HUMAN GENETICS

INVESTIGATION OF GENETIC SUSCEPTIBILITY TO LATE-ONSET ALZHEIMER DISEASE THROUGH GENOMIC CONVERGENCE

XUEYING LIANG

Dissertation under the direction of Professor Jonathan L. Haines

With the exception of ApoE gene, no universally accepted genetic association has been identified with the complex Late-onset Alzheimer Disease (LOAD). A broad region of chromosome 10 has engendered continued interest generated from both preliminary genetic linkage and candidate gene studies.

To better examine this region, we applied the genomic convergence approach by combining unbiased genetic linkage with candidate gene association studies. We genotyped 36 SNPs across 80.2 Mb in 567 multiplex families to narrow the peak region of linkage using both covariate and subset analyses. Simultaneously, we examined seven functional candidate genes that also fell within the broad area of linkage. Although a two point LOD score of 2.69 was obtained in the linkage analysis, the associated candidate genes were not under the linkage peak, suggesting a more extensive heterogeneity on chromosome10 than previously expected.

We then converged linkage analysis and gene expression data to identify genes that were under linkage peaks and also differentially expressed in AD cases and controls based on the rationale that genes showing positive results in multiple studies are more

likelihood to be involved in AD. We identified and examined 28 genes on chromosome 10 for the association with AD. Both single marker and haplotypic associations were tested in overall and eight subsets that were stratified by age, gender, ApoE status and clinical diagnosis. Gene-gene interaction was tested to detect important genes in this complex disease. PTPLA gene showed allelic, genotypic and haplotypic association in the overall dataset. The SORCS1 gene showed very significant association in the female dataset (allelic association p=0.00002, a 3-locus haplotype has p=0.00098). Two SNPs in CACNB2 gene showed gene-gene interaction in overall dataset using Multifactor Dimensionality Reduction (MDR).

The work presented in this dissertation applied a multifactorial, multistep approach, genomic convergence, which combined linkage analysis, gene expression data, and candidate gene association analysis to identify and prioritize candidate susceptibility genes for AD. This study suggests that genetic variations in PTPLA, SORCS1 and CACNB2 genes might alter the risk for Alzheimer disease by affecting multiple pathways.

Approved ______________________________________ Date______________________

INVESTIGATION OF GENETIC SUSCEPTIBILITY TO LATE-ONSET ALZHEIMER DISEASE THROUGH GENOMIC CONVERGENCE

By

Xueying Liang

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Human Genetics

May, 2007

Nashville, Tennessee

Approved:

Professor Jonathan L. Haines

Professor Marylyn D. Ritchie

Professor Scott M. Williams

Professor Harry E. Gwirtsman

Professor Douglas P. Mortlock

Copyright © 2007 by Xueying Liang

All Rights Reserved

To my parents for always believing in me
Jun Liang and Yuzhi Zhao

To my son, the most wonderful part of my life
Andrew Wang

And most importantly to my loving husband who is infinitely supportive
Yongqing Wang

ACKNOWLEDGEMENTS

The work presented here was supported by NIH grants AG019757, AG021547, AG20135, and a grant from the Alzheimer Association (IIRG-00-2006). None of this work would have been possible without the generous participation of families and the collaborating clinics and clinicians that referred patients to our studies.

I am greatly indebted to the members of my thesis committee (Dr. Jonathan L. Haines, Dr. Marylyn D. Ritchie, Dr. Scott M. Williams, Dr. Harry E. Gwirtsman, and Dr. Douglas, P. Mortlock) for their invaluable time, professional guidance, and tremendous support. I would especially like to thank my Ph. D. mentor, Dr. Jonathan L. Haines. I am extremely fortunate to have found an advisor who conducts science with ingenuity and integrity. Without his direction, none of my accomplishments would have been possible. I am forever grateful of his constant unwavering support.

I wish to thank the members of the Haines lab, present and past (Dr. Nathalie Schnetz-Boutaud, Brent Anderson, Melissa Allen, Kylee Spencer, Dr. Jake McCauley, Becca Zuvich, Dr. Shannon Kenealy, and Dr. Tricia Thornton-Wells), for their contributions and dedication to all the ongoing projects in the lab and for the friendships that made the journey enjoyable.

I also acknowledge the members of the Center for Human Genetics Research (CHGR), including the Computational Genomics Core (Jackie Bartlett, Kim Luci, Chara Gibson, Justin Giles, and Lana Olson), Family Ascertainment Core (Benita Lynch), DNA Resources Core (Cara Sutcliff and Ping Mayo), and Scott Dudek in Ritchie lab.

I also like to thank fellow graduate students Will Bush, Todd Edwards, Digna Velez, and Kelli Ryckman for their support and friendship. I would like to express my

gratitude to my Cousin, Xiuping Yu, and friend, Larry Raper, for their help in preparing the dissertation.

No one has been more important to me in the pursuit of this project than the members of my family. I would like to thank my parents, Jun Liang and Yuzhi Zhao, and my sister, Xueming Liang, who love and believe in me whatever I pursue. Most importantly, huge thanks go to my loving husband, Yongqing, and my wonderful son, Andrew, who provide continuous support and motivation for the timely completion of this dissertation.

TABLE OF CONTENTS

Page

DEDICATION ... ii

ACKNOWLEDGEMENTS ... iii

LIST OF TABLES ... ix

LIST OF FIGURES ... xi

LIST OF ABBREVIATIONS ... xiii

Chapter

I. OVERVIEW ... 1

II. INTRODUCTION AND BACKGROUND ... 4

Clinical Aspects of Alzheimer Disease (AD) ... 4
Alzheimer Disease ... 4
Diagnosis of Alzheimer Disease ... 5
Genetic Epidemiology of Alzheimer Disease ... 8
Population Prevalence ... 8
Familial Aggregation ... 9
Genetic Analysis of Alzheimer Disease ... 11
Approaches ... 11
Known Genes Involved in Alzheimer Disease ... 15
The Search for New AD Susceptibility Genes ... 18

III. HYPOTHESIS AND SPECIFIC AIMS ... 24

Specific Aim I ... 24
Specific Aim II ... 25
Specific Aim III ... 25

IV. IDENTIFICATION OF THE MINIMUM CANDIDATE REGION (MCR) FOR ALZHEIMER DISEASE ON CHROMOSOME 10 USING LINKAGE ANALYSIS ... 27

Introduction ... 28
Materials and Methods ... 29
Study Populations ... 29

SNP Selection and Genotyping..31
Statistical Methods..33
Results..35
Two-point Linkage Analysis..35
Multipoint Linkage Analysis 39
Discussion 44

V. EXAMINATION OF FIVE CANDIDATE GENES FOR ALZHEIMER DISEASE 47

Introduction..47
Materials and Methods...47
Study Populations 47
SNP Selection and Genotyping..49
Statistical Methods...53
Results..55
Candidate Gene Association Analysis 55
Linkage Analysis in Candidate Genes 58
Genomic Convergence of Association and Linkage Analysis.......59
Discussion 60

VI. GENETICS OF CDC2 LOCUS IN ALZHEIMER DISEASE 64

Introduction..64
Materials and Methods...66
Study Populations 66
SNP and Genotyping..68
Statistical Methods...70
Power Calculation..72
Results..73
Family-based Analysis...73
Case-control Analysis 76
Discussion 79

VII. EXAMINATION OF THE ASSOCIATION BETWEEN RS498055 IN LOC439999 GENE ON CHROMOSOME 10 AND ALZHEIMER DISEASE ...81

Introduction..81
Materials and Methods...82
Study Populations 82
Genotyping...83
Family-based Data Set Analysis 84
Case-control Data Set Analysis 84
Results..85
Discussion 87

VIII. GENOMIC CONVERGENCE TO IDENTIFY CANDIDATE GENES FOR ALZHEIMER DISEASE 89

Introduction 89
Materials and Methods 90
Human Brain Samples 90
Construction of SAGE Libraries 91
SAGE Data Analyses 92
Gene Selection 92
SNP Selection 93
Genotyping 94
HWE and Linage Disequilibrium Test 94
Results 95
Genomic Convergence Identified Genes 95
Quality Control of Genotyped SNPs 97
Linkage Disequilibrium Patterns of the 28 Genes in the Overall Data Set 99
Example of Haplotype Blocks in Genes 100
Discussion 103

IX. ASSOCIATION ANALYSIS OF GENOMIC CONVERGENCE GENES FOR ALZHEIMER DISEASE 107

Introduction 107
Materials and Methods 110
Study Populations 110
Power Study 113
Statistical Methods 120
Results 122
Power Study 122
Allelic Association Analysis in the Overall and Subsets 123
Genotypic Association Analysis in the Overall and Subsets 127
Effect Sizes of the Most Significant SNPs in Two Genes 130
Haplotypic Association Analysis 135
Analytical Convergence of the Genetic Association Tests 137
Follow Up of SNPs in SORCS1 and PTPLA Genes 138
Discussion 141

X. GENE-GENE INTERACTION AMONG GENOMIC CONVERGENCE GENES IN ALZHEIMER DISEASE 148

Introduction 148
Materials and Methods 151
Study Population 151
Genotyping 151
Statistical Analysis 152

Results 155
Interaction Among Genes on Chromosome 10 155
Interaction Among All Genomic Convergence Genes in AD 158
Discussion 160

XI. CONCLUSIONS AND FUTURE DIRECTIONS 163

Summary and Conclusions 163
Future Directions 166

APPENDIX 169

REFERENCES 203

LIST OF TABLES

Table Page

2-1 NINCDS-ADRDA criteria for Alzheimer Disease 6

4-1 Study populations for Alzheimer Disease..30

5-1 Genotyped SNPs in five candidate genes 51

5-2 Candidate genes and SNPs showing association in at least one data set 56

6-1 CDC2 SNP allele and genotype frequencies for family-based data set.................74

6-2 Linkage disequilibrium between SNPs in CDC2 (D' and r^2) 75

6-3 Linkage and association results for family-based sample.....................................75

6-4 Allelic and genotypic association for case-control samples. 77

6-5 Association of Age-at-onset and CDC2 gene in case-control samples..................78

7-1 Association and linkage results for family-based sample.....................................85

7-2 Allelic and genotypic association for case-control samples 86

8-1 Summary of linkage study results on chromosome 10 92

8-2 Genomic convergence identified Genes 96

8-3 Quality control of 667 genotyped SNPs 98

8-4 Haplotype blocks and tag SNPs in each gene...99

9-1 Alzheimer disease case control data set..111

9-2 Overall and subsets of study populations for Alzheimer disease 111

9-3 Family-based data sets for Alzheimer disease 112

9-4 Genetic models..114

9-5 Genotype frequency table for 2-disease loci models 115

9-6 Penetrances for three inheritance models 116

9-7 Marginal and joint ORs in dominant model 118

9-8 Marginal and joint ORs in recessive model 118

9-9 Marginal and joint ORs in additive model 118

9-10 Parameters for genetic models 119

9-11 The allelic association results (nominal P-values) in subsets for the SNPs showing significant association (p<0.05) in the overall data set 126

9-12 The genotypic association in subsets for the SNPs showing significant association (p<0.05) in overall data set 129

9-13 Effect size of SORCS1 (rs17277986) in the overall data set 130

9-14 Effect size of SORCS1 (rs17277986) in the female subset 130

9-15 Effect size of SORCS1 (rs17277986) adjusted by covariates 131

9-16 Effect size of PTPLA in the overall data set 133

9-17 Effect size of PTPLA adjusted by covariates 134

9-18 Haplotypic association in the overall data set 136

9-19 Family-based association tests for rs17277986 in SORCS1 in validation data sets 138

9-20 Allelic and genotypic association of all genotyped SNPs in PTPLA in the overall data set 140

9-21 Associated genes in the overall and subsets after the FDR correction for multiple comparisons (q=0.2) 142

10-1 Summary of MDR results for genes on chromosome 10 156

10-2 The frequency of 3 genes in the top 10 MDR models from 5 cross-validations 158

10-3 Summary of MDR results in the female subset 158

10-4 Summary of MDR results among genomic convergence genes in AD 159

LIST OF FIGURES

Figure Page

4-1 The position of previously reported candidate genes and 36 genotyped SNPs for linkage analysis on chromosome 10....32

4-2 Original two-point (FASTLINK, shown in dots) and multipoint (MERLIN, shown in line) LOD scores in overall and three subsets....36

4-3 Two point analyses in the overall data set....37

4-4 Two point and multipoint linkage analyses using the D12S368 LOD score as a covariate to order families from low to high in the overall and three subsets....38

4-5 Multipoint analyses using covariates to order families from low to high in overall and three subsets....40

4-6 OSA multipoint analyses (D12S368 LOD Low to High)....41

4-7 Multipoint analyses using covariates to order families from high to low in the overall and three subsets....43

5-1 Overall suggestive linkage and association data on chromosome 10....59

6-1 Gene structure and relevant features of CDC2 and surrounding sequences....68

8-1 Functional classification of genotyped SNPs....98

8-2 Example of the linkage disequilibrium structure of genes with one haplotype block (eg. HELLS)....100

8-3 Linkage disequilibrium structure of TNFRSF6 gene....101

8-4 Linkage disequilibrium structure of EIF4EBP2 gene....102

8-5 Example of linkage disequilibrium structure of genes with more than eight haplotype blocks (eg. SORCS1)....103

9-1 Power of the data set with 500 cases and 500 controls at $p<0.05$....123

9-2 Allelic association test in the overall data set....124

9-3 Allelic association analysis in the female subset....127

9-4 Genotypic association test in the overall data set 128

9-5 Genotypic association analysis in the female subset 128

9-6 Haplotype analysis in SORCS1 in the female subset 137

9-7 Linkage disequilibrium between all nine genotyped SNPs in PTPLA 139

9-8 Haplotypic association analysis using 3-SNP sliding window in PTPLA...........140

9-9 Genomic DNA comparisons of intron 1 in SORCS1 between Human and other Mammals..144

10-1 Summary of the general steps to implement the MDR method 154

11-1 Analytical convergence in Alzheimer disease 164

11-2 Hypothesized pathways involved in Alzheimer Disease 165

LIST OF ABBREVIATIONS

3MS	Modified Mini-Mental State Examination
AAE	age at examination
AAO	age at onset
AD	Alzheimer Disease
ApoE	Apolipoprotein E
APP	β-amyloid precursor protein
ASP	affected sibpairs
Aβ	amyloid β
bp	Base Pair
CAP	Collaborative Alzheimer Project
CBC	complete blood count
CDC2	Cell Division Cycle 2
cDNA	Complementary Deoxyribonucleic Acid
CEPH	Centre d'Etude du Polymorphism Humain
CHGR	Center for Human Genetics Research
cM	CentiMorgan
CMP	comprehensive metabolic panel
CT	computed tomography
DNA	Deoxyribonucleic Acid
DSP	discordant sib pairs
EEG	Electroencephalography
Ensembl	Genome Browser

EOAD	early-onset Alzheimer Disease
FBAT	Family based association test
FDG	Fluorodeoxyglucose
FDR	False Discovery Rate
GLM	generalized liner model
GOLD	Graphical Overview of Linkage Disequilibrium
Haploview	Java-based Tool for Visualizing LD blocks
HLOD	heterogeneity LOD score
htSNPs	haplotype tagged SNPs
HWE	Hardy-Weinberg equilibrium
IDE	insulin degrading enzyme
kb	Kilobase
LD	Linkage disequilibrium
LOAD	late-onset Alzheimer Disease
LOD	Logarithm of the Odds
LRRTM3	Leucine-rich repeat transmembrane neuronal 3 protein
Mb	Megabase
MDR	Multifactor Dimensionality Reduction
μg	Microgram
MMSE	Mini-Mental State Exam
MRI	magnetic resonance imaging
mRNA	Messenger Ribonucleic Acid
NCRAD	National Cell Repository for Alzheimer Disease

NIMH	National Institute of Mental Health repository
NINCDS-ADRDA	National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer disease and Related Disorders Association
OPA	Oligo Pool Assay
OR	Odds Ratio
OSA	Ordered Subset Analysis
PCR	Polymerase Chain Reaction
PDT	pedigree disequilibrium test
PET	positron emission tomography
PLAU	urokinase-type plasminogen activator
PS1	presenilin 1
PS2	presenilin 2
QC	quality control
RPR	rapid plasma regain
SAGE	Serial Analysis of Gene Expression
SNP	Single nucleotide polymorphism
Taq	*Thermus Aquaticus* Polymerase
TFBS	transcription factor binding sites
TNFRSF6	tumor necrosis factor receptor superfamily, member 6
UTR	Untranslated Region
VISTA	Visualization Tools for Alignment
VR22	alpha-T-catenin

CHAPTER I

OVERVIEW

Alzheimer disease (AD) is a devastating neurodegenerative disease of elderly people and one of the most serious public health problems facing an increasingly aging society. A lot of effort has been put into identifying genes that underlie the etiology of Alzheimer disease. However, like many other common diseases, Alzheimer disease is a complex disorder involving multiple genes together with non-genetic factors. Therefore, the likelihood of a single method giving us an answer in dissecting complex disease traits is very low. Thus, I proposed a genomic convergence and peak-wide association approach to identify and characterize genes involved in the susceptibility of Alzheimer disease. I converged linkage, gene expression and candidate gene studies to identify a subset of genes. These genes were investigated for association in not only the overall data set, but also in more homogeneous subsets, and then main effects and gene-gene interactions were examined in each of these subsets.

The introduction and background for Alzheimer disease is presented in chapter II. The clinical aspects and genetic epidemiology of Alzheimer disease are presented. Current approaches for genetic analysis of complex diseases, known genes (APP, PS1, PS2 for early-onset AD, ApoE for late-onset AD), and the progress and efforts to find new susceptibility genes for AD are introduced.

Chapter III lays out my hypothesis and three specific aims.

