan

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Doctor of Philosophy
in the Department of Medical and Molecular Genetics,
Indiana University

August 2012

Accepted by the Faculty of Indiana University, in partial
fulfillment of the requirements for the Degree of Doctor of Philosophy.

Andrew J. Saykin, PsyD, Chair

Tatiana Foroud, PhD

Doctoral Committee

John I. Nurnberger, Jr, MD, PhD

May 23, 2012

Li Shen, PhD

To my parents and sister

ACKNOWLEDGEMENTS

This work would not have been possible without the help and support of a lot of people. I would like to thank Dr. Andrew Saykin, my thesis advisor, for providing me the opportunity to be a part of his lab and be able to pursue research in the field of imaging genetics. Dr. Saykin has always encouraged me to be pursue new information and be up to date with the current developments in the field. He has been a constant source of support and guidance and was a person whom I could always turn to when I have had problems. His feedback and suggestions have been very helpful. I would also like to thank the members of my research committee: Drs. Tatiana Foroud, John Nurnberger and Li Shen for their help in this work. They encouraged me to pursue this work and offered feedback and suggestions which were very helpful. They also made sure that I was on track with the program requirements and was making smooth progress towards the PhD degree.

I would like to thank other members of the lab including Dr. Li Shen, Dr. Sungeun Kim, Dr. Kwangsik Nho, Dr. Shannon Risacher, Susan Conroy, John West, Jing Wan, Kelley Holohan and Vijay Ramanan for their help with this work. I was fortunate to work with them on a number of projects, which enabled me to learn some of the methods used in imaging analyses and genetic analyses. I am also grateful to Dr. Mark Inlow for his help and guidance on the statistical analyses. I would also like to thank the faculty and staff at the Center for Neuroimaging.

I express my gratitude to Dr. Kenneth Cornetta, Dr. Brittney-Shea Herbert, Dr. Terry Reed and Peggy Knople for their guidance on the program requirements and for making sure I was completing the requirements in a timely manner. I also thank the faculty and

staff at the Department of Medical and Molecular Genetics as well as the staff of the IUPUI Graduate Office and the Graduate Division of the Indiana University School of Medicine for their help during the course of my study.

Finally, I would like to offer my heartfelt thanks to my family. They encouraged me to pursue my interests in performing research and were there with me at every step. They follow my progress keenly and are very interested in learning about the various projects I am involved in as well the Genetics field in general. Their motivation and support has helped me progress along smoothly and I would like to express my sincere gratitude to them.

ABSTRACT

Shanker Swaminathan

ROLE OF GENOMIC COPY NUMBER VARIATION IN ALZHEIMER'S DISEASE AND MILD COGNITIVE IMPAIRMENT

Alzheimer's disease (AD) is the most common form of dementia defined by loss in memory and cognitive abilities severe enough to interfere significantly with daily life activities. Amnestic mild cognitive impairment (MCI) is a clinical condition in which an individual has memory deficits not normal for the individual's age, but not severe enough to interfere significantly with daily functioning. Every year, approximately 10-15% of individuals with MCI will progress to dementia. Currently, there is no treatment to slow or halt AD progression, but research studies are being conducted to identify causes that can lead to its earlier diagnosis and treatment.

Genetic variation plays a key role in the development of AD, but not all genetic factors associated with the disease have been identified. Copy number variants (CNVs), a form of genetic variation, are DNA regions that have added genetic material (duplications) or loss of genetic material (deletions). The regions may overlap one or more genes possibly affecting their function. CNVs have been shown to play a role in certain diseases.

At the start of this work, only one published study had examined CNVs in late-onset AD and none had examined MCI. In order to determine the possible involvement of CNVs in AD and MCI susceptibility, genome-wide CNV analyses were performed in participants from three cohorts: the ADNI cohort, the NIA-LOAD/NCRAD Family Study cohort, and a

unique cohort of clinically characterized and neuropathologically verified individuals. Only participants with DNA samples extracted from blood/brain tissue were included in the analyses. CNV calls were generated using genome-wide array data available on these samples. After detailed quality review, case (AD and/or MCI)/control association analyses including candidate gene and genome-wide approaches were performed.

Although no excess CNV burden was observed in cases compared to controls in the three cohorts, gene-based association analyses identified a number of genes including the AD candidate genes *CHRFAM7A*, *RELN* and *DOPEY2*. Thus, the present work highlights the possible role of CNVs in AD and MCI susceptibility warranting further investigation. Future work will include replication of the findings in independent samples and confirmation by molecular validation experiments.

Andrew J. Saykin, PsyD, Chair

TABLE OF CONTENTS

List of Tables x

List of Figures xii

List of Abbreviations xiv

I. Introduction

A. Alzheimer's disease 1

B. Role of genetic variation in Alzheimer's disease 4

C. Copy number variation 11

D. Mechanisms for formation of copy number variations 12

E. Methods for copy number variation detection 16

F. Copy number variation detection algorithms 20

G. Copy number variation in selected neuropsychiatric disorders 23

H. Copy number variation in Alzheimer's disease 33

I. Statement of purpose 35

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

A. Introduction 37

B. Methods 40

C. Results 47

D. Discussion 59

III. Analysis of Copy Number Variation in Alzheimer's Disease: the NIA-LOAD/NCRAD Family Study

A. Introduction 65

B. Methods 68

C. Results 75

D. Discussion....88

IV. Analysis of Copy Number Variation in Alzheimer's Disease in a Cohort of Clinically Characterized and Neuropathologically Verified Individuals

A. Introduction....97

B. Methods....99

C. Results....104

D. Discussion....116

V. Summary and Future Directions....120

References....127

Curriculum Vitae

LIST OF TABLES

1. Genes associated with early-onset familial Alzheimer's disease 6

2. Summary of published genome-wide association studies in Alzheimer's disease 7

3. Examples of copy number variations and selected neuropsychiatric disorders 24

4. Sample demographics of participants in the ADNI cohort 48

5. Characteristics of copy number variation calls in the three diagnostic groups from participants in the ADNI cohort 49

6. Participants in the ADNI cohort grouped by copy number variation call size 50

7. Genes that had copy number variation calls from at least one Alzheimer's disease and/or one mild cognitive impairment participant and no healthy controls in the ADNI cohort using the candidate gene approach 54

8. Significant (uncorrected $p<0.05$, relative to healthy controls) genes present in either Alzheimer's disease and/or mild cognitive impairment participants, but not healthy controls in the ADNI cohort using the whole genome approach 56

9. Sample demographics of participants in the NIA-LOAD/NCRAD Family Study cohort 76

10. Characteristics of copy number variation calls from participants in the NIA-LOAD/NCRAD Family Study cohort 77

11. Participants in the NIA-LOAD/NCRAD Family Study cohort grouped by copy number variation call size 78

12. Genes overlapped by copy number variation calls from at least one Alzheimer's disease participant and no controls in the NIA-LOAD/NCRAD Family Study cohort using the candidate gene approach.....82
13. Sample demographics in the TGen cohort.....105
14. Characteristics of copy number variation calls from samples in the TGen cohort 106
15. TGen samples grouped by copy number variation call size 107
16. Genes overlapped by copy number variation calls from at least one case and no control samples in the TGen cohort using the candidate gene approach 109
17. Meta-analysis of the *CHRFAM7A* gene 114

LIST OF FIGURES

1. Representative images of a neuritic plaque and neurofibrillary tangles.....3

2. Representation of the recurrent and nonrecurrent genomic rearrangements that play a role in genomic disorders 15

3. Representative image of B Allele Frequency and Log R Ratio of the participant in the ADNI cohort who had multiple deletions and duplications on chromosome 18.....45

4. Representative image of B Allele Frequency and Log R Ratio of the Alzheimer's disease participant in the ADNI cohort who had a deletion >2 Mb on chromosome 4.....51

5. Representative image of B Allele Frequency and Log R Ratio of the Alzheimer's disease participant in the ADNI cohort who had a duplication >2 Mb on chromosome 11.....52

6. Candidate genes *CHRFAM7A* and *LRRTM3*, overlapped by copy number variation calls from at least one Alzheimer's disease and/or mild cognitive impairment participant, but no healthy controls in the ADNI cohort.....57

7. Representative image of B Allele Frequency and Log R Ratio of the participant in the ADNI cohort who had a duplication at 16p11.2 58

8. Representative image of B Allele Frequency and Log R Ratio of the Alzheimer's disease participant in the NIA-LOAD/NCRAD Family Study cohort who had a deletion >2 Mb on chromosome 11 including chromosomal coordinates and the genes in the corresponding regions.....79

9. Representative image of B Allele Frequency and Log R Ratio of the Alzheimer's disease participant in the NIA-LOAD/NCRAD Family Study cohort who had a duplication >2 Mb on chromosome 3 including chromosomal coordinates and the genes in the corresponding regions.....80
10. Representative UCSC Genome Browser (March 2006 (NCBI36/hg18) assembly) plots of deletions overlapping the *CHRFAM7A* and *IMMP2L* genes in participants from the NIA-LOAD/NCRAD Family and ADNI studies.....85
11. Representative image of B Allele Frequency and Log R Ratio of the Alzheimer's disease participant in the NIA-LOAD/NCRAD Family Study cohort who had a duplication at 16p13.11 along with chromosomal coordinates and genes in corresponding regions are shown 86
12. Representative image of B Allele Frequency and Log R Ratio of the Alzheimer's disease participant in the NIA-LOAD/NCRAD Family Study cohort who had a deletion at 17p12 along with chromosomal coordinates and genes in corresponding regions are shown.....87
13. Representative UCSC Genome Browser (March 2006 (NCBI36/hg18) assembly) plots of duplications overlapping the candidate genes *APP* and *DOPEY2* in samples of the TGen cohort.....111
14. Representative UCSC Genome Browser (March 2006 (NCBI36/hg18) assembly) plots of deletions and duplications overlapping the *CHRFAM7A* gene in samples of the TGen cohort 112
15. Forest plot of the *CHRFAM7A* gene 115

LIST OF ABBREVIATIONS

α7 nAChR	α7 nicotinic acetylcholine receptor
AD	Alzheimer's disease
ADNI	Alzheimer’s Disease Neuroimaging Initiative
ADRDA	Alzheimer’s Disease and Related Disorders Association
APOE	Apolipoprotein E
APP	Amyloid beta (A4) precursor protein
ASD	Autism spectrum disorder
ATXN1	Ataxin 1
BAF	B Allele Frequency
BIN1	Bridging integrator 1
BOAT1	Brother of *ATXN1*
bp	Base pair
CERAD	The Consortium to Establish a Registry for Alzheimer’s Disease
CGH	Comparative genome hybridization
CHRFAM7A	*CHRNA7* (cholinergic receptor, nicotinic, alpha 7, exons 5-10) and *FAM7A* (family with sequence similarity 7A, exons A-E) fusion
Chr	Chromosome
CI	Confidence interval
CLU	Clusterin
CNV	Copy number variation
CR1	Complement component (3b/4b) receptor 1 (Knops blood group)
CSMD1	CUB and Sushi multiple domains 1
Ct	Cycle threshold
dbGaP	The database of Genotypes and Phenotypes

DECIPHER	Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources
Del	Deletions
DNA	Deoxyribonucleic acid
DOPEY2	Dopey family member 2
Dup	Duplications
ECARUCA	European Cytogeneticists Association Register of Unbalanced Chromosome Alterations
EOFAD	Early-onset familial Alzheimer's disease
ERBB4	V-erb-a erythroblastic leukemia viral oncogene homolog 4 (avian)
EXOC3L2	Exocyst complex component 3-like 2
FoSTeS	Fork stalling and template switching
GBE1	Glucan (1,4-alpha-), branching enzyme 1
GSTT1	Glutathione S-transferase theta 1
GWAS	Genome-wide association study
hAPP695	Human amyloid precursor protein 695
HC	Healthy control
HLA	Human leukocyte antigen
HLA-DPB1	Major histocompatibility complex, class II, DP beta 1
HLA-DRA	Major histocompatibility complex, class II, DR alpha
HMM	Hidden Markov model
HNRNPCL1	Heterogeneous nuclear ribonucleoprotein C-like 1
IMMP2L	Inner mitochondrial membrane peptidase-like (*S. cerevisiae*)
kb	Kilobase
LCR	Low-copy repeats
LRR	Log R Ratio

LUZP2	Leucine zipper protein 2
Mb	Megabase
MCI	Mild cognitive impairment
MIR1973	MicroRNA 1973
MMBIR	Microhomology-mediated break-induced replication
MRI	Magnetic resonance imaging
N/A	Not available
NAHR	Nonallelic homologous recombination
NCRAD	The National Cell Repository for Alzheimer's Disease
NDST4	N-deacetylase/N-sulfotransferase 4
NHEJ	Nonhomologous end joining
NIA-LOAD	National Institute of Aging-Late Onset Alzheimer's Disease
NINCDS	National Institute of Neurological and Communicative Disorders
NRG1	Neuregulin 1
NRXN1	Neurexin 1
NXPH1	Neurexophilin 1
OR	Odds ratio
PCR	Polymerase chain reaction
PD	Parkinson disease
PET	Positron emission tomography
PICALM	Phosphatidylinositol binding clathrin assembly protein
PSEN1	Presenilin 1
PSEN2	Presenilin 2 (Alzheimer disease 4)
QC	Quality control
qPCR	Quantitative real-time polymerase chain reaction
RELN	Reelin

SD	Standard deviation
SNP	Single nucleotide polymorphism
SLC35F2	Solute carrier family 35, member F2
TRAM1L1	Translocation-associated membrane protein 1-like 1
TS	Tourette syndrome
WF	Waviness Factor

I. Introduction

A. Alzheimer's disease

Alzheimer's disease (AD) is the most common form of dementia accounting for 50-80% of dementia cases. Dementia is a general term which includes symptoms such as loss in memory and cognitive abilities severe enough to interfere significantly with daily life. AD is a progressive disease in which dementia symptoms gradually worsen over time. The most common early symptom is difficulty in remembering new information. Difficulty in remembering names and recent events, apathy and depression are other early symptoms of the disease. The individual's cognitive and functional abilities decline as the disease progresses. Impaired judgment, behavior changes, confusion, disorientation, and speaking, swallowing, and walking difficulties are later symptoms of the disease.

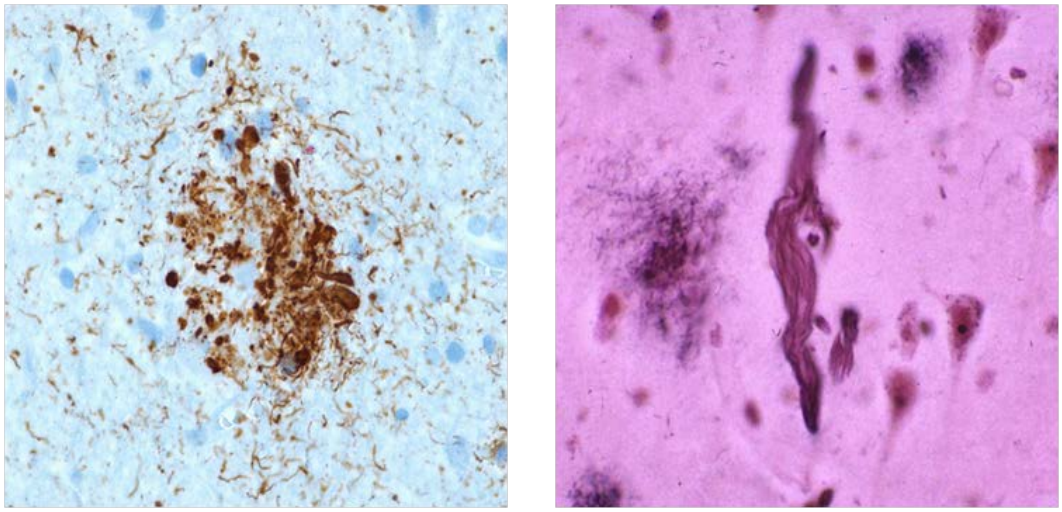
The hallmark abnormalities of AD are the accumulation of beta-amyloid protein fragments (amyloid plaques) in between nerve cells (neurons) in the brain and the accumulation of twisted strands of tau protein (neurofibrillary tangles) inside neurons (Figure 1). Synapses are specialized connections between neurons that enable information flow between individual neurons. It is believed that in AD, the accumulation of beta-amyloid outside the synapses interferes with neuron-to-neuron communication and leads to cell death. Abnormally high levels of tau inside the neuron form neurofibrillary tangles, which block transport of nutrient and other essential molecules throughout the cell possibly leading to cell death.

An estimated 5.4 million Americans of all ages, including 5.2 million individuals aged ≥65 years, and 200000 individuals aged <65 years have AD [1]. It is the sixth leading cause

of death across all ages in the United States and the fifth leading cause of death in Americans aged ≥65 years. Amnestic mild cognitive impairment (MCI) is a clinical condition in which an individual has memory deficits 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 in the population have MCI, and 10-15% of these individuals are likely to progress to dementia, particularly AD each year [2].

Although there are no current treatments that can slow or halt the progression of AD, there are treatments that can temporarily slow the worsening of dementia symptoms and possibly improve the quality of life in individuals with AD and their caregivers. The U.S. Food and Drug Administration has approved two types of medications to treat the cognitive symptoms of AD: cholinesterase inhibitors (Aricept, Exelon, Razadyne, Cognex) and memantine (Namenda) (Alzheimer's Association: http://www.alz.org/alzheimers_disease_standard_prescriptions.asp). A large worldwide research effort is also underway to identify the causes of AD that can lead to its earlier diagnosis and better treatment.

Figure 1. Representative images of a neuritic plaque (A) and neurofibrillary tangles (B). Adapted from Castellani et al. (2010) [3].



A **B**

B. Role of genetic variation in Alzheimer's disease

Genetic variation is known to 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 [4]. Based on the age of onset, AD can be classified into two subtypes: early-onset AD and late-onset AD [5]. Early-onset AD with an age at onset roughly ranging from 30 to 60 or 65 years accounts for approximately one to six percent of all cases. Among the early-onset AD cases, approximately 60% have multiple AD cases within their families, and of these familial early-onset AD cases, 13% have an autosomal dominant inheritance pattern with at least three generations affected [5, 6]. Around several hundred families carry mutations in three genes: *APP* (amyloid beta (A4) precursor protein) on chromosome 21q, *PSEN1* (presenilin 1) on chromosome 14q and *PSEN2* (presenilin 2 (Alzheimer disease 4)) on chromosome 1q (Table 1), but these accounts for less than 1% of cases. A review of the role of these genes can be found in Bekris et al. (2010) [5]. A number of mutations have been identified in these genes and information regarding these mutations can be found in Alzheimer Disease and Frontotemporal Dementia Mutation Database (http://www.molgen.ua.ac.be/ADMutations/) [7].

Late-onset AD, the more common form of AD (>90% of all cases), has an age at onset later than 60 or 65 years. A number of genetic association studies have been performed in AD and these have been catalogued in the AlzGene database (http://www.alzgene.org/) [8]. The leading genetic risk factor for late-onset AD is the *APOE* (apolipoprotein E) ε4 allele on chromosome 19q. The ε4 allele, a member of a three allele haplotype composed of ε2, ε3 and ε4 alleles confers a dose-dependent increase in AD risk of approximately four-fold in carriers compared to non-carriers [9-11].

Large case-control genome-wide association studies (GWASs) have identified and replicated other AD risk loci including: *CLU*, *CR1*, *PICALM*, *BIN1*, *EXOC3L2*, *MTHFD1L*, *MS4A4A/MS4A6E*, *CD2AP*, *CD33*, *ABCA7* and *CUGBP2* [12-19]. A summary of published GWASs in AD can be found in Table 2. However, for the strongest SNPs at each of *CR1*, *CLU*, *PICALM*, *BIN1*, *EPHA1*, *MS4A*, *CD33*, *CD2AP* and *ABCA7* loci, the population attributable fractions (the proportional reduction in mortality or population decrease if a risk factor exposure were reduced to an alternative ideal exposure scenario) were estimated to between 2.72% and 5.97%. Furthermore, the cumulative population-attributable fraction for these non-*APOE* loci is estimated to be as much as 35% [17]. Thus, the identified loci do not account for all the genetic variation associated with the disease. It is possible that other forms of genetic variation such as copy number variations (CNVs) may play a role.

Table 1. Genes associated with early-onset familial Alzheimer's disease. Adapted from Bekris et al. (2010) [5] and Alzheimer Disease and Frontotemporal Dementia Mutation Database: http://www.molgen.ua.ac.be/ADMutations/default.cfm?MT=1&ML=6&Page=StatPerGene AD-Alzheimer's disease; EOFAD-Early-onset familial Alzheimer's disease.

AD Loci	Gene Symbol	Gene Name	Chromosome	% of EOFAD	Number of mutations	Number of families
AD1	*APP*	amyloid beta (A4) precursor protein	21q21.3	10-15	32	89
AD3	*PSEN1*	presenilin 1	14q24.3	18-50	185	405
AD4	*PSEN2*	presenilin 2 (Alzheimer disease 4)	1q31-q42	Rare	13	22

Table 2. Summary of published genome-wide association studies in Alzheimer's disease. Adapted from the AlzGene database: http://www.alzgene.org/largescale.asp; updated 18 April 2011. AD-Alzheimer's disease; GWAS-Genome-wide association study; SNP-Single nucleotide polymorphism.

Study	Design	Number of SNPs	Number of AD cases in GWAS (follow-up)	Number of controls in GWAS (follow-up)	Featured genes
Abraham et al. (2008) [20]	Case-control	561494	1082 (-)	1239 (1400)	*APOE*, *LRAT*
Beecham et al. (2009) [21]	Case-control	532000	492 (238)	496 (220)	*APOE*, *FAM113B*
Bertram et al. (2008) [22]	Family-based	484522	941 (1767)	404 (838)	*APOE*, *ATXN1*, *CD33*, GWA_14q31.2
Carrasquillo et al. (2009) [23]	Case-control	313504	844 (1547)	1255 (1209)	*APOE*, *PCDH11X*
Coon et al. (2007) [24]	Case-control	502627	664 (-)	422 (-)	*APOE*

Study	Design	Number of SNPs	Number of AD cases in GWAS (follow-up)	Number of controls in GWAS (follow-up)	Featured genes
Grupe et al. (2007) [25]	Case-control	17343	380 (1428)	396 (1666)	*ACAN*, *APOE*, *BCR*, *CTSS*, *EBF3*, *FAM63A*, *GALP*, *GWA_14q32.13*, *GWA_7p15.2*, *LMNA*, *LOC651924*, *MYH13*, *PCK1*, *PGBD1*, *TNK1*, *TRAK2*, *UBD*
Harold et al. (2009) [12]	Case-control	529205	3941 (2023)	7848 (2340)	*APOE*, *CLU*, *PICALM*
Heinzen et al. (2009) [26]	Case-control	Not available	331 (-)	368 (-)	*APOE*, *CHRNA7*
Hollingworth et al. (2011) [18]	Case-control	496763	6688 (13182)	13685 (26261)	*ABCA7*, *BIN1*, *CD2AP*, *CD33*, *CR1*, *EPHA1*, *MS4A4E*, *MS4A6A*
Hu et al. (2011) [27]	Case-control	509376	1831 (751)	1764 (751)	*APOE*, *BIN1*
Lambert et al. (2009) [13]	Case-control	537029	2032 (3978)	5328 (3297)	*APOE*, *CLU*, *CR1*

Study	Design	Number of SNPs	Number of AD cases in GWAS (follow-up)	Number of controls in GWAS (follow-up)	Featured genes
Lee et al. (2011) [28]	Case-control	627380	549 (2449)	544 (1390)	*DGKB*, *GWA_10q23.1*, *GWA_18q23*, *GWA_3q25.2*, *HPCAL1*
Li et al. (2008) [29]	Case-control	469438	753 (418)	736 (249)	*APOE*, *GOLM1*, *GWA_15q21.2*, *GWA_9p24.3*
Naj et al. (2010) [16]	Case-control	483399	931 (1338)	1104 (2003)	*APOE*, *MTHFD1L*
Naj et al. (2011) [17]	Case-control	2324889 (imputed)	8309 (3531)	7366 (3565)	*APOE*, *BIN1*, *CD2AP*, *CD33*, *CLU*, *CR1*, *EPHA1*, *MS4A4A*, *PICALM*
Poduslo et al. (2009) [30]	Case-control and Family-based	489218	9 (199)	10 (225)	*TRPC4AP*
Potkin et al. (2009) [31]	Case-control and quantitative trait	516645	172 (-)	209 (-)	*APOE*, *ARSB*, *CAND1*, *EFNA5*, *MAGI2*, *PRUNE2*, *TOMM40*
Reiman et al. (2007) [32]	Case-control	312316	446 (415)	290 (260)	*GAB2*
Seshadri et al. (2010) [15]	Case-control	2540000 (imputed)	3006 (6505)	22604 (13532)	*APOE*, *BIN1*, *CLU*, *EXOC3L2*, *PICALM*

Study	Design	Number of SNPs	Number of AD cases in GWAS (follow-up)	Number of controls in GWAS (follow-up)	Featured genes
Sherva et al. (2011) [33]	Case-control	2540000 (imputed)	124 (-)	142 (-)	*AGPAT1*, *ATP6V0A4*, *GLOD4*, *RGS6*, *TMEM132C*
Wijsman et al. (2011) [19]	Case-control and Family-based	565336	1848 (617)	1991 (573)	*APOE*, *CELF2*

C. Copy number variation

Recent advances in genome-wide technologies such as comparative genome hybridization (CGH), single nucleotide polymorphism (SNP) microarrays, and genome sequencing have led to identification of structural variants termed CNVs, ranging in size from one kilobase (kb) to several megabases (Mb) not previously identifiable by chromosome banding. These structural variants are present in variable copy number on comparing two or more genomes, and can include simple addition (copy number gains or duplications) or loss (copy number losses or deletions) of genetic material, or more complex rearrangements [34, 35]. The present work focuses only on deletions and duplications. CNVs can be inherited or sporadic; and may encompass one or more genes possibly affecting their function. The phenotypic effects of CNVs depend mainly on whether the genomic rearrangement affects dosage-sensitive genes or regulatory sequences. CNVs have been catalogued in a number of databases such as the Toronto Database of Genomic Variants (http://projects.tcag.ca/variation/) [36], DECIPHER (Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources; http://decipher.sanger.ac.uk/) [37] and ECARUCA (European Cytogeneticists Association Register of Unbalanced Chromosome Alterations; http://umcecaruca01.extern.umcn.nl:8080/ecaruca/ecaruca.jsp) [38].

D. Mechanisms for formation of copy number variations

The processes that could be involved in the formation of CNVs have been described in Stankiewicz and Lupski (2010) [35] and are summarized below. Figure 2 shows a representation of the recurrent and nonrecurrent rearrangements associated with genomic disorders.

1. Nonallelic Homologous Recombination

Low-copy repeats (LCRs) or segmental duplications are deoxyribonucleic acid (DNA) fragments that are >1 kb in size and have >90% DNA sequence identity. Many LCRs have complex structure and those that are >10 kb and have ~97% sequence identity can result in local genomic instability. Misalignment of chromosomes or chromatids that mediate nonallelic homologous recombination (NAHR) can result in unequal crossing-over [39], with recombination hotspots, gene conversion, and apparent minimal efficient processing segments. NAHR between directly oriented LCRs can result in deletions or reciprocal duplications of the genomic segment between them. When LCRs are inverted, NAHR leads to an inversion of the intervening genomic segment. When LCRs have a complex structure consisting of both direct and inverted subunits, they may be substrates for NAHR leading to genomic deletions/duplications and inversions respectively. The vast majority of the common-sized recurrent rearrangements, i.e. reciprocal deletions and duplications, or inversions, have been shown to be caused by this molecular mechanism.

2. Nonhomologous End Joining

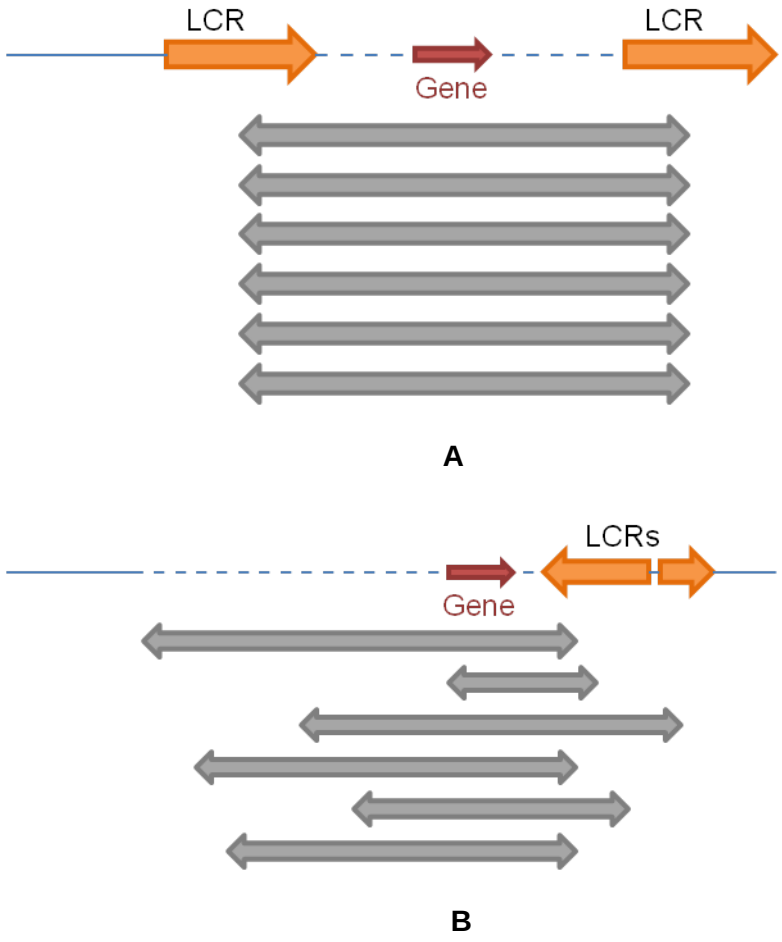
Nonrecurrent genomic rearrangements have been thought to arise by nonhomologous end joining (NHEJ), a recombination-based mechanism responsible for repair of DNA double strand breaks. In this mechanism, double strand breaks are detected; then both broken DNA ends undergo bridging, modification, and ligation [40]. Unlike NAHR, LCRs or minimal efficient processing segments are not required by NHEJ to mediate recombination. The process may also be stimulated by the genomic architecture. Often, additional nucleotides at the DNA end junction are contained in the product of repair leaving a “molecular scar” [41].

3. Replication-Error Mechanisms

A mechanism that has recently been shown to be involved in the origin of genomic-disorder-associated non-recurrent rearrangements that have a complex structure is fork stalling and template switching (FoSTeS), a mechanism based on DNA replication error [42, 43]. Here, the DNA replication fork can stall and the lagging strand is disengaged from the original template. It switches to another replication fork in physical proximity and by priming the new fork reinitiates DNA synthesis on the new fork. Hastings et al. (2009) based on experimental observations in human, yeast and other model organisms, proposed a generalized replicative template-switch model termed the microhomology-mediated break-induced replication (MMBIR) model, that may account for many of the structural variations seen in genomes and genes [44]. Based on the direction of fork progression and whether the lagging or leading strand in the new fork was used as a template and copied, the erroneously incorporated fragment from the new replication fork could be in the same direction or inverted relative to its original position. Also, the

template switching results in either a deletion or duplication depending on whether the new fork is located upstream or downstream of the original fork. FoSTeS/MMBIR has been proposed to play a major role in generating structural variation including nonrecurrent CNVs and complex genomic rearrangements in the human genome, and even in the formation of LCRs [45, 46].

Figure 2. Representation of the recurrent (A) and nonrecurrent (B) genomic rearrangements that play a role in genomic disorders. (A) For the more common recurrent rearrangements, the same-sized deletions and duplications have both breakpoints mapping (clustering) within the low-copy repeats (LCRs) that are directly oriented. (B) For the nonrecurrent rearrangements, the breakpoints are scattered. The breakpoints group in the vicinity of a LCR if it is present. Adapted from Stankiewicz and Lupski (2010) [35].



E. Methods for copy number variation detection

A number of methods have been developed to assess copy number within genomic CNVs. A review of these methods can be found in Fanciulli et al. (2010) [47] and are briefly summarized below. For the present work, the SNP arrays: Illumina Human610-Quad BeadChip and Affymetrix Genome-Wide Human SNP 6.0 Array were used.

1. Comparative genome hybridization arrays

CGH arrays have been widely used for CNV identification. To fabricate the arrays, a solid support (usually glass) is spotted with genomic DNA sequences. Different fluorescent markers are used to label the test and reference DNAs, which are then hybridized simultaneously to the array. The respective signal intensities are compared to assess the copy number. Constructing the arrays involves using DNA sequences ranging in size from 200 kb (bacterial artificial chromosome arrays) to 25 base pair (bp) (oligonucleotide array).

2. Single nucleotide polymorphism arrays

Arrays designed for the detection of SNPs can also be used to identify CNVs. This can be performed by determining differences in signal intensity independent of genotype. The current generation of SNP arrays are designed to provide greater genome-wide coverage and also to include non-polymorphic probes that have been optimized for measurement of copy number [48].

3. Quantitative real-time polymerase chain reaction

Quantitative real-time polymerase chain reaction (qPCR) is a commonly used method to screen targeted genomic regions for CNVs. It offers the advantages of being an efficient method for identification of deletions or duplications at single loci and also being a relatively high-throughput and technically straightforward assay. However, qPCR cannot be used for simultaneous amplification of many targets of interest in a single reaction, and as copy number estimates from these assays form a continuous distribution, it is also not possible to get precise integer measurements of gene copy number [49-51].

4. Paralogue ratio test

The paralogue ratio test offers an accurate and relatively high-throughput method to obtain gene copy number at targeted single loci [50, 51]. The target element whose copy number is being determined and another unlinked reference locus are simultaneously amplified in the same reaction tube by the same primer pairs. The two regions are distinguished by internal differences such as amplicon length or restriction digest, quantified and compared to determine the target copy number compared to the reference. Simultaneous amplification of target and control in the same reaction helps reduce experimental variability and enables better precision of the measurements, generating copy number estimates close to integer values.

5. Multiplex ligation-dependent probe amplification and multiple amplifiable probe hybridization

Simultaneous analysis of multiple genomic regions (up to 40 target sequences) can be performed using multiplex ligation-dependent probe amplification and multiple amplifiable probe hybridization, which are alternative target polymerase chain reaction (PCR)-based approaches [52]. In these approaches, oligonucleotide probes are used to generate locus-specific amplicons which can be resolved by capillary electrophoresis. Duplications are indicated by enhanced peak signals while deletions are indicated by reduced peaks. The techniques offer the advantages of being sensitive and simple to apply, and they can also be used for screening multiple targets in addition to their more common use for locus-specific studies.

6. Sequencing and genome assembly comparison

Structural variants have been successfully identified by alignment of DNA sequences from different sources. The development of high-throughput DNA sequencing protocols or so-called next-generation sequences has led to the generation of new assembles of complete genome sequences from single individuals enabling more robust and reliable genome comparisons and CNV identification when compared to earlier approaches [53, 54]. Next-generation sequencing platforms have led to the development of new computational methods for identifying structural variations. The paired-end read mapping is a commonly used approach [54, 55]. A library of fragments (typically 300-500 bp of genomic DNA) is generated, followed by massively parallel deep sequencing, generating millions of short sequence reads from a given sample to determine the 'paired-end spans'. Structural variations are identified by comparing the size of the paired-end spans

to a reference genome. The method permits the identification of inversions and provides CNV boundary resolution at the single nucleotide level. However, it can detect only insertions smaller than the average insert size of the library, and variants located within complex genomic regions cannot be reliably identified by the method [56]. A new approach that incorporates a novel CNV calling algorithm (Event Wise Testing) designed for ad hoc analysis of read depth has been recently proposed [57]. Estimation of coverage is done in non-overlapping intervals across the genome providing a quantitative measure of copy number. Deletion or duplication events are indicated by a decrease or increase in read depth across multiple consecutive genomic windows respectively. However, a limitation of the method is that it cannot determine balanced rearrangements or structural variations involving highly repetitive sequences. It also shows limitations in identifying novel insertions or their precise location. Paired-end mapping and read depth offer different and complementary advantages, thus using both approaches together in next-generation sequence data may enhance detection of different structural variations.

F. Copy number variation detection algorithms

A number of algorithms have been proposed for CNV detection using CGH and SNP arrays. These can be classified into several models such as smoothing methods, clustering methods, maximum likelihood procedures including Hidden Markov models (HMMs) and expectation-maximization algorithms. A review of these models can be found in Koike et al. (2011) [58]. A brief summary of the different models is given below.

For smoothing methods, the simplest method is to use a moving average for smoothing $\log_2$ratio profiles and detecting duplicated or deleted regions over the specified thresholds [59]. The intensity of the target probe divided by that of the reference probe is the $\log_2$ratio. A quantile smoothing method based on L1 norm (the sum of absolute values) penalty minimization [60] and a wavelet de-noising method [61] have been proposed as more sophisticated smoothing methods. The cluster along with chromosomes method was developed as a clustering method, in which calculation of hierarchical clustering trees along each chromosome arm (or chromosome) is performed and the 'interesting' clusters are selected after considering the false discovery rate [62]. Although effective in simulation data, smoothing and clustering methods do not achieve a CNV detection performance comparable with other methods in CGH array experimental data [63].

A number of maximum likelihood-related approaches have been proposed to date. Genetic local search algorithms (memetic algorithms) for maximizing the likelihood by considering the penalty function of breakpoints were introduced by Jong et al. (2003) [64]. An adaptive method for estimating the penalty constant was developed by Picard et al. (2005) to avoid selecting a very large segmentation number for over fitting the given

data. In this method, the probe intensity profile (log_2ratio) is assumed to be a Gaussian distribution, and maximizing the likelihood is used to estimate the number of segments [65]. Venkatraman and Olshen (2007) proposed a circular binary segmentation method, which also assumes the average probe intensity to have a Gaussian distribution [66]. This method introduces the likelihood ratio statistic for testing the null hypothesis where there is no change, and the alternative hypothesis where there is exactly one change at an unknown location. A permutation test is used and the hypothetical change-points are adopted if the null hypothesis is rejected. The change-points are recursively searched using overlapping windows [66].

An HMM is a statistical model where the system is thought to follow a Markov process [67-69]. In most HMM models, it is assumed that the probe intensity values or Log R Ratio (LRR, $\log_2(R_{observed}/R_{expected})$, where R is the sum of probe intensities and $R_{expected}$ is obtained from linear interpolation of canonical genotype clusters) and B Allele Frequency (BAF, a normalized measure of the relative signal intensity ratio of the B and A alleles on the SNP array) or genotypes are independent. The copy number states of the probes are assigned to be hidden states with certain transition probabilities. The copy number state of each probe is obtained by maximizing the likelihood of observed data (probe intensity, LRR and BAF, or genotypes).

Previous studies have compared different platforms and algorithms [58, 70-73]. However, each algorithm has its own set of parameters that need to be fine-tuned according to the data to obtain the best possible results [71]. Unfortunately, most software developers provide little or no guidance for evaluating and choosing optimal parameter settings for their algorithms. Even when identical raw data is used as the input, the quantity and quality of CNV calls can be different depending on the algorithm

used [72]. Thus, it has been suggested to try multiple algorithms and take the union of copy number regions for downstream association analyses, which may improve sensitivity [71-73]. It has also been suggested to use software designed specifically for the platform that was used to generate the data, as it has been shown that algorithms typically perform better when they have been developed specifically for a certain data type as compared to algorithms that are platform-independent or software that has been readapted for newer versions of an array.

G. Copy number variation in selected neuropsychiatric disorders

CNVs have been shown to play a role in various neuropsychiatric disorders (Table 3). Representative studies for some of these disorders are summarized below.

Table 3. Examples of copy number variations and selected neuropsychiatric disorders.

Disorder	Possible gene/loci affected by copy number variation	Reference(s)
Mendelian disorders		
Williams-Beuren syndrome	7q11.23	Peoples et al. (2000) [74]
Williams-Beuren-region duplication syndrome	7q11.23	Berg et al. (2007) [75]
Smith-Magenis syndrome	17p11.2	Chen et al. (1997) [76]
Potocki-Lupski syndrome	17p11.2	Potocki et al. (2007) [77]
22q11.2 deletion syndrome (DiGeorge syndrome and velocardiofacial syndrome)	22q11.2	Edelmann et al. (1999) [78]
22q11.2 duplication syndrome	22q11.2	Ensenauer et al. (2003) [79]
Miller-Dieker lissencephaly syndrome	17p13.3	Cardoso et al. (2003) [80]
Complex disorders		
Alzheimer's disease	*APP*	Rovelet-Lecrux et al. (2006) [81] Kasuga et al. (2009) [82]
Attention-deficit hyperactivity disorder	*ASTN2*, *TRIM32* *A2BP1*, *AUTS2*, *CNTNAP2*, *IMMP2L*	Lionel et al. (2011) [83] Elia et al. (2010) [84]
Autism	*NRXN1* 16p11.2	Autism Genome Project Consortium et al. (2007) [85] Weiss et al. (2008) [86]
Bipolar disorder	*GRM7*, *CNTNAP2*, *COMT*, *GNB1L*	Zhang et al. (2009) [87]
Epilepsy	*AUTS2*, *CNTNAP2*	Mefford et al. (2010) [88]

Disorder	Possible gene/loci affected by copy number variation	Reference(s)
Parkinson disease	*SNCA* *PARK2*	Singleton et al. (2003) [89] Pankratz et al. (2011) [90]
Schizophrenia	*NRXN1* 16p13.1	Rujescu et al. (2009) [91] Ingason et al. (2011) [92]
Tourette syndrome	*NRXN1*, *CTNNA3*	Sundaram et al. (2010) [93]

1. Autism

Autism spectrum disorder (ASD) is a group of complex neurodevelopmental disorders characterized by social impairments, communication difficulties, and restricted, repetitive, and stereotyped patterns of behavior. The most severe form of ASD is autistic disorder, also called autism or classical ASD. Other conditions include a milder form known as Asperger syndrome, and childhood disintegrative disorder and pervasive developmental disorder not otherwise specified (National Institute of Neurological Disorders and Stroke Autism Fact Sheet: http://www.ninds.nih.gov/disorders/autism/detail_autism.htm).

A number of studies have examined the role of CNVs in ASD. The Autism Genome Project Consortium (2007) performed linkage and CNV analyses in 1181 families with at least two affected participants [85]. The authors were able to implicate candidate loci including the 11p12-p13 region and neurexins. Marshall et al. (2008) in their genome-wide assessment for structural abnormalities identified novel loci at *DPP6-DPP10-PCDH9* (synapse complex), *ANKRD11*, *DPYD*, *PTCHD1* and the 15q24 regions among others to play a possible role in ASD susceptibility [94]. Their results further implicated the *SHANK3-NLGN4-NRXN1* postsynaptic genes. Microdeletions and microduplications at 16p11.2 [86, 95] and genomic rearrangements involving the *NRXN1* gene [96, 97] are thought to play a role in autism susceptibility. Glessner et al. (2009) performed a whole-genome CNV study on a cohort of 859 ASD cases and 1409 healthy children of European ancestry, and evaluated positive findings in an independent cohort of 1336 ASD cases and 1110 controls of European ancestry [98]. The authors were able to identify new susceptibility genes encoding neuronal cell-adhesion molecules including the *NLGN1* and *ASTN2* genes enriched with CNVs in ASD cases compared to controls.

They also observed CNVs within or surrounding genes involved in the ubiquitin pathways including the *UBE3A, PARK2, RFWD2 and FBXO40* genes not observed in controls.

In a CNV analysis in 912 multiplex families from the Autism Genetics Resource Exchange collection and 1488 healthy controls, Bucan et al. (2009) were able to observe rare variants including exonic deletions at the *NRXN1* gene and whole gene duplications encompassing *UBE3A* and several other genes in the 15q11-q13 region [99]. Other genes such as the *BZRAP1* and *MDGA2* genes were also identified. In their cohort of 996 ASD participants of European ancestry and 1287 matched controls, Pinto et al. (2010) were able to identify numerous de novo and inherited CNVs implicating many novel ASD genes such as the *SHANK2*, *SYNGAP1* and *DLGAP2* genes, and the X-linked *DDX53-PTCHD1* locus [100]. They also observed an enrichment of CNVs that may play a disruptive role in functional gene sets involved in cellular proliferation, projection and motility, and GTPase/Ras signaling. In 42 extended families with ASD, Salyakina et al. (2011) identified regions on 7p21.3, 15q24.1, 3p26.3 and 12q24.32 that could be associated with ASD [101]. In a recent rare genome-wide CNV analysis in 1124 autism families, Sanders et al. (2011) were able to find de novo duplications of the 7q11.23 region to be significantly associated with ASD [102]. They were also able to identify rare de novo CNVs at additional regions including 1q21.1, 15q13.2-13.3, 16p13.2 (encompassing the *USP7* and *C16orf72* genes) and the *CDH13* locus. Cumulative data provided evidence for the association of rare de novo events at 7q11.23, 15q11.2-13.1, 16p11.2 and *NRXN1* regions with ASD. Rare microdeletions overlapping *NRXN3* exons have also been observed in ASD-affected individuals [103]. Thus, these studies highlight the potential role of CNVs in ASD.

2. Bipolar disorder

Bipolar disorder, also known as manic-depressive illness, is a brain disorder that results in unusual changes in mood, energy, activity levels and the ability to carry out day-to-day tasks. The symptoms of bipolar disorder can be severe resulting in damaged relationships or affecting job or school performance (National Institute of Mental Health Bipolar Disorder: http://www.nimh.nih.gov/health/publications/bipolar-disorder/complete-index.shtml).

A few studies have examined CNVs in bipolar disorder. Wilson et al. (2006) in a CNV analysis of post-mortem brain DNA from bipolar disorder cases, schizophrenia cases and controls (35 individuals in each category), were able to identify aberrations at four loci [104]. The aberrant loci contained the genes encoding the *EFNA5*, *GLUR7*, *CACNG2* and *AKAP5* proteins. All of these proteins are expressed in the brain and have potential roles in neuronal function. In a different study, a CNV in the *GSK3β* locus on 3q13.3 was found with an increased frequency in bipolar disorder participants compared to controls [105]. The gene codes for glycogen synthase kinase, a key component of the Wnt signaling pathway. Zhang et al. (2009) performed a genome-wide CNV analysis in 1001 bipolar and 1033 control participants [87]. Genes disrupted by singleton deletions in their cases were found to be significantly overrepresented in pathways categorized as important for psychological disorders and behaviors. The authors identified *GRM7*, *CNTNAP2*, *COMT* and *GNB1L* as genes of interest. Recently, Priebe et al. (2011) performed a genome-wide CNV analysis in 882 participants with bipolar disorder and 872 population-based controls [106]. Two common CNVs on the 10q11 and 6q27 regions were found to be overrepresented in bipolar disorder participants who had an early age-at-onset (≤21 years) compared with controls. The authors suggested an

influence of CNVs on the development of early-onset, but not late-onset bipolar disorder, and provided support for the hypothesis of an etiological difference between early-onset and late-onset bipolar disorder. In a genome-wide analysis of de novo CNVs in a cohort of 788 trios, Malhotra et al. (2011) observed a significant enrichment of de novo CNVs in bipolar disorder and schizophrenia cases compared to controls [107]. De novo CNVs were found to be enriched in early-onset bipolar disorder cases (age-at-onset≤18 years).