Numerous studies have tested hundreds of candidate genes and many such genes were reported as associating with AD. However, none of the genes has been universally accepted as an AD risk gene except ApoE. Multiple genomic screens showed evidence that chromosome 10 may harbor susceptibility genes for AD. However, the reported linkage region scatters across an 80 Mb region on chromosome 10 and provides an impractically large area for a molecular genetic search. To examine this region in more detail, we tested several recently reported candidate genes within this broad region of linkage, VR22, LRRTM3, PLAU, TNFRSF6, IDE, CDC2, and LOC439999. Simultaneously, we examined the linkage signal in more detail to see if we could narrow the peak region of linkage and converge the candidate gene and linkage data.

Chapter IV presents the linkage analysis approach to identify the minimum candidate region for AD on chromosome 10. Chapter V presents the candidate gene studies on the genes VR22, LRRTM3, PLAU, TNFRSF6, and IDE. There is at least one SNP showing association with AD in each gene in at least one of our data sets. Chapter VI details the examination of a separate candidate gene, CDC2, with AD. Due to the property of the SNPs in CDC2, we used a special genotyping method to test for association with this gene. However, our results didn't show evidence of association between CDC2 and AD. Chapter VII presents the association study of a polymorphism rs498055 in gene LOC439999 on chromosome 10. This SNP was identified from a study of 1412 functional SNPs in 667 genes under the previously identified linkage peak on chromosome 10. However, we didn't confirm this association in three independent family-based and one independent case-control datasets.

Chapter VIII details the genomic convergence approach to identify candidate genes for Alzheimer disease and the genomic analysis of the identified candidate genes.

Chapter IX presents the comprehensive association analysis of genomic convergence genes with AD including allelic, genotypic and haplotypic association in the overall and more homogenous subsets. Genetic variations in PTPLA and SORCS1 demonstrated association with AD with modest effect size. The replication of the effect of these genes in different study populations and the search for susceptible variants and functional studies of these genes are necessary to get a better understanding of the roles of the genes in Alzheimer disease.

Chapter X presents the examination of potential gene-gene interaction involved in Alzheimer disease. Epistasis, or gene-gene interaction, is crucial in detecting polymorphisms associated with an increased risk of disease. In this chapter, we applied Multifactor Dimensionality Reduction (MDR) to the 1536 SNPs in genomic convergence candidate genes on a population data set (506 cases and 558 controls) to identify potential gene-gene interactions involved in Alzheimer disease.

Chapter XI describes the conclusions and future directions.

CHAPTER II

INTRODUCTION AND BACKGROUND

Clinical Aspects of Alzheimer Disease (AD)

Alzheimer Disease

Although various forms of dementia have always been part of the human experience, the efforts toward a systematic description of the clinical-pathologic characteristics of the phenomenon gained traction in the late 1890s, roughly paralleling advances in histology, light microscopy, neuroanatomy, and other areas of biology (Berchtold and Cotman, 1998; Beach, 1987).

The 1907 report by Alois Alzheimer became the index case for "Alzheimer Disease" (Khachaturian, 1985; Khachaturian, 2006). Alzheimer disease (AD) is the leading cause of dementia in the elderly and the most common form of dementia occurring after the age of 40 (Rocca et al., 1991; Schoenberg et al., 1987). It is a devastating neurodegenerative disorder of later life with a complex inheritance (Growden, 1995; Khachaturian, 1985). Alzheimer disease targets specific brain regions early in its course, especially the cholinergic basal forebrain and medial temporal lobe structures including the hippocampus, amygdala, and entorhinal cortex (Khachaturian, 1985; Braak and Braak, 1995; Braak and Braak, 1997; Nagy et al., 1999; Beach et al., 2000).

Clinically AD progresses slowly, resulting in memory loss and alterations of higher intellectual function and cognitive abilities (Guttman et al., 1999). Pathologically

AD is characterized by neurofibrillary tangles in the neurons of the cerebral cortex and hippocampus and the deposition of amyloid within senile plaques and cerebral blood vessels (Wisniewski et al., 1993). Both characteristics are prerequisites for a confirmed diagnosis of AD.

Diagnosis of Alzheimer Disease

Criteria for Alzheimer Disease. The clinical diagnosis of AD requires both physical and neuropsychological examinations. The standardized AD diagnostic criteria was developed by the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer disease and Related Disorders Association (NINCDS-ADRDA) (McKhann et al., 1984a), shown in Table 2-1. These criteria divide patients into definite AD, probable AD, and possible AD. Definite AD is the most accurate diagnosis, but it can only be made after death. It requires that the patient meets the criteria for probable AD as well as histopathological evidence of AD by autopsy, which is the presence of the plaques and neurofibrillary tangles. There are several criteria for probable AD. If a patient meets three of these, he/she is diagnosed in this category. For example, the patient has progressive deficits in two or more areas of cognition, including memory, which is tested by mini-mental state exam. The onset should be between the ages of 40 and 90 years. There should be absence of systemic or other brain diseases capable of producing a dementia syndrome. A clinical diagnosis of possible AD may be made on the basis of the dementia syndrome, in the absence of other neurologic, psychiatric or systemic disorders sufficient to cause dementia.

Table 2-1 NINCDS-ADRDA criteria for Alzheimer Disease

Criteria
Definite AD
Clinical criteria for probable AD;
Histopathological evidence obtained from autopsy.
Probable AD
Dementia established by Mini-Mental State Exam or Blessed Dementia Scale or similar exams and confirmed by neuropsychological tests;
Deficits in two or more areas of cognition;
Progressive worsening of memory and other cognitive functions;
No disturbance of consciousness;
Onset between ages 40 and 90, most often after age 65;
Absence of systemic disorders or other brain diseases that account for the symptoms.
Possible AD
Atypical onset, presentation, or clinical course of dementia;
Absence of other disorders sufficient to cause dementia;
Another illness capable of producing dementia is present but is not considered to be the primary cause.

Diagnostic evaluation. Alzheimer disease is not simply a diagnosis by exclusion. Standard methods of examination (medical and family history; neurologic, psychiatric and clinical examinations; neuropsychological tests, laboratory studies, and brain imaging) are needed to fulfill inclusionary and exclusionary criteria for the diagnosis of Alzheimer disease.

History and physical examination. Alzheimer disease has typically gradual onset. The symptoms begin with memory loss and extend to other cognitive deficits including anomia, constructional apraxia and often anosognosia. Both medical and family histories are important for the diagnosis. Family history of dementia in a parent or other relative is common. A medical history should be taken from the patient and from an informant who is well acquainted with the affected individual. This is essential to establish a history of progressive deterioration and for identifying tasks that the patient can no longer perform adequately. Mental status can be evaluated by Mini-Mental State

Examination or Modified Mini-Mental State Examination (3MS) (Folstein et al., 1975; Teng and Chui, 1987), that includes questions related to orientation, attention, learning, memory, language, and constructional praxis. The Blessed Dementia Scale (Blessed et al., 1968) and other scales are also available for measuring different aspects of the disease.

Neuropsychological testing. Neuropsychological tests may provide additional information for the diagnosis of dementia, particularly in assessing mild cases, medicolegally contested cases, and cases where major lifestyle changes will likely be imposed. They are used primarily to provide confirmatory evidence for the diagnosis of dementia. They are also valuable for determining patterns of impairment, for assessing changes in impairment over time and after drug treatment or rehabilitation, and for establishing correlations of abnormal performance with laboratory and neuropathologic examinations.

Laboratory assessment. There are no widely accepted, highly reliable, commercially available biomarkers for AD; therefore diagnosis rests on clinical recognition and exclusion of potentially contributory medical factors. The lab tests include thyroid, B12, Folate, rapid plasma regain (RPR), electrolytes, complete blood count (CBC), comprehensive metabolic panel (CMP), sedimentation rate, et al. Doppler test can be used for excluding cerebrovascular disease.

Brain imaging. Structural brain imaging with magnetic resonance imaging (MRI) or computed tomography (CT) is essential to assess relevant structural pathology such as brain tumors, vascular lesions, subdural hematomas, hydrocephalus, and other problems. In patients with AD, MRI and CT typically reveal nonspecific, mild to moderate atrophy that may be most pronounced in the most symptomatic regions, especially the medial

temporal lobe. Fluorodeoxyglucose (FDG) positron emission tomography (PET) may be helpful in distinguishing AD from frontotemporal dementia and other less common forms of dementia. Patients with AD have characteristic and progressive patterns of decline in regional glucose metabolism, beginning in posterior locations (the precuneus and posterior cingulated, posterior parietal, and temporal cortex), and subsequently affecting prefrontal cortex and the whole brain (Hoffman et al., 2000).

Genetic Epidemiology of Alzheimer Disease

Population prevalence

Epidemiological studies provide strong evidence for both environmental and genetic risk factors in AD. Population and family-based studies have been conducted to assess disease prevalence and aggregation in an attempt to identify and elucidate genetic contribution to the disease.

Prevalence studies suggest that the number of persons with Alzheimer disease in the United States was 4.5 million in 2000 (Hebert et al., 2003). This number is projected to triple over the next 50 years as the population ages. Owing to the rapid growth of the oldest age groups of the US population, the number who are 85 years and older will more than quadruple to 8.0 million. The prevalence of Alzheimer disease increases with advancing age. One percent of 60-year-olds and up to forty percents of 85-year-olds have the disease (Evans et al., 1989; Mullan et al., 1994). Also, the incidence increases with aging. The estimated annual incidence of Alzheimer disease is approximately 2.4

per 100,000 people age 40 to 60, 1 per 1,000 person-years in individuals aged 60-64 years, and 25 per 1,000 person-years in individuals older than 85 years (van Duijn, 1996).

Familial aggregation

Genetic factors play a role in the etiology of Alzheimer disease. Familial clustering has long been recognized (Akesson, 1969), and a positive family history of dementia is one of the most consistent risk factors (van Duijn et al., 1991). The gene mapping strategies applied to early-onset families have identified three early-onset AD loci: APP on chromosome 21 (Goate et al., 1991), PS1 on chromosome 14 (St George-Hyslop et al., 1992; Sherrington et al., 1995), and PS2 on chromosome 1 (Levy-Lahad et al., 1995; Rogaev et al., 1995).

However, identifying genetic mechanisms for the more common late-onset AD (LOAD) has been more elusive. This is due to several properties specific to AD. For example, diagnostic uncertainty, using life-to-date information of at-risk individuals, small effective family size, and incomplete family histories are the factors that hamper the family studies of late-onset Alzheimer disease. Despite these problems, data from the risk to relatives, segregation analysis and twin studies have suggested a strong genetic influence in LOAD (Bergen, 1994; Bonney, 1984; Breitner and Murphy, 1992; Farrer et al., 1991; Rao et al., 1994).

Studies that examined incidence of disease in relatives of AD probands, using survival analysis (Chase et al., 1983; Cupples et al., 1989; Cupples et al., 1991), showed that first degree relatives of a patient have higher risk of disease than unrelated individuals. The relationship between risk of disease, age and sex indicated that risk of

the disease increases as age increases. Cumulative risk increases by an average of 1.90 ± 0.79% per year between the ages of 80 and 90 (Cupples et al., 1991). The risk of developing AD was significantly higher in female relatives than male relatives at all ages. By age 93, women have a 13% higher risk than men of developing AD. The survival distribution showed a 39% risk of AD by age 96 years among first-degree relatives, which suggested that AD has a strong genetic component (Lautenschlager et al., 1996).

Segregation analysis is one method to test for the existence of inheritance. In Alzheimer disease, it is limited by changes in the definition of the disorder over time and by difficulty in accessing pedigree and medical history information for a disproportionately large number of deceased persons. However, data simulation using several computer programs (Bonney, 1984; Farrer et al., 1991; Rao et al., 1994; van Duijn et al., 1993; Morton and MacLean, 1974; Bonney, 1986) suggested that the most general (unrestricted) genetic model gave a significantly better fit than models in which the etiology was sporadic, environmental, or not due to any major gene. Apolipoprotein E (ApoE) genotypes have been identified as a significant genetic risk factor for late-onset Alzheimer disease. It was identified through an early use of genomic convergence combining linkage analysis results with biological information (Corder et al., 1993; Strittmatter et al., 1993). It is the only universally accepted susceptibility gene for late-onset AD. However, studies that stratified AD families by ApoE genotype of the probands revealed that there are likely other major genes for the common form of the disorder (Rao et al., 1996; Jarvik et al., 1996). These findings support the computer simulations of segregation analysis.

Adoption studies have not been done in AD because of its very late age at onset, which also makes it impossible to trace the histories of most of biological relatives. Twin studies also have not greatly clarified the role of genes in AD. However, the concordance rates among twins of monozygotic probands with AD are generally higher (22%-83%) than between dizygotic twins (0%-50%) (Bergen, 1994; Bergen et al., 1997; Breitner and Murphy, 1992; Breitner et al., 1993; Breitner et al., 1995b). All studies examined small numbers of twins, and this may introduce bias. Therefore, caution should be taken in interpreting the data, considering the sample size used in these studies. Nonetheless, the concordance rates suggest a strong, but not absolute, genetic influence (Breitner et al., 1995a; St George-Hyslop et al., 1989; Cook et al., 1981; Nee et al., 1987; Rapoport et al., 1991).

Taken together, these epidemiological studies provide evidence in support of a strong genetic component in AD. The data also suggest that, like most common complex diseases, AD susceptibility is the result of multiple genetic and non-genetic factors (Myers and Goate, 2001).

Genetic Analysis of Alzheimer Disease

Approaches

Understanding the genetics of AD is critical to developing new treatments. More and more evidence demonstrates that AD is a complex and heterogeneous disease with strong genetic and environmental determinants. It is considered to be *complex* because there is no single (or simple) mode of inheritance that accounts for its heritability. It is

considered to be *heterogeneous* because mutations and polymorphisms in multiple genes are involved together with non-genetic factors and because different individuals with AD may have distinct combinations of genetic and non-genetic factors. Despite these complexities, tremendous progress has been made over the past two decades in deciphering AD genetics, which has laid the foundation for our current understanding of the etiological and pathogenic mechanisms underlying AD and for the development of novel approaches for treatment and prevention. As therapeutic strategies become more effective and our grasp of AD genetics is strengthened, reliable and comprehensive genetic risk profiling will eventually enable systematic early-prediction/early-prevention procedures.

There are three main approaches to finding susceptibility loci: one candidate gene studies and two non-candidate gene approaches. In candidate gene studies, genes are selected on the basis of the known biology of the disease, and are assessed individually to determine whether variants in each candidate are associated with disease. Alternative approaches include the genomic screen and whole genome wide association (WGA) studies. These two non-candidate gene approaches do not select genes *a priori*, but instead tests for linkage or association between the disease and polymorphic markers that are spaced evenly throughout the genome.

Genetic linkage analysis is a biologically unbiased method of detecting regions of the genome more frequently inherited in common by affected individuals in a family than expected by chance. Linkage analysis uses extended families or affected sibling pairs to test for regions of the genome in which there is increased allele sharing between affected relatives. The amount of allele-sharing within and across families is measured in the

LOD (logarithm of odds) score. This is calculated based on the relationship between recombinant and non-recombinant alleles at a given marker and across nearby markers (Morton, 1995). Linkage data is analyzed using either parametric methods, which specify a mode of transmission, allele frequencies and penetrances of the susceptibility locus, or by nonparametric methods, which are for the most part model-free.

Microsatellite markers and single nucleotide polymorphisms (SNPs) are widely applied in genome-wide analyses. Microsatellite markers are highly polymorphic. An average genome-wide spacing of microsatellite markers at 10 cM or less can generally provide sufficient coverage and information content to identify broad regions of linkage for more detailed analysis. Because most SNPs are bi-allelic, a denser SNP map is needed to provide the same level of coverage and information content as microsatellites. However, the high density of available SNPs across the genome, their ease of genotyping, high-throughput, and high quality of the resulting genotypes currently make SNPs more widely used in genetic studies.

Genetic linkage study is a powerful method for narrowing down a chromosomal candidate region to 10~30 Mb. Such regions of interest may harbor dozens to hundreds of genes, i.e. locational candidate genes, and hundreds or thousands of SNPs. However, the power of linkage analysis decreases in narrowing genetic effects to smaller regions.

Candidate gene association studies test genes likely to be involved in disease etiology that are identified based on the altered physiology or pathology seen in patients with the disease. It requires the use of unrelated case-control samples or siblings who are discordant or concordant for disease, and tests whether a common and specific allele is either present more or less often in cases than controls, or is more frequently transmitted

to affected individuals within a family. Therefore, there are two basic frameworks for association analysis: case-control and family-based studies. The case-control approach is generally regarded as more powerful and sensitive than family-based tests (Morton and Collins, 1998). However, caution needs to be taken when using a case-control dataset because it can produce spurious associations due to population stratification (Ewens and Spielman, 1995). Failure to match cases and controls for the same underlying population, ethnicity, and other factors, can lead to false positive or false negative results due to admixture and selection or drift between unlinked loci (Weeks and Lathrop, 1995; Lander and Schork, 1994; Khoury and Beaty, 1994).

Association tests depend on the availability of numerous polymorphic markers in a given candidate region and the availability of abundant SNPs to make this approach viable. It is estimated that SNPs occur on average every 500-1,000 base pairs and have a low mutation rate. These two properties of SNPs are advantageous in association studies (Kruglyak, 1997; Cooper et al., 1985; Li and Sadler, 1991). To avoid false negative results from an association study, it is important to use multiple polymorphisms, and consider linkage disequilibrium patterns and haplotype structures to adequately cover the gene of interest (Wall, 2001; Conrad et al., 2006; Wall and Pritchard, 2003a).

Linkage disequilibrium (LD) is the non-random association of alleles at different sites. The consideration of LD is becoming more and more important in the discovery of genes involved in complex human diseases (Wall and Pritchard, 2003b). A common structure used to represent LD in the human genome is haplotype blocks, defined as sets of consecutive sites between which there is little or no evidence of historical recombination. Similar to ApoE, there may be more than one intragenic polymorphism

determining the haplotypes conferring disease risk or protection. A single gene may cover 1 to >100kb and contain in addition to neutral polymorphisms functional ones with or without significant pairwise LD. Considering this and the high number of theoretically possible candidate genes and haplotypes it seems unlikely that one would detect a true risk allele for LOAD by genotyping a single SNP of a positional candidate gene. By testing a sufficiently large collection of SNPs within and near a gene, it should be possible to define the common haplotypes underlying blocks of LD. It should also be possible to select an optimal reference set of haplotype tagged SNPs (htSNPs) for any subsequent genotyping study. The haplotype structure provides a crisp approach for testing the association of genome segments with disease. Once the haplotype blocks are identified, they can be treated as alleles and tested for LD.

Known genes involved in Alzheimer disease

There are two forms of Alzheimer disease. If the age-of-onset is less than 60, it is early onset AD (EOAD); if the age-of-onset is greater than 60, it is late onset AD (LOAD). Most early onset AD results from the rare Mendelian subform of Alzheimer disease that exhibits autosomal dominant inheritance and both locus and allelic heterogeneity. Genetic studies have identified three genes underlying early onset AD. This form can be caused by over 120 mutations in three genes encoding β-amyloid precursor protein (APP), presenilin 1 (PS1), and presenilin 2 (PS2) (Goate et al., 1991; Levy-Lahad et al., 1995; Rocchi et al., 2003; Sherrington et al., 1995). These mutations all affect APP metabolism such that more Aβ42 peptide, found in senile plaques, is

produced. This underscores the pathogenetic importance of Aβ in the evolution of AD (Hardy and Higgins, 1992).

The discovery of APP as the source of Aβ has led to an extensive characterization of the protein and the mechanisms by which the processing leads to the deposition of amyloid. APP is a ubiquitously expressed type 1 membrane glycoprotein (Kang et al., 1987) and is encoded by a single gene on chromosome 21. Mapping of the APP gene to chromosome 21, together with observation of plaques and tangles in most elderly individuals with Down's syndrome (trisomy 21), suggests an important role for amyloid β (Goedert and Spillantini, 2006; Oliver and Holland, 1986). In the late 1980s, it was speculated that mutations in the APP gene would also be found in familial AD, some cases of which had been linked to chromosome 21 (Goate et al., 1991; Siemers et al., 2006; Geling et al., 2002). Twenty missense mutations gene and gene dosage variation have been described in APP.

Mutations in the APP gene account for only a minority of familial AD cases. The vast majority of all known familial AD mutations have been found in the gene encoding presenilin 1 (PS1). PS1 was mapped to chromosome 14 by linkage studies (St George-Hyslop et al., 1992). Positional cloning identified mutations in PS1 that encodes a polytopic membrane protein (Pfeifer et al., 2002; Sherrington et al., 1995). The name of the gene is taken from its role in familial pre-senile dementia. More than 150 reported mutations throughout PS1 (McGowan et al., 2006) suggested that this protein is important in the pathological events causing AD.

PS2, a homologue of PS1, was found on chromosome 1 by homology searches of databases for amino acid sequences with high similarity to PS1 (Levy-Lahad et al., 1995;

Rogaev et al., 1995). Compared to PS1, 20 mutations have been found in PS2 (McGowan et al., 2006). Presenilins are central components of the atypical aspartyl protease complexes responsible for the γ–secretase cleavage of APP (Pfeifer et al., 2002; Weiner et al., 2000). Mutations in presenilin increase the ratio of amyloid β42 to amyloid β40, and this appears to result from a change in function that manifests itself in reduced γ–secretase activity (Gervais et al., 2007). Studies from 16 major transgenic mouse models support the functions of these genes (McGowan et al., 2006).

In contrast, the vast majority of AD, late-onset Alzheimer disease (LOAD) does not show a classical Mendelian inheritance, although epidemiological, family and twin studies suggest a significant proportion of LOAD can be attributed to genetic factors (Bergen, 1994; Bergen et al., 1997; Bonney, 1984; Breitner and Murphy, 1992; Chase et al., 1983; Cupples et al., 1989; Cupples et al., 1991; Farrer et al., 1991). The only genetic locus universally accepted as an important risk factor for late-onset AD is the apolipoprotein E (Apo E) locus on chromosome 19 (Pericak-Vance et al., 1991). ApoE, Apolipoprotein E, is a plasma lipoprotein involved in lipid transport and metabolism. ApoE, with three common alleles (ApoE-2, ApoE-3, and ApoE-4) that encode three different isoforms (ApoE ε2, ApoE ε3, and ApoE ε4, respectively) (Weisgraber et al., 1982; Rall, Jr. et al., 1982). ApoE ε4 is the risk-promoting allele and acts in a dose-dependent manner to increase risk and decrease age of onset in both late-onset familial and sporadic and early-onset sporadic AD (Corder et al., 1993). The ApoE ε2 allele is the protective (Corder et al., 1994), but its effect in early-onset sporadic AD is controversial (Scott et al., 1997). The ApoE ε3 allele is the neutral (and most common) allele in AD.

There are several potential roles for ApoE in increasing AD risk. The ApoE ε4 allele causes hypercholesterolemia, which increases the production of amyloid β (Sanan et al., 1994; Evans et al., 1995; Ma et al., 1994; Castano et al., 1995) and stabilizes amyloid β (Rebeck et al., 1993; Rebeck et al., 1995). The ApoE ε4 allele can also induce cerebrovascular pathology and oxidative stress and impair neuronal plasticity (Olichney et al., 1996; Poirier et al., 1995; Soininen et al., 1995). However, more than 50 percent of Alzheimer's disease cases do not carry an ApoE ε 4 allele (Farrer et al., 1997; Jarvik et al., 1996; Pericak-Vance and Haines, 1995; Bennett et al., 1995), suggesting that other risk factors exist.

The search for new Alzheimer disease susceptibility genes

Genomic Screens

To date, five Late-Onset AD genome-wide screens have been reported. In the first study, Pericak-Vance and colleagues (Pericak-Vance et al., 1997; Pericak-Vance et al., 1998) performed a two-stage screen, first genotyping markers at 10-cM intervals in 16 families, and then examining regions with a LOD score of >1.0 or $P<0.05$ in an additional 38 families. Two years later, this group rescreened the entire genome in 466 families (Pericak-Vance et al., 2000). These studies used both parametric and nonparametric linkage methods. The third study, reported by Kehoe (Kehoe et al., 1999), used nonparametric linkage analysis in affected sibling pairs. These investigators genotyped 292 affected sibling pairs with markers spaced 20 cM apart. Four years later, Myers et al. (Myers et al., 2002) did a full genome screen using 451 affected sibling pairs and 328 microsatellite markers. This was stage II of the Kehoe 1999 genome screen. All

four of these studies stratified the data by ApoE genotype; however, methods of stratification differed between the studies. In the fifth study, Blacker and colleagues (Blacker et al., 2003) performed a 9 cM genome screen of 437 families within the National Institute of Mental Health (NIMH) sample, using both parametric and non-parametric linkage analyses. Various genome screens have implicated chromosomes 2, 9, 10 and 12 as potential locations of additional AD loci (Kehoe et al., 1999; Pericak-Vance et al., 2000).

In 2000, three laboratories reported significant linkage of LOAD to two regions of the long arm of chromosome 10 (Bertram et al., 2000; Ertekin-Taner et al., 2000; Myers et al., 2000). Myers et al. followed up "suggestive" (Lander and Kruglyak, 1995) linkage results on this chromosome, using an affected sibling pair method, resulting in significant linkage near ~ 67 Mb. Using a different methodology, a second study (Ertekin-Taner et al., 2000) performed multipoint linkage analyses in five late-onset AD families using Aβ plasma levels as a quantitative phenotype and found significant linkage very close to the same chromosomal region. The third report (Bertram et al., 2000) was a candidate gene-driven linkage analysis of 435 families using six genetic markers in a region of chromosome 10 that lies ~40 Mb distal to the one implicated by the former two studies. A candidate gene of interest in this region encodes the insulin-degrading enzyme (*IDE*) that, together with neprilysin and other proteases, has been previously suggested as playing a major role in the degradation and clearance of Aβ in brain (Bertram et al., 2000; Selkoe, 2001). Two of the six markers in this region showed significant linkage to AD. In addition, one marker, *D10S583*, displayed significant association with AD through an

under-representation of one of its alleles in affecteds versus unaffecteds, possibly indicating linkage disequilibrium with a disease-modifying DNA variant nearby.

Combining all the positive genome screen results in these three papers, however, the region of interest on chromosome 10 is still quite broad (approximately 80Mb total) and provides an impractically large area for a molecular genetic search. Additional studies of linkage and association/linkage disequilibrium are needed to identify a minimum candidate region and the putative underlying AD gene(s) in the region. This will also require the identification of sequence variants in the candidate genes in families that are strongly linked to this region and those linked to the more proximal region of chromosome 10 implicated by the two other linkage studies. Systematic linkage-disequilibrium mapping using high-throughput SNP genotyping should effectively narrow down these candidate gene regions and eventually lead to the identification of potentially pathogenic DNA variants in the linked region of this chromosome.

Candidate gene studies

Complex disease involves multiple genes. It is believed that AD susceptibility is the result of multiple genes acting either independently or interactively in their contribution to overall risk (Pericak-Vance and Haines, 1995). There are many ways to identify genes to be tested as candidates. The traditional functional candidate gene approach assesses genes that are selected based on their potential biological relevance to a disease. Genes involved in the amyloid β (Aβ) generation and clearance, oxidative stress, inflammation, and apoptosis have potential roles in the etiology of AD (Citron, 2004; Forero et al., 2006; Wyss-Coray, 2006).

Numerous studies have tested hundreds of candidate genes for involvement in AD. Some examples of the candidate genes include α-antichymotrypsin (AACT) (Kamboh et al., 1995), low-density lipoprotein receptor-related protein (LRP) (Lendon et al., 1997), α_2-Macroglobulin (A2M) (Blacker et al., 1998), ubiquitin (Bertram et al., 2005), the HLA complex (Curran et al., 1997; Renvoize, 1984; Small et al., 1991), and mitochondrial mutations (Shoffner et al., 1993).

In 2000, Bertram *et al.* found linkage in a region more downstream of D10S1125 on chromosome 10 and focused on the insulin-degrading enzyme (IDE) as a possible candidate gene (Bertram et al., 2000). There are other candidate genes on chromosome 10, such as PLAU (the gene encoding uPA, urokinase-type plasminogen activator), TNFRSF6 (FAS receptor), CDC2 (cell division cycle 2), VR22 (also known as α-T catenin (CTNNA3)), which have been tested by association analysis. However, there isn't a single LOAD locus that has been generally replicated in multiple study populations and thus generally accepted, except ApoE. Most associations have not been replicated at all and the involvement of these gene in LOAD has been rejected.

Based on the AlzGene database (http://www.alzgene.org), 875 studies have tested 355 functional candidate genes representing 1,055 polymorphisms as of August 15, 2006. Recently, Bertram and colleagues published a systematic meta-analysis of AD genetic association studies based on a 'data freeze' of the AlzGene database on December 1, 2005. This analysis covered 789 publications reporting on 802 different polymorphisms in 277 genes after screening ~23,500 titles and abstracts (Bertram et al., 2007). They analyzed 127 polymorphisms across 69 different putative Alzheimer disease risk genes with available genotype data in at least three case-control samples. In addition to *APOE*

ε4 and four other probably ε4 related effects, they discovered 20 polymorphisms in 13 genes (*ACE, CHRNB2, CST3, ESR1, GAPDHS, IDE, MTHFR, NCSTN, PRNP, PSEN1, TF, TFAM* and *TNF*) that yielded significant allelic summary ORs (ranging from 1.11–1.38 for risk alleles and 0.92–0.67 for protective alleles).

Genomic convergence

No single method can give us an answer in dissecting complex disease traits. A combination of methods is the most promising way to understand the etiology of a complex disease. The genome screens identify AD gene locations independent of any presumed knowledge of function. By utilizing linkage analysis, the region of interest can be narrowed down. Database mining using bioinformatics tools can find the functional candidate genes in this region. Therefore, combining the locational candidate gene and functional candidate gene approaches is a good way to identify genes of interest.

Genomic convergence is another approach to narrowing down the large pool of candidate genes to a few select choices. It is a multifactor approach that combines different kinds of genetic data analysis to identify and prioritize susceptibility genes for a complex disease (Hauser et al., 2003a). One of the growing problems in the study of complex disease is how to prioritize research and make sense of the immense amount of data. The best approach may be to take advantage of the strengths of all available means (Noureddine et al., 2005).

The following dissertation presents a multifactorial, multistep approach, genomic convergence, which combines linkage analysis, gene expression analysis, and candidate gene association analysis to identify and prioritize candidate susceptibility genes for AD.

This genomic convergence approach led to the investigation of several candidate genes, with the promising evidence for genes involved in Aβ and Tau pathways.

CHAPTER III

HYPOTHESIS AND SPECIFIC AIMS

Hypothesis: Chromosome 10 harbors susceptibility gene(s) involved in Alzheimer Disease.

Specific aim 1: Identify and define the minimum candidate region (MCR) for Alzheimer disease (AD) on chromosome 10 using linkage analysis.

Although multiple lines of evidence suggest linkage between chromosome 10 loci and Alzheimer disease, the reported linkage region on chromosome 10 is very broad (~80Mb). After combining results from previous publications, the most consistent region will be prioritized and selected. Genotyping markers in this region will be selected at 2 Mb spacing and genotyped in 587 multiplex families. LOD scores will be calculated to determine a narrowed peak region. Considering the likely heterogeneity of AD, ordered subset analysis (OSA) will be utilized to obtain a LOD score with a more precise localization of a peak. Stratification of subsets, weighting schemes, and conditional analysis will also be used to identify regions of interest masked in the initial analysis. Results of these studies will define a minimum candidate region (MCR) on chromosome 10 most likely to contain a gene related to AD.

Specific aim 2: Genomic analysis of candidate genes for AD.

a. Identify candidate genes on chromosome 10 for AD.

b. Identify a subset of candidate genes for association analysis.

c. Identify and genotype SNPs within each of the genes.

d. Test linkage disequilibrium (LD) within genes, confirm and possibly construct gene-specific HapMaps.

Genes under linkage peaks and genes that are differentially expressed between AD cases and controls will be converged to select candidate genes for association analysis. After the list of potential candidates is achieved, SNPs within each gene will be identified and selected based on several databases (e.g. dbSNP, Ensembl, UCSC, and HapMap). SNP detection and genotyping will be primarily accomplished via the Illumina Goldengate Oligo Pool Assay (OPA). Linkage Disequilibrium (LD) will be assessed for the SNPs using the Haploview program and characterized using the r^2 statistic. The haplotype blocks will be characterized and evaluated. This will test for the existence of the blocks defined by the HapMap project. If the blocks are not confirmed, we will construct our own haplotype map.

Specific aim 3: Test candidate genes for association with AD.

a. Association analysis on the entire AD dataset will be done by testing SNPs in the candidate genes for association with AD.

b. Complete analyses on different subsets likely to be genetically more homogeneous.

c. Select the two most interesting candidate genes for follow up.

A case-control dataset will be used to test the association between SNPs in the candidate genes and risk of AD. I will perform allelic, genotypic and haplotypic association analyses on both the overall dataset and subsets. The subsets will include those identified by age-at-onset, gender, phenotypic subgroups, ApoE genotype, and linkage to other known genes. Epistasis, or gene-gene interaction, will be tested using Multifactor Dimensionality Reduction (MDR) and logistic regression to detect potential polymorphisms associated with the AD risk.

CHAPER IV

IDENTIFICATION OF THE MINIMUM CANDIDATE REGION (MCR) FOR ALZHEIMER DISEASE ON CHROMOSOME 10 USING LINKAGE ANALYSIS

Overview of chapter IV- VII

Numerous studies have tested hundreds of functional candidate genes and many such genes were reported as having association with AD. However, none of the genes has been universally accepted as an AD risk gene except ApoE. Genomic screens have shown evidence that chromosome 10 may harbor susceptibility genes for AD. However, the reported linkage region scatters across 80 Mb region on chromosome 10 and provides an impractically large area for a molecular genetic search. To examine this region in more detail, we tested several recently reported candidate genes within this broad region of linkage: VR22, LRRTM3, PLAU, TNFRSF6, IDE, CDC2, and LOC439999. Simultaneously, we examined the linkage signal in more detail to see if we could narrow the peak region of linkage and converge the candidate gene and linkage data. Chapter IV presents the linkage analysis to identify the minimum candidate region for AD on chromosome 10. Chapter V presents the candidate gene studies on VR22, LRRTM3, PLAU, TNFRSF6, IDE. There is at least one SNP in each gene showing association with AD in at least one of our data sets. Chapter VI details the examination of the CDC2 gene with AD. Due to the properties of the SNPs in CDC2, we used a special genotyping method to characterize the polymorphisms of this gene. However, our results failed to show evidence of association between CDC2 and AD. Chapter VII presents the association study of a polymorphism rs498055 in gene LOC439999 on chromosome 10.

This SNP was identified from a study of 1412 potentially functional SNPs in 667 genes under previously identified linkage peak on chromosome 10 (Grupe et al., 2006). However, we didn't see the association in our independent three family and one case-control datasets.