3. Parkinson disease

Parkinson disease (PD) is part of a group of conditions called motor system disorders caused by the loss of dopamine-producing brain cells. It generally affects people over the age of 50. The primary symptoms of the disease are tremor (trembling in hands, arms, legs, jaws and face), rigidity (stiffness of the limb and trunk), bradykinesia (slowness of movement), and postural instability (impaired balance and coordination) (National Institute of Neurological Disorders and Stroke Parkinson's Disease Information Page: http://www.ninds.nih.gov/disorders/parkinsons_disease/parkinsons_disease.htm).

A causal association for PD has been obtained for at least five genes: *SNCA*, *PARK2*, *PINK1*, *DJ-1* and *LRRK2*. Simple mutations (missense, nonsense, silent, splice site, and untranslated region mutations), small insertions and deletions, and CNVs of these five genes have been shown to result in PD. A review of the role of these genes can be found in Nuytemans et al. (2010) [108] and Crosiers et al. (2011) [109]. Duplications and triplications of the *SNCA* gene locus have been described in individuals affected with PD. A more severe, early-onset form of PD is associated with triplications of the *SNCA* gene [89, 110], whereas the phenotype associated with *SNCA* gene duplication resembles the typical late-onset idiopathic PD [111, 112]. In a genome-wide study of

CNVs in 273 PD participants and 275 controls, Simon-Sanchez et al. (2008) identified CNVs within the *PARK2* locus in both PD participants and controls [113]. In a recent CNV study in 816 cases and 856 controls, Pankratz et al. (2011) were able to replicate the association of CNVs of the *PARK2* gene with PD [90]. They also identified genome-wide significant CNVs in two novel genes (*DOCK5* and *USP32*) associated with an increase in PD risk. These studies suggest a role of CNVs in PD susceptibility.

4. Schizophrenia

Schizophrenia is a chronic, severe, and disabling brain disorder that affects approximately one percent of Americans. The symptoms of schizophrenia include hallucinations, delusions, thought disorders, and cognitive symptoms (National Institute of Mental Health Schizophrenia: http://www.nimh.nih.gov/health/publications/schizophrenia/complete-index.shtml).

A number of studies have been performed to determine the possible role of CNVs in schizophrenia and a few of these studies are mentioned here. Steffanson et al. (2008) identified three deletions on 1q21.1, 15q12.2 and 15q13.3 significantly associated with schizophrenia in two large samples of 1433 cases and 33250 controls, and 3285 cases and 7951 controls [114]. In a genome-wide survey of rare CNVs in 3391 participants with schizophrenia and 3181 controls, the International Schizophrenia Consortium (2008) observed deletions within the 22q11.2 region critical for velo-cardial-facial syndrome, as well as deletions on 15q13.3 and 1q21.1 [115]. Deletions in the 22q11.2 region have also been reported in other studies [116, 117], suggesting a possible role of this region in schizophrenia. Kirov et al. (2008) in a study of 93 schizophrenia and 372 control participants identified a deletion in the 2p16.3 region disrupting the *NRXN1* gene and a

duplication in the 15q13.1 region spanning the *APBA2* gene; both genes encode proteins that play a role in synaptic development and function [118]. A larger CNV study in 2977 schizophrenia participants and 33746 controls also identified the disruption of the *NRXN1* gene to be associated with schizophrenia [91]. Support for the role of CNVs at the 16p13.1, 1q21.1 and *NRXN1* regions in schizophrenia were also obtained in a Japanese study of 575 schizophrenia and 564 control participants [119]. CNVs in the 16p13.1 region which includes the candidate genes *NTAN1* and *NDE1* have been recently shown to confer risk of schizophrenia in a study of 4345 schizophrenia and 35079 control participants [92].

Another region identified to be a schizophrenia risk region is the 3q29 region. Deletions in this region have been identified in two studies: one study involving 245 schizophrenia and 490 control participants [120], and the other study involving 3945 schizophrenia or schizoaffective disorder participants and 3611 screened comparison participants [117]. The later study also confirmed the association of deletions in the 1q21.1, 15q13.3, and 22q11.1 regions, duplications in the 16p11.2 region, and *NRXN1* gene deletions with schizophrenia. Duplications in the vasoactive intestinal peptide receptor gene *VIPR2* were also identified in this study as well as in a different study of 8290 cases and 7432 controls [121]. Grozeva et al. (2012) conducted a study to examine the frequencies of nine schizophrenia-associated CNV loci in 10259 individuals from the UK Wellcome Trust Case Control Consortium with non-psychiatric disorders [122]. The authors found a significantly higher frequency of deletions at 1q21.1, 3q29, 15q11.2, 15q13.1, 22q11.2, and duplications at 16p11.2 in schizophrenia cases compared to the non-psychiatric controls. These studies show that regions affected by CNVs may play an important role in schizophrenia susceptibility. A further review can be found in Lee et al. (2011) [123].

5. Tourette syndrome

Tourette syndrome (TS) is a neurodevelopmental disorder characterized by repetitive, stereotyped, involuntary movements and vocalizations called tics. Approximately 200000 Americans have the most severe form of TS, and as much as one in 100 show milder and less complex symptoms such as chronic motor or vocal tics (National Institute of Neurological Disorders and Stroke Tourette Syndrome Fact Sheet: http://www.ninds.nih.gov/disorders/tourette/detail_tourette.htm).

Sundaram et al. (2010) performed a genome-wide CNV analysis in 111 participants with TS and 73 ethnically matched controls [93]. They were able to identify five exon-affecting rare CNVs that were either de novo or recurrent in ten TS participants. Genes/loci in three of the five CNVs have been implicated by studies in other neurodevelopmental disorders including schizophrenia, autism, and attention-deficit hyperactivity disorder. In a recent case-control CNV analysis of 460 individuals with TS and 1131 controls, Fernandez et al. (2012) observed an enrichment of genes within histamine receptor signaling pathways as well as axon guidance, cell adhesion, nervous system development, and synaptic structure and function processes [124]. The authors also identified three large de novo events that they thought were likely pathogenic, including one disrupting multiple gamma-aminobutyric acid receptor genes. A significant overlap of genes mapping within rare CNVs in TS was observed with those identified in autism spectrum disorders. The two studies thus showed an overlap of genes within rare CNVs in TS with those identified in other neurodevelopmental disorders.

H. Copy number variation in Alzheimer's disease

1. Early-onset Alzheimer's disease

Rovelet-Lecrux et al. (2006) identified a duplication of the *APP* locus on chromosome 21 in five French families with autosomal dominant early-onset AD and cerebral amyloid angiopathy [81]. A Dutch family having an autosomal dominant segregation pattern, and neuropathology compatible with AD and cerebral amyloid angiopathy was also identified to have an *APP* locus duplication [125]. Kasuga et al. (2009) examined the occurrence of the *APP* locus duplication in a Japanese AD cohort consisting of familial and early-onset sporadic cases [82]. *APP* locus duplications were identified in two unrelated early-onset familial AD families. A significantly higher *APP* mRNA expression level was observed in participants with the *APP* locus duplication in the peripheral blood when compared to age- and sex-matched controls. Although *APP* locus duplications were not identified in a screen of Swedish and Finnish participants with early-onset AD [126], a Swedish patient with early-onset AD was recently reported carrying an *APP* locus duplication [127]. Thus, a possible role for duplications at the *APP* locus exists in early-onset AD warranting further examination. In order to determine if rare CNVs could play a role in autosomal dominant early-onset AD families without mutations in the *APP*, *PSEN1* and *PSEN2* genes as well as in rare sporadic young-onset AD cases, Rovelet-Lecrux et al. (2011) performed a genome-wide CNV study in 21 unrelated autosomal dominant early-onset AD cases and 12 sporadic AD cases, with an onset age younger than 55 years [128]. Their analysis identified seven singleton CNVs, four of which target genes (*KLK6*, *SLC30A3*, *MEOX2*, and *FPR2*) that encode proteins relating to amyloid-beta peptide signaling or metabolism. The authors suggest that the results of their study provide novel support for the amyloid cascade hypothesis.

2. Late-onset Alzheimer's disease and Mild cognitive impairment

To our knowledge, only four studies other than the present work have investigated the role of CNVs in late-onset AD. Heinzen et al. (2010) performed a genome-wide scan of AD in 331 cases evaluated with a clinical diagnosis of dementia (>80% had a clinical diagnosis of AD) and 368 controls [26]. Although no CNVs were found to be significant, the authors identified a duplication in the *CHRNA7* gene which they thought warranted further investigation. In a follow-up analysis of the *CR1* region in a Flanders-Belgian cohort, Brouwers et al. (2011) identified a low-copy repeat associated CNV in the *CR1* region, producing different CR1 isoforms, CR1-F and CR1-S [129]. Significant association was obtained in carriers of CR1-S, and the authors were able to replicate this finding in a French cohort. In a case-only genome-wide CNV association study, Shaw et al. (2011) identified a chromosomal region on 14q11.2 encompassing a cluster of olfactory receptors to be associated with age of onset of AD [130]. Ghani et al. (2012) conducted a genome-wide scan for large CNVs among Caribbean Hispanics and identified a ~470 kb duplication on 15q11.2 to be nominally associated with AD [131]. To our knowledge, no study has investigated the role of CNVs in MCI.

I. Statement of purpose

AD is the most common form of dementia and genetic variation represents one of the major risk factors in its development and progression. Recent genetic studies have identified a number of AD risk loci, but these do not account for all of the genetic variation associated with the disease. It is possible that other forms of genetic variation, such as CNVs, may be involved in disease susceptibility. CNVs, which are DNA regions present in variable copy number, have been implicated in a number of neuropsychiatric disorders such as autism, Parkinson disease and schizophrenia. However, at the time of starting of this work, only one published study had examined the role of CNVs in AD [26] and there were no published studies in MCI. The overall goal of this work was to determine if CNVs could be possible genetic risk factors in the development of AD and MCI. This would be accomplished through the following specific aims:

I. Perform an initial CNV analysis in non-Hispanic Caucasian individuals from the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort who had DNA samples extracted from peripheral blood.

A. Compare the CNV burden between cases (AD and MCI participants) and controls.

B. Identify through case/control association analyses genomic regions where CNVs were detected in cases but not in controls.

II. Perform an initial CNV analysis using similar approaches as the ADNI study in non-Hispanic Caucasian individuals from the National Institute of Aging-Late Onset

AD/National Cell Repository for AD (NIA-LOAD/NCRAD) Family Study who had DNA samples extracted from blood or brain tissue.

A. Compare the CNV burden between cases (AD participants) and controls.

B. Perform case/control association analyses to characterize genomic regions where CNVs were detected in cases but not in controls.

C. Identify genomic regions also reported in the ADNI study to determine potential candidate AD regions.

III. Perform an initial CNV analysis using similar approaches as the ADNI and NIA-LOAD/NCRAD Family studies in Caucasian individuals from a unique cohort of clinically characterized and neuropathologically defined cases (AD participants) and controls who had DNA samples extracted from brain tissue.

A. Compare the CNV burden between cases and controls.

B. Identify through case/control association analyses genomic regions overlapped by CNVs in cases but not in controls.

C. Identify genomic regions reported in the ADNI and NIA-LOAD/NCRAD Family studies that replicate in this study to determine potential AD candidate regions.

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

A. 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 beta-amyloid protein fragments (plaques) and twisted strands of the tau protein (tangles). Amnestic 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 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% [2].

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 [4]. Early-onset AD (onset earlier than 60 or 65 years) accounts for a small percentage (one to six percent) of cases and is primarily caused by mutations in three genes that affect the cerebral levels of beta-amyloid peptide: *APP* (amyloid beta (A4) precursor protein) on chromosome 21, *PSEN1* (presenilin 1) on chromosome 14 and *PSEN2* (presenilin 2 (Alzheimer disease 4)) on chromosome 1 [132]. Late-onset AD accounts for the majority of AD cases, but only the ε4 allele of the *APOE* (apolipoprotein E) gene on chromosome 19 has been consistently replicated across studies. At the time of performing this work, three large genome-wide association studies (GWASs) identified

five additional loci: *CLU* (clusterin), *CR1* (complement component (3b/4b) receptor 1 (Knops blood group)), *PICALM* (phosphatidylinositol binding clathrin assembly protein), *BIN1* (bridging integrator 1), and *EXOC3L2* (exocyst complex component 3-like 2) to be strongly associated with AD [12, 13, 15]. These loci also showed strong association in replication studies [14], 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 [34]. At the time of performing this work, the role of CNVs in late-onset AD has only been examined in one study [26]. 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 cases, were significant at genome-wide threshold, Heinzen et al. (2010) were able to identify a duplication in the *CHRNA7* gene warranting further investigation. At the time of performing this work, 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 late-onset AD. 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.

B. Methods

1. Alzheimer's Disease Neuroimaging Initiative

The ADNI was launched in 2003 by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies, and nonprofit organizations, as a $60 million, multiyear public-private partnership. The Principal Investigator of this initiative is Michael W. Weiner, MD, 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 three years, 400 people diagnosed with MCI to be followed for three years, and 200 people diagnosed with early AD to be followed for two years [133]. Longitudinal imaging, including structural 1.5 Tesla MRI scans collected on the full sample and [^{11}C]Pittsburgh Compound-B- and [^{18}F]fluorodeoxyglucose-PET imaging on a subset, and performance on neuropsychological and clinical assessments were collected at baseline and at follow-up visits in six-to-twelve intervals. Other biomarkers are also available including *APOE* and whole genome genotyping on the full ADNI sample, and longitudinal cerebrospinal fluid 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 [134] and at http://www.adni-info.org.

2. Participants

Participants in the present analysis 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 [135, 136]. Current diagnoses were downloaded from the ADNI database as 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 this analysis is publicly available on the ADNI website (http://adni.loni.ucla.edu). The focus of ADNI is on incident late-onset AD. To our knowledge, no participants in the present study carry a known causal mutation [5].

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 (Phoenix, AZ) using the Illumina Human610-Quad BeadChip as previously described [31, 137]. As indicated by the manufacturer's documentation, the Human610-Quad BeadChip contains 620901 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 collection) population analyzed here with a median marker spacing of 2.7 kb. In addition, 27635 markers are included in "unSNPable" regions likely to contain CNVs that are not easily assessed by SNPs. Coverage is provided for 3938 CNV regions (184064 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 standard deviation (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 the Translational Genomics Research Institute using BeadStudio software (Illumina, Inc., CA). In January 2010, we reprocessed the array data using GenomeStudioV2009.1, and this dataset will be made available on the ADNI website in a follow-up data release.

4. Inference of the Log R Ratio and B Allele Frequency

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 $R=X_{norm}+Y_{norm}$ and composition (copy angle) $\theta=(2/pi)*\arctan(Y_{norm}/X_{norm})$, where X_{norm} and Y_{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=$\log_2$(normalized R value/expected R value) for the SNP.

Linear interpolation of the R value at the SNP's θ 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 θ 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:

BAF=0 if $\theta<\theta_{AA}$

$=0.5*(\theta-\theta_{AA})/(\theta_{AB}-\theta_{AA})$ if $\theta<\theta_{AB}$

$=0.5+0.5*(\theta-\theta_{AB})/(\theta_{BB}-\theta_{AB})$ if $\theta<\theta_{BB}$

$=1$ if $\theta\geq\theta_{BB}$,

where θ_{AA}=mean θ value of all genotypes in AA cluster plotted in polar normalized coordinates, θ_{AB}=mean θ value of all genotypes in AB cluster plotted in polar normalized coordinates, and θ_{BB}=mean θ value of all genotypes in BB cluster plotted in polar normalized coordinates (GenomeStudio Genotyping Module v1.0 User Guide).

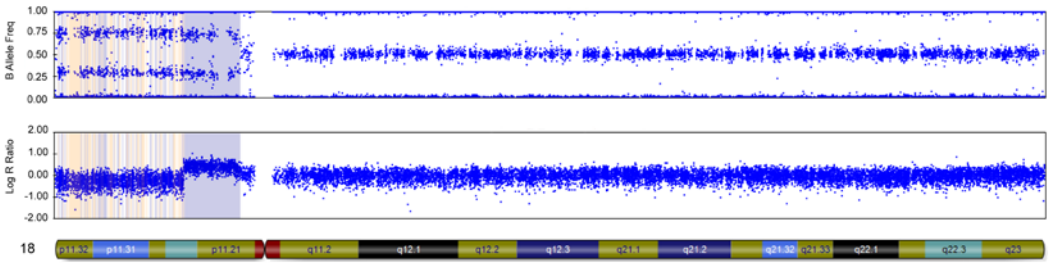
5. Generation of copy number variation calls and quality control

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) model [68], 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 [138], 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 (Figure 3) and was excluded from further analysis as it may be a mosaic sample [139]. 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 hemizygosity 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 more than 50% overlap with centromeric, telomeric, and immunoglobulin regions as defined in Need et al. (2009) [140] were excluded. 501 participants (AD=222, MCI=136, HC=143) passed all QC checks and included in further CNV analyses.

Figure 3. Representative image of B Allele Frequency and Log R Ratio of the participant in the ADNI cohort who had multiple deletions and duplications on chromosome 18. The orange shaded portion indicates regions with deletions and purple shaded portion indicates regions with duplications (Human Genome Build 36.1).



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 (http://pngu.mgh.harvard.edu/~purcell/plink/) [141] 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 (http://www.alzgene.org) [8] 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 17938 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 analyzed.

Representative plots of CNV calls (Figure 6) were generated in UCSC Genome Browser (http://genome.ucsc.edu/) [142] (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 (Figures 3 to 5 and 7) were generated using the Illumina Genome Viewer plugin within GenomeStudio (Human Genome Build 36.1).

C. Results

1. Description of copy number variation 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 4 and 5.

A total of 6737 CNV calls (4746 deletions and 1991 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 ($p<0.05$) when evaluated by permutation. We also evaluated whether CNV burden was higher in cases than controls in the *APOE* ε4 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 ($p<0.05$; data not shown). A large proportion of deletions and duplications were found in the 0.1-0.5 Mb size range (Table 6). Two AD participants were found to have very large CNV calls (>2 Mb) (Figures 4 and 5). One AD participant had a deletion on chromosome 4 (Figure 4), which includes the following genes: *NDST4* (N-deacetylase/N-sulfotransferase 4), *TRAM1L1* (translocation-associated membrane protein 1-like 1), and *MIR1973* (microRNA 1973). The other AD participant had a duplication on chromosome 11 (Figure 5), which includes the gene *LUZP2* (leucine zipper protein 2).

Table 4. Sample demographics of participants in the ADNI cohort.

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±SD)	75.10±7.27	75.88±7.17	75.83±5.32	not significant
Years of education (Mean±SD)	15.30±3.05	15.85±3.01	16.24±2.62	0.009
APOE group (ε4 negative/ε4 positive)	73/149	70/66	108/35	<0.001
Age at onset (Mean±SD)	74.08±7.73	-	-	-

Table 5. Characteristics of copy number variation calls in the three diagnostic groups from participants in the ADNI cohort. CNV-Copy number variation.

	Alzheimer's disease (n=222)	Mild cognitive impairment (n=136)	Healthy controls (n=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

Table 6. Participants in the ADNI cohort grouped by copy number variation call size. Del-Deletions; Dup-Duplications.

Call size	Alzheimer's disease (n=222)		Mild cognitive impairment (n=136)		Healthy controls (n=143)	
	Del n (%)	Dup n (%)	Del n (%)	Dup n (%)	Del n (%)	Dup n (%)
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)

Figure 4. Representative image of B Allele Frequency and Log R Ratio of the Alzheimer's disease participant in the ADNI cohort who had a deletion >2 Mb on chromosome 4. The orange shaded portion indicates the deleted region (Human Genome Build 36.1).

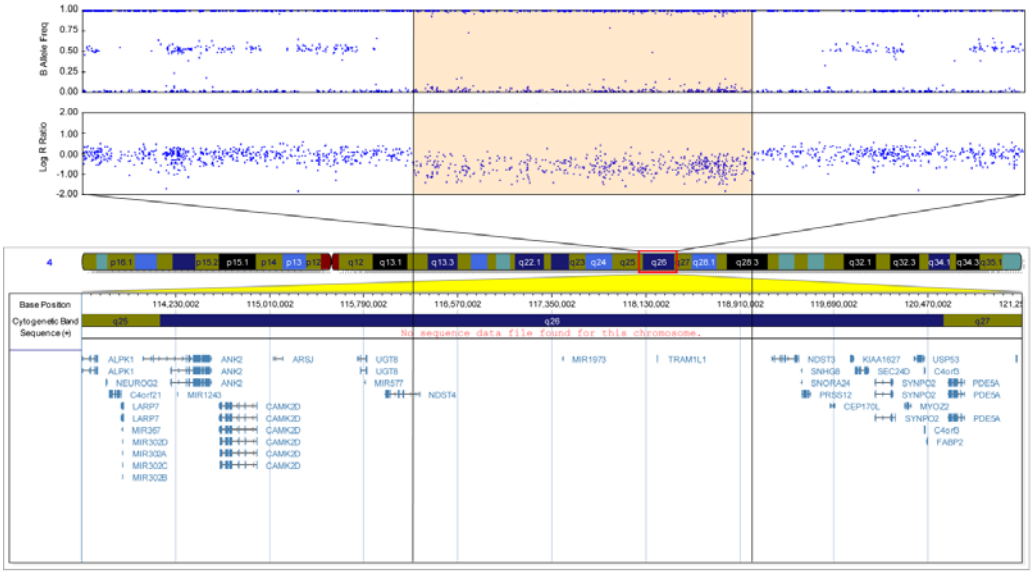
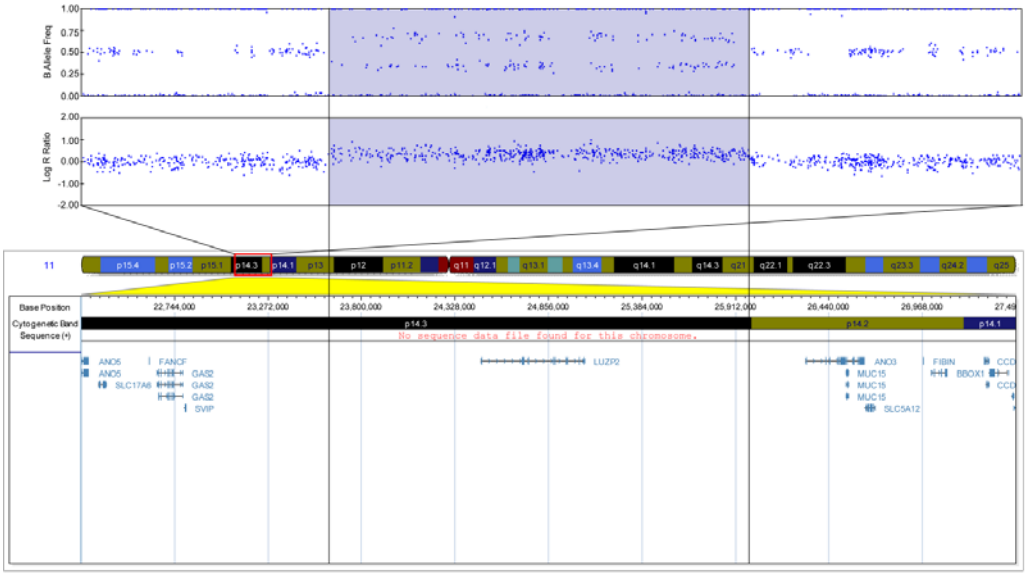


Figure 5. Representative image of B Allele Frequency and Log R Ratio of the Alzheimer's disease participant in the ADNI cohort who had a duplication >2 Mb on chromosome 11. The purple shaded portion indicates the duplicated region (Human Genome Build 36.1).



2. Case/control association analyses

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 7 for reference although these did not meet conventional significance ($p<0.05$). 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 6 shows representative plots of two of these genes (*CHRFAM7A* and *LRRTM3*).