Introduction

Alzheimer disease (AD, MIM#104300) is a progressive neurodegenerative disorder of later life with a complex etiology and a strong genetic component. Mutations in three genes encoding the β-amyloid precursor protein (APP, OMIM#104760), presenilin 1 (PS1, OMIM#104311), and presenilin 2 (PS2, OMIM#600759) genes cause the rare early-onset AD (EOAD). ApoE is the only gene universally accepted as an important risk factor for late-onset AD (LOAD). However, more than 50 percent of Alzheimer disease cases do not carry an ApoE ε4 allele, suggesting that other genetic risk factors exist (Farrer et al., 1997; Jarvik et al., 1996; Pericak-Vance and Haines, 1995; Bennett et al., 1995).

Numerous studies have tested hundreds of functional candidate genes and many genes for AD have reported positive associations in at least one study group (Bertram et al., 2007). However, there is no single locus that has been sufficiently replicated to be widely accepted as an AD risk gene, except ApoE. Genome screening is an alternative approach that identifies AD gene locations independent of any presumed knowledge of function. These have suggested that the loci for late-onset AD lie on several chromosomes, including chromosomes 9, 10, and 12 (Pericak-Vance et al., 1997; Kehoe et al., 1999; Myers et al., 2002; Blacker et al., 2003). Perhaps the most interesting region

is chromosome 10 since two independent linkage studies identified the same chromosome region using different methods (Myers et al., 2000; Ertekin-Taner et al., 2000), and a candidate gene driven linkage analysis found a linked region at 113 cM on chromosome 10q, near insulin-degrading enzyme (IDE) (Bertram et al., 2000). Recently, the microsatellite marker D10S1208 (63.3 cM) was linked to Alzheimer disease among families with affected mothers (Bassett et al., 2002; Bassett et al., 2005).

The multiple reports of chromosome 10 linkage point to linkage peaks scattered across an 80Mb region and provide an impractically large area for a molecular genetic search. All the notable differences in the location of the linkage signals call attention to the potential for locus heterogeneity and highlight the need for strategies to increase homogeneity. One strategy is to examine phenotypic or genotypic subgroups, e.g., based on ApoE genotype. Another strategy is to apply new statistical techniques, such as Ordered Subset Analysis (OSA). We performed a linkage study at 2 Mb spacing across this 80 Mb region to identify a minimum candidate region on chromosome 10.

Materials and methods

Study populations

We used a family sample consisting of 730 pedigrees with a total of 567 multiplex families with a total of 922 affected sibpairs (ASP) (Table 4-1). All subjects are late-onset AD (LOAD) patients (minimum age at onset (AAO) $\geq$ 60 years).

Table 4-1 Study populations for Alzheimer Disease

Family	Overall	NIMH	NCRAD	CAP
Total family				
Total pedigrees	730	352	154	224
Affected individuals	1521	807	315	390
Unaffected individuals	974	331	162	481
Discordant Sib Pairs (DSP)	1337	629	269	439
Independent Discordant Sib Pairs	674	283	129	262
Pedigrees with at least one DSP	406	165	75	166
Affected Relative Pairs (ARP)	188	66	26	96
Pedigrees with at least one ARP	64	31	11	22
Multiplex family				
Total pedigrees	567	352	122	96
Affected Sib Pairs (ASP)	922	517	190	225
Independent Affected Sib Pairs	707	418	150	139
Pedigrees with at least one ASP	534	331	118	85

NIMH = National Institute of Mental Health repository; NCRAD= National Cell Repository for Alzheimer Disease at Indiana University; CAP = Collaborative Alzheimer Project

All individuals included in this study are Caucasian. Written consent was obtained from all participants in agreement with protocols approved by the institutional review board at each contributing center. AD was diagnosed according to the NINCDS-ADRDA criteria (McKhann et al., 1984b). All controls had results within the normal range in the Mini-Mental State Exam (MMSE) or Modified Mini-Mental State Exam (3MS). Samples were ascertained by the following centers: the NCRAD repository at Indiana University (NCRAD); the Collaborative Alzheimer Project (CAP), including Duke and Vanderbilt Universities and University of California at Los Angeles; and the National Institute of Mental Health repository (NIMH). All data sets were independent. Age at onset was recorded as that age at which the first symptoms were noted by the participant or a family member. If the affected individuals were early in the disease process, we included their report of age-at-onset as part of the determination. If the

disease was more advanced, we only used information as collected from multiple family members (such as, spouse and children). Mean ± SD of age-at-onset (AAO) in affected individuals in the family-based sample was 72.9±6.4 years. The mean age-at-exam (AAE) ± SD was 80.1±7.1 and 69.9±11.2 in affected and unaffected individuals, respectively. The AD affected group was 67.3% female, while the unaffected group was 56.2% female.

SNP selection and genotyping

Following informed consent, blood samples were collected from each individual. Genomic DNA was obtained from the repositories (NIMH, NCRAD) or extracted from whole blood (CAP) by use of the Puregene system (Gentra Systems, Minneapolis, MN). All samples were coded and stored at 4°C until used.

A map of the chromosome 10 region and previously reported candidate genes is shown in Figure 4-1.

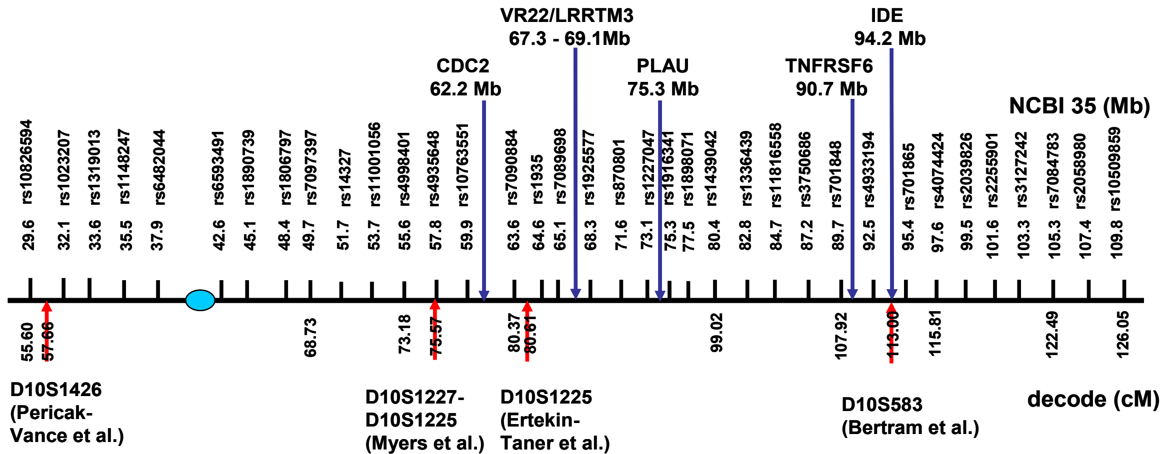


Figure 4-1 The position of previously reported candidate genes and 36 genotyped SNPs for linkage analysis on chromosome 10

The physical location (Mb) of 36 SNPs for linkage analysis according to NCBI build 35 are shown in the middle; the lower panel shows the DECODE genomic map location (cM) of linked markers shown in the literature. Upper panel shows the candidate genes within this region.

NCBI (http://www.ncbi.nih.gov), Ensembl (http://www.ensembl.org) and Applied Biosystems web sites (http://home.appliedbiosystems.com) were mined to select SNPs according to their location relative to other selected SNPs, high minor allele frequency (≥ 0.25), and availability of quality assays. We carefully selected 36 SNPs at 2 Mb spacing for linkage analysis (Figure 4-1).

SNP genotyping was performed using Assays-On-Demand™ or Assays-By-Design™ with the ABI PRISM® 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA). Amplification was performed in a 384-well DNA Engine Tetrad® 2 Peltier Thermal Cycler (MJ Research, Waltham, MA) with the following conditions: 94°C-10 min; 92°C-15 sec, 60°C-1 min (50 cycles); 4°C-hold. Systematic genotyping errors were minimized by use of a system of quality control (QC) checks with duplicated samples (Rimmler et al., 1998). Linkage disequilibrium (LD) and deviations

from Hardy-Weinberg equilibrium (HWE) were assessed using Haploview (Barrett et al., 2005).

Statistical methods

Two-point linkage analyses

Two-point heterogeneity LOD score (HLOD) analyses were computed using FASTLINK and HOMOG (Ott, 1999). Because the mode of inheritance for AD is unknown, affected-only parametric analyses were performed using both autosomal dominant and autosomal recessive models with disease allele frequencies of 0.001 and 0.20, respectively, to model the susceptibility allele. Greenberg & Hodge (Greenberg et al., 1998) suggested a critical factor in LOD-score analysis is the Mode of inheritance (MOI) at the linked locus, not that of the disease or trait per se. Thus, a limited set of simple genetic models in LOD-score analysis can work well in testing for linkage. If a signal is picked up in either one of the models (most typically dominant and recessive), there is increased power to detect true linkage using this approach.

Because it is likely that there is genetic heterogeneity in Alzheimer disease, we applied Ordered Subset Analysis (OSA) (Hauser et al., 2004) in our linkage analysis to test for a subset of linked families. In OSA, the statistical significance of increased evidence for linkage in a subset of the data relative to evidence for linkage in the entire sample is assessed via random permutation of the order of inclusion of the families to estimate empirical p values. Since the ApoE ε4 allele is the only known risk allele for AD, we used the ApoE LOD score (at theta=0) and ApoE weight (the proportion of affecteds within a family who carry at least one ApoE ε4 allele) as covariates to order

families from low to high and high to low. A locus on chromosome 12 has been suggested to be linked with AD (Liang et al., 2005; Mayeux et al., 2002), so the linked marker D12S368 was also used as a covariate to order families in OSA analysis. The same parametric models were applied to D12S368. OSA was also performed using mean age of onset as a covariate.

Multipoint linkage analyses

Multipoint LOD scores were calculated by MERLIN (Abecasis et al., 2002) and OSA multipoint LOD scores were calculated using the four covariates as described above. Conditional linkage analysis using ApoE as a covariate was done in Allegro (Gudbjartsson et al., 2000). The analysis was weighted by the proportion of affecteds within a family who carry at least one ApoE ε4 allele and by proportion of affecteds within a family who don't carry the ApoE ε4 allele (inverse ApoE- ε4 weighted). There is no linkage disequilibrium (LD) between 34 out of the 36 linkage analysis SNPs and very low LD between the other two linkage SNPs (rs1935 and rs7089698, r^2=0.46). We used the LD option of MERLIN to perform the multipoint linkage analysis to take the LD into account. An r^2>0.1 was used as criteria to define SNPs in LD. In the OSA multipoint analysis, we excluded the SNP (rs7089698) in LD with the other SNP (rs1935) as it had the lower minor allele frequency.

Family-based association analysis

Family based association analysis was conducted using the pedigree disequilibrium test (PDT, for allelic effect) and Genotype-PDT (GenoPDT, for genotypic effect) for single-locus tests to assess association between genotypes and risk of AD in the family data (Martin et al., 2000b; Martin et al., 2002).

Results

Two-point linkage analyses

36 SNPs evenly spaced across the 80.2 Mb region on chromosome 10 were genotyped in 567 multiplex families (922 affected sib pairs). The reliability of duplicate genotyping across plates was >99% and the average genotyping efficiency was 98%. All analyses were applied to the overall data set and three stratified subsets (autopsy-confirmed, ApoE ε4+, ApoE ε4-).

Using the 36 linkage SNPs panel, we conducted both FASTLINK/HOMOG two-point analyses and OSA two-point analyses using different covariates. In the overall data set, only SNP rs11816558 generated a LOD>1.0 (LOD=1.31, Figure 4-2 A).

Several SNPs yielded LOD >1.0 in the subsets. SNPs rs6482044 and rs870801 had LOD=1.29 and 1.05, respectively, in autopsy subset (Figure 4-2 B); rs6482044 had LOD=1.41 in the ApoE ε4-positive subset (Figure 4-2 C); rs3750686 had LOD=1.54 in ApoE ε4-negative subset (Figure 4-2 D).

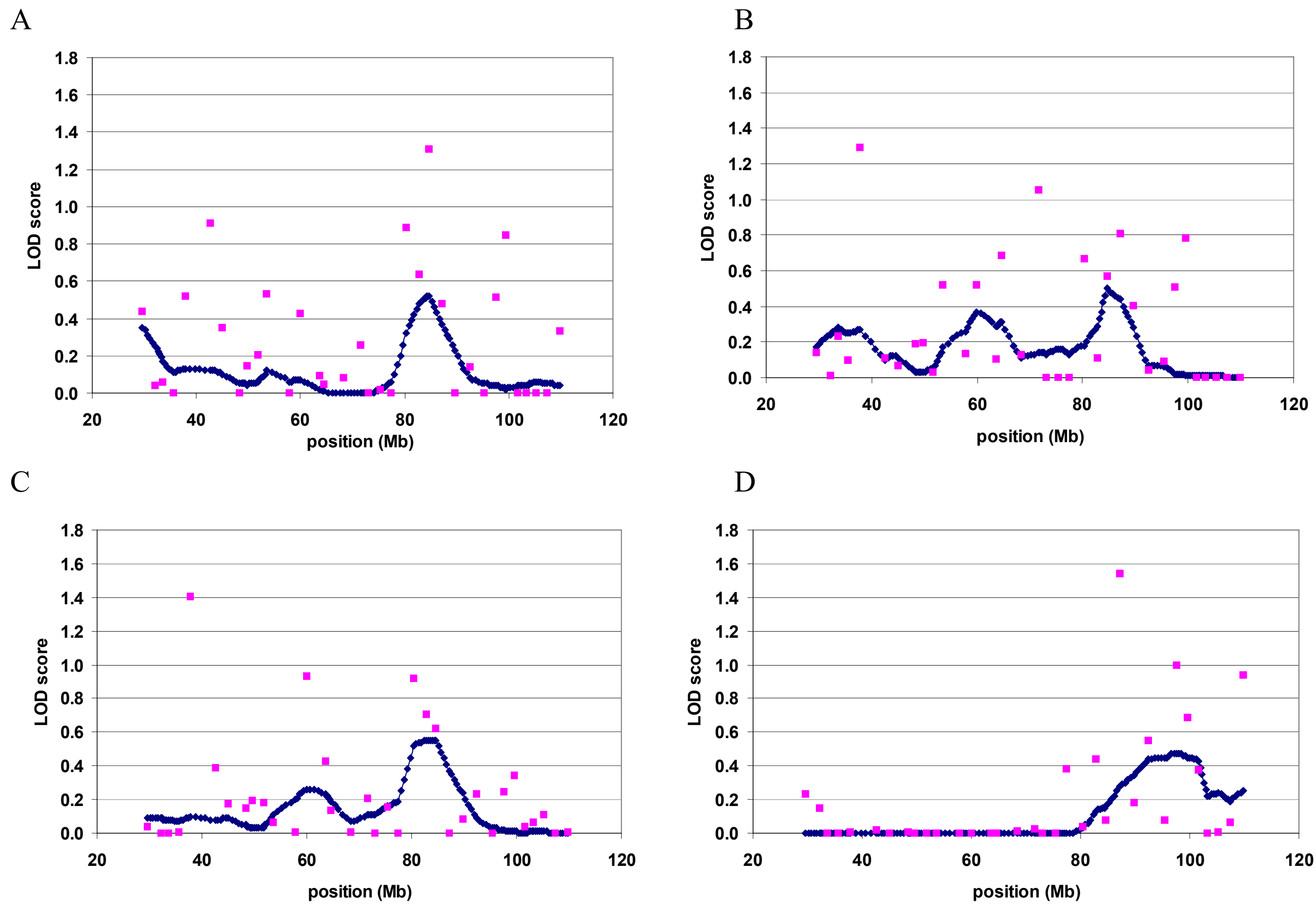


Figure 4-2 Original two-point (FASTLINK, shown in dots) and multipoint (MERLIN, shown in line) LOD scores in overall and three subsets

A. Overall data set B. Autopsy-confirmed subset C. ApoE ε4-positive subset D. ApoE ε4-negative subset

In the OSA analysis using the ApoE_LOD score as a covariate, a two point LOD score of 2.69 was obtained at SNP rs1890739 at 45.1 Mb (p=0.03 in 21% families, low to high), rs1319013 at 33.6 Mb generated LOD=2.50, and rs11816558 at 84.7Mb generated LOD=1.70 (Figure 4-3).

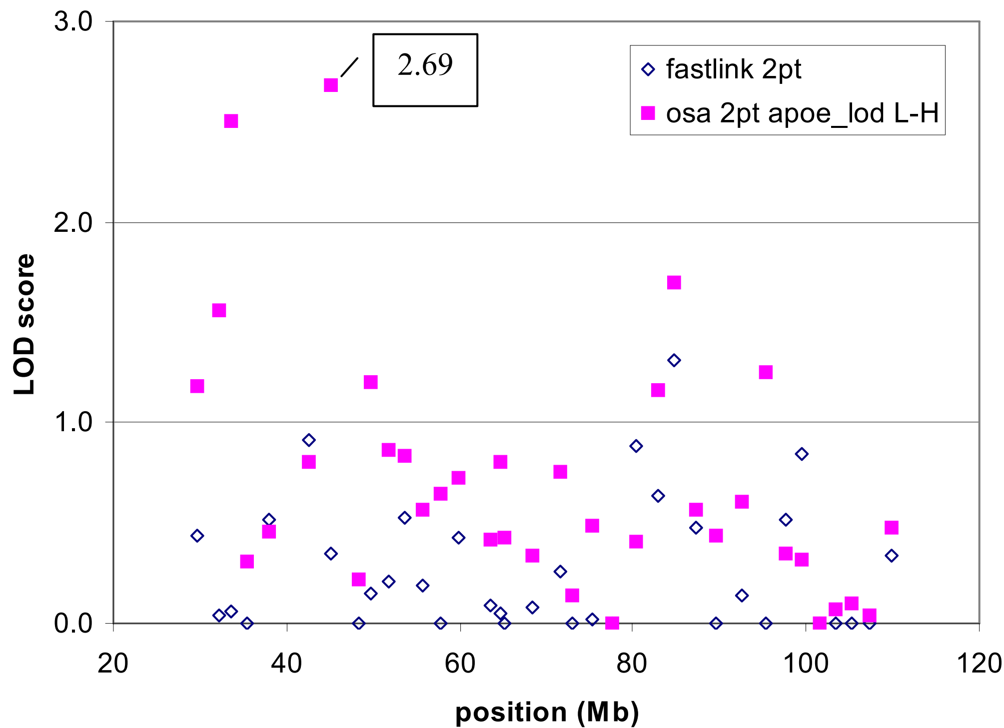


Figure 4-3 Two-point analyses in the overall data set
Ordered subset two point analysis in the overall data set obtained a two point LOD score of 2.69 at 45.1 Mb (p=0.03 in 21% families) using ApoE_LOD score to order families from low to high.

SNP rs14327 (51.7 Mb) generated a two point LOD score of 1.94 in 13% of families when the families were ordered from low to high using covariate of the linkage signal on chromosome 12, p=0.009 (Figure 4-4 A). An OSA two point LOD score of 2.30 was generated in 93% of families when the families were ordered from low to high using the same covariate (Figure 4-4D), however, the permutation p value was not significant (p=0.22).

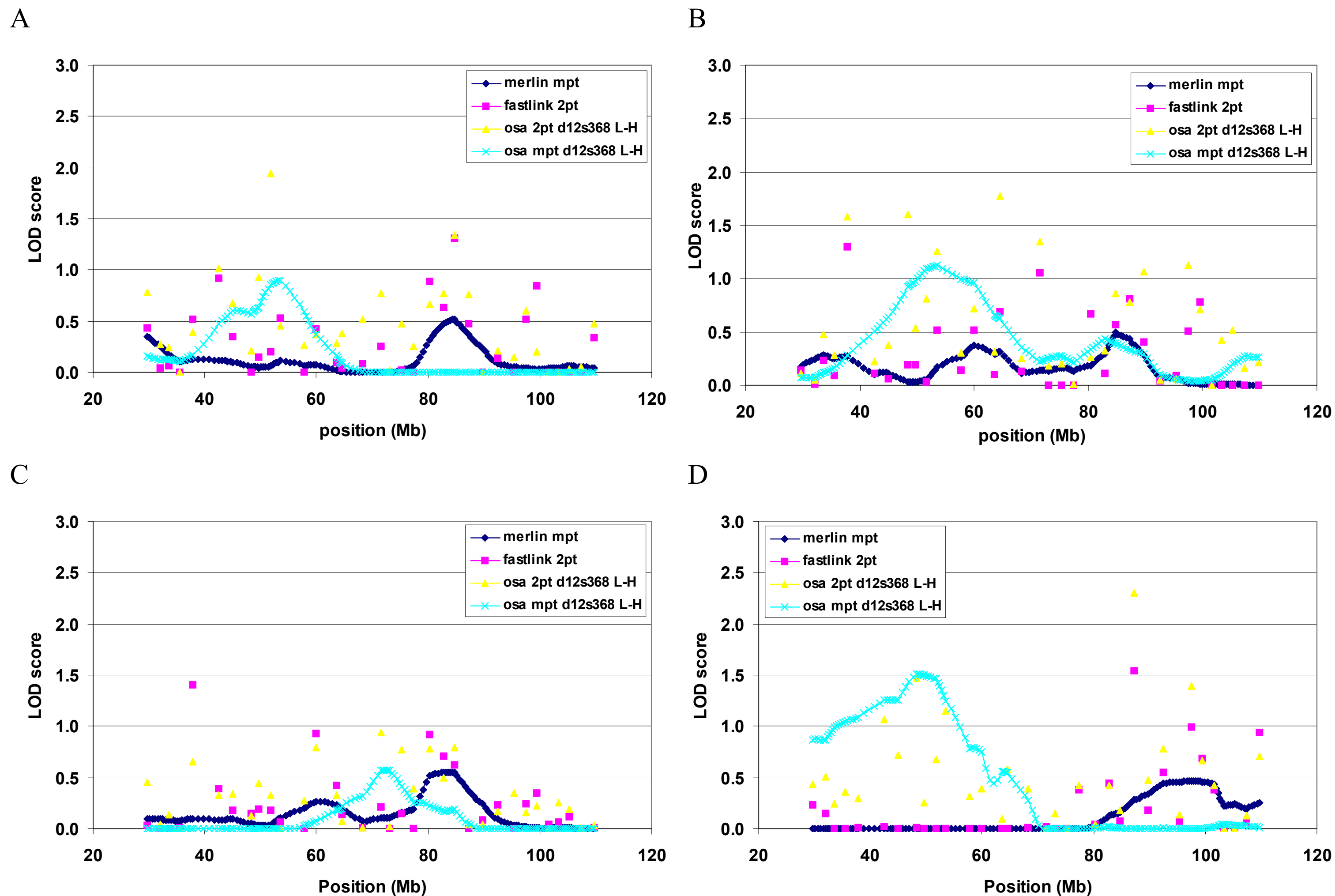


Figure 4-4 Two point and multipoint linkage analyses using the D12S368 LOD score as a covariate to order families from low to high in the overall and three subsets

A. Overall data set B. Autopsy-confirmed subset C. ApoE ε4-positive subset D. ApoE ε4-negative subset

Multipoint linkage analyses

Multipoint analysis using the 36 linkage SNP panel on the overall data set didn't show any suggestive LOD scores, shown as the line of overall multipoint analysis in Figure 4-2A. When we applied the ApoE LOD score as covariate to order families from low to high in Ordered Subset Analysis, the highest LOD score was 1.85 (rs10826594). Other SNPs, rs1023207 and rs1319013, also had LOD scores higher than 1.0 (Figure 4-5 A). SNP rs14327 (51.7 Mb) had LOD of 1.03 using the same covariate in OSA (Figure 4-5 A). There is no SNP showing linkage with the disease in the ApoE ε4+ subset (Figure 4-5 C). However, rs1806797 (48.3 Mb) showed a peak LOD score of 1.52 in ApoE ε4- subset when we applied D12S368 as a covariate to order families from low to high in OSA (Figure 4-5 D). In the autopsy-confirmed subset, there was no effect of the number of ApoE ε4 alleles as a covariate. However, a peak of LOD=1.91 was generated at 49.7 Mb between rs7097397 and rs14327 in OSA multipoint analysis when the covariate ApoE LOD score ranked families from low to high (Figure 4-5B).

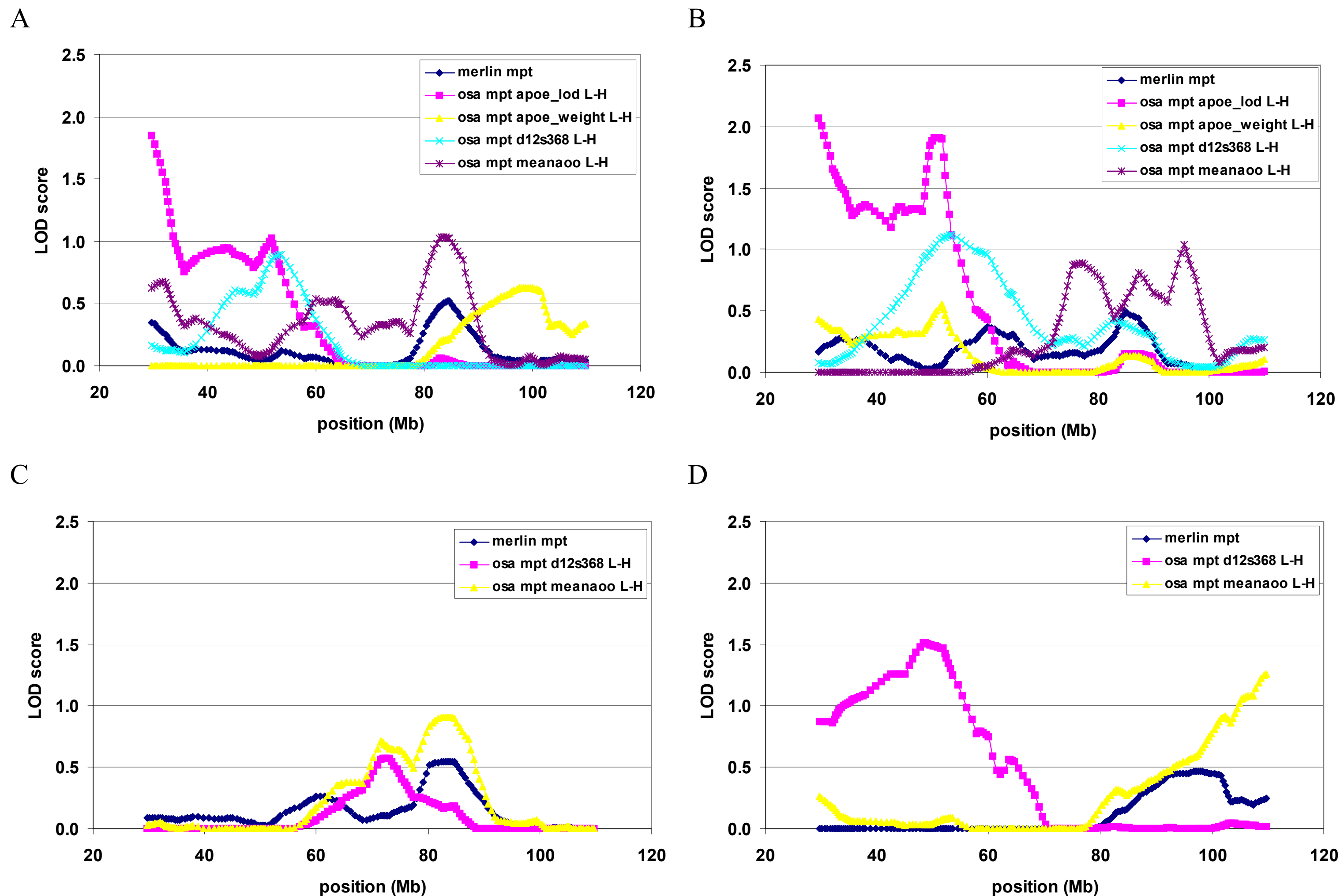


Figure 4-5 Multipoint analyses using covariates to order families from low to high in the overall and three subsets

A. Overall data set B. Autopsy-confirmed subset C. ApoE ε4-positive subset D. ApoE ε4-negative subset

This peak was also seen in the ApoE ε4- subset with LOD=1.52 around 49Mb between rs1806797 and rs7097397 when the families were ordered from low to high by the peak linkage signal on chromosome 12 (D12S368 LOD score) (Liang et al., 2005) with p=0.04 (Figure 4-6).

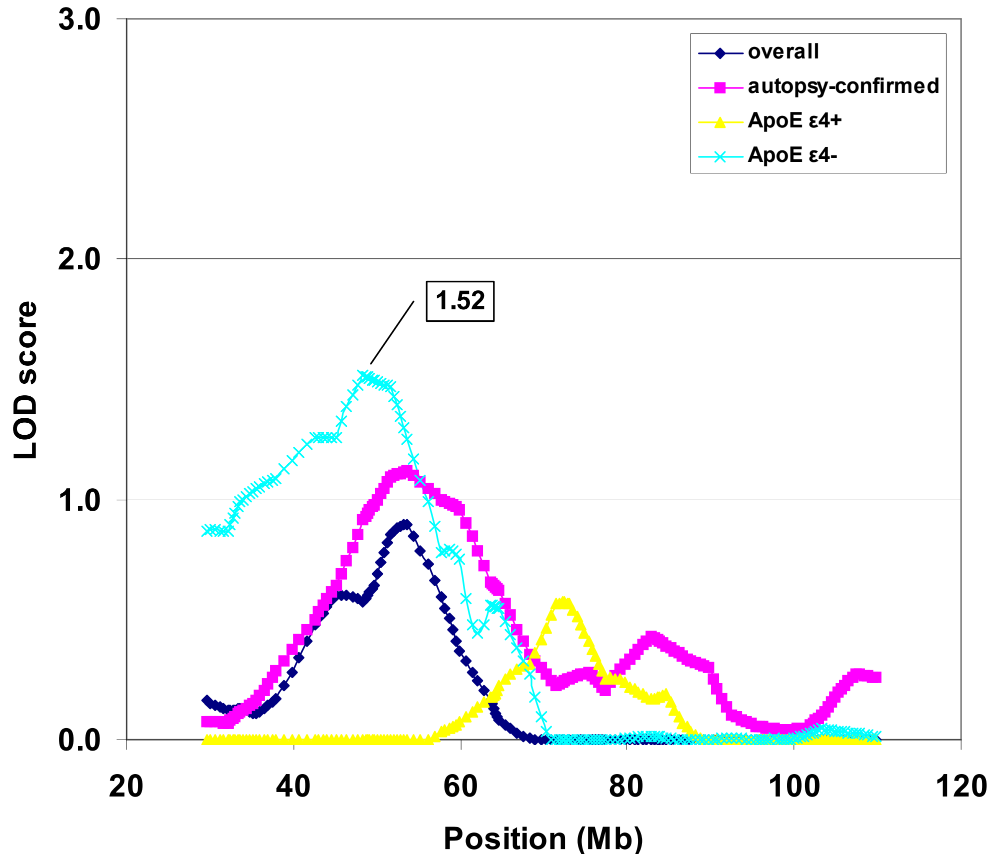


Figure 4-6 OSA multipoint analyses (D12S368 LOD Low to High)

The peak LOD=1.52 around 49 Mb between rs1806797 and rs7097397 was seen in the ApoEε4- subset when the families were ordered from low to high by the peak linkage signal on chromosome 12 (D12S368 LOD score).

In the ApoE ε4- subset, SNP rs3750686 at 87.2 Mb generated a parametric two point LOD score of 1.54. This SNP also had a two point LOD of 2.30 in the OSA analysis using the linkage signal on chromosome 12 from low to high as a covariate

(Figure 4-4 D). Thus families that are not linked to ApoE or the locus on chromosome 12 have suggestive LOD scores around 49.7Mb on chromosome 10.

When we used the covariates to order families from high to low (families with higher LOD scores at the covariate were weighted more than the families with lower LOD scores at that covariate), a peak LOD of 1.16 was seen in overall data set (rs1439042, 80 Mb, Figure 4-7A). When the LOD score of marker D12S368 was used as a covariate to order families from high to low, the ApoE ε4-negative subset showed a peak at SNP rs701865 (95 Mb) with LOD=1.32 (Figure 4-7D). The autopsy-confirmed subset showed a peak at SNP rs1227047 (73Mb) with LOD=1.64 (Figure 4-7B). Thus a locus around 80 Mb might interact with the locus on chromosome 12 or the ApoE locus. We also performed analyses to look at the data in different prospects with the combinations of covariates and subsets, see appendix (Figure A.1-A.6).

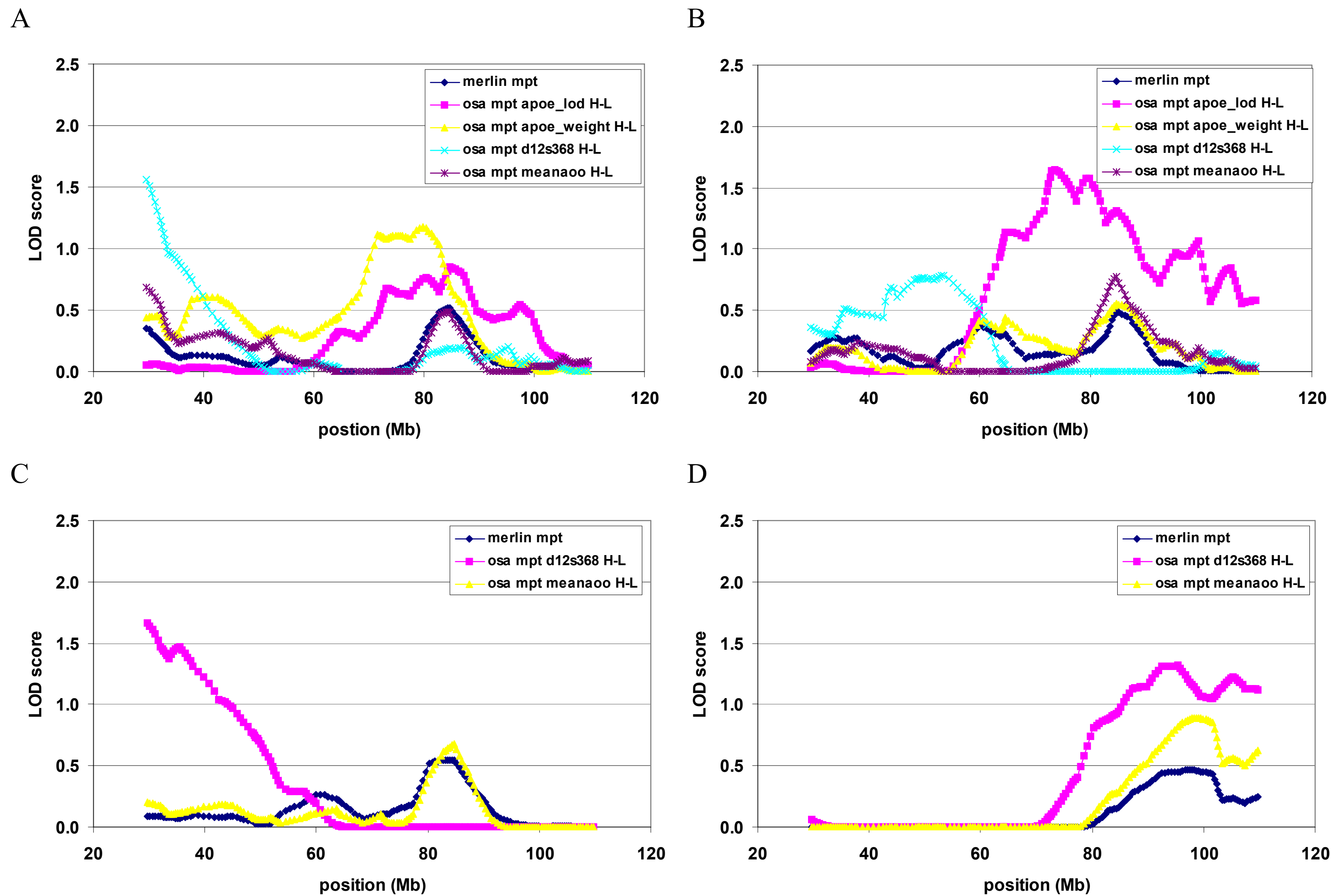


Figure 4-7 Multipoint analyses using covariates to order families from high to low in the overall and subsets
A. Overall data set B. Autopsy-confirmed subset C. ApoE ε4-positive subset D. ApoE ε4-negative subset

Discussion

In 2000, three groups reported significant linkage of LOAD to two regions of the long arm of chromosome 10 (Bertram et al., 2000; Ertekin-Taner et al., 2000; Myers et al., 2000). One study (Myers et al., 2000) followed up "suggestive" (Lander and Kruglyak, 1995) linkage results on this chromosome using an affected sibpair method resulting in significant linkage near ~ 58 Mb. Using a different methodology, a second study (Ertekin-Taner et al., 2000) performed multipoint linkage analyses in five late-onset AD families using Aβ plasma levels as a quantitative phenotype and found significant linkage very close to the same chromosomal region. The third report (Bertram et al., 2000) was a candidate gene-driven linkage analysis of 435 families using six genetic markers in a region of chromosome 10 that lies ~40 Mb distal to the one implicated by the former two studies. The candidate gene of interest encodes the insulin-degrading enzyme (*IDE*) that, together with neprilysin and other proteases, has been previously suggested to play a major role in the degradation and clearance of Aβ in brain (Selkoe, 2001). Two of the six markers in this region showed significant linkage to AD. In addition, one marker, *D10S583*, displayed modest evidence for linkage disequilibrium with AD and one of its alleles using a diallelic test (p=0.04 after correction for multiple testing). Across this broad region on chromosome 10q, there are many candidate genes.