Table 7. Genes that had copy number variation calls from at least one Alzheimer's disease and/or one mild cognitive impairment participant and no healthy controls in the ADNI cohort using the candidate gene approach. AD-Alzheimer's disease; Chr-Chromosome; MCI-Mild cognitive impairment; N/A-Not available; [a]Age at onset of AD symptoms, available only for participants with a baseline diagnosis of AD; [b]The same participant had CNV calls overlapping the two genes.

Chr	Region	Start (bp)	End (bp)	AD (n)	*APOE* status	Age at onset[a]	MCI (n)	*APOE* status
5	*PPP2R2B*	145949260	146441226	1	ε3/ε3	N/A	0	-
6	*ATXN1*	16407321	16869700	1	ε3/ε4	83 years	0	-
7	*MAGI2*	77484309	78920826	1	ε2/ε4	N/A	0	-
7	*RELN*	102899472	103417198	0	-	-	1	ε4/ε4
9	*GRIN3A*	103371455	103540683	1	ε3/ε3	74 years	0	-
10	*LRRTM3*	68355797	68530873	1	ε3/ε3	55 years	0	-
10	*LIPA*	90963305	91001640	0	-	-	1[b]	ε3/ε3
12	*PPM1H*	61324030	61614932	1	ε2/ε3	N/A	1	ε2/ε3
15	*CHRFAM7A*	28440734	28473156	2	ε3/ε3 ε3/ε4	N/A N/A	2	ε3/ε3 ε3/ε4
15	*ADAM10*	56675801	56829469	1	ε3/ε3	N/A	1[b]	ε3/ε3
21	*DNAJC28*	33782107	33785893	1	ε3/ε4	74 years	0	-
21	*DOPEY2*	36458708	36588442	0	-	-	1	ε3/ε4
22	*GSTT1*	22706138	22714284	1	ε3/ε3	59 years	0	-

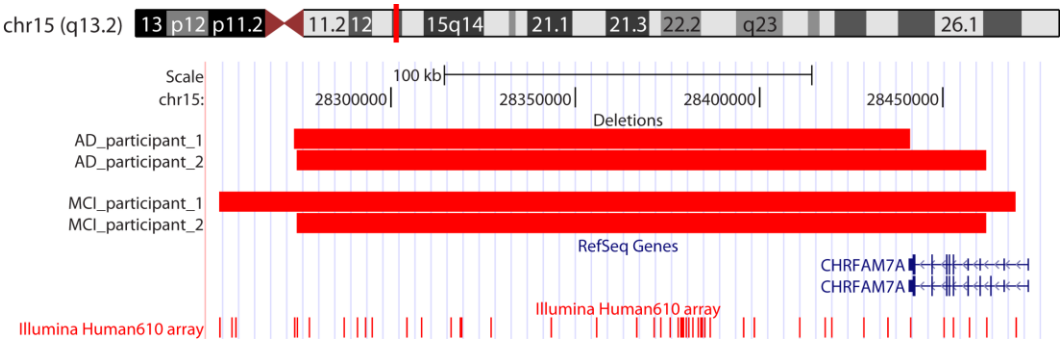
2.2. Whole genome approach

We also identified CNV calls present in cases (AD and/or MCI) but not controls (HC) within regions overlapping 17938 genes. There was no significant ($p<0.05$) gene after correction for multiple testing. We, therefore, focused on genes that had an uncorrected $p<0.05$. 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 8). We also observed CNVs overlapping two genes associated with neuropsychiatric disorders: *NRXN1* (neurexin 1) [85, 91] and *ERBB4* (v-erb-a erythroblastic leukemia viral oncogene homolog 4 (avian)) [143], but these did not reach significance ($p<0.05$). 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 7).

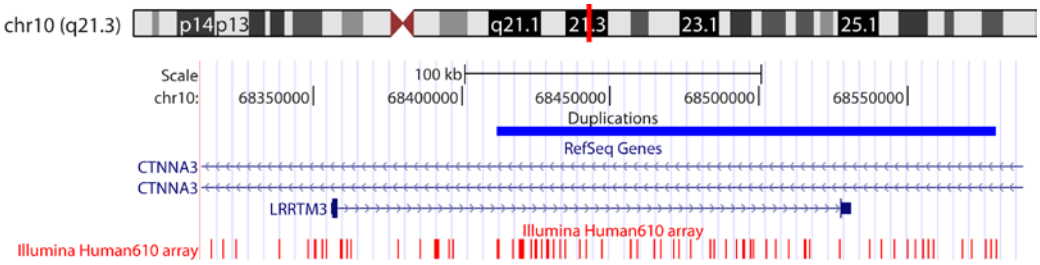
Table 8. Significant (uncorrected p<0.05, relative to healthy controls) genes present in either Alzheimer's disease and/or mild cognitive impairment participants, but not healthy controls in the ADNI cohort using the whole genome approach. AD-Alzheimer's disease; Chr-Chromosome; MCI-Mild cognitive impairment.

Chr	Region	Start (bp)	End (bp)	AD calls (n)	AD calls (p value)	MCI calls (n)	MCI calls (p value)
8	*CSMD1*	2780281	4839736	9	0.0114	4	0.0556
1	*HNRNPCL1*	12829847	12831165	6	0.0493	4	0.0549
11	*SLC35F2*	107166926	107234864	5	0.0820	6	0.0120

Figure 6. Candidate genes *CHRFAM7A* (A) and *LRRTM3* (B), overlapped by copy number variation calls from at least one Alzheimer's disease and/or mild cognitive impairment participant, but no healthy controls in the ADNI cohort. The red rectangles represent deletions, and the blue rectangles represent duplications (March 2006 (NCBI36/hg18) assembly).

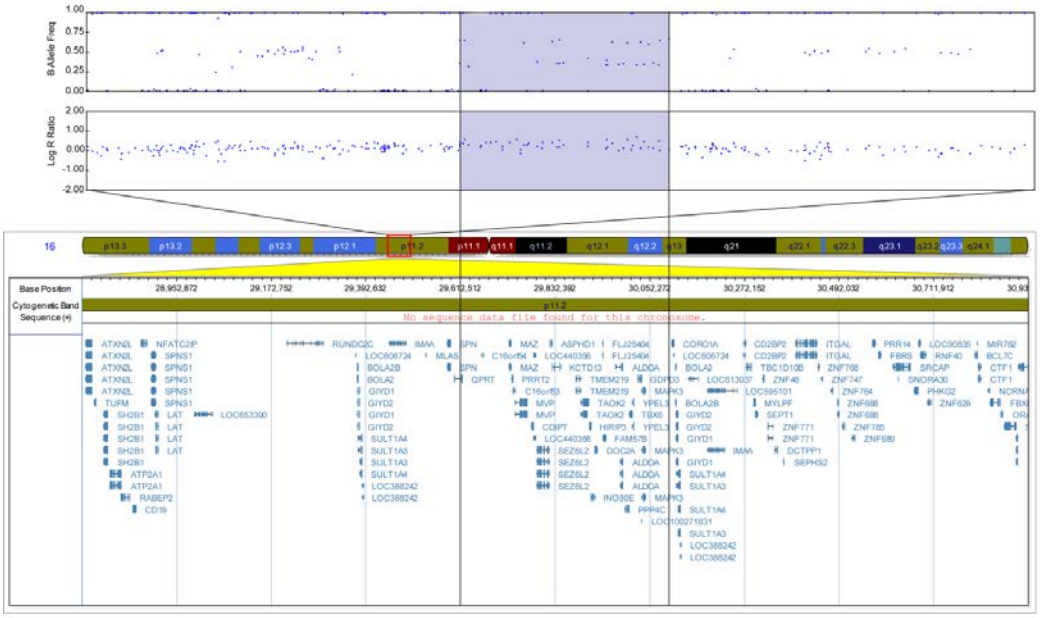


A



B

Figure 7. Representative image of B Allele Frequency and Log R Ratio of the participant in the ADNI cohort who had a duplication at 16p11.2. The purple shaded portion indicates the duplicated region (Human Genome Build 36.1).



D. Discussion

The present report represents an initial analysis of CNVs in the ADNI dataset and is the first CNV analysis of participants 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 [26]. Two AD participants were found to have CNV calls >2Mb. One AD participant had a duplication on chromosome 11 (Figure 5) 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 [144]. The authors of this study also found this gene to be deleted in some participants with Wilms tumor-aniridia-genitourinary anomalies-mental retardation syndrome. Another AD participant had a deletion on chromosome 4 (Figure 4), which includes the following genes: *NDST4* (N-deacetylase/N-sulfotransferase 4), *TRAM1L1* (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 or molecular biology techniques such as quantitative real-time polymerase chain reaction (qPCR) 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 7 and Figure 6). 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) [145]. The *CHRFAM7A* gene contains a polymorphism consisting of a 2-bp deletion at position 497-498 of exon 6. This 2-bp polymorphism has been associated with schizophrenia [146]. The *CHRFAM7A* genotype without the 2-bp allele has also been shown to be significantly over-represented in AD ($p=0.011$), dementia with Lewy bodies ($p=0.001$), and Pick's disease ($p<0.0001$) participants [147].

Heinzen et al. (2010) 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 [26]. 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 [26], 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 $p<0.05$, as shown in Table 8. *CSMD1* (CUB and Sushi multiple domains 1) has been shown to be primarily synthesized in the developing central nervous system and epithelial tissues [148]. It is enriched in the nerve growth cone, suggesting that it may be an important regulator of complement activation and inflammation in the developing central nervous system. *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 (avian)) 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 than binds neuroligins. The Ca^{2+}-dependent neurexin-neuroligin complex is present in the central nervous system at synapses and is required for efficient neurotransmission and formation of synaptic contacts [149]. This gene has been found to have reduced expression with AD severity [150], and its disruption has been shown to be associated with schizophrenia [91, 119, 140] and autism [85, 98]. 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 [151]. Interestingly, an SNP (rs6463843) flanking the *NXPH1* (neurexophilin 1) gene was identified by our group in a GWAS of neuroimaging phenotypes in the ADNI cohort [152]. The *NXPH1* 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 [153]. One AD participant had a large duplication that included 23 genes in the 16p11.2 region (Figure 7). CNVs in this region have previously been associated with autism [86, 94, 95], schizophrenia [154], cognitive impairment and speech/language delay [155], and obesity [156, 157], 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 cerebrospinal fluid 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 study were generated from DNA samples derived only from peripheral blood-78 participants

whose DNAs were derived from lymphoblastoid cell lines were excluded. Lymphoblastoid cell lines are generated by transforming peripheral B lymphocytes by the Epstein-Barr virus. Epstein-Barr virus-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 [136]. 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 the 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 SD 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. (2010). 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) [70]. 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 cerebrospinal fluid and Pittsburgh Compound-B-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 cerebrospinal fluid and amyloid PET data, enabling further examination of this issue.

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.

III. Analysis of Copy Number Variation in Alzheimer's Disease: the NIA-LOAD/NCRAD Family Study

A. 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. Amnestic 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 [4]. Early-onset AD (onset earlier than 60 or 65 years), which accounts for one to six percent 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 (onset later than 60 or 65 years) is the *APOE* ε4 allele [158]. A member of a three allele haplotype (composed of ε2, ε3 and ε4 alleles), the ε4 allele shows a dose-dependent increase in AD risk of approximately four-fold in carriers as compared to noncarriers [9-11].

Recently, other AD risk loci have been identified and replicated including: *CLU*, *CR1*, *PICALM*, *BIN1*, *EXOC3L2*, *MTHFD1L*, *MS4A4A/MS4A6E*, *CD2AP*, *CD33*, *ABCA7* and *CUGBP2* [12-19]. 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 one 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 [34]. At the time of performing this work, only four studies have investigated the role of CNVs in late-onset AD [26, 129, 130, 159]. Heinzen et al. (2010) performed a case-control genome-wide scan of AD and identified a duplication in the *CHRNA7* gene that they thought warranted further investigation [26]. In an in-depth analysis of the *CR1* region, Brouwers et al. (2011) observed a low-copy repeat associated CNV in *CR1*, that produced different CR1 isoforms, CR1-F and CR1-S [129]. 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. (2011) demonstrated that a chromosomal region on 14q11.2, encompassing a cluster of olfactory receptors, is associated with age at onset of AD [130].

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 (described in Chapter II and [159]). 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-Late Onset AD/National Cell Repository for AD (NIA-LOAD/NCRAD) Family Study [19] 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.

B. 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.

1. Participants

Recruitment information for participants in the NIA-LOAD Family Study and NCRAD has been previously described [19]. 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 late-onset AD in a first degree relative, whereas the University of Kentucky controls were not excluded if they had a family history of late-onset AD. Genome-wide genotyping for all samples was performed at the Center for Inherited Disease Research (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.

2. Alzheimer's Disease Neuroimaging Initiative

The 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 [137, 159]. 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.

3. 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.

3.1. 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 were excluded because cell line transformation may bias CNV results [135, 136]. 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.

3.2. 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 the National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria for probable or possible AD [160], or meeting NINCDS-ADRDA criteria for definite AD when clinical and pathological information was available, or the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) pathological criteria [161] 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 [162]. 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.

3.3. 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.

4. Generation of copy number variation 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. (2011) [159]. 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 (2010Jun22 version) (http://www.openbioinformatics.org/penncnv/) [68], 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 (SDs) of the LRR. A comparison of these methods has been performed by Dellinger et al. (2010) [70]. 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 [138] 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 90^{th} 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 hemizygosity 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. (2009) [140] 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.

5. Case/control association analyses

Case/control association analyses were performed using CNV calls generated for the AD participants and controls using PLINK v1.07 (http://pngu.mgh.harvard.edu/~purcell/plink/) [141] 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 (Updated 5 January 2011) (http://www.alzgene.org/) [8] 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 (Figure 10) were created in UCSC Genome Browser (http://genome.ucsc.edu/) [142] (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 (Figures 8, 9, 11 and 12).

C. Results

1. Sample demographics and copy number variation call characteristics

The sample demographics of the 882 participants (711 AD and 171 controls) are shown in Table 9. 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* ε4 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 10.

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 11). Two AD participants were observed to have very large CNV calls (>2 Mb) (Figures 8 and 9). The first AD participant had a 2.4 Mb deletion on chromosome 11 (Figure 8), which includes many genes. The second AD participant had a 3.2 Mb duplication on chromosome 3 (Figure 9), which includes the *GBE1* gene.

Table 9. Sample demographics of participants in the NIA-LOAD/NCRAD Family Study cohort.

	Alzheimer's disease	Controls	p (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
APOE group (ε4 negative/ ε4 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 10. Characteristics of copy number variation calls from participants in the NIA-LOAD/NCRAD Family Study cohort. CNV-Copy number variation.

		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 11. Participants in the NIA-LOAD/NCRAD Family Study cohort grouped by copy number variation 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)

Figure 8. Representative image of B Allele Frequency and Log R Ratio of the Alzheimer's disease participant in the NIA-LOAD/NCRAD Family Study cohort who had a deletion >2 Mb on chromosome 11 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 single nucleotide polymorphism. The orange shaded portion indicates the deleted region (Human Genome Build 36.1).

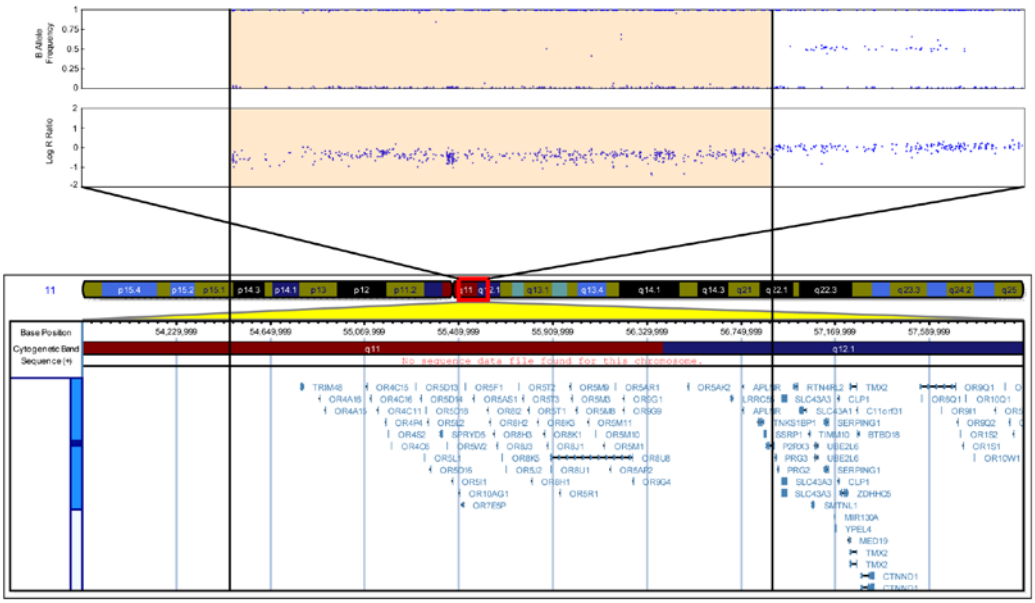
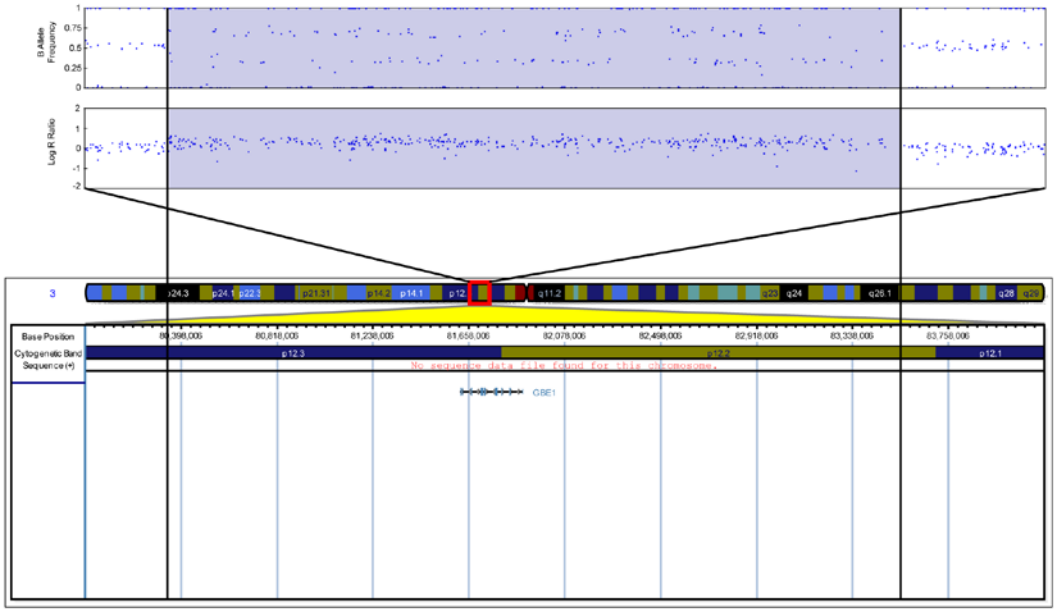


Figure 9. Representative image of B Allele Frequency and Log R Ratio of the Alzheimer's disease participant in the NIA-LOAD/NCRAD Family Study cohort who had a duplication >2 Mb on chromosome 3 including chromosomal coordinates and the genes in the corresponding regions. Each blue dot in the B Allele Frequency and Log R Ratio plots represents a single nucleotide polymorphism. The blue shaded portion indicates the duplicated region (Human Genome Build 36.1).



2. Case/control association analyses

2.1. 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 presented in Table 12 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 [159] was also identified in 12 AD participants and one control in the present study (Figure 10).

Table 12. Genes overlapped by copy number variation calls from at least one Alzheimer's disease participant and no controls in the NIA-LOAD/NCRAD Family Study cohort using the candidate gene approach. [a]Genes also identified in the Alzheimer's Disease Neuroimaging Initiative Study as being overlapped by copy number variation calls in Alzheimer's disease and/or mild cognitive impairment participants, but not controls; [b]One Alzheimer's disease participant had copy number variation calls overlapping the two genes; [c]A different Alzheimer's disease participant had copy number variation calls overlapping the two genes.

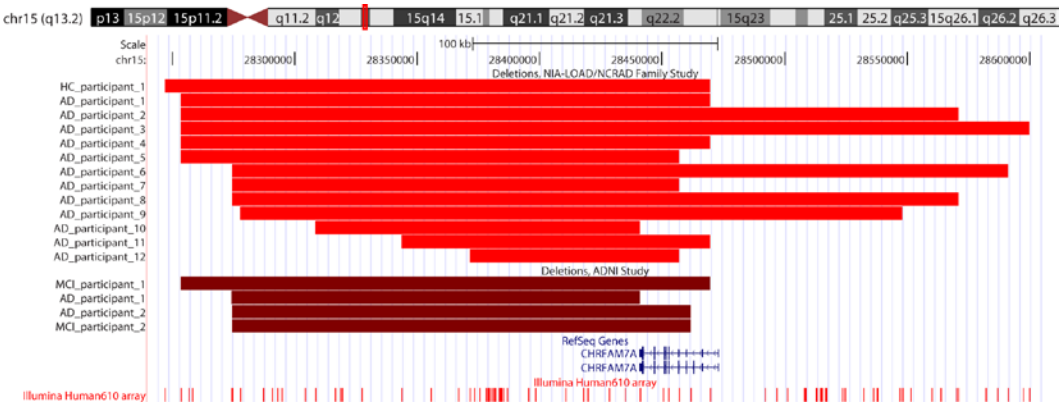
Chromosome	Region	Start (bp)	End (bp)	Number of Alzheimer's disease participants
1	*CFH*	194887630	194983257	1
6	*ATXN1*[a]	16407321	16869700	3
6	*HLA-A*	30018309	30021633	2
6	*MICA*	31479349	31491069	4
6	*HLA-DQA1*	32713160	32719407	5[b]
6	*HLA-DOA*	33079937	33085367	1[c]
6	*HLA-DPB1*[a]	33151737	33162954	1[c]
7	*CD36*	80069439	80146529	1
7	*RELN*[a]	102899472	103417198	1
9	*APBA1*	71235021	71477042	4
9	*ABCA1*	106583104	106730257	3
9	*RXRA*	136358230	136472252	1
10	*ABCC2*	101532452	101601652	1
11	*PICALM*	85346132	85457756	1
15	*CYP19A1*	49287545	49418087	1
15	*CHRNA3*	76674705	76700377	1
15	*MEF2A*	97956184	98071524	1

Chromosome	Region	Start (bp)	End (bp)	Number of Alzheimer's disease participants
17	*TP53*	7512444	7531588	1
17	*COX10*	13913443	14052721	1
17	*SREBF1*	17656110	17681050	3[b]
17	*CCL3*	31439715	31441619	2
17	*KIF18B*	40358973	40380608	1
18	*DSC1*	26963211	26996817	1
21	*NCAM2*	21292503	21833085	1
21	*APP*	26174731	26465003	1
21	*DOPEY2*[a]	36458708	36588442	2
21	*KCNJ6*	37918656	38210566	2
22	*COMT*	18309308	18336530	2
22	*BCR*	21852551	21990224	1
22	*GSTT1*[a]	22706138	22714284	1

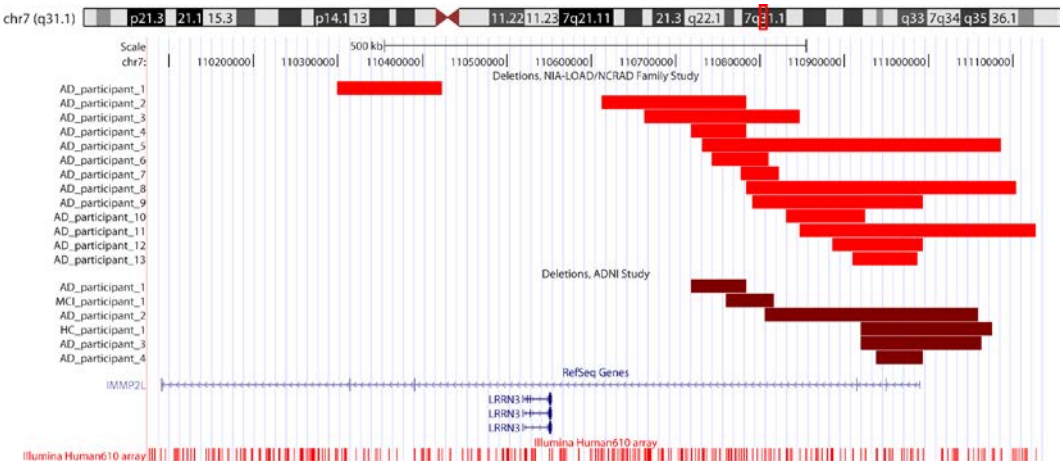
2.2. Genome-wide approach

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) (Figure 10). 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 (Figure 11) and another AD participant was observed to have a deletion in the 17p12 region (Figure 12).

Figure 10. 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.



A



B

Figure 11. Representative image of B Allele Frequency and Log R Ratio of the Alzheimer's disease participant in the NIA-LOAD/NCRAD Family Study cohort who had a duplication at 16p13.11 along with chromosomal coordinates and genes in corresponding regions are shown. Blue dots in the B Allele Frequency and Log R Ratio plots represent single nucleotide polymorphisms. The blue shaded portion indicates the duplicated region (Human Genome Build 36.1).

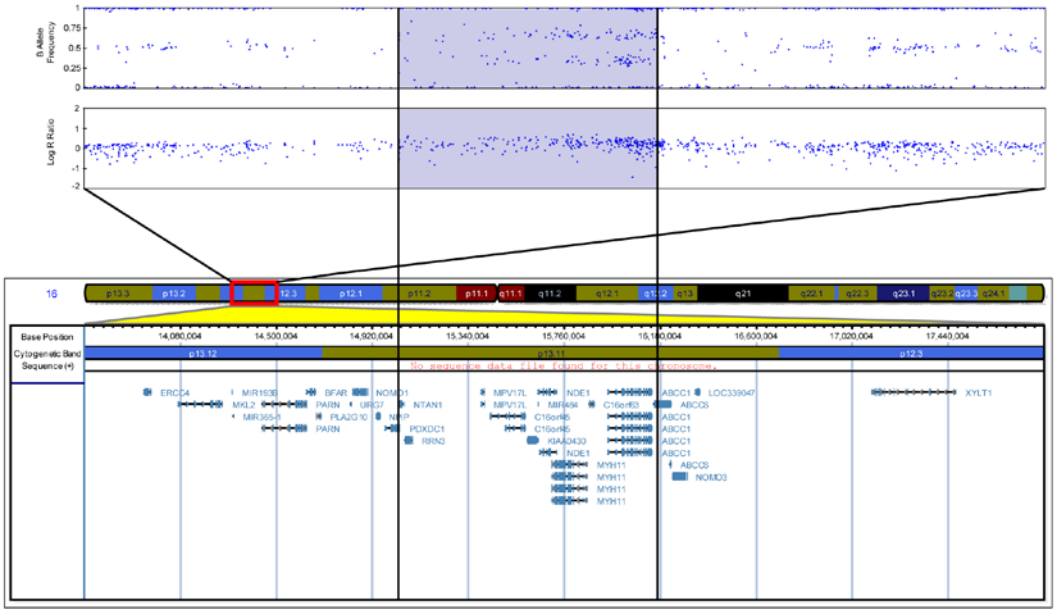
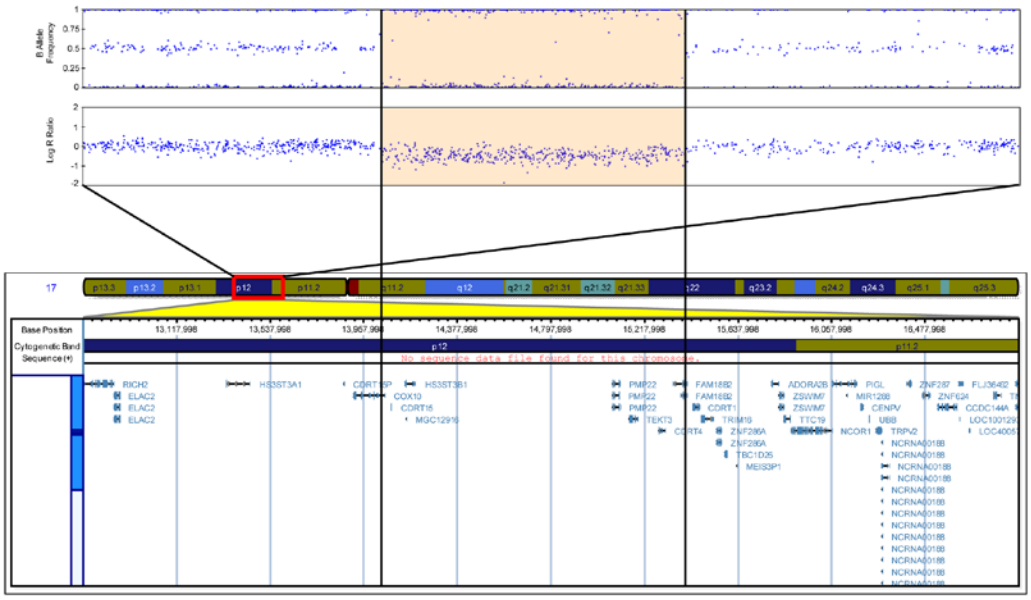


Figure 12. Representative image of B Allele Frequency and Log R Ratio of the Alzheimer's disease participant in the NIA-LOAD/NCRAD Family Study cohort who had a deletion at 17p12 along with chromosomal coordinates and genes in corresponding regions are shown. Blue dots in the B Allele Frequency and Log R Ratio plots represent single nucleotide polymorphisms. The orange shaded portion indicates the deleted region (Human Genome Build 36.1).



D. 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) (Figures 8 and 9). The first AD participant had a 2.4 Mb deletion on chromosome 11 (Figure 8), 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 [163]. 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 [164] and an increase in odor identification deficits has been shown in MCI participants compared to participants without MCI [165]. 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 [130]. The second AD participant had a 3.2 Mb duplication on chromosome 3 (Figure 9), 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 [166]. 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 12). 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 (Figure 10). 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 [145, 167]. 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 [146], 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 [147]. Although *CHRFAM7A* is transcribed, its translation and the possible role of the resulting

protein is uncertain. Recently, de Lucas-Cerrillo et al. (2011) cloned and expressed the full-length coding sequence of the *CHRFAM7A* transcript in pituitary-derived GH4C1 cells and oocytes [168]. On performing a functional study of the protein in oocytes, they observed a dominant negative regulatory function of the protein on α7 nicotinic acetylcholine receptors (α7 nAChRs) activity through reduction in the number of functional α7 nAChRs incorporated into the oocyte surface. Based on these and other results, they suggested that the *CHRFAM7A* gene product could possibly modulate α7 subunit receptor-mediated synaptic transmission and cholinergic anti-inflammatory response. Another recent study by Araud et al. (2011) suggests *CHRFAM7A* to be a dominant negative modulator of *CHRNA7* function and important for receptor regulation in humans [169]. In order to determine how α7 nAChR activation would affect *APP* processing, Nie et al. (2010) constructed a SH-EP1-α7 nAChR-hAPP695 cell line model co-expressing α7 nAChR gene and human amyloid precursor protein 695 (hAPP695) gene [170]. Their results demonstrated that, by regulating γ-secretase activity, activation of α7 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 α7 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-DPB1*, *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 and amyloid-beta 42 levels by potentiating beta-secretase processing of amyloid precursor protein [171]. *ATXN1* and the related Brother of *ATXN1* (*BOAT1*) 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 [172]. An evidence of association has been suggested between an intronic SNP (rs179943) in this gene and AD [173]. The *HLA-DPB1* (major histocompatibility complex, class II, DP beta 1) also on chromosome 6 is a member of the human leukocyte antigen (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 [174, 175]. 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 [176]. 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 [177]. 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 candidate gene [178]. The gene has been shown to be differentially

expressed and overexpressed in Down syndrome brains and it is thought that its overexpression could play an important role in the neurological phenotypes and mental retardation in Down syndrome 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 [179, 180].

The genome-wide approach revealed CNV calls overlapping the *IMMP2L* (IMP2 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 (Figure 10). Located on chromosome 7, *IMMP2L* is a catalytic subunit of the mitochondrial inner membrane peptidase complex [181]. The mitochondrial inner membrane peptidase 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 (TS) who carried a de novo inverted duplication of a segment of the long arm of chromosome 7 [182]. Recently, it has been identified as being disrupted in another patient with TS who had a de novo translocation between chromosomes 2 and 7 [183]. 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 [184]. 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 [185]. 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 [149]. 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 [150]. The gene has also been associated with autism [85], autism spectrum disorders [98] and schizophrenia [91]. 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 [186]. 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 [187]. 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 [188]. 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 [189]. 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 (Figure 11). Duplications at 16p13.11 have been observed to be significantly enriched in individuals with neurocognitive disease [190], attention-deficit hyperactivity disorder [191] and schizophrenia [92] 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 (Figure 12). Although this region is a possible schizophrenia candidate loci [116, 192], 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 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-transformed peripheral B lymphocytes, in the postimmortal stage, develop strong telomerase activity and aneuploidy, as well as gene mutations and reprogramming [136]. These changes may lead to biological artifacts [135, 193], 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.

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-DPB1*, *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.

IV. Analysis of Copy Number Variation in Alzheimer's Disease in a Cohort of Clinically Characterized and Neuropathologically Verified Individuals

A. Introduction

Alzheimer's disease (AD) is the most common form of dementia characterized by loss of memory and other cognitive abilities, severe enough to disrupt 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]. No treatments at present can slow or halt its progression. Amnestic mild cognitive impairment (MCI) is a clinical condition in which a person has memory problems not normal for his/her age, but not severe enough to interfere significantly with daily functioning. Approximately 14-18% of individuals aged 70 years and older have MCI, and every year 10-15% of these individuals will likely progress to dementia, particularly AD [2].

Genetic factors play a key role in AD development accounting for approximately 58-79% of the phenotypic variation [4]. Mutations in *APP*, *PSEN1* and *PSEN2* primarily cause early-onset AD (age at onset<60 or 65 years) [5]. The leading genetic risk factor for the more common late-onset AD (age at onset>60 or 65 years) is the *APOE* ε4 allele. Large case-control genome-wide association studies (GWASs) have identified and replicated other AD risk loci including: *CLU*, *CR1*, *PICALM*, *BIN1*, *EXOC3L2*, *MTHFD1L*, *MS4A4A/MS4A6E*, *CD2AP*, *CD33*, *ABCA7* and *CUGBP2* [12-19]. However it is estimated that the *APOE* ε4 allele accounts for approximately 20% and the non-*APOE* loci cumulatively account for as much as 35% of the AD risk [17, 194]. Thus the loci do not explain all the genetic variation associated with AD, and other forms of genetic variation such as copy number variations (CNVs) may play a role.

CNVs are DNA regions (one kilobase (kb) to several megabases (Mb) in size) that have differences in copy number. These can result in the addition (copy number gains or duplications) or loss (copy number losses or deletions) of genetic material. CNVs often encompass a single gene or multiple genes and may affect their function [34]. The role of CNVs in late-onset AD has been investigated in prior studies [26, 129-131]. Previously, we analyzed the role of CNVs in AD and MCI using data from participants in the Alzheimer's Disease Neuroimaging Initiative (ADNI) study (described in Chapter II and [159]) and the National Institute of Aging-Late Onset AD/National Cell Repository for AD (NIA-LOAD/NCRAD) Family Study (described in Chapter III and [195]). For both studies, DNA extracted either from peripheral blood or brain tissue were used. Case/control association analyses including candidate gene and genome-wide approaches were performed to determine genes overlapped by CNVs only in cases (AD and/or MCI) but not controls. A number of genes were identified in the two studies including *ATXN1*, *CHRFAM7A*, *CSMD1*, *DOPEY2*, *ERBB4*, *GSTT1*, *HLA-DPB1*, *HNRNPCL1*, *IMMP2L*, *NRXN1*, *RELN* and *SLC35F2*.

In the present report, we analyzed the role of CNVs in AD using data from a unique cohort of clinically characterized and neuropathologically defined cases (AD) and controls (TGen cohort) [196]. All DNA samples were extracted from brain tissue. Case/control association analyses similar to the two previous studies were performed to determine the CNV burden in cases relative to controls and genes overlapped by CNVs detected in cases but not controls.

B. Methods

1. Samples

The TGen cohort included samples extracted from brain tissue of 1617 Caucasian individuals (1022 AD cases and 595 controls). Recruitment information for the participants has been previously described [196]. Briefly, the United States cohort was obtained from 21 National Institute on Aging-supported Alzheimer's Disease Center brain banks and from the Miami Brain Bank [197, 198]. Cohorts from other brain banks in the United States, United Kingdom, and the Netherlands were obtained similar to the original United States cohort. De-identification of samples was done before receipt, and the study met human studies institutional review board and the Health Insurance Portability and Accountability Act of 1996 regulations. The present work is declared not human-subjects research and is institutional review board exempt under regulation 45 CFR 46. Genome-wide genotyping for all samples was performed using the Affymetrix Genome-Wide Human SNP 6.0 Array (Santa Clara, CA, USA) as previously described [196]. *APOE* genotyping was done using Crook et al.'s (1994) method [199] or using a fluorescence-based allele-specific polymerase chain reaction (PCR), also called PCR Amplification of Specific Alleles, on array tape [200] by PreventionGenetics (Marshfield, WI, USA).

The ADNI data used in the preparation of the present report were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.ucla.edu) as described in Chapter II and [159]. ADNI's primary goal is to test whether imaging markers, genetic markers, other biological markers, and clinical and neuropsychological assessments can be combined to measure progression of MCI and early AD. More

information on ADNI can be found on http://www.adni-info.org/. The Illumina Human610-Quad BeadChip (San Diego, CA, USA) was used to perform genome-wide genotyping of the ADNI sample as previously described [137, 159]. The *APOE* polymorphisms (rs429358 and rs7412) were genotyped separately.

The NIA-LOAD/NCRAD Family Study data used in the present report 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 (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000168.v1.p1) website as described in Chapter III and [195]. Recruitment information for NIA-LOAD Family Study and NCRAD participants has been previously described [19]. Genome-wide genotyping for all samples was performed using the Illumina Human610-Quad BeadChip at the Center for Inherited Disease Research (Baltimore, MD, USA). The *APOE* polymorphisms (rs429358 and rs7412) were genotyped at PreventionGenetics.

2. Generation of copy number variation calls and quality control

CNV calls were generated for the 1617 TGen samples using PennCNV (2011Jun16 version; http://www.openbioinformatics.org/penncnv/) [68], a Hidden Markov model based program. The PennCNV-Affy protocol (http://www.openbioinformatics.org/penncnv/penncnv_tutorial_affy_gw6.html) for the Affymetrix Genome-Wide Human SNP 6.0 Array was first performed to transform raw CEL files into a signal intensity file containing the Log R Ratio (LRR) and B Allele Frequency (BAF) values used by PennCNV to generate CNV calls. The Hidden Markov

model “affygw6.hmm”, population frequency of B allele “affygw6.hg18.pfb” and gcmodel “affygw6.hg18.gcmodel” files were used. Extensive quality control (QC) was performed on all samples. A genomic wave adjustment procedure using PennCNV’s gcmodel file was applied as samples that have below optimal genomic wave QC values can be considered unreliable [138]. Frequency distribution plots of the number of CNV calls, LRR standard deviation (SD), BAF Drift and Waviness Factor (WF) were made. A sample was excluded if at least one of the above measures for the sample was greater than 90th percentile of the frequency distribution, i.e. the sample had >56 CNV calls, LRR SD>0.38, BAF Drift>0.01 or WF>0.02.

Due to complications of hemizygosity in males and X-chromosome inactivation in females, analyses were restricted to autosomes. To ensure we were including only high-confidence CNVs 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 <10, CNVs called based on data <10 SNPs, and CNVs that had >50% overlap with centromeric, telomeric, and immunoglobulin regions as defined in Need et al. (2009) [140] were excluded. CNV calls were not filtered for size because both large and small variants could be of potential significance. A case sample observed to have a very large (~8.4 Mb) deletion on chromosome 19, and a control sample observed to have a very large (~22.4 Mb) duplication on chromosome 1, were excluded from the analyses as they may be possible outliers. The ~8.4 Mb deletion on chromosome 19 encompassed both sides of the centromere, but did not overlap any RefSeq or UCSC Genes according to the UCSC Genome Browser (http://genome.ucsc.edu/) [142]. 1166 samples (728 cases, 438 controls) with 31045 CNV calls remained after all QC measures and were entered into case/control association analyses.

3. Case/control association analyses

Case/control association analyses in the TGen study were performed similar to the ADNI [159] and NIA-LOAD/NCRAD Family [195] studies. PLINK v1.07 (http://pngu.mgh.harvard.edu/~purcell/plink/) [141] was used to investigate CNV call differences between cases (AD) and controls. Two approaches were used: a candidate gene approach including 317 AD genes identified from the AlzGene database (Updated 5 January 2011) (http://www.alzgene.org/) [8] as having a positive association with AD in at least one study, and a genome-wide approach using 17938 genes from PLINK's gene list (hg18 coordinates). The AlzGene database is a publicly available online resource that provides a comprehensive and regularly updated catalog of genetic case/control and family association studies in AD. In both approaches, CNV segments either partially or completely overlapping genes were analyzed. The analyses included both deletions and duplications.

4. Meta-analysis

We performed a meta-analysis for the *CHRFAM7A* gene using results from the ADNI, NIA-LOAD/NCRAD Family and TGen studies to determine differences in frequency of CNV calls overlapping the gene between cases (AD or MCI) and controls. A fixed-effects model was run and a summary odds ratio (OR) was calculated using the Mantel Haenszel method. MetaAnalyst Beta 3.13 (http://tuftscaes.org/meta_analyst/index.html) [201] and Comprehensive Meta-Analysis Version 2 [202] were used for the meta-analysis and generation of the forest plot (Figure 15). The UCSC Genome Browser (http://genome.ucsc.edu/) [142] (March 2006 (NCBI36/hg18) assembly) was used to create representative plots of the CNV calls (Figures 13 and 14). The Genome Browser

track for the Affymetrix Genomewide 6.0 array was obtained from the PennCNV website (http://www.openbioinformatics.org/penncnv/penncnv_download.html).

C. Results

1. Sample demographics and copy number variation call characteristics

The sample demographics and CNV call characteristics of the 728 cases and 438 controls who passed all QC measures are shown in Tables 13 and 14. Significant ($p<0.05$; two-sided) differences in gender, absence or presence of the *APOE* ε4 allele, age at death, Braak stage and the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) score were observed between cases and controls. 31045 CNV calls (24188 deletions and 6857 duplications) were observed with an average of 45 SNPs per CNV call and an average CNV call length of 64.76 kb. A higher CNV call rate and a lower average CNV call size were observed in deletions compared to duplications. Rates of deletions and duplications did not significantly differ between cases and controls. There were no significant differences in the rates of deletions and rates of duplications when males and females were analyzed separately (data not shown). A large proportion of deletions and duplications were found in the 0.1-0.5 Mb size range (Table 15).

Table 13. Sample demographics in the TGen cohort. CERAD-The Consortium to Establish a Registry for Alzheimer's Disease.

	Cases	Controls	p (two-sided)
Number of samples	728	438	-
Gender (Males/Females)	180/548	274/164	<0.001
APOE group (ε4 negative/ ε4 positive)	93/173	138/45	<0.001
Age at onset	72.84±6.14 (n=60)	-	-
Age at death	82.46±7.58 (n=266)	80.80±9.17 (n=183)	0.037
Braak stage	5.09±0.83 (n=168)	1.62±0.76 (n=96)	<0.001
CERAD score	2.31±0.84 (n=26)	0.83±0.64 (n=47)	<0.001

Table 14. Characteristics of copy number variation calls from samples in the TGen cohort. CNV-Copy number variation.

	Cases (n=728)	Controls (n=438)
Deletions:		
Number of CNVs	15177	9011
Rate per sample	20.85	20.57
Average size (kb)	45.34	46.75
Duplications:		
Number of CNVs	4334	2523
Rate per sample	5.95	5.76
Average size (kb)	140.8	131.1

Table 15. TGen samples grouped by copy number variation call size.

Call size	Cases (n=728)		Controls (n=438)	
	Deletions n (%)	Duplications n (%)	Deletions n (%)	Duplications n (%)
0.1-0.5 Mb	647 (88.87)	625 (85.85)	387 (88.36)	368 (84.02)
0.5-1.0 Mb	95 (13.05)	164 (22.53)	61 (13.93)	101 (23.06)
1.0-1.5 Mb	10 (1.37)	25 (3.43)	7 (1.60)	12 (2.74)
1.5-2.0 Mb	4 (0.55)	7 (0.96)	0 (0.00)	2 (0.46)
>2.0 Mb	2 (0.27)	11 (1.51)	5 (1.14)	1 (0.23)

2. Case/control association analyses

2.1. Candidate gene approach

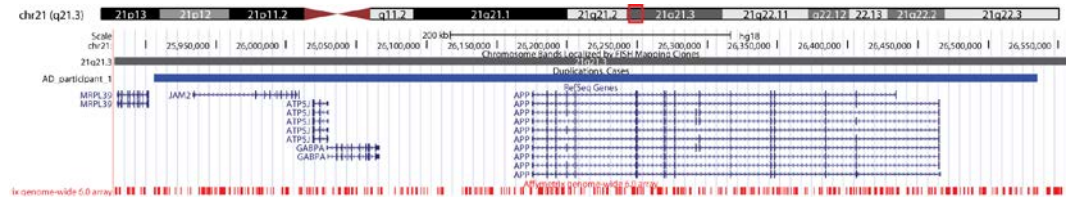
We identified 32 candidate genes in the TGen study overlapped by CNV calls from at least one case (AD) but no controls (Table 16). Representative plots of two genes (*APP* and *DOPEY2*) are shown in Figure 13. The *HLA-DRA* gene was overlapped by deletions in nine cases (uncorrected p=0.0140; one-sided). This gene was also found to be overlapped by deletions in two controls in the ADNI study. Two genes (*RELN* overlapped by deletions in two cases and *DOPEY2* overlapped by duplications in four cases) identified in this study were also reported from only cases (AD and/or MCI) in the ADNI and NIA-LOAD/NCRAD Family studies. One AD sample (*APOE* ε2/ε3 genotype, age at death=67) had a novel *APP* gene duplication supported by 443 sequential SNP and CNV probes. The *CHRFAM7A* gene reported in the ADNI and NIA-LOAD/NCRAD Family studies was overlapped by deletions in 10 cases and two controls, and duplications in 12 cases and one control (corrected p=0.0198; one-sided) in this study (Figure 14).

Table 16. Genes overlapped by copy number variation calls from at least one case and no control samples in the TGen cohort using the candidate gene approach. [a]A case sample had copy number variation calls overlapping the *AGT*, *LHCGR* and *BAT1* genes; [b,c,d]Three different case samples had copy number variation calls overlapping the *HLA-DRA* and *HLA-E*, *HLA-DRA* and *HLA-DQB1*, and *HLA-DRA* and *NGB* genes respectively; [e]A different case sample had copy number variation calls overlapping the *MYH13* and *MYH8* genes.