We hypothesized that locus heterogeneity might be one reason for the inconsistent linkage results. To increase homogeneity, we stratified our linkage data set by autopsy

confirmation and ApoE status. We also applied the OSA method that has potential to find a more homogeneous subset by incorporating covariates into analysis.

Our data suggest that applying covariates to the analysis increases the homogeneity in the data set. Before the covariate ApoE_LOD score was used in the analysis, the highest two point LOD score was 1.31. However, we found a major linkage peak and a minor linkage peak in the overall data set after we applied the ApoE_LOD score as covariate. One is between 40 and 60 Mb with peak LOD of 2.69, at SNP rs1890739 (45.1 Mb); the other one is between 80 and 100 Mb with peak LOD of 1.70, at SNP rs11816558 (84.7 Mb), as shown in the Figure 4-3. The p values for the increases of LOD scores were 0.03 and 0.43 with 21% and 94% of the families, respectively. Furthermore, by combining subsetting with covariate analysis we were able to decrease the heterogeneity and refine the linkage peak. Without applying covariates, the multipoint analysis in overall data set failed to identify any SNP with a LOD>1.0. However, in the autopsy-confirmed subset, the single peak is located between 40 and 60 Mb when we used ApoE LOD score as covariate to order families from low to high (Figure 4-6). The peak LOD is 1.91, at SNP rs7097397 (49.7 Mb). These data suggest that the families without ApoE ε4 alleles may have a gene at this location that is linked to the disease. When we use D12S368, a surrogate for linkage to chromosome 12, as a covariate to order families from low to high, we also see the peak between 40 and 60 Mb in the autopsy confirmed and ApoE ε4- subsets. Analysis using ApoE and D12S368 LOD scores greatly increased evidence of linkage in areas showing no linkage in the

overall analyses. For example, rs1890739 at 45.1 Mb reached LOD of 2.69 among 21% of families by using the covariate of ApoE_LOD score to order families from low to high in OSA, whereas the same SNP showed a LOD of 0.35 when all families were analyzed together.

Our results show evidence for linkage to an LOAD gene on chromosome 10 (near 45.1 Mb) and this locus is independent of the chromosome 12 locus and ApoE gene. We also see a minor linkage peak between 80-100 Mb. Further studies on the candidate genes around 45 Mb location and possible roles of the functional candidate genes on chromosome 10 in AD pathology will hopefully lead to a better understanding of the etiology of Alzheimer disease.

CHAPTER V

EXAMINATION OF FIVE CANDIDATE GENES FOR ALZHEIMER DISEASE

Introduction

Other than the ApoE locus on Chromosome 19, there is no single locus that has been consistently replicated as a genetic risk factor for AD. Recently, five candidate genes within the broad region of linkage on chromosome 10 have been reported as associated with Alzheimer disease. These genes, VR22 (also known as CTNNA3, OMIM#607667), LRRTM3 (OMIM#610869), PLAU (OMIM#191840), TNFRSF6 (OMIM#134637) and IDE (OMIM#146680) (Bertram et al., 2000; Ertekin-Taner et al., 2003; Feuk et al., 2000; Finckh et al., 2003; Martin et al., 2005), may all be relevant to the Alzheimer disease pathophysiological process based on their function. Thus we examined these genes in more detail by performing both family-based and case-control association studies.

Materials and methods

Study populations

We used a case-control sample consisting of 483 unrelated cases with probable or definite AD and 879 unrelated cognitively normal elderly controls who were either the

spouses of AD patients or subjects recruited from the outpatient clinics of the participating institutions. Cases and controls were collected through the Center for Human Genetics Research (CHGR) at Vanderbilt University and the Center for Human Genetics (CHG) at Duke University. All individuals included in this study are Caucasian. Written consent was obtained from all participants in agreement with protocols approved by the institutional review boards at each contributing center. AD was diagnosed according to the NINCDS-ADRDA criteria (McKhann et al., 1984b). Controls had no obvious signs of cognitive or neurological impairment when enrolled in the study as determined by personal interview by clinical personnel of the Vanderbilt CHGR and Duke CHG. All controls had results within the normal range in the Mini-Mental State Exam (MMSE) or Modified Mini-Mental State Exam (3MS). Age at onset was recorded as that age at which the first symptoms were noted by the participant or a family member. If the affected individuals were early in the disease process, we included their report of age-at-onset as part of the determination. If the disease was more advanced, we only used information as collected from multiple family members (such as, spouse and children). Mean ± SD (standard deviation) of age at onset (AAO) in cases was 71.8±6.0 years. The control group's mean age at examination (AAE) ± SD was 72.0±6.3 years, while the case group's mean AAE ± SD was 76.5±6.53 years. The AD cases were 63.3% female, while the controls were 58.5% female.

Our independent family data set consists of 730 pedigrees with a total of 1337 affected discordant sib pairs (DSP) with a total of 1521 late-onset AD (LOAD) patients

(minimum age at onset (AAO) ≥ 60 years) and 974 unaffected individuals. Samples were ascertained by the following centers: the NCRAD repository at Indiana University (NCRAD); the Collaborative Alzheimer Project (CAP), including Duke and Vanderbilt Universities and University of California at Los Angeles; and the National Institute of Mental Health repository (NIMH). Age at onset was recorded as that age at which the first symptoms were noted by the participant or a family member. If the affected individuals were early in the disease process, we included their report of age-at-onset as part of the determination. If the disease was more advanced, we only used information as collected from multiple family members (such as, spouse and children). Mean ± SD of age-at-onset (AAO) in affected individuals in the family-based sample was 72.9±6.4 years. The mean age-at-exam (AAE) ± SD was 80.1±7.1 and 69.9±11.2 in affected and unaffected individuals, respectively. The AD affected group was 67.3% female, while the unaffected group was 56.2% female.

SNP selection and genotyping

Following informed consent, blood samples were collected from each individual. Genomic DNA was obtained from the repositories (NIMH, NCRAD) or extracted from whole blood (CAP) by use of the Puregene system (Gentra Systems, Minneapolis, MN). All samples were coded and stored at 4°C until used.

NCBI (http://www.ncbi.nih.gov), Ensembl (http://www.ensembl.org) and Applied Biosystems web sites (http://home.appliedbiosystems.com) were mined to select SNPs

according to location relative to other selected SNPs, high minor allele frequency, and availability of quality assays. First, we selected the SNPs that showed association in previous reports. Second, any known functional SNPs were included. Third, HapMap data were mined to select TagSNPs using Haploview (Barrett et al., 2005). Finally, SNPs with high heterozygosity and available high quality assays were used to fill any remaining large genomic gaps. Fifty SNPs in five genes were selected (Table 5-1).

Table 5-1 Genotyped SNPs in five candidate genes

SNP	Alternative ID	Location (bp)*	Allele variation	Function
VR22				
RS1786927	HCV1380042	67,352,267	A>G	Intron
RS2126750	HCV3096510	67,507,709	A>T	Intron
RS7911820	HCV3096479 (4825)	67,534,145	G>T	Intron
RS12357560	4783	67,534,187	C>T	Intron
RS7070570	10P0002VR22 (4360)	67,534,610	A>G	Intron
RS7074454	CV3096478	67,534,965	C>T	Intron
RS10822719	HCV3096477	67,535,076	C>T	Intron
RS6480140	CV11295092	67,538,887	A>C	Intron
RS922347	HCV8934629	67,652,964	C>T	Intron
RS4463744	HCV378461	67,778,486	A>T	Intron
RS2441718	HCV11848786	67,806,967	A>G	Intron
RS2939947	HCV26676598	67,808,364	A>G	Intron
RS2456737	HCV16158312	67,825,340	A>G	Intron
RS997225	HCV8040922	67,952,976	A>G	Intron
RS4746606	HCV1635544	68,061,108	C>T	Intron
RS7909676	HCV8040157	68,104,803	A>C	Intron
RS11593235	HCV1903928	68,546,044	C>T	Intron
RS10997591	HCV1853797	68,671,884	C>T	Intron
RS7903421	HCV29243978	68,951,738	A>G	boundary
RS3096244	HCV1395441	69,080,192	C>T	Intron
LRRTM3				
RS1001016	HCV8934595	68,347,044	C>G	Promoter
RS12769870	HCV1974488	68,347,401	A>G	Promoter
RS1925583	HCV11735884	68,349,950	G>T	Promoter
RS2394314	HCV26813816	68,350,254	A>G	Promoter
RS10762122	HCV1974468	68,386,380	C>T	Intron
RS942780	HCV8934576	68,406,547	A>G	Intron
RS1925617	HCV11735825	68,434,823	G>T	Intron
RS1925622	HCV1974419	68,439,644	A>G	Intron
RS1925632	HCV11735814	68,469,620	A>C	Intron
RS1952060	HCV11735812	68,472,940	C>T	Intron
RS2147886	HCV1903965	68,488,649	A>G	Intron
RS2251000	HCV1903957	68,494,777	A>G	Intron
RS2764807	HCV16064467	68,498,938	A>G	Intron
RS10762136	HCV1903945	68,513,538	A>C	Intron

Table 5-1 Genotyped SNPs in five candidate genes (cont'd)

SNP	Alternative ID	Location (bp)*	Allele variation	Function
PLAU				
RS1916341	HCV11458608	75,341,168	A>C	Exon
RS2227564	HCV26165616	75,343,107	C>T	Coding exon
RS2227566	HCV3155395	75,343,737	C>T	boundary
RS2227568	HCV26165619	75,343,885	C>T	Coding exon
RS4065	HCV3155393	75,346,470	C>T	Exon
TNFRSF6				
RS1800682	HCV9578811	90,739,943	C>T	Promoter
RS1324551	HCV9578820	90,741,496	A>G	Intron
RS2031612	HCV8984582	90,756,960	A>G	Intron
RS2296600	HCV16184452	90,760,419	C>G	boundary
IDE				
RS2251101	HCV27104906 (IDE-7)	94,201,284	C>T	3' UTR
RS2251101	HCV27104906 (IDE-7(2))	94,201,284	C>T	3' UTR
RS7076966	HCV1819856	94,315,491	C>T	Intron
RS4646954	HCV1819861	94,323,807	A>G	Exon
RS3758505	HCV1819863	94,324,758	G>T	Promoter
RS7099761	HCV31982773 (IDEU-4)	94,325,779	A>G	Promoter
RS1544210	HCV7480091 (HHEX-2)	94,477,781	A>G	unknown

* SNP locations were based on NCBI build 35

SNP genotyping was performed using Assays-On-Demand™ or Assays-By-Design™ with the ABI PRISM® 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA). Amplification was performed in a 384-well DNA Engine Tetrad® 2 Peltier Thermal Cycler (MJ Research, Waltham, MA) with the following conditions: 94°C-10 min; 92°C-15 sec, 60°C-1 min (50 cycles); 4°C-hold. Systematic genotyping errors were minimized by use of a system of quality control (QC) checks with duplicated samples (Rimmler et al., 1998). Linkage disequilibrium (LD) and deviations from Hardy-Weinberg equilibrium (HWE) were assessed using Haploview (Barrett et al., 2005).

Statistical methods

Association analyses

Case-control association analyses for single alleles and for genotypes were conducted using logistic regression (SAS Institute, Cary, N.C., USA, version 8.1). We tested three different models. Model I assumed an additive effect on the log scale for the alleles (e.g., if A is the minor allele, having no A alleles=0, having one A allele=1, having two A alleles=2). Model II dichotomized genotype according different alleles (having allele A =0, not having allele A =1). Model III categorized genotypes into three groups using the most frequent genotype as a reference genotype and other groups were compared with the reference group separately. Statistical significance was declared at $\alpha=0.05$.

To adjust for potential confounders, we included gender and age at examination (AAE) as covariates in the regression analysis. In addition, we included ApoE status in the model to exclude its possible confounding effect. We considered ApoE status as 3 categories; those that have two ApoE ε4 alleles, those that have at least one ApoE ε4 allele and those that have no ApoE ε4 allele. We analyzed the data using models with and without covariates because the data set had 7.4% missing data after we adjusted for the confounders.

Haplotype analyses in the case-control data set were conducted using haplo.stats (Schaid et al., 2002). An EM algorithm is used to generate haplotype frequency estimates under the null hypothesis. We estimated haplotype frequencies and tested association of each haplotype with a frequency of at least 1% in our case-control data set with age and gender adjusted score statistics. Haplotype logistic regression was modeled using a GLM algorithm including age and gender as covariates. The most frequent haplotype occurring in a similar percentage of cases and controls was selected as the baseline haplotype. To evaluate the association of subsets of alleles from the full haplotype, a sliding window of three SNPs was used and the global score statistics were reported.

Family based association analysis was conducted using the pedigree disequilibrium test (PDT, for allelic effect) and Genotype-PDT (GenoPDT, for genotypic effect) for single-locus tests to assess association between genotypes and risk of AD in

the family data (Martin et al., 2000b; Martin et al., 2002). The Haplotype Family Based Association Test (Horvath et al., 2001) was used for haplotype analysis in family data set.

Linkage analysis in candidate genes

Two-point and multipoint linkage analyses were performed as previously described (Chapter IV). Because LD between SNPs within the five candidate genes is high, we used $r^2>0.1$ as criteria in MERLIN to define SNPs in LD to be able to perform the multipoint linkage analysis in candidate genes.

Results

Candidate gene association analyses

We selected a number of representative SNPs in the five candidate genes, VR22 (N = 20 SNPs), LRRTM3 (N = 14 SNPs), PLAU (N = 5 SNPs), TNFRSF6 (N= 4 SNPs) and IDE (N = 7 SNPs) on chromosome 10, respectively (Table 5-1). The SNPs that showed nominal association with the disease using logistic regression are listed in Table 5-2. The complete sets of results are in the Appendix (Table A-1 to A-13). For Table A-1, the results were based on χ^2 analysis of allele distribution.

Table 5-2 Candidate genes and SNPs showing association in at least one data set

Gene	SNP	Allele variation	Nucleotide position*	Case-control [1]		Family[2]
				P	OR (95%CI)	P
VR22	**RS7911820**	G>T	67,534,145	0.27[a]	0.91 (0.77, 1.08)	***0.03[g]***
	RS7070570	A>G	67,534,610	0.06[e]	1.69 (0.97, 2.95)	***0.03[h]***
	RS7074454	C>T	67,534,965	0.14[f]	0.64 (0.35, 1.16)	***0.01[g]***
	RS2441718	A>G	67,806,967	***0.03[c]***	***1.29 (1.02, 1.62)***	***0.03[g]***
	RS2456737	A>G	67,825,340	***0.03[c]***	***1.33(1.04, 1.72)***	***0.05[h]***
	RS7909676	A>C	68,104,803	0.33[c]	1.14 (0.88, 1.48)	***0.00[h]***
	RS3096244	C>T	69,080,192	0.15[b]	1.16 (0.95, 1.42)	***0.05[g]***
LRRTM3	**RS12769870**	A>G	68,347,401	***0.01[e]***	***0.70 (0.53, 0.92))***	0.08[h]
	RS2394314	A>G	68,350,254	***0.04[d]***	***0.72 (0.54, 0.98)***	0.14[h]
	RS10762122	C>T	68,386,380	***0.02[b]***	***1.24 (1.02, 1.50)***	0.09 [h]
	RS942780	A>G	68,406,547	0.48[d]	1.11 (0.83, 1.48)	***0.03[h]***
	RS1925617	G>T	68,434,823	0.17[c]	0.82 (0.61, 1.09)	***0.00[h]***
	RS1925622	A>G	68,439,644	0.48[e]	0.90 (0.66, 1.22)	***0.01[h]***
	RS1952060	C>T	68,492,940	0.24[c]	0.86 (0.67, 1.11)	***0.00[h]***
PLAU	**RS1916341**	A>C	75,341,168	0.20[e]	0.85 (0.66, 1.09)	***0.02[g]***
	RS2227566	C>T	75,343,737	0.35[d]	0.87(0.65, 1.16)	***0.01[g]***
	RS2227568	C>T	75,343,885	0.11[b]	0.80 (0.61,1.05)	***0.01[h]***
	RS4065	C>T	75,346,470	0.22[e]	0.85 (0.67, 1.10)	***0.02[g]***
TNFRSF6	**RS1800682**	C>T	90,739,943	0.27[d]	1.19 (0.87, 1.61)	***0.03[h]***
	RS2031612	A>G	90,756,960	***0.02[a]***	***1.23 (1.04, 1.46)***	0.87[g]
	RS2296600	C>G	90,760,419	***0.02[a]***	***0.82 (0.69, 0.97)***	0.26[h]
IDE	**RS7099761**	A>G	94,325,779	0.50[e]	0.91 (0.68, 1.20)	***0.02[g]***

*The map location is from NCBI dbSNP, build 35.