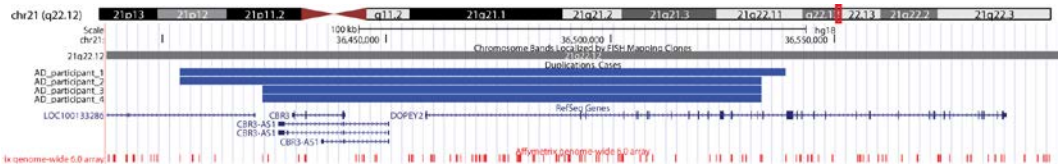
Chromosome	Region	Start (bp)	End (bp)	Number of cases
1	*FAM63A*	149234172	149245957	1
1	*SOAT1*	177529639	177591076	1
1	*AGT*	228904891	228916959	1[a]
2	*LHCGR*	48767416	48836384	4[a]
6	*HLA-G*	29902734	29906878	1
6	*HLA-A*	30018309	30021633	2
6	*HLA-E*	30565249	30569072	3[b]
6	*BAT1*	31605974	31618204	1[a]
6	*HLA-DRA*	32515624	32520802	9[b,c,d]
6	*HLA-DQB1*	32735634	32742444	1[c]
7	*MAGI2*	77484309	78920826	1
7	*CD36*	80069439	80146529	1
7	*RELN*	102899472	103417198	2
8	*NAT2*	18293034	18303003	1
10	*ALDH18A1*	97355675	97406557	1
10	*EBF3*	131523536	131652081	1
11	*PICALM*	85346132	85457756	1
12	*C12orf41*	47333261	47362302	3
12	*ALDH2*	110688728	110732167	2

Chromosome	Region	Start (bp)	End (bp)	Number of cases
14	*PSEN1*	72672931	72756862	1
14	*NGB*	76801586	76807408	1[d]
17	*SERPINF2*	1592879	1605309	1
17	*MYH13*	10144907	10217047	2[e]
17	*MYH8*	10234366	10265992	1[e]
17	*MAPT*	41327543	41461546	1
19	*GALP*	61379200	61388956	1
21	*APP*	26174731	26465003	1
21	*DOPEY2*	36458708	36588442	4
21	*CBS*	43346369	43369493	1
21	*S100B*	46842958	46849463	1
22	*COMT*	18309308	18336530	1
22	*BCR*	21852551	21990224	1

Figure 13. Representative UCSC Genome Browser (March 2006 (NCBI36/hg18) assembly) plots of duplications overlapping the candidate genes *APP* (A) and *DOPEY2* (B) in samples of the TGen cohort. The chromosomal location of the gene and probes on the Affymetrix SNP 6.0 Structural Variation array are shown. The region with the duplication for each sample relative to the gene is represented by a blue rectangle.

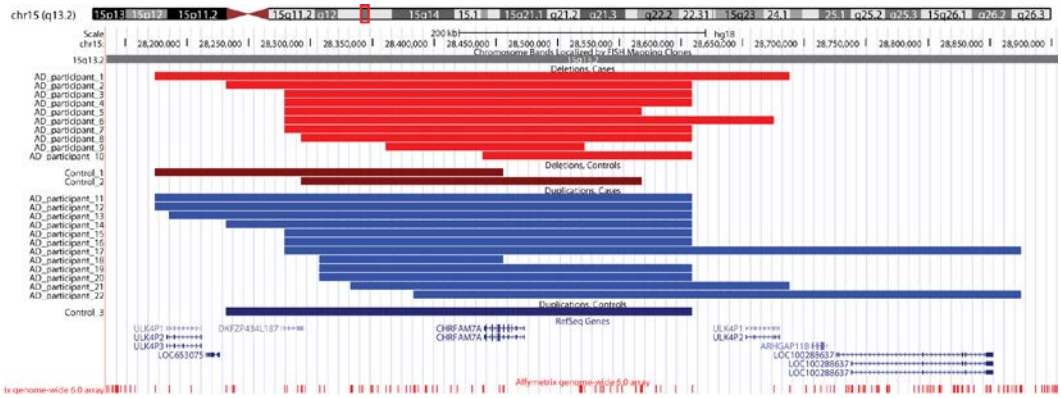


A



B

Figure 14. Representative UCSC Genome Browser (March 2006 (NCBI36/hg18) assembly) plots of deletions and duplications overlapping the *CHRFAM7A* gene in samples of the TGen cohort. The chromosomal location of the gene and probes on the Affymetrix SNP 6.0 Structural Variation array are shown. The region with the deletion for each sample relative to the gene is highlighted by a red rectangle and the region with the duplication for each sample relative to the gene is highlighted by a blue rectangle.



2.2. Genome-wide approach

We also identified genes across the genome overlapped by CNV calls only in cases (AD) but not controls in the TGen study. The *HLA-DRA* gene identified in the candidate gene approach was also found from this approach (uncorrected p=0.0144; one-sided). The *CHRFAM7A* gene was also identified from this approach (uncorrected p=0.0046; one-sided). Genes reported in the ADNI and NIA-LOAD/NCRAD Family studies were overlapped by CNV calls in this study including: *CSMD1* (deletions in 65 cases and 32 controls, duplications in one case and one control), *ERBB4* (deletions in 71 cases and 35 controls, duplication in one control), *HNRNPCL1* (deletions in 19 cases and eight controls, duplications in one case and two controls), *IMMP2L* (deletions in six cases and five controls, duplication in one control), *NRXN1* (deletions in two cases and three controls), and *SLC35F2* (duplications in two cases and one control).

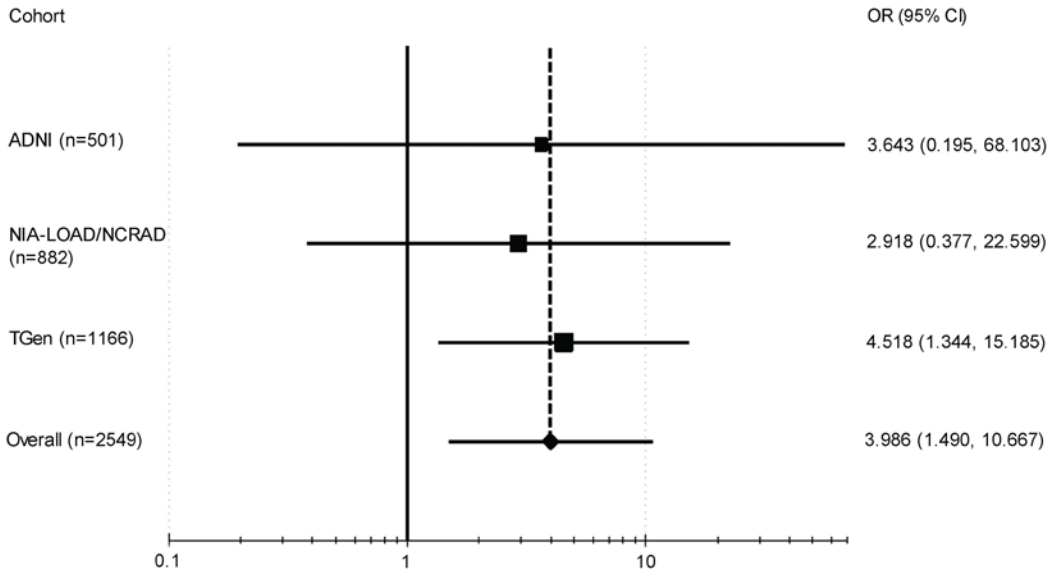
3. Meta-analysis

A meta-analysis for the *CHRFAM7A* gene was performed using findings from the ADNI, NIA-LOAD/NCRAD Family and TGen studies (Table 17 and Figure 15). Overall, the gene was overlapped by CNV calls in 38 of 1797 (2.115%) cases (AD or MCI) and four of 752 (0.532%) controls. A significant association was observed for the gene with AD or MCI risk (summary OR=3.986; 95% confidence interval (CI), 1.490-10.667; p=0.006). On repeating the analysis with only the AD cases and not the MCI cases in the ADNI cohort, a significant association for the gene with AD risk (summary OR=3.952; 95% CI, 1.471-10.617) was observed.

Table 17. Meta-analysis of the *CHRFAM7A* gene. CNV-Copy number variation.

Cohort	ADNI (n=501)	NIA-LOAD/NCRAD (n=882)	TGen (n=1166)	Overall (n=2549)
Cases:				
Number of samples	358	711	728	1797
CNV overlapping gene (n)	4	12	22	38
No CNV overlapping gene (n)	354	699	706	1759
Controls:				
Number of samples	143	171	438	752
CNV overlapping gene (n)	0	1	3	4
No CNV overlapping gene (n)	143	170	435	748
Odds ratio	3.643	2.918	4.518	3.986
95% confidence interval	0.195-68.103	0.377-22.599	1.344-15.185	1.490-10.667
p	0.387	0.305	0.015	0.006

Figure 15. Forest plot of the *CHRFAM7A* gene. The plot represents the meta-analysis of the *CHRFAM7A* gene using results from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study, the National Institute of Aging-Late Onset AD/National Cell Repository for AD (NIA-LOAD/NCRAD) Family Study and the TGen study. The odds ratio and 95% confidence interval for the odds ratio for each study are represented by black squares and horizontal lines. The summary odds ratio is depicted as a black diamond. OR-Odds ratio; CI-Confidence interval.



D. Discussion

The present report represents an initial CNV analysis in the TGen cohort, a unique cohort of clinically characterized and neuropathologically verified individuals. After extensive QC, case (AD)/control association analyses using candidate gene and genome-wide approaches were performed. Genes enriched in cases relative to controls were determined, suggesting possible involvement of these genes in AD susceptibility.

Rates of deletions and duplications did not significantly differ between cases and controls. This is different from findings in the ADNI and NIA-LOAD/NCRAD Family studies and could be due to different participant selection criteria, random sampling variation, different QC criteria, and that the NIA-LOAD/NCRAD Family Study and TGen study analyses included only AD and control samples, whereas the ADNI study analyses included MCI samples in addition to AD and control samples.

The candidate gene approach revealed a number of interesting genes (Table 16 and Figure 13). The *HLA-DRA* (major histocompatibility complex, class II, DR alpha) gene on chromosome 6 is a human leukocyte antigen (HLA) class II alpha chain paralogue which plays an important role in the immune system by presenting peptides derived from extracellular proteins. Variants in *HLA-DRA* have been associated with Parkinson disease [203, 204] and multiple sclerosis [205, 206], but not with AD. Other *HLA* alleles however have been investigated for a possible role in AD [174, 175, 207, 208].

The *CHRFAM7A* (*CHRNA7* (cholinergic receptor, nicotinic, alpha 7, exons 5-10) and *FAM7A* (family with sequence similarity 7A, exons A-E) fusion) gene (Figure 14) is located on chromosome 15. It is formed as a hybrid of a partially duplicated *CHRNA7*

gene and the *FAM7A* gene [145, 167]. It is highly polymorphic and individuals with and without this gene have been identified. A 2-bp deletion polymorphism at position 497-498 in exon 6 of this gene has been observed to be significantly over-represented in participants with AD, dementia with Lewy bodies and Pick's disease compared to controls [147].

Although *CHRFAM7A* is transcribed, its translation and possible function of the resulting protein is uncertain. The gene is expressed in the hippocampus, a brain region known to be first affected in AD. Recently, it has been suggested to possibly modulate α7 subunit receptor-mediated synaptic transmission and cholinergic anti-inflammatory response [168]. It may also be a dominant negative modulator of *CHRNA7* function and important for receptor regulation in humans [169]. A meta-analysis using gene results from the ADNI, NIA-LOAD/NCRAD Family and TGen studies revealed a significant association of the gene with AD or MCI risk (Table 17 and Figure 15). On repeating the analysis with only the AD cases and not the MCI cases in the ADNI cohort, a significant association for the gene with AD risk was observed.

Two genes (*RELN* and *DOPEY2*) in the three studies were found to be overlapped by CNV calls from cases (AD and/or MCI) but not controls. The *RELN* (reelin) gene on chromosome 7 encodes the glycoprotein reelin, which activates a signaling pathway required for proper positioning of neurons within laminated nervous system parenchyma. Gene variants have been associated with AD [176] and the protein has been observed to have increased expression in pyramidal neurons of the hippocampus in AD individuals and in cognitively intact controls with AD-associated pathology [177]. The *DOPEY2* (dopey family member 2, also known as *C21orf5*) gene located on chromosome 21 in the Down syndrome critical region is a potential Down syndrome candidate gene [178,

209]. Overexpression of the gene may be associated with the neurological phenotypes and mental retardation observed in Down syndrome patients. One AD sample was identified with an *APP* gene duplication, representing a novel finding since *APP* duplications have been associated with early-onset [81, 125, 210] but not late-onset AD. The genes identified from the candidate gene approach have been previously investigated in AD studies and thus represent potential candidate genes. Replication in independent samples and laboratory validation can help confirm the role of these genes in AD susceptibility.

The genome-wide approach revealed the candidate genes: *HLA-DRA* and *CHRFAM7A* as well as identified genes reported in the two previous studies. The *CSMD1*, *HNRNPCL1*, *IMMP2L* and *SLC35F2* genes have not been previously associated with AD. The *NRXN1* gene has been associated with autism [85], schizophrenia [91], and has been shown to have reduced expression with increasing AD severity [150]. The *ERBB4* gene may play a possible role in the progression of AD pathology [187-189].

It is important to note the limitations of the present report. Although the same software (PennCNV) was used in the three studies, different QC criteria were used for the selection of samples. The ADNI and NIA-LOAD/NCRAD Family study samples were genotyped on the Illumina Human610-Quad BeadChip and used similar QC criteria. The TGen study samples were genotyped on the Affymetrix Genome-Wide Human SNP 6.0 Array and thus we used a slightly different QC criteria. To our knowledge, there does not appear to be consensus on a well defined set of QC criteria for inclusion of the most appropriate samples in CNV analyses. The QC criterion applied in the present study may have been too stringent, leading to samples possibly having informative CNV data being excluded. A direct comparison of CNV calls from the three studies would be

difficult as two different genotyping platforms were used. Probes from the two platforms may not correspond with each other with respect to their location. Replication in additional independent data sets and future molecular studies will help confirm the findings.

In sum, we have conducted an initial CNV analysis in samples from a cohort of clinically characterized and neuropathologically verified individuals. Rates of deletions and duplications did not significantly differ between cases and controls. Gene-based association analysis identified a number of genes including those reported in the ADNI and NIA-LOAD/NCRAD Family studies (*CHRFAM7A*, *RELN* and *DOPEY2*) as well as a new gene (*HLA-DRA*). Meta-analysis from the three studies revealed a significant association for *CHRFAM7A* with AD and/or MCI risk. Replication in independent samples will be necessary to confirm these findings. Targeted analyses of the identified regions will help determine the biological role of these variants. Overall, there appears to be some consistency of CNVs across AD cohorts and this variation holds promise for revealing novel risk factors and disease mechanisms.

V. Summary and Future Directions

Genetic variation is an important risk factor in the development of Alzheimer's disease (AD). The disease has been shown to have a high heritability with 58-79% of phenotypic variation estimated to be caused by genetic factors [4]. Mutations in the *APP*, *PSEN1* and *PSEN2* genes cause early-onset AD (onset earlier than 60 or 65 years), but these account for approximately one to six percent of all AD cases. The leading genetic risk factor for the more common late-onset AD (onset later than 60 or 65 years) is the *APOE* ε4 allele. Recent genome-wide association studies (GWASs) have identified and replicated other AD risk loci including: *CLU*, *CR1*, *PICALM*, *BIN1*, *EXOC3L2*, *MTHFD1L*, *MS4A4A/MS4A6E*, *CD2AP*, *CD33*, *ABCA7* and *CUGBP2* [12-19]. However, not all genetic variation associated with AD is explained by these loci, and other forms of genetic variation such as copy number variations (CNVs) may play a role.

CNVs are DNA regions ranging from one kilobase (kb) to several megabases (Mb) in size that have added genetic material (copy number gains or duplications) or loss of genetic material (copy number losses or deletions). They play a role in various neuropsychiatric disorders such as autism [85, 86, 102], Parkinson disease [89, 90, 113] and schizophrenia [91, 92, 114]. At the time of starting this work, only one published study had examined the role of CNVs in late-onset AD [26]. The authors of the study identified a duplication in the *CHRNA7* gene which they thought warranted further investigation. The role of CNVs in mild cognitive impairment (MCI) had not been investigated at the time.

We thus performed a genome-wide CNV analyses in 288 AD, 183 MCI, and 184 control participants in the Alzheimer's Disease Neuroimaging Initiative (ADNI) study [134] who

had DNA samples extracted from peripheral blood. Participants whose DNA samples were extracted from cell lines were excluded from the analyses as CNV results could be influenced by cell line transformation [135, 136]. Case/control association analyses were performed to compare the CNV burden between cases (AD and/or MCI) and controls, and to identify genomic regions where CNVs could be detected in cases but not controls. A trend towards a higher CNV call rate for deletions and a lower CNV rate for duplications was observed in cases compared to controls. Case/control association analyses identified a number of genes including: *CHRFAM7A*, an AD candidate gene; *NRXN1* and *ERBB4*, genes associated with neuropsychiatric disorders [85, 143]; and *CSMD1*, *HNRNPCL1* and *SLC35F2*, genes not previously associated with AD.

We then performed a similar genome-wide CNV analyses in 794 AD and 196 neurologically evaluated control non-Hispanic Caucasian participants in the National Institute of Aging-Late Onset AD/National Cell Repository for AD (NIA-LOAD/NCRAD) Family Study [19] who had DNA samples extracted from peripheral blood or brain tissue. The controls were unrelated to AD participants and had no family history of AD. Case/control association analyses were performed using similar methods and quality control (QC) criteria as in the ADNI study. We observed a trend for lower CNV call rate for deletions as well as duplications in cases compared to controls. Case/control association analyses identified genes reported in the ADNI study (*ATXN1*, *CHRFAM7A*, *DOPEY2*, *ERBB4*, *GSTT1*, *HLA-DPB1*, *NRXN1* and *RELN*) as well as a new gene (*IMMP2L*). This gene is disrupted in patients with Tourette syndrome [182, 183], but has not been previously associated with AD.

We then performed a similar genome-wide CNV analyses in a third cohort of 1022 AD and 595 control Caucasian participants who had DNA samples extracted from brain

tissue. This is a unique cohort of clinically characterized and neuropathologically verified cases (AD) and controls [196]. As samples in this study were genotyped using the Affymetrix Genome-Wide Human SNP 6.0 Array, and samples in the ADNI and NIA-LOAD/NCRAD Family studies were genotyped using the Illumina Human610-Quad BeadChip, we applied a different set of QC criteria for samples in this study. However, we used similar methods for the case/control association analyses. Rates of deletions and duplications did not significantly differ between cases and controls. A number of genes were identified from the case/control association analyses including those reported in the ADNI and NIA-LOAD/NCRAD Family studies (*CHRFAM7A*, *RELN* and *DOPEY2*) as well as a new gene (*HLA-DRA*). The *HLA-DRA* gene has been associated with Parkinson disease [203, 204] and multiple sclerosis [205, 206], but has not been associated with AD. A meta-analysis of the *CHRFAM7A* findings from the three studies revealed a significant association for the *CHRFAM7A* gene with AD or MCI risk (summary OR=3.986; 95% CI, 1.490-10.667). On repeating the analysis with only the AD cases and not the MCI cases in the ADNI cohort, a significant association for the gene with AD risk (summary OR=3.952; 95% CI, 1.471-10.617) was observed.

Thus, the present work includes genome-wide CNV analyses in three independent cohorts of AD, MCI and control participants. A number of possible candidate genes were identified warranting further investigation. Future directions of this work include replication in independent datasets and molecular validation to confirm the findings. Other cohorts of AD, control and possibly MCI participants who have DNA samples extracted from peripheral blood and/or brain tissue and genotype data can be used for replication analyses. Gene-based association analyses using similar approaches as the present work can be performed. Genes reported in the present work can be analyzed in these cohorts to determine if a higher proportion of cases have CNVs overlapping the

gene compared to controls thus replicating the findings in the present work. Also, if the identified CNVs overlapping the gene have similar length and location in multiple individuals, these may help identify specific regions within the gene that may be disrupted by the CNVs possibly affecting gene function. These regions would be potential targets for further investigation.

After possible candidate regions have been identified using bioinformatics approaches, the regions need to be validated using molecular techniques to confirm the copy number of these regions. Few techniques used for molecular validation of CNV results are: quantitative real-time polymerase chain reaction (qPCR), multiplex ligation-dependent probe amplification and multiple amplifiable probe hybridization, comparative genome hybridization (CGH) arrays or array CGH, and targeted resequencing.

qPCR with TaqMan copy number assays is a commonly used technique to validate CNVs. The assays are run simultaneously in the qPCR along with a TaqMan copy number reference assay. The target gene or genomic sequence of interest is amplified by the copy number assay. A sequence known to exist in a diploid genome in two copies such as the RNase P H1 RNA gene is amplified by the reference assay. The target sequence copy number in each test sample is determined by relative quantitation using the comparative Ct method. Cycle threshold (Ct) determines the qPCR cycles needed to obtain sufficient product to detect fluorescent signal in this assay. The Ct difference between target and reference sequences is measured, after which a comparison of the Ct difference values of test samples to a calibrator sample known to have two copies of the target sequence is performed [211]. The target copy number is calculated as two times the relative quantity.

Two other techniques which can be used for validation are multiplex ligation-dependent probe amplification and multiple amplifiable probe hybridization. These are targeted PCR-based approaches that can be used to analyze multiple genomic regions up to 40 target sequences simultaneously. In these approaches, oligonucleotide probes are used to generate locus-specific amplicons that can be resolved by capillary electrophoresis. Duplications are indicated by enhanced peak signals and deletions are indicated by reduced peak signals [47, 52].

Array CGH is another technique that can be used to perform validation experiments. In this technique, two fluorescently labeled samples (test and reference) are competitively hybridized to a known target DNA sequence immobilized on a solid glass substrate. The signal ratio between the test and reference sample after normalization and conversion to a $\log_2$ ratio acts as a proxy for copy number. A gain in copy number in the test sample compared with the reference sample is indicated by an increased $\log_2$ ratio; conversely, a loss in copy number is indicated by a decrease in $\log_2$ ratio [212]. Custom arrays are available from manufacturers such as Roche NimbleGen and Agilent Technologies that can be used for targeted analyses.

Targeted resequencing of candidate regions can also be used to perform validation experiments. In this technique, DNA is extracted from the sample and purified. The DNA is then enriched using one of the following methods: PCR, long-range PCR, array hybridization, in-solution hybridization, or chromosome sorting. DNA libraries (fragment libraries and mate-paired libraries) are constructed and the DNA is amplified using standard emulsion PCR. Sequencing is performed on specialized systems such as the Applied Systems SOLiD System and the data generated is analyzed. Results can then be validated using systems such as the Applied Systems VariantSEQr Resequencing

System (http://www.appliedbiosystems.com/absite/us/en/home/applications-technologies/solid-next-generation-sequencing/targeted-resequencing.html). Sequencing can be performed on the whole gene or just the protein coding portion of the gene (the exome). Systems such as the Applied Biosystems TargetSeq Exome Enrichment System, which is based on an in-solution hybridization method for exome capture, can be used for performing exome sequencing. Methods to detect CNVs from whole genome or exome sequencing data are being developed [213, 214]. We will be receiving whole genome sequencing data for DNA samples in the ADNI cohort which will enable us to interrogate the regions identified in the CNV analyses. We also have proteomic data for samples in this cohort collected using the Human DiscoveryMAP panel from Myriad RBM and are planning RNA expression studies, thus enabling us to determine possible downstream effects of the CNVs on transcription and translation of the overlapped regions.

After the CNVs have been molecularly validated, their clinical significance will need to be determined. One approach would be to determine if the CNV is inherited or de novo in nature. If the CNV is inherited from a healthy parent or similar to a CNV in a healthy relative, it is likely that the CNV is benign [215, 216]. If the CNV is inherited from an affected parent or similar to a CNV in an affected relative, it is likely that the CNV is pathogenic. Databases such as the Toronto Database of Genomic Variants (http://projects.tcag.ca/variation/) [36], DECIPHER (Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources; http://decipher.sanger.ac.uk/) [37] and ECARUCA (European Cytogeneticists Association Register of Unbalanced Chromosome Alterations; http://umcecaruca01.extern.umcn.nl:8080/ecaruca/ecaruca.jsp) [38] can be used to determine the clinical relevance of de novo CNVs. Targeted analysis of the overlapped

regions in a large number of case and control samples can help determine the utility of these regions for genetic testing and counseling.

In sum, AD represents an important disease in society for which there is an urgent need to find a cure. The number of people surviving into the oldest ages (80's and 90's) is expected to dramatically increase because of medical advances as well as better social and economic conditions [1]. With the growing number of older people, it is anticipated that there would be an increase in the number of existing and new cases of AD and dementia. Ongoing research work focused on determining the causes of AD may lead to its earlier diagnosis and better treatment. Genetic variation plays a key role in AD development and progression, but recent GWASs have not identified all the genetic factors associated with the disease. Other forms of genetic variation such as CNVs which play a role in neuropsychiatric disorders [34] have not been extensively studied in AD and MCI. CNV analyses conducted in the present work have revealed a number of possible candidate regions warranting further investigation. The findings in the present work suggest the possible involvement of other forms of structural genetic variation in disease susceptibility. Further studies including targeted resequencing and other approaches may help determine this.

REFERENCES

1. Alzheimer's Association, W. Thies, and L. Bleiler, *2011 Alzheimer's disease facts and figures.* Alzheimers Dement, 2011. **7**(2): p. 208-44.
2. Petersen, R.C., et al., *Mild cognitive impairment: ten years later.* Arch Neurol, 2009. **66**(12): p. 1447-55.
3. Castellani, R.J., R.K. Rolston, and M.A. Smith, *Alzheimer disease.* Dis Mon, 2010. **56**(9): p. 484-546.
4. Gatz, M., et al., *Role of genes and environments for explaining Alzheimer disease.* Arch Gen Psychiatry, 2006. **63**(2): p. 168-74.
5. Bekris, L.M., et al., *Genetics of Alzheimer disease.* J Geriatr Psychiatry Neurol, 2010. **23**(4): p. 213-27.
6. Campion, D., et al., *Early-onset autosomal dominant Alzheimer disease: prevalence, genetic heterogeneity, and mutation spectrum.* Am J Hum Genet, 1999. **65**(3): p. 664-70.
7. Cruts, M. and C. Van Broeckhoven, *Molecular genetics of Alzheimer's disease.* Ann Med, 1998. **30**(6): p. 560-5.
8. Bertram, L., et al., *Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database.* Nat Genet, 2007. **39**(1): p. 17-23.
9. Corder, E.H., et al., *Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families.* Science, 1993. **261**(5123): p. 921-3.
10. Farrer, L.A., 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, 1997. **278**(16): p. 1349-56.
11. Saunders, A.M., et al., *Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease.* Neurology, 1993. **43**(8): p. 1467-72.
12. Harold, D., et al., *Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease.* Nat Genet, 2009. **41**(10): p. 1088-93.
13. Lambert, J.C., et al., *Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease.* Nat Genet, 2009. **41**(10): p. 1094-9.
14. Jun, G., et al., *Meta-analysis confirms CR1, CLU, and PICALM as alzheimer disease risk loci and reveals interactions with APOE genotypes.* Arch Neurol, 2010. **67**(12): p. 1473-84.
15. Seshadri, S., et al., *Genome-wide analysis of genetic loci associated with Alzheimer disease.* JAMA, 2010. **303**(18): p. 1832-40.
16. Naj, A.C., et al., *Dementia revealed: novel chromosome 6 locus for late-onset Alzheimer disease provides genetic evidence for folate-pathway abnormalities.* PLoS Genet, 2010. **6**(9).
17. Naj, A.C., et al., *Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease.* Nat Genet, 2011. **43**(5): p. 436-41.
18. Hollingworth, P., et al., *Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease.* Nat Genet, 2011. **43**(5): p. 429-35.
19. Wijsman, E.M., 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, 2011. **7**(2): p. e1001308.

20. Abraham, R., et al., *A genome-wide association study for late-onset Alzheimer's disease using DNA pooling.* BMC Med Genomics, 2008. **1**: p. 44.
21. Beecham, G.W., et al., *Genome-wide association study implicates a chromosome 12 risk locus for late-onset Alzheimer disease.* Am J Hum Genet, 2009. **84**(1): p. 35-43.
22. Bertram, L., et al., *Genome-wide association analysis reveals putative Alzheimer's disease susceptibility loci in addition to APOE.* Am J Hum Genet, 2008. **83**(5): p. 623-32.
23. Carrasquillo, M.M., et al., *Genetic variation in PCDH11X is associated with susceptibility to late-onset Alzheimer's disease.* Nat Genet, 2009. **41**(2): p. 192-8.
24. Coon, K.D., et al., *A high-density whole-genome association study reveals that APOE is the major susceptibility gene for sporadic late-onset Alzheimer's disease.* J Clin Psychiatry, 2007. **68**(4): p. 613-8.
25. Grupe, A., et al., *Evidence for novel susceptibility genes for late-onset Alzheimer's disease from a genome-wide association study of putative functional variants.* Hum Mol Genet, 2007. **16**(8): p. 865-73.
26. Heinzen, E.L., et al., *Genome-wide scan of copy number variation in late-onset Alzheimer's disease.* J Alzheimers Dis, 2010. **19**(1): p. 69-77.
27. Hu, X., et al., *Meta-analysis for genome-wide association study identifies multiple variants at the BIN1 locus associated with late-onset Alzheimer's disease.* PLoS One, 2011. **6**(2): p. e16616.
28. Lee, J.H., et al., *Identification of novel loci for Alzheimer disease and replication of CLU, PICALM, and BIN1 in Caribbean Hispanic individuals.* Arch Neurol, 2011. **68**(3): p. 320-8.
29. Li, H., et al., *Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease.* Arch Neurol, 2008. **65**(1): p. 45-53.
30. Poduslo, S.E., et al., *Genome screen of late-onset Alzheimer's extended pedigrees identifies TRPC4AP by haplotype analysis.* Am J Med Genet B Neuropsychiatr Genet, 2009. **150B**(1): p. 50-5.
31. Potkin, S.G., et al., *Hippocampal atrophy as a quantitative trait in a genome-wide association study identifying novel susceptibility genes for Alzheimer's disease.* PLoS One, 2009. **4**(8): p. e6501.
32. Reiman, E.M., et al., *GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers.* Neuron, 2007. **54**(5): p. 713-20.
33. Sherva, R., et al., *Identification of novel candidate genes for Alzheimer's disease by autozygosity mapping using genome wide SNP data.* J Alzheimers Dis, 2011. **23**(2): p. 349-59.
34. Cook, E.H., Jr. and S.W. Scherer, *Copy-number variations associated with neuropsychiatric conditions.* Nature, 2008. **455**(7215): p. 919-23.
35. Stankiewicz, P. and J.R. Lupski, *Structural variation in the human genome and its role in disease.* Annu Rev Med, 2010. **61**: p. 437-55.
36. Zhang, J., et al., *Development of bioinformatics resources for display and analysis of copy number and other structural variants in the human genome.* Cytogenet Genome Res, 2006. **115**(3-4): p. 205-14.
37. Firth, H.V., et al., *DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources.* Am J Hum Genet, 2009. **84**(4): p. 524-33.
38. Feenstra, I., et al., *European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations (ECARUCA); an online database for rare chromosome abnormalities.* Eur J Med Genet, 2006. **49**(4): p. 279-91.

39. Stankiewicz, P. and J.R. Lupski, *Genome architecture, rearrangements and genomic disorders.* Trends Genet, 2002. **18**(2): p. 74-82.
40. Weterings, E. and D.C. van Gent, *The mechanism of non-homologous end-joining: a synopsis of synapsis.* DNA Repair (Amst), 2004. **3**(11): p. 1425-35.
41. Gu, W., F. Zhang, and J.R. Lupski, *Mechanisms for human genomic rearrangements.* Pathogenetics, 2008. **1**(1): p. 4.
42. Lee, J.A., C.M. Carvalho, and J.R. Lupski, *A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders.* Cell, 2007. **131**(7): p. 1235-47.
43. Zhang, F., et al., *The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans.* Nat Genet, 2009. **41**(7): p. 849-53.
44. Hastings, P.J., G. Ira, and J.R. Lupski, *A microhomology-mediated break-induced replication model for the origin of human copy number variation.* PLoS Genet, 2009. **5**(1): p. e1000327.
45. Zhang, F., C.M. Carvalho, and J.R. Lupski, *Complex human chromosomal and genomic rearrangements.* Trends Genet, 2009. **25**(7): p. 298-307.
46. Hastings, P.J., et al., *Mechanisms of change in gene copy number.* Nat Rev Genet, 2009. **10**(8): p. 551-64.
47. Fanciulli, M., E. Petretto, and T.J. Aitman, *Gene copy number variation and common human disease.* Clin Genet, 2010. **77**(3): p. 201-13.
48. McCarroll, S.A., *Extending genome-wide association studies to copy-number variation.* Hum Mol Genet, 2008. **17**(R2): p. R135-42.
49. McCarroll, S.A., *Copy-number analysis goes more than skin deep.* Nat Genet, 2008. **40**(1): p. 5-6.
50. Armour, J.A., et al., *Accurate, high-throughput typing of copy number variation using paralogue ratios from dispersed repeats.* Nucleic Acids Res, 2007. **35**(3): p. e19.
51. Hollox, E.J., et al., *Psoriasis is associated with increased beta-defensin genomic copy number.* Nat Genet, 2008. **40**(1): p. 23-5.
52. Sellner, L.N. and G.R. Taylor, *MLPA and MAPH: new techniques for detection of gene deletions.* Hum Mutat, 2004. **23**(5): p. 413-9.
53. Levy, S., et al., *The diploid genome sequence of an individual human.* PLoS Biol, 2007. **5**(10): p. e254.
54. Wheeler, D.A., et al., *The complete genome of an individual by massively parallel DNA sequencing.* Nature, 2008. **452**(7189): p. 872-6.
55. Korbel, J.O., et al., *Paired-end mapping reveals extensive structural variation in the human genome.* Science, 2007. **318**(5849): p. 420-6.
56. Tuzun, E., et al., *Fine-scale structural variation of the human genome.* Nat Genet, 2005. **37**(7): p. 727-32.
57. Yoon, S., et al., *Sensitive and accurate detection of copy number variants using read depth of coverage.* Genome Res, 2009. **19**(9): p. 1586-92.
58. Koike, A., et al., *Comparative analysis of copy number variation detection methods and database construction.* BMC Genet, 2011. **12**: p. 29.
59. Pollack, J.R., et al., *Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors.* Proc Natl Acad Sci U S A, 2002. **99**(20): p. 12963-8.
60. Eilers, P.H. and R.X. de Menezes, *Quantile smoothing of array CGH data.* Bioinformatics, 2005. **21**(7): p. 1146-53.
61. Hsu, L., et al., *Denoising array-based comparative genomic hybridization data using wavelets.* Biostatistics, 2005. **6**(2): p. 211-26.

62. Wang, P., et al., *A method for calling gains and losses in array CGH data.* Biostatistics, 2005. **6**(1): p. 45-58.
63. Lai, W.R., et al., *Comparative analysis of algorithms for identifying amplifications and deletions in array CGH data.* Bioinformatics, 2005. **21**(19): p. 3763-70.
64. Jong, K., et al., *Chromosomal Breakpoint Detection in Human Cancer.* Lecture Notes in Computer Science, 2003. **2611**: p. 107-116.
65. Picard, F., et al., *A statistical approach for array CGH data analysis.* BMC Bioinformatics, 2005. **6**: p. 27.
66. Venkatraman, E.S. and A.B. Olshen, *A faster circular binary segmentation algorithm for the analysis of array CGH data.* Bioinformatics, 2007. **23**(6): p. 657-63.
67. Korn, J.M., et al., *Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs.* Nat Genet, 2008. **40**(10): p. 1253-60.
68. Wang, K., et al., *PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data.* Genome Res, 2007. **17**(11): p. 1665-74.
69. Colella, S., et al., *QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data.* Nucleic Acids Res, 2007. **35**(6): p. 2013-25.
70. Dellinger, A.E., et al., *Comparative analyses of seven algorithms for copy number variant identification from single nucleotide polymorphism arrays.* Nucleic Acids Res, 2010. **38**(9): p. e105.
71. Eckel-Passow, J.E., et al., *Software comparison for evaluating genomic copy number variation for Affymetrix 6.0 SNP array platform.* BMC Bioinformatics, 2011. **12**: p. 220.
72. Pinto, D., et al., *Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants.* Nat Biotechnol, 2011. **29**(6): p. 512-20.
73. Winchester, L., C. Yau, and J. Ragoussis, *Comparing CNV detection methods for SNP arrays.* Brief Funct Genomic Proteomic, 2009. **8**(5): p. 353-66.
74. Peoples, R., et al., *A physical map, including a BAC/PAC clone contig, of the Williams-Beuren syndrome--deletion region at 7q11.23.* Am J Hum Genet, 2000. **66**(1): p. 47-68.
75. Berg, J.S., et al., *Speech delay and autism spectrum behaviors are frequently associated with duplication of the 7q11.23 Williams-Beuren syndrome region.* Genet Med, 2007. **9**(7): p. 427-41.
76. Chen, K.S., et al., *Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome.* Nat Genet, 1997. **17**(2): p. 154-63.
77. Potocki, L., et al., *Characterization of Potocki-Lupski syndrome (dup(17)(p11.2p11.2)) and delineation of a dosage-sensitive critical interval that can convey an autism phenotype.* Am J Hum Genet, 2007. **80**(4): p. 633-49.
78. Edelmann, L., et al., *A common molecular basis for rearrangement disorders on chromosome 22q11.* Hum Mol Genet, 1999. **8**(7): p. 1157-67.
79. Ensenauer, R.E., et al., *Microduplication 22q11.2, an emerging syndrome: clinical, cytogenetic, and molecular analysis of thirteen patients.* Am J Hum Genet, 2003. **73**(5): p. 1027-40.

80. Cardoso, C., et al., *Refinement of a 400-kb critical region allows genotypic differentiation between isolated lissencephaly, Miller-Dieker syndrome, and other phenotypes secondary to deletions of 17p13.3.* Am J Hum Genet, 2003. **72**(4): p. 918-30.
81. Rovelet-Lecrux, A., et al., *APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy.* Nat Genet, 2006. **38**(1): p. 24-6.
82. Kasuga, K., et al., *Identification of independent APP locus duplication in Japanese patients with early-onset Alzheimer disease.* J Neurol Neurosurg Psychiatry, 2009. **80**(9): p. 1050-2.
83. Lionel, A.C., et al., *Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD.* Sci Transl Med, 2011. **3**(95): p. 95ra75.
84. Elia, J., et al., *Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes.* Mol Psychiatry, 2010. **15**(6): p. 637-46.
85. Autism Genome Project Consortium, et al., *Mapping autism risk loci using genetic linkage and chromosomal rearrangements.* Nat Genet, 2007. **39**(3): p. 319-28.
86. Weiss, L.A., et al., *Association between microdeletion and microduplication at 16p11.2 and autism.* N Engl J Med, 2008. **358**(7): p. 667-75.
87. Zhang, D., et al., *Singleton deletions throughout the genome increase risk of bipolar disorder.* Mol Psychiatry, 2009. **14**(4): p. 376-80.
88. Mefford, H.C., et al., *Genome-wide copy number variation in epilepsy: novel susceptibility loci in idiopathic generalized and focal epilepsies.* PLoS Genet, 2010. **6**(5): p. e1000962.
89. Singleton, A.B., et al., *alpha-Synuclein locus triplication causes Parkinson's disease.* Science, 2003. **302**(5646): p. 841.
90. Pankratz, N., et al., *Copy number variation in familial Parkinson disease.* PLoS One, 2011. **6**(8): p. e20988.
91. Rujescu, D., et al., *Disruption of the neurexin 1 gene is associated with schizophrenia.* Hum Mol Genet, 2009. **18**(5): p. 988-96.
92. Ingason, A., et al., *Copy number variations of chromosome 16p13.1 region associated with schizophrenia.* Mol Psychiatry, 2011. **16**(1): p. 17-25.
93. Sundaram, S.K., et al., *Tourette syndrome is associated with recurrent exonic copy number variants.* Neurology, 2010. **74**(20): p. 1583-90.
94. Marshall, C.R., et al., *Structural variation of chromosomes in autism spectrum disorder.* Am J Hum Genet, 2008. **82**(2): p. 477-88.
95. Kumar, R.A., et al., *Recurrent 16p11.2 microdeletions in autism.* Hum Mol Genet, 2008. **17**(4): p. 628-38.
96. Kim, H.G., et al., *Disruption of neurexin 1 associated with autism spectrum disorder.* Am J Hum Genet, 2008. **82**(1): p. 199-207.
97. Hedges, D.J., et al., *Evidence of novel fine-scale structural variation at autism spectrum disorder candidate loci.* Mol Autism, 2012. **3**(1): p. 2.
98. Glessner, J.T., et al., *Autism genome-wide copy number variation reveals ubiquitin and neuronal genes.* Nature, 2009. **459**(7246): p. 569-73.
99. Bucan, M., et al., *Genome-wide analyses of exonic copy number variants in a family-based study point to novel autism susceptibility genes.* PLoS Genet, 2009. **5**(6): p. e1000536.
100. Pinto, D., et al., *Functional impact of global rare copy number variation in autism spectrum disorders.* Nature, 2010. **466**(7304): p. 368-72.

101. Salyakina, D., et al., *Copy number variants in extended autism spectrum disorder families reveal candidates potentially involved in autism risk.* PLoS One, 2011. **6**(10): p. e26049.
102. Sanders, S.J., et al., *Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism.* Neuron, 2011. **70**(5): p. 863-85.
103. Vaags, A.K., et al., *Rare deletions at the neurexin 3 locus in autism spectrum disorder.* Am J Hum Genet, 2012. **90**(1): p. 133-41.
104. Wilson, G.M., et al., *DNA copy-number analysis in bipolar disorder and schizophrenia reveals aberrations in genes involved in glutamate signaling.* Hum Mol Genet, 2006. **15**(5): p. 743-9.
105. Lachman, H.M., et al., *Increase in GSK3beta gene copy number variation in bipolar disorder.* Am J Med Genet B Neuropsychiatr Genet, 2007. **144B**(3): p. 259-65.
106. Priebe, L., et al., *Genome-wide survey implicates the influence of copy number variants (CNVs) in the development of early-onset bipolar disorder.* Mol Psychiatry, 2012. **17**(4): p. 421-32.
107. Malhotra, D., et al., *High frequencies of de novo CNVs in bipolar disorder and schizophrenia.* Neuron, 2011. **72**(6): p. 951-63.
108. Nuytemans, K., et al., *Genetic etiology of Parkinson disease associated with mutations in the SNCA, PARK2, PINK1, PARK7, and LRRK2 genes: a mutation update.* Hum Mutat, 2010. **31**(7): p. 763-80.
109. Crosiers, D., et al., *Parkinson disease: insights in clinical, genetic and pathological features of monogenic disease subtypes.* J Chem Neuroanat, 2011. **42**(2): p. 131-41.
110. Polymeropoulos, M.H., et al., *Mapping of a gene for Parkinson's disease to chromosome 4q21-q23.* Science, 1996. **274**(5290): p. 1197-9.
111. Chartier-Harlin, M.C., et al., *Alpha-synuclein locus duplication as a cause of familial Parkinson's disease.* Lancet, 2004. **364**(9440): p. 1167-9.
112. Ibanez, P., et al., *Causal relation between alpha-synuclein gene duplication and familial Parkinson's disease.* Lancet, 2004. **364**(9440): p. 1169-71.
113. Simon-Sanchez, J., et al., *Genomewide SNP assay reveals mutations underlying Parkinson disease.* Hum Mutat, 2008. **29**(2): p. 315-22.
114. Stefansson, H., et al., *Large recurrent microdeletions associated with schizophrenia.* Nature, 2008. **455**(7210): p. 232-6.
115. International Schizophrenia Consortium, *Rare chromosomal deletions and duplications increase risk of schizophrenia.* Nature, 2008. **455**(7210): p. 237-41.
116. Kirov, G., et al., *Support for the involvement of large copy number variants in the pathogenesis of schizophrenia.* Hum Mol Genet, 2009. **18**(8): p. 1497-503.
117. Levinson, D.F., et al., *Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications.* Am J Psychiatry, 2011. **168**(3): p. 302-16.
118. Kirov, G., et al., *Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia.* Hum Mol Genet, 2008. **17**(3): p. 458-65.
119. Ikeda, M., et al., *Copy number variation in schizophrenia in the Japanese population.* Biol Psychiatry, 2010. **67**(3): p. 283-6.
120. Mulle, J.G., et al., *Microdeletions of 3q29 confer high risk for schizophrenia.* Am J Hum Genet, 2010. **87**(2): p. 229-36.
121. Vacic, V., et al., *Duplications of the neuropeptide receptor gene VIPR2 confer significant risk for schizophrenia.* Nature, 2011. **471**(7339): p. 499-503.

122. Grozeva, D., et al., *Independent estimation of the frequency of rare CNVs in the UK population confirms their role in schizophrenia.* Schizophr Res, 2012. **135**(1-3): p. 1-7.
123. Lee, K.W., et al., *Genome wide association studies (GWAS) and copy number variation (CNV) studies of the major psychoses: what have we learnt?* Neurosci Biobehav Rev, 2012. **36**(1): p. 556-71.
124. Fernandez, T.V., et al., *Rare copy number variants in tourette syndrome disrupt genes in histaminergic pathways and overlap with autism.* Biol Psychiatry, 2012. **71**(5): p. 392-402.
125. Sleegers, K., et al., *APP duplication is sufficient to cause early onset Alzheimer's dementia with cerebral amyloid angiopathy.* Brain, 2006. **129**(Pt 11): p. 2977-83.
126. Blom, E.S., et al., *Low prevalence of APP duplications in Swedish and Finnish patients with early-onset Alzheimer's disease.* Eur J Hum Genet, 2008. **16**(2): p. 171-5.
127. Thonberg, H., et al., *Mutation screening of patients with Alzheimer disease identifies APP locus duplication in a Swedish patient.* BMC Res Notes, 2011. **4**(1): p. 476.
128. Rovelet-Lecrux, A., et al., *A genome-wide study reveals rare CNVs exclusive to extreme phenotypes of Alzheimer disease.* Eur J Hum Genet, 2011.
129. Brouwers, N., et al., *Alzheimer risk associated with a copy number variation in the complement receptor 1 increasing C3b/C4b binding sites.* Mol Psychiatry, 2012. **17**(2): p. 223-33.
130. Shaw, C.A., et al., *Olfactory copy number association with age at onset of Alzheimer disease.* Neurology, 2011. **76**(15): p. 1302-9.
131. Ghani, M., et al., *Genome-wide survey of large rare copy number variants in Alzheimer's disease among Caribbean hispanics.* G3 (Bethesda), 2012. **2**(1): p. 71-8.
132. Sleegers, K., et al., *The pursuit of susceptibility genes for Alzheimer's disease: progress and prospects.* Trends Genet, 2010. **26**(2): p. 84-93.
133. Petersen, R.C., et al., *Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization.* Neurology, 2010. **74**(3): p. 201-9.
134. Weiner, M.W., et al., *The Alzheimer's disease neuroimaging initiative: progress report and future plans.* Alzheimers Dement, 2010. **6**(3): p. 202-11 e7.
135. Wellcome Trust Case Control Consortium, et al., *Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls.* Nature, 2010. **464**(7289): p. 713-20.
136. Sie, L., S. Loong, and E.K. Tan, *Utility of lymphoblastoid cell lines.* J Neurosci Res, 2009. **87**(9): p. 1953-9.
137. Saykin, A.J., et al., *Alzheimer's Disease Neuroimaging Initiative biomarkers as quantitative phenotypes: Genetics core aims, progress, and plans.* Alzheimers Dement, 2010. **6**(3): p. 265-73.
138. Diskin, S.J., et al., *Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms.* Nucleic Acids Res, 2008. **36**(19): p. e126.
139. Conlin, L.K., et al., *Mechanisms of mosaicism, chimerism and uniparental disomy identified by single nucleotide polymorphism array analysis.* Hum Mol Genet, 2010. **19**(7): p. 1263-75.
140. Need, A.C., et al., *A genome-wide investigation of SNPs and CNVs in schizophrenia.* PLoS Genet, 2009. **5**(2): p. e1000373.
141. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses.* Am J Hum Genet, 2007. **81**(3): p. 559-75.