** SNPs showing association (P<0.05) are bold and italic and those showing association in both case-control and family-based data sets are shaded.

[1.] We tested a total of 6 models for case-control data set and only most significant results are shown in the table.

a) model I with genotype-only; b) model I with both genotype and covariates; c) model II with genotype-only; d) model II with both genotype and covariates; e) model III with genotype-only; f) model III with both genotype and covariates;

[2.] We applied two methods to the family-based association test and only most significant results are shown in the table.

g) PDT; h) geno-PDT

There is at least one SNP showing association in each of these five candidate genes, VR22 (N = 7 SNPs), LRRTM3 (N = 7 SNPs), PLAU (N = 4 SNPs), TNFRSF6 (N= 3 SNPs) and IDE (N = 1 SNP), in case-control, family-based, or both data sets. SNPs rs2441718 and rs2456737 in VR22 showed association in both family-based and case-control data sets, with an odds ratio (OR) in the case-control data set of 1.29 (95% CI=1.02-1.62, p=0.03) and 1.33 (95% CI=1.04-1.72, p=0.03), respectively, when we used the common allele to dichotomize genotypes. However, the association went away after we adjusted for the confounders (age, gender and ApoE status). SNP rs2031612 in TNFRSF6 showed association in the case-control data set with an OR of 1.23 (95% CI=1.04-1.46, p=0.02 with the common allele A under an additive model). SNP rs10762122 in LRRTM3 also showed association in the case-control data set after we adjusted for confounders, with an OR of 1.24 (95% CI=1.02-1.50, p=0.02 with the common allele T). The complete sets of results are in the appendix.

Using haplo.stats to calculate age and gender adjusted score statistics, two haplotypes in PLAU were associated with AD in the overall case-control data set. All five genotyped SNPs (rs1916341, rs2227564, rs2227566, rs2227568, and rs4065) in PLAU were used to define the haplotype. The haplotype GTCCC was associated with AD at p=0.01 and haplotype GCCTC had p=0.02 for their global score statistics. These two haplotypes remained significant when we applied FDR for multiple comparisons. The most frequent haplotype TCTCT was used as the referent haplotype (p=0.43). Haplotype GTCCC was also borderline significant (p=0.05) in the logistic regression

after taking into account age and gender. A sliding window analysis (n=3) across the five genotyped SNPs in PLAU generated three haplotypes and all of those three were significant ($p<0.05$ for age and gender adjusted score statistics). Haplotype analyses on the other candidate genes didn't show any significant results at $p<0.05$ (data not shown). Although we also see a significant effect for PLAU haplotypes in the family-based data set, the association is in the opposite direction. The most common haplotype, TCTCT, was significantly associated with AD (P=0.008) in the haplotype FBAT analysis, while the haplotype GTCCC was not associated with AD (p=0.70). Haplotype GCCTC had a marginal effect (P=0.05).

Linkage analysis in candidate genes

We did linkage analyses on the SNPs in the five candidate genes. With these SNPs, we applied the same approaches as for the SNPs in the linkage panel on the overall data set. Two SNPs (rs7070570 and rs2441718) in VR22 had LOD>1.0 (1.30 and 1.02, respectively) in FASTLINK/HOMOG two-point analyses. In the OSA analysis using the ApoE_LOD score as a covariate, a two point LOD score of 2.30 was obtained at SNP rs2456737 in VR22 (p=0.05 in 51% families, high to low). A two point LOD score of 1.62 was obtained at SNP rs942780 in LRRTM3 (p=0.02 in 22% families, low to high).

We also performed multipoint analysis using SNPs in the candidate genes on overall data set. Since there is high LD between SNPs in genes, we used $r^2 > 0.1$ to define SNPs in LD in program MERLIN. Only one SNP (rs2456737) in VR22 had a

LOD score of 1.05 (data not shown). This SNP also showed association in both case-control and family-based data sets ($p<0.05$).

Genomic convergence of association and linkage analyses

Combined with the linkage analysis results from the previous chapter, both the linkage and association studies gave us suggestive results, shown in Figure 5-1. However, these results did not converge.

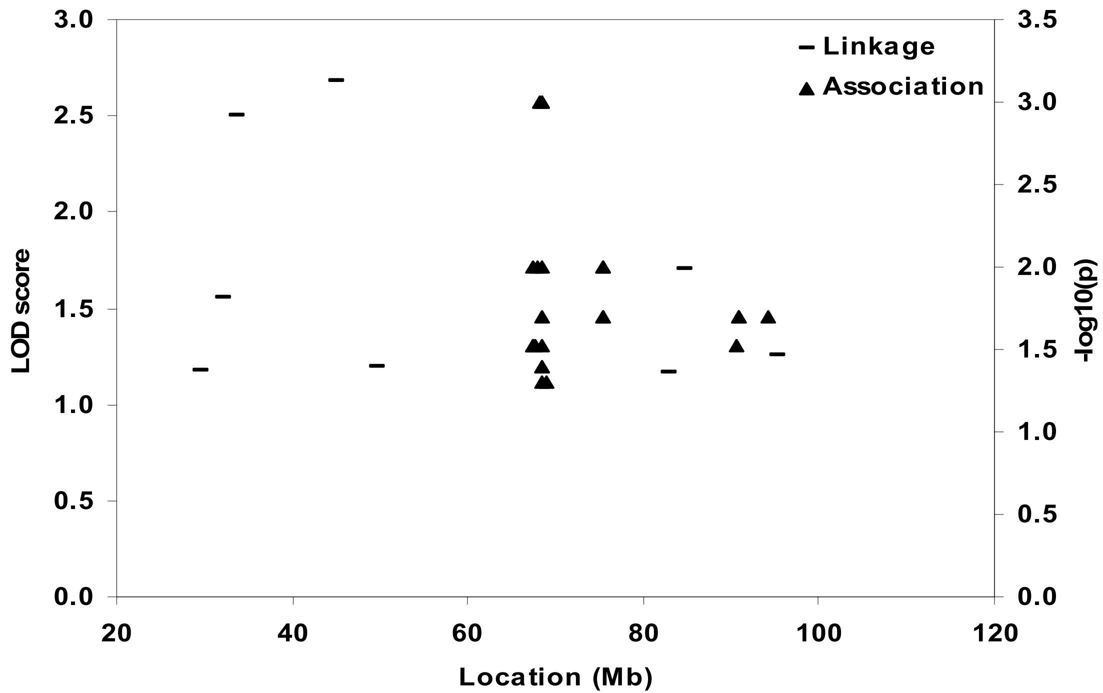


Figure 5-1 Overall suggestive linkage and association data on chromosome 10
Triangles are $-\log_{10}$ (p) from the candidate gene association study. The dashes are two-point LOD score in Ordered Subset Analysis (OSA) from the linkage study.

The linkage study revealed a major peak between 40-60 Mb (LOD=2.69) and a minor peak between 80-100 Mb (LOD=1.70). In the association analysis, VR22 (67.8

Mb), LRRTM3 (68.3 Mb) and PLAU (75.3 Mb) showed evidence of association. VR22 showed association in both the family and case-control data sets. LRRTM3 demonstrated association in the case-control data set after adjusting for confounders, and this gene survived both Bonferroni and FDR correction for multiple comparisons in the family data set. PLAU showed haplotypic association in both the family and case-control data sets. The SNPs in TNFRSF6 remained associated after FDR correction for multiple testing (q=0.2). Only TNFRSF6 (90.7 Mb) falls near one of the observed linkage peaks (80-100 Mb).

Discussion

Across this broad region of linkage on chromosome 10q, there are many candidate genes that may be involved in AD pathophysiology. In our study, we focused on VR22 (α-catenin), LRRTM3, PLAU, TNFRSF6 and IDE. VR22 is a large (1.8 Mb) gene located at 67.3 Mb that encodes alpha-T catenin, which is a binding partner of beta catenin. This makes VR22 an attractive candidate gene because beta catenin interacts with presenilin 1, which has many mutations that elevate Aβ42 and can cause early onset familial AD. LRRTM3 is a nested gene in an intron of VR22 which also has plausible biological function of promoting APP processing by β-secretase (BACE1). Ertekin-Taner et al. reported two intronic SNPs in VR22 showing highly significant association (P=0.0001 and 0.0006) with plasma Aβ42 in 10 extended LOAD families (Ertekin-Taner et al., 2003). Martin et al. suggested significant evidence of association between VR22

and AD in both the families and the unrelated cases and controls, and that the effect is dependent on ApoE status (Martin et al., 2005).

PLAU encodes urokinase-type plasminogen activator (uPA) which converts plasminogen to plasmin. Plasmin is involved in the processing of amyloid precursor protein and degrades secreted and aggregated amyloid-beta, a hallmark of Alzheimer disease (AD). Finckh et al. found a coding SNP in PLAU showing association ($P=0.00039$) in 347 patients with LOAD and 291 control subjects (Finckh et al., 2003).

The TNFRSF6 gene encodes FAS, a cell-surface receptor involved in apoptosis initiation. Elevated levels of FAS have been reported in the brains of Alzheimer's disease (AD) patients (Feuk et al., 2000). Despite these tantalizing reports, none of the associations to these genes has been widely replicated in multiple studies and thus generally accepted as an AD risk gene.

Although we found that there is at least one SNP in each gene showing association with Alzheimer disease, the results are not strikingly significant, especially in consideration of the problem of multiple comparisons. We adjusted for multiple comparisons in both the family and case-control data sets by the number of genotyped SNPs in each gene. Using the overly conservative Bonferroni correction, none of the 50 genotyped SNPs was significant at global $p<0.05$ in the case-control data set. The only SNP remaining significant after Bonferroni correction was SNP (rs1952060) in LRRTM3 within the family data set. When we genotype a lot of SNPs that are not independent with each other in a gene, Bonferroni correction is too conservative for detecting a

positive result. Perhaps a more appropriate approach is to control for the false discovery rate. When we applied a false discovery rate of q=20%, three SNPs in LRRTM3 (rs1925617, rs1925622 and rs1952060) and one SNP (rs2227568) in PLAU remained significant in the family data set. In the case-control data set, two SNPs (rs2031612 and rs2296600) in TNFRSF6 were significant.

PLAU is the only gene showing haplotypic association with AD in both the family-based and case-control data sets. Although the associated haplotype is different between two data sets, considering the strong single marker effect (one SNP survived FDR multiple comparison correction), these results remain very interesting. More detailed examination of this gene is needed to explore its role in AD.

Martin et al., 2005 tested 11 SNPs in VR22 and LRRTM3 in an overlapping data set, and we included them in our SNP list. We extended their analyses by genotyping 12 more SNPs in VR22 and 11 more SNPs in LRRTM3. These additional SNPs were selected because they were functional SNPs or tag SNPs or filled large genomic gaps. Although the overlap with Martin et al.'s data set is extensive, the current data set includes more updated clinical information. We found the same associated SNPs (rs7911820 and rs7074454 in VR22; rs1925617 in LRRTM3) at $p<0.05$ in the family-based data set, but also identified two SNPs (rs2441718 and rs2456737 in VR22) showing association in both family and case-control data sets.

The inconsistencies between our study and other groups' findings may reflect differences in the samples, study designs or analytical techniques. They may also reflect

heterogeneity within the samples or more complicated genetic mechanisms such as the interplay of genetic and epigenetic factors. Also, we selected several SNPs in each gene based on the previous reports, spacing and heterozygosity. But the genes haven't been exhaustively examined. There is a possibility that association could be found if we genotype all tagSNPs to cover each haplotype block and do an exhaustive test of these tagSNPs in these genes. In addition, the candidate gene selection strategy is biased by the known biological function and the function of most genes is poorly understood. This prompted us to combine linkage and candidate gene association studies to try to locate the susceptibility genes underlying Alzheimer disease.

However, the linkage analysis did not converge with the candidate gene association studies, in which VR22 (67.8 Mb) showed association in both case-control and family-based data sets (Figure 5-1). Since we applied linkage and candidate gene association approaches simultaneously, we could not predict the results in advance. Our results support the role of the tested candidate genes in AD, but it is also important to study the candidate genes around the 45 Mb location. Our results suggest a more complicated heterogeneity background on chromosome 10 with LOAD. Further experiments on the possible roles of the genes on chromosome 10 in AD pathology will hopefully lead to a better understanding of the etiology of Alzheimer disease.

CHAPTER VI

GENETICS OF CDC2 LOCUS IN ALZHEIMER DISEASE

Introduction

Alzheimer disease (AD) is the most common form of dementia among the elderly. This progressive neurodegenerative disorder accounts for more than half of all cases of dementia among people over 65 years of age (Francis et al., 1999). Clinically AD is slowly progressive, resulting in memory loss and alterations of higher intellectual function and cognitive abilities (Guttman et al., 1999). Pathologically, AD is characterized by neurofibrillary tangles in the neurons of the cerebral cortex and hippocampus and the deposition of amyloid within senile plaques and cerebral blood vessels (Wisniewski et al., 1993).

Alzheimer disease is a complex neurodegenerative disorder resulting from multiple genetic and nongenetic factors (Myers and Goate, 2001). The only well-established genetic susceptibility factor for non-mendelian late-onset AD is the ε4 allele of the apolipoprotein E (*APOE*) gene (Corder et al., 1993). However, the presence of the APOE-4 allele is neither necessary nor sufficient to cause AD, indicating that additional genetic or non-genetic factors influencing AD risk are yet to be identified. To discover additional susceptibility genes, genome scans were initiated resulting in the identification of genomic regions of interest, predominantly on chromosomes 2, 9, 10, 12 and 15 (Liang

et al., 2005; Mayeux et al., 2002; Myers et al., 2002; Pericak-Vance et al., 2000). In 2000, three quite different approaches led to convincing evidence that there is at least one other susceptibility gene for Alzheimer disease on chromosome 10q. Two linkage studies gave strong evidence of a locus at almost exactly the same location (D10S1225): one using plasma levels of the amyloid beta brain-deposited fragment of the amyloid precursor protein as a continuous phenotype, and the other using a categorical disease phenotype. A third candidate gene linkage and association analysis approach interestingly found a maximum signal ~35–60cM distal to the previous studies (D10S583). (Bertram et al., 2000; Myers et al., 2000; Ertekin-Taner et al., 2000).

Some of the over 240 genes in the broad region spanned by these results may be considered as positional candidates because they are also hypothesized to interfere with AD-related biochemical pathways.

The cell division cycle 2 (CDC2) gene is located within this linkage peak (2 Mb from the marker D10S1225) and is thought to be one of the main candidate kinases involved in the abnormal phosphorylation of tau, which is in the hallmark neurofibrillary tangles of AD. CDC2 is also designated as CDK1 or $p34^{CDC2}$ and its neuron-specific form is CDK5. CDC2 is involved in the critical event at the point of convergence of the mitosis and neurodegeneration pathways in the AD. Therefore, CDC2 is a reasonable positional and functional candidate for association with AD. Johansson et al. (Johansson et al., 2003) reported that a variation in CDC2 gene is associated with AD at odds ratio of 1.78 (95% CI 1.18-2.68) using a data set of 272 Caucasian AD cases and 160 controls.

To test whether allelic variation in this potentially important candidate gene confers susceptibility to AD risk, we genotyped six single nucleotide polymorphisms (SNPs) across CDC2 in two independent samples; family-based and case-control. Our data suggest a lack of association of the variations in CDC2 gene with AD.

Materials and methods

Study populations

Family sample

We used a total of 1337 affected discordant sib pairs (DSP) defined by 1521 late-onset AD (LOAD) patients (minimum age at onset (AAO) ≥ 60 years) and 974 unaffected individuals in these families, as described in the previous chapters. Family data was ascertained by the following centers: the NCRAD repository at Indiana University (NCRAD); the Collaborative Alzheimer Project (CAP), including Duke and Vanderbilt Universities and University of California at Los Angeles; and the National Institute of Mental Health repository (NIMH). All participants were Caucasian Americans. Written consent was obtained from all participants in agreement with protocols approved by the institutional review board at each contributing center. AD was diagnosed according to the NINCDS-ADRDA criteria (McKhann et al., 1984b). Age-at-onset (AAO) was recorded as that age at which the first symptoms were noted by the participant or a family member. Mean AAO ± SD (standard deviation) in affected

individuals in the family-based sample was 72.9±6.4 years. The mean age-at examination (AAE) ± SD was 80.1±7.1 and 69.9±11.2 in affected and unaffected individuals, respectively. The AD affected group was 67.3% female, while the unaffected group was 56.2% female.