142. Kent, W.J., et al., *The human genome browser at UCSC.* Genome Res, 2002. **12**(6): p. 996-1006.
143. Lu, C.L., et al., *Support for the involvement of the ERBB4 gene in schizophrenia: a genetic association analysis.* Neurosci Lett, 2010. **481**(2): p. 120-5.
144. Wu, M., E.J. Michaud, and D.K. Johnson, *Cloning, functional study and comparative mapping of Luzp2 to mouse chromosome 7 and human chromosome 11p13-11p14.* Mamm Genome, 2003. **14**(5): p. 323-34.
145. Riley, B., et al., *A 3-Mb map of a large Segmental duplication overlapping the alpha7-nicotinic acetylcholine receptor gene (CHRNA7) at human 15q13-q14.* Genomics, 2002. **79**(2): p. 197-209.
146. Sinkus, M.L., 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, 2009. **1291**: p. 1-11.
147. Feher, A., et al., *Association between a genetic variant of the alpha-7 nicotinic acetylcholine receptor subunit and four types of dementia.* Dement Geriatr Cogn Disord, 2009. **28**(1): p. 56-62.
148. Kraus, D.M., et al., *CSMD1 is a novel multiple domain complement-regulatory protein highly expressed in the central nervous system and epithelial tissues.* J Immunol, 2006. **176**(7): p. 4419-30.
149. Reissner, C., et al., *Mutational analysis of the neurexin/neuroligin complex reveals essential and regulatory components.* Proc Natl Acad Sci U S A, 2008. **105**(39): p. 15124-9.
150. Gomez Ravetti, M., et al., *Uncovering molecular biomarkers that correlate cognitive decline with the changes of hippocampus' gene expression profiles in Alzheimer's disease.* PLoS One, 2010. **5**(4): p. e10153.
151. Ching, M.S., et al., *Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders.* Am J Med Genet B Neuropsychiatr Genet, 2010. **153B**(4): p. 937-47.
152. Shen, L., 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, 2010. **53**(3): p. 1051-63.
153. Buonanno, A., *The neuregulin signaling pathway and schizophrenia: from genes to synapses and neural circuits.* Brain Res Bull, 2010. **83**(3-4): p. 122-31.
154. McCarthy, S.E., et al., *Microduplications of 16p11.2 are associated with schizophrenia.* Nat Genet, 2009. **41**(11): p. 1223-7.
155. Shinawi, M., et al., *Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size.* J Med Genet, 2010. **47**(5): p. 332-41.
156. Walters, R.G., et al., *A new highly penetrant form of obesity due to deletions on chromosome 16p11.2.* Nature, 2010. **463**(7281): p. 671-5.
157. Bochukova, E.G., et al., *Large, rare chromosomal deletions associated with severe early-onset obesity.* Nature, 2010. **463**(7281): p. 666-70.
158. Bertram, L., C.M. Lill, and R.E. Tanzi, *The genetics of Alzheimer disease: back to the future.* Neuron, 2010. **68**(2): p. 270-81.
159. Swaminathan, S., et al., *Genomic Copy Number Analysis in Alzheimer's Disease and Mild Cognitive Impairment: An ADNI Study.* Int J Alzheimers Dis, 2011. **2011**: p. 729478.
160. McKhann, G., et al., *Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease.* Neurology, 1984. **34**(7): p. 939-44.

161. Mirra, S.S., 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, 1991. **41**(4): p. 479-86.
162. Price, A.L., et al., *Principal components analysis corrects for stratification in genome-wide association studies.* Nat Genet, 2006. **38**(8): p. 904-9.
163. Buettner, J.A., et al., *Organization and evolution of olfactory receptor genes on human chromosome 11.* Genomics, 1998. **53**(1): p. 56-68.
164. Serby, M., P. Larson, and D. Kalkstein, *The nature and course of olfactory deficits in Alzheimer's disease.* Am J Psychiatry, 1991. **148**(3): p. 357-60.
165. Devanand, D.P., et al., *Olfactory identification deficits and MCI in a multi-ethnic elderly community sample.* Neurobiol Aging, 2010. **31**(9): p. 1593-600.
166. Bruno, C., et al., *Neuromuscular forms of glycogen branching enzyme deficiency.* Acta Myol, 2007. **26**(1): p. 75-8.
167. Gault, J., et al., *Genomic organization and partial duplication of the human alpha7 neuronal nicotinic acetylcholine receptor gene (CHRNA7).* Genomics, 1998. **52**(2): p. 173-85.
168. de Lucas-Cerrillo, A.M., et al., *Function of partially duplicated human alpha77 nicotinic receptor subunit CHRFAM7A gene: potential implications for the cholinergic anti-inflammatory response.* J Biol Chem, 2011. **286**(1): p. 594-606.
169. Araud, T., et al., *The chimeric gene CHRFAM7A, a partial duplication of the CHRNA7 gene, is a dominant negative regulator of alpha7*nAChR function.* Biochem Pharmacol, 2011. **82**(8): p. 904-14.
170. Nie, H.Z., et al., *Activation of alpha7 nicotinic receptor affects APP processing by regulating secretase activity in SH-EP1-alpha7 nAChR-hAPP695 cells.* Brain Res, 2010. **1356**: p. 112-20.
171. Zhang, C., et al., *Loss of function of ATXN1 increases amyloid beta-protein levels by potentiating beta-secretase processing of beta-amyloid precursor protein.* J Biol Chem, 2010. **285**(12): p. 8515-26.
172. Tong, X., et al., *Ataxin-1 and Brother of ataxin-1 are components of the Notch signalling pathway.* EMBO Rep, 2011. **12**(5): p. 428-35.
173. Bettens, K., et al., *Follow-up study of susceptibility loci for Alzheimer's disease and onset age identified by genome-wide association.* J Alzheimers Dis, 2010. **19**(4): p. 1169-75.
174. Ma, S.L., et al., *Association between HLA-A alleles and Alzheimer's disease in a southern Chinese community.* Dement Geriatr Cogn Disord, 2008. **26**(5): p. 391-7.
175. Lehmann, D.J., et al., *Replication of the association of HLA-B7 with Alzheimer's disease: a role for homozygosity?* J Neuroinflammation, 2006. **3**: p. 33.
176. Seripa, D., et al., *The RELN locus in Alzheimer's disease.* J Alzheimers Dis, 2008. **14**(3): p. 335-44.
177. Kramer, P.L., et al., *Alzheimer disease pathology in cognitively healthy elderly: a genome-wide study.* Neurobiol Aging, 2011. **32**(12): p. 2113-22.
178. Rachidi, M., et al., *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, 2009. **27**(4): p. 393-8.
179. Pinhel, M.A., 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, 2008. **46**(4): p. 439-45.

180. Spalletta, G., et al., *Glutathione S-transferase P1 and T1 gene polymorphisms predict longitudinal course and age at onset of Alzheimer disease.* Am J Geriatr Psychiatry, 2007. **15**(10): p. 879-87.
181. Burri, L., et al., *Mature DIABLO/Smac is produced by the IMP protease complex on the mitochondrial inner membrane.* Mol Biol Cell, 2005. **16**(6): p. 2926-33.
182. Petek, E., et al., *Disruption of a novel gene (IMMP2L) by a breakpoint in 7q31 associated with Tourette syndrome.* Am J Hum Genet, 2001. **68**(4): p. 848-58.
183. Patel, C., et al., *Translocation breakpoint at 7q31 associated with tics: further evidence for IMMP2L as a candidate gene for Tourette syndrome.* Eur J Hum Genet, 2011. **19**(6): p. 634-9.
184. Lu, B., et al., *A mutation in the inner mitochondrial membrane peptidase 2-like gene (Immp2l) affects mitochondrial function and impairs fertility in mice.* Biol Reprod, 2008. **78**(4): p. 601-10.
185. George, S.K., et al., *Mitochondrial peptidase IMMP2L mutation causes early onset of age-associated disorders and impairs adult stem cell self-renewal.* Aging Cell, 2011. **10**(4): p. 584-94.
186. Srinivasan, R., et al., *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, 1998. **185**(3): p. 236-45.
187. Chaudhury, A.R., et al., *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, 2003. **62**(1): p. 42-54.
188. Woo, R.S., et al., *Expression of ErbB4 in the apoptotic neurons of Alzheimer's disease brain.* Anat Cell Biol, 2010. **43**(4): p. 332-9.
189. Woo, R.S., et al., *Expression of ErbB4 in the neurons of Alzheimer's disease brain and APP/PS1 mice, a model of Alzheimer's disease.* Anat Cell Biol, 2011. **44**(2): p. 116-27.
190. Mefford, H.C., et al., *A method for rapid, targeted CNV genotyping identifies rare variants associated with neurocognitive disease.* Genome Res, 2009. **19**(9): p. 1579-85.
191. Williams, N.M., et al., *Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis.* Lancet, 2010. **376**(9750): p. 1401-8.
192. Magri, C., et al., *New copy number variations in schizophrenia.* PLoS One, 2010. **5**(10): p. e13422.
193. Awadalla, P., et al., *Direct measure of the de novo mutation rate in autism and schizophrenia cohorts.* Am J Hum Genet, 2010. **87**(3): p. 316-24.
194. Slooter, A.J., et al., *Risk estimates of dementia by apolipoprotein E genotypes from a population-based incidence study: the Rotterdam Study.* Arch Neurol, 1998. **55**(7): p. 964-8.
195. Swaminathan, S., et al., *Analysis of Copy Number Variation in Alzheimer's Disease: the NIA-LOAD/NCRAD Family Study.* Curr Alzheimer Res, 2012.
196. Corneveaux, J.J., et al., *Association of CR1, CLU and PICALM with Alzheimer's disease in a cohort of clinically characterized and neuropathologically verified individuals.* Hum Mol Genet, 2010. **19**(16): p. 3295-301.
197. Webster, J.A., et al., *Genetic control of human brain transcript expression in Alzheimer disease.* Am J Hum Genet, 2009. **84**(4): p. 445-58.
198. Myers, A.J., et al., *A survey of genetic human cortical gene expression.* Nat Genet, 2007. **39**(12): p. 1494-9.
199. Crook, R., J. Hardy, and K. Duff, *Single-day apolipoprotein E genotyping.* J Neurosci Methods, 1994. **53**(2): p. 125-7.

200. Hawkins, J.R., et al., *Miniaturized sealed-tube allele-specific PCR.* Hum Mutat, 2002. **19**(5): p. 543-53.
201. Wallace, B.C., et al., *Meta-Analyst: software for meta-analysis of binary, continuous and diagnostic data.* BMC Med Res Methodol, 2009. **9**: p. 80.
202. Borenstein, M., et al., *Comprehensive Meta-analysis Version 2, Biostat, Englewood NJ.* 2005.
203. Hamza, T.H., et al., *Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease.* Nat Genet, 2010. **42**(9): p. 781-5.
204. Guo, Y., et al., *HLA rs3129882 variant in Chinese Han patients with late-onset sporadic Parkinson disease.* Neurosci Lett, 2011. **501**(3): p. 185-7.
205. International Multiple Sclerosis Genetics Consortium, et al., *Risk alleles for multiple sclerosis identified by a genomewide study.* N Engl J Med, 2007. **357**(9): p. 851-62.
206. Hoppenbrouwers, I.A., et al., *Replication of CD58 and CLEC16A as genome-wide significant risk genes for multiple sclerosis.* J Hum Genet, 2009. **54**(11): p. 676-80.
207. Guerini, F.R., et al., *HLA-A*01 is associated with late onset of Alzheimer's disease in Italian patients.* Int J Immunopathol Pharmacol, 2009. **22**(4): p. 991-9.
208. Listi, F., et al., *Association between the HLA-A2 allele and Alzheimer disease.* Rejuvenation Res, 2006. **9**(1): p. 99-101.
209. Rachidi, M., et al., *C21orf5, a human candidate gene for brain abnormalities and mental retardation in Down syndrome.* Cytogenet Genome Res, 2006. **112**(1-2): p. 16-22.
210. McNaughton, D., et al., *Duplication of amyloid precursor protein (APP), but not prion protein (PRNP) gene is a significant cause of early onset dementia in a large UK series.* Neurobiol Aging, 2012. **33**(2): p. 426 e13-21.
211. Mayo, P., et al., *CNV analysis using TaqMan copy number assays.* Curr Protoc Hum Genet, 2010. **Chapter 2**: p. Unit2 13.
212. Alkan, C., B.P. Coe, and E.E. Eichler, *Genome structural variation discovery and genotyping.* Nat Rev Genet, 2011. **12**(5): p. 363-76.
213. Krumm, N., et al., *Copy number variation detection and genotyping from exome sequence data.* Genome Res, 2012.
214. Abyzov, A., et al., *CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing.* Genome Res, 2011. **21**(6): p. 974-84.
215. Lee, C., A.J. Iafrate, and A.R. Brothman, *Copy number variations and clinical cytogenetic diagnosis of constitutional disorders.* Nat Genet, 2007. **39**(7 Suppl): p. S48-54.
216. Miller, D.T., et al., *Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies.* Am J Hum Genet, 2010. **86**(5): p. 749-64.

CURRICULUM VITAE

Shanker Swaminathan

Education

08/2007-08/2012 **PhD in Medical and Molecular Genetics**
Indiana University, Indianapolis, IN
GPA: 3.94
Thesis title: 'Role of genomic copy number variation in Alzheimer's disease and Mild cognitive impairment'
Thesis advisor: Andrew J. Saykin, PsyD

08/2003-08/2007 **BTech in Biomedical Engineering**
Sathyabama University, Chennai, India
Overall grade: 81%

Honors and Awards

2011 2011 Semifinalist for Trainee Research Award, American Society of Human Genetics for abstract submission to the 12th International Congress of Human Genetics

Research Experience

2009-Present **Indiana University School of Medicine, Indianapolis, IN**
Projects:
1. Role of copy number variation in AD and MCI using participants in the ADNI study, the NIA-LOAD/NCRAD Family Study and a unique cohort of clinically characterized and neuropathologically verified individuals
2. Genome-wide association study of cerebrospinal fluid measures from participants in the ADNI cohort
3. Gene-based association analysis of PiB-PET data from participants in the ADNI cohort
4. Characteristics of bipolar patients grouped by externalizing disorders

12/2006-03/2007 **Frontier Lifeline Hospitals, International Center for Cardio Thoracic and Vascular Diseases, Chennai, India**
Project: DNA Polymorphism (Pro12Ala) of the Peroxisome Proliferator Activated Receptor (γ) Gene and Obesity among Indians

12/2005 **Madras Medical Mission, Chennai, India**
Project: Cytogenetic Analysis of Syndromes of Children Suspected with Chromosomal Abnormalities

Related Professional Experience

01/2008-05/2008	Math tutor for undergraduate students, Mathematics Assistance Center, Indiana University-Purdue University, Indianapolis, IN
08/2008	Math mentor for high school students, Summer Bridge Academy 2008, Indiana University-Purdue University, Indianapolis, IN

Membership of Professional Organization

2010-Present	The American Society of Human Genetics (ASHG)

Publications and Abstracts

Publications

Swaminathan S, Shen L, Kim S, Inlow M, West JD, Faber KM, Foroud T, Mayeux R, Saykin AJ, the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the NIA-LOAD/NCRAD Family Study Group. Analysis of Copy Number Variation in Alzheimer's Disease: the NIA-LOAD/NCRAD Family Study. Current Alzheimer Research, 2012.

Swaminathan S, Shen L, Risacher SL, Yoder KK, West JD, Kim S, Nho K, Foroud T, Inlow M, Potkin SG, Huentelman MJ, Craig DW, Jagust WJ, Koeppe RA, Mathis CA, Jack CR Jr, Weiner MW, Saykin AJ and the Alzheimer's Disease Neuroimaging Initiative (ADNI). Amyloid pathway-based candidate gene analysis of [^{11}C]PiB-PET in the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort. Brain Imaging and Behavior, 2012. 6(1): 1-15.

Swaminathan S, Kim S, Shen L, Risacher SL, Foroud T, Pankratz N, Potkin SG, Huentelman MJ, Craig DW, Weiner MW, Saykin AJ, and the Alzheimer's Disease Neuroimaging Initiative. Genomic Copy Number Analysis in Alzheimer's Disease and Mild Cognitive Impairment: An ADNI Study. International Journal of Alzheimer's Disease, 2011. 2011: 729478.

Kim S*, **Swaminathan S***, Shen L, Risacher SL, Nho K, Foroud T, Shaw LM, Trojanowski JQ, Potkin SG, Huentelman MJ, Craig DW, DeChairo BM, Aisen PS, Petersen RC, Weiner MW, Saykin AJ, For the Alzheimer's Disease Neuroimaging Initiative. Genome-wide association study of CSF biomarkers Aβ1-42, t-tau and p-tau181p in the ADNI cohort. Neurology, 2011. 76(1): 69-79.

Saykin AJ, Shen L, Foroud TM, Potkin SG, **Swaminathan S**, Kim S, Risacher SL, Nho K, Huentelman MJ, Craig DW, Thompson PM, Stein JL, Moore JH, Farrer LA, Green RC, Bertram L, Jack CR Jr, Weiner MW, and Alzheimer's Disease Neuroimaging Initiative. Alzheimer's Disease Neuroimaging Initiative biomarkers as quantitative phenotypes: Genetics core aims, progress, and plans. Alzheimer's and Dementia, 2010. 6(3): 265-273.

Shen L, Kim S, Risacher SL, Nho K, **Swaminathan S**, West JD, Foroud T, Pankratz N, Moore JH, Sloan CD, Huentelman MJ, Craig DW, DeChairo BM, Potkin SG, Jack CR Jr, Weiner MW, Saykin AJ, and the Alzheimer's Disease Neuroimaging Initiative. 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, 2010. 53(3): 1051-1063.

Abstracts

Swaminathan S, Shen L, Kim S, Inlow M, West JD, Foroud T, Faber KM, Mayeux R, Saykin AJ, The NIA-LOAD/NCRAD Family Study Group. Analysis of copy number variation in Alzheimer's disease: the NIA-LOAD/NCRAD Family Study. 12th International Congress of Human Genetics/61st Annual Meeting of the American Society of Human Genetics (ASHG); October 11-15, 2011; Montreal, Canada.

Swaminathan S, Risacher SL, Yoder KK, West JD, Shen L, Kim S, Inlow M, Foroud T, Jagust WJ, Koeppe RA, Mathis CA, Shaw LM, Trojanowski JQ, Soares H, Aisen PS, Petersen RC, Weiner MW, Saykin AJ and the Alzheimer's Disease Neuroimaging Initiative (ADNI). *APOE* ε4 status modulates the association of plasma and cortical Aβ and [^{11}C]PiB PET in the ADNI cohort. Alzheimer's Association International Conference (AAIC) 2011; July 16-21, 2011; Paris, France.

Swaminathan S, Shen L, Risacher SL, Kim S, Nho K, West JD, Foroud T, Yoder KK, Potkin SG, Huentelman MJ, Craig DW, Koeppe RA, Jagust WJ, Mathis CA, Weiner MW, Alzheimer's Disease Neuroimaging Initiative. Amyloid pathway-based candidate gene analysis of [^{11}C]PiB-PET in the Alzheimer's Disease Neuroimaging Initiative cohort. Presented at the 60th Annual Meeting of The American Society of Human Genetics (ASHG); November 2-6, 2010; Washington, DC.

Swaminathan S, Kim S, Shen L, Risacher SL, Foroud T, Pankratz N, Potkin SG, Huentelman MJ, Craig DW, Weiner MW, Saykin AJ, The Alzheimer's Disease Neuroimaging Initiative. Preliminary analysis of copy number variation in the ADNI cohort. Presented at the Alzheimer's Association International Conference on Alzheimer's Disease (AAICAD) 2010; July 10-15, 2010; Honolulu, HI.

Swaminathan S, Koller DL, Foroud TM, Xuei X, Edenberg HJ, Niculescu AB, Nurnberger JI Jr, BIGS Consortium. Characteristics of bipolar patients grouped by externalizing disorders. Presented at the 2009 Annual Indiana University School of Medicine and Clinical and Translational Sciences Institute (CTSI) Scientific Poster Session; September 16, 2010; Indianapolis, IN.

Nho K, Shen L, Kim S, **Swaminathan S**, Risacher SL, Saykin AJ, and the Alzheimer's Disease Neuroimaging Initiative (ADNI). The Effect of Reference Panels and Software Tools on Genotype Imputation. AMIA 2011: Annual Symposium on Biomedical and Health Informatics; October 22-26, 2011; Washington, DC.

Shen L, Kim S, Qi Y, Inlow M, **Swaminathan S**, Nho K, Wan J, Risacher SL, Shaw LM, Trojanowski JQ, Weiner MW, Saykin AJ, and ADNI. Identifying Neuroimaging and Proteomic Biomarkers for MCI and AD via the Elastic Net. MBIA 2011: International Workshop on Multimodal Brain Image Analysis; September 18, 2011; Toronto, Canada.

Wan J, Kim S, Inlow M, Nho K, **Swaminathan S**, Risacher SL, Fang S, Weiner M, Beg F, Wang L, Saykin AJ, Shen L, ADNI. Hippocampal surface mapping of genetic risk factors in AD via sparse learning models. MICCAI 2011: the 14th International Conference on Medical Image Computing and Computer Assisted Intervention; September 18-22, 2011; Toronto, Canada.

Kim S, **Swaminathan S**, Inlow M, Risacher SL, Shen L, Foroud T, Shaw LM, Trojanowski JQ, Soares H, Weiner MW, Saykin AJ and the Alzheimer's Disease Neuroimaging Initiative (ADNI). Influence of genetic variation on plasma proteomics in AD, MCI and controls: Pairwise gene-protein analysis in the ADNI-1 cohort. Alzheimer's Association International Conference (AAIC) 2011; July 16-21, 2011; Paris, France.

Wan J, Kim S, Nho K, Risacher S, **Swaminathan S**, Bertram L, Jack Jr C, Weiner M, Beg F, Wang L, Saykin A, Shen L. Influence of Candidate AlzGene SNPs on Hippocampal Shape: A Study of the ADNI Cohort. 17th Annual Meeting of the Organization for Human Brain Mapping; June 26-30, 2011; Quebec City, Canada.

Deters K, Nho K, **Swaminathan S**, Risacher SL, Kim S, Foroud T, Shen L, Saykin AJ, and the Alzheimer's Disease Neuroimaging Initiative (ADNI). Association of Haplotypes of the *MAPT* Gene with Alzheimer's Disease and Quantitative Neuroimaging Phenotypes. Annual Biomedical Research Conference for Minority Students; November 10-13, 2010; Charlotte, NC.

Kim S, **Swaminathan S**, Shen L, Risacher SL, Nho K, Foroud T, Shaw LM, Trojanowski JQ, Potkin SG, Huentelman MJ, Craig DW, DeChairo BM, Weiner MW, Saykin AJ, Alzheimer's Disease Neuroimaging Initiative. Genome-wide association study of CSF biomarkers amyloid-beta 1-42, tau and tau phosphorylated at threonine 181 in the ADNI cohort. Alzheimer's Association International Conference on Alzheimer's Disease (AAICAD) 2010; July 10-15, 2010; Honolulu, HI.

Shen L, Wang J, Kim S, McCullough K, Nho K, **Swaminathan S**, West JD, Fang S, McHugh T, Flashman LA, Wishart HA, Rabin LA, Rhodes CA, Guerin SJ, Moore JH, Santulli RB, Saykin AJ. Association analysis of candidate SNPs on hippocampal volume and shape in mild cognitive impairment and older adults with cognitive complaints. Alzheimer's Association International Conference on Alzheimer's Disease (AAICAD) 2010; July 10-15, 2010; Honolulu, HI.