Case-control sample

Our case-control sample consists of 745 unrelated cases with probable or definite AD and 998 unrelated cognitively normal elderly controls who were either the spouses of AD patients or subjects recruited from the outpatient clinics of the participating institutions. Cases and controls were collected through the Center for Human Genetics Research (CHGR) at Vanderbilt University and the Center for Human Genetics (CHG) at Duke University. Criteria for AD diagnosis and screening of unaffected relatives were the same as described above. Controls had no obvious signs of cognitive or neurological impairment when enrolled in the study as determined by personal interview by clinical personnel of Vanderbilt CHGR and Duke CHG. All individuals included in this study were Caucasian. The control group's mean AAE ± SD was 72.0±6.3 years, while the case group's mean AAE ± SD was 76.5±6.53 years. The AD cases were 63.3% of female, while the controls were 58.5% female.

SNP and genotyping

Following informed consent, blood samples were collected from each individual. Genomic DNA was obtained from the repositories (NIMH, NCRAD) or extracted from whole blood (CAP) by use of the Puregene system (Gentra Systems, Minneapolis, MN). All samples were coded and stored at 4°C until used.

A map of the CDC2 region and SNPs studied is shown in Figure 6-1.

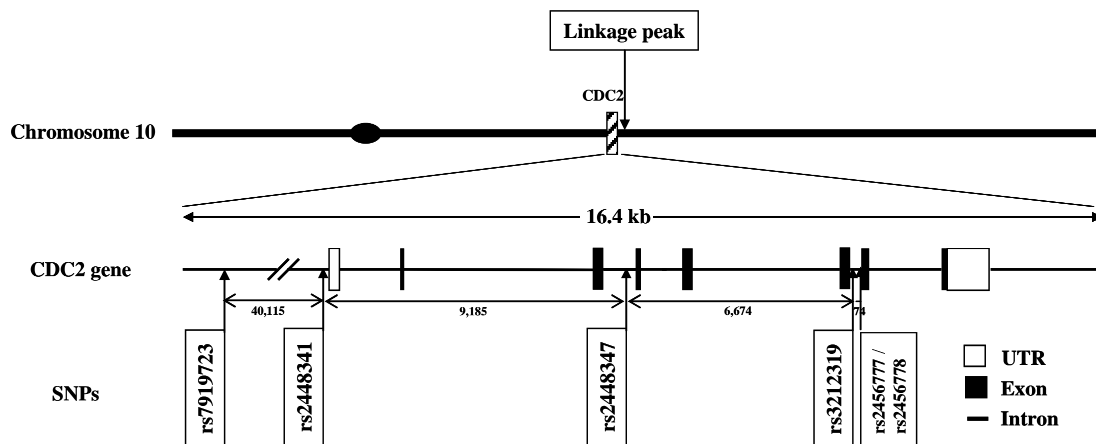


Figure 6-1 Gene structure and relevant features of CDC2 and surrounding sequences

The relevant SNPs are labeled with the dbSNP rs number. The distances between SNPs are shown in base pairs.

The CDC2 gene structure and transcript information were extracted from the Ensembl database (http://www.ensembl.org) according to the NCBI human genome sequence assembly build 35. An insertion/deletion SNP (rs3212319) which was associated with AD by Johansson et al. was selected. Two other adjacent SNPs that were discovered by that same group were also selected. Applied Biosystems (http://home.appliedbiosystems.com) web sites were mined to select all available SNPs

according to location relative to other selected SNPs, high minor allele frequency ($\geq$ 0.25), and availability of quality assays during the time when we did the genotyping. SNP genotyping was performed using Assays-On-Demand™ or Assay-By-Design™ with the ABI PRISM® 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA). Amplification was performed in a 384-well DNA Engine Tetrad® 2 Peltier Thermal Cycler (MJ Research, Waltham, MA) with the following conditions: 94°C-10 min; 92°C-15 sec, 60°C-1 min (50 cycles); 4°C-hold.

SNPs rs2456777 and rs2456778 reside right next to each other ([A/G], [A/T]), and a standard Taqman probe cannot detect adjacent SNPs. Thus we defined the most common haplotypes by sequencing 20 CEPH individuals. The two most common haplotypes for these two SNPs are AA and GT, allowing us to develop a single assay to test for the haplotype. Then we designed probes and primers oriented 5' to 3'. The probes are aagaaattttt AA/GTtttcctgtttt and forward primer is 5'-AAAGGTAACATATATGTAACAATGAGATTACATTTA-3', reverse primer is 5'-TGGGAGTGCCCAAAGCTCTA-3'. We used a Taqman assay for the genotyping. For those individuals with uncertain haplotype status, sequencing was performed to obtain the genotype. The forward primer is 5'-GAACTAGCAACTAAGAAACCACTTTTCC-3' and reverse primer is 5'-CCTTACCGAGAGCAAATCCA-3'. The genotypes were transcribed to be consistent with the TaqMan output file. Systematic genotyping errors were minimized by use of a system of quality control (QC) checks with duplicated samples (Rimmler et al., 1998).

Statistical methods

Descriptive analyses

We tested for deviations from Hardy-Weinberg equilibrium (HWE) using the program Genetic Data Analysis (GDA) (Zaykin et al., 1995). The linkage disequilibrium (LD) measures D' and r^2 was calculated using the GOLD program (Abecasis and Cookson, 2000).

Association analyses

Case-control association analyses for single alleles and for genotypes were conducted using logistic regression (SAS Institute, Cary, N.C., USA, version 8.1). We tested three different models. The first is an allele-based model that assumes an additive effect on the log scale for the alleles (e.g., having no A alleles=0, having one A allele=1, having two A alleles=2). The second is an allele-based model that dichotomizes genotypes according to different alleles (having allele=0, not having allele=1). The third is a genotype-based model where the most frequent genotype served as a reference genotype and other two genotypes were compared with the reference separately. Statistical significance was declared at $\alpha=0.05$.

To adjust for potential confounders, we included gender and age at examination (AAE) as covariates in the regression analysis. In addition, we considered ApoE status as two different models: "ApoE 4_2" model which breaks the analysis into 2 parts; those that have no ApoE 4 allele and those that have at least one ApoE 4 allele and the

“ApoE_3” model which breaks the analysis into 3 parts; those that have both ApoE 4 alleles, those that have at least one ApoE 4 allele and those that have no ApoE 4 allele.

Age-at-onset (AAO) was also analyzed as a dependent variable using a generalized liner model (GLM) to test the association between the age-at-onset and the gene and four models were used. Geno_linear is the model where different genotypes were considered as a linear progression; Geno_group is the model where genotypes were put into groups without linear progression; Geno_A and Geno_B are the models to group the genotypes into 2 groups according to the two alleles (A and B) of the SNP (one is the homozygote of one of the two alleles; the heterozygote and homozygote of the other allele were in the 2nd group). ApoE status and gender were used as covariates.

Family based association analysis was conducted using the pedigree disequilibrium test (PDT, for allelic effect) and Genotype-PDT (GenoPDT, for genotypic effect) for single-locus tests to assess association between genotypes and risk of AD in the family data (Martin et al., 2000b; Martin et al., 2002).

Linkage analysis of family data set

Two-point heterogeneity LOD score (HLOD) analyses were computed using FASTLINK and HOMOG (Ott, 1999). Because the mode of inheritance for AD is unknown, affected-only parametric analyses were performed using both autosomal dominant and autosomal recessive models with disease allele frequencies of 0.001 and 0.20, respectively, to model the susceptibility allele. Because it is likely that there is genetic heterogeneity in Alzheimer disease, we applied Ordered Subset Analysis (OSA)

(Hauser et al., 2004) in our linkage analysis to test for a set of families in which the LOD score in a particular region is higher than in the overall data set. The statistical significance of the increased evidence for linkage relative to evidence for linkage in the entire sample is assessed via random permutation of the order of inclusion of the families to estimate empirical p values. Families were ranked based on APOE LOD score at theta=0 (low to high or high to low) and family-specific APOE-4 allele weights (low to high or high to low). OSA analysis was also performed using mean age of onset as covariate (low to high or high to low) to test for a subset generating a significantly increased LOD score relative to the overall sample.

Power calculation

The PS Power and Sample Size calculations program (Dupont and Plummer, Jr., 1990) was used to calculate the power to detect a CDC2 association given the 745 cases and 998 controls sample. We used the odds ratio of 1.78 and the probability of exposure in controls of 0.35 as given in the original report of association (Johansson et al., 2003).

Results

Family-based analyses

CDC2 SNP allele and genotype frequencies for family-based data set are listed in Table 6-1. The LD measure (D' and r^2) between SNPs are shown in Table 6-2. There was no significant evidence for deviation from HWE for any SNP (Table 6-3). The PDT was conducted in the overall family sample (Table 6-3). None of these tests indicated association between these six SNPs and Alzheimer disease.

Table 6-1 CDC2 SNP allele and genotype frequencies for family-based data set

SNP	Location (Mb)	variation	allele	Allele frequency	genotype	Affected (frequency)	Unaffected (frequency)
			A	0.70	AA	676 (0.48)	403 (0.47)
RS7919724	62,165,848	A/G	G	0.30	AG	623 (0.44)	366 (0.43)
					GG	117 (0.08)	86 (0.10)
			C	0.66	CC	632 (0.45)	373 (0.44)
RS2448341	62,205,963	C/T	T	0.34	CT	628 (0.44)	383 (0.45)
					TT	157 (0.11)	99 (0.12)
			A	0.43	AA	281 (0.20)	164 (0.19)
RS2448347	62,215,148	A/G	G	0.57	AG	647 (0.47)	427 (0.49)
					GG	462 (0.33)	282 (0.32)
			C	0.68	CC	701 (0.47)	409 (0.45)
RS3212319	62,221,822	C/-	-	0.32	C-	632 (0.43)	392 (0.44)
					--	156 (0.10)	101 (0.11)
					AA/AA	100 (0.07)	46 (0.05)
			AA	0.25	AA/GT	497 (0.34)	299 (0.33)
RS2456777	62,221,895	A/G	GT	0.74	AA/AT	5 (0)	4 (0)
RS2456778	62,221,896	A/T	AT	0.01	GT/GT	836 (0.58)	540 (0.60)
			GA	0	GT/AT	14 (0.01)	5 (0.01)
					AT/AT	0 (0)	0 (0)

Table 6-2 Linkage disequilibrium between SNPs in CDC2 (D' and r^{2*})

	rs7919724	rs2448341	rs2448347	rs3212139	rs2456777, rs2456778
rs7919724		0.52	0.49	0.39	0.53
rs2448341	0.24		0.61	0.54	0.65
rs2448347	0.14	0.30		0.71	0.73
rs3212139	0.03	0.07	0.14		0.73

* D' is shown in the upper right half (shaded), and r^2 is shown in the lower left half. Because r^2 is only defined for bi-allelic markers in GOLD, there isn't r^2 value for the rs2456777/rs2456778 haplotype.

Table 6-3 Linkage and association results for family-based sample

	Linkage analysis		Association analysis	HWE (p value)	
SNP (rs#)	Dominant model HLOD score	Recessive model HLOD score	Pedigree disequilibrium test (PDT) p value	Affected	Unaffected
RS7919724	0.00	0.00	0.49	0.37	0.66
RS2448341	0.02	0.11	0.90	0.87	0.82
RS2448347	0.00	0.06	0.19	0.12	0.47
RS3212319	0.08	0.04	0.60	0.46	0.72
RS2456777 RS2456778	0.00	0.00	0.42	0.18	0.20

In the linkage analysis, the highest two-point heterogeneity LOD score (HLOD) was generated at SNP rs2448341 (0.11) under recessive model and allowing for heterogeneity. The OSA method also did not reveal any evidence of linkage in any subset of the data using either ApoE or mean age of onset as the covariate (data not shown).

Case-control analyses

Table 6-4 shows both allelic and genotypic association and the adjusted odds ratios in the overall dataset of cases and controls. The highest odds ratio (OR) of 1.38 (95% CI = 0.79 -- 2.43; p=0.26) was generated by the rs2456777/rs2456778 haplotype when we used the most frequent genotype (GT/GT) as the baseline to compare with the haplotypes that do not contain the GT haplotype. Age-at-onset, gender and ApoE status were included in the logistic regression as covariates. None of these results are significant. SNP rs2448341 had a marginal effect on Age-at-onset when the genotype was the only risk factor in the model (p=0.01, Table 6-5), but the effect became non-significant after the models were adjusted by covariates of ApoE status and gender (Table 6-5).

Table 6-4 Allelic and genotypic association for case-control samples

	Allelic Association						Genotypic Association			
SNP (RS #)	Allele	Cases	Controls	Odds ratio	95% CI	p-value	genotype	Odds ratio	95% CI	p-value
RS7919724	A G	693 331	928 464	0.96	(0.80,1.13)	0.60	GG vs AA AG vs AA	1.10 0.89	(0.69,1.73) (0.67,1.20)	0.69 0.45
RS2448341	C T	660 376	897 509	1.00	(0.85,1.19)	0.96	TT vs CT CC vs CT	1.27 1.07	(0.84,1.92) (0.80,1.44)	0.27 0.65
RS2448347	A G	468 568	644 756	1.03	(0.88,1.21)	0.69	GG vs AG AA vs AG	1.25 1.22	(0.91,1.72) (0.85,1.73)	0.16 0.28
RS3212319	C -	767 502	1076 327	0.92	(0.77,1.08)	0.29	C- vs CC -- vs CC	0.82 0.83	(0.62,1.09) (0.53,1.31)	0.18 0.42
RS2456777 **RS2456778**	AA GT AT	245 799 4	305 1106 3	-	-	0.31	AA/AA,AA/AT,AT/AT vs GT/GT AA/GT,GT/ATvs GT/GT	1.38 1.10	(0.79,2.43) (0.81,1.48)	0.26 0.53

Table 6-5 Association of Age-at-onset and CDC2 gene in case-control samples

	Models: genotype only				Models: genotype and covariates (ApoE status and gender)			
SNP (RS #)	**Geno_linear**	**Geno_group**	**Geno_A**	**Geno_B**	**Geno_linear**	**Geno_group**	**Geno_A**	**Geno_B**
RS7919724	0.26	0.53	0.47	0.28	0.39	0.65	0.55	0.43
RS2448341	**0.01**	**0.04**	0.11	**0.02**	0.08	0.24	0.12	0.18
RS2448347	0.94	0.95	0.93	0.80	0.66	0.87	0.74	0.68
RS3212319	0.84	0.72	0.67	0.61	0.44	0.47	0.99	0.31
RS2456777 **RS2456778**	N/A	0.39	N/A	N/A	N/A	0.46	N/A	N/A

* P values in the generalized linear model are shown in the table. The rs2456777/rs2456778 haplotype has more than two alleles and was applied to the Geno_group model only.

Discussion

CDC2 is thought to be one of the main candidate kinases involved in paired helical filaments (PHFs)-tau formation (Baumann et al., 1993). Previously, it has been shown that the occurrence of hyperphosphorylated tau correlates with cell division, differentiation and mitosis. The selective accumulation of CDC2 in neurofibrillary tangles (NFTs)-bearing neurons indicates that a mitotic post-translational mechanism might contribute to the PHF-tau formation in AD (Vincent et al., 1997). Johansson et al. reported that an insertion/deletion polymorphism influences cerebrospinal fluid (CSF) total tau protein levels (Johansson et al., 2005) using continuous traits analysis.

We have conducted a comprehensive analysis of the association between six CDC2 gene polymorphisms spanning the entire length of the gene and the disease risk in both case-control and family-based samples. Johansson et al. sequenced all coding exons, flanking intronic sequences and the promoter region in 10 AD cases and 10 controls and found 3 ploymorphisms (EX6+7I/D (rs3212319), rs2456777 and rs2456778), all within 74 bp of each other. We genotyped these three SNPs and three additional SNPs to cover the region. These SNPs demonstrated LD with each other (D' varies from 0.39 ~ 0.73). Although the CDC2 locus is an obvious functional and positional candidate locus for AD, the detailed genetic study presented here did not find evidence for association with Alzheimer disease. Our study has the advantage of having both a large set of unrelated cases and controls and a large, well-characterized family sample. To explore our statistical power to detect allele frequency differences between cases and controls, we estimated the power of our datasets. Given the odds ratio of 1.78 (95% CI 1.18-2.68) and the probability of exposure of 0.35 in controls (Johansson et al., 2003), we had 99%

power to detect the disease-associated variation. Given our sample size, we still had 80% power to detect the variation even when the OR=1.38.

Although these six SNPs are in LD with each other, there are no obvious LD blocks. We cannot exclude the possibility that there are variants (rare or common) associated with AD in regulatory elements outside the CDC2 coding region and that are not in strong LD with any of the SNPs that we have genotyped. The detection of such alleles, particularly when they are of low frequency, remains a challenge for molecular genetic studies. However, even if this is the case, it does not explain the discrepancy of our results with the earlier report. Although the gene showed the effect on age-at onset at p=0.01, the effect became non-significant after the confounders were considered.

In conclusion, we did not find the association with CDC2 polymorphisms in our large case-control and family datasets and this suggests that the variations in the CDC2 gene do not have a significant effect on the risk of LOAD. Although CDC2 may be pathophysiologically related to AD, the contribution of common genetic variants of this gene to the risk for developing AD is likely to be low in Caucasian Americans.

CHAPTER VII

EXAMINATION OF THE ASSOCIATION BETWEEN RS498055 IN LOC439999 GENE ON CHROMOSOME 10 AND ALZHEIMER DISEASE

Introduction

SNP rs498055 in the predicted gene LOC439999 on chromosome 10 was identified by Grupe et al. (Grupe et al., 2006) as strongly associated with late-onset Alzheimer disease (AD, MIM#104300). This SNP falls within a chromosomal region that has been implicated in several non-overlapping data sets and in studies that applied different analytical approaches, including linkage analysis using affected sib pairs, a genome screen using amyloid β levels as a quantitative trait, and a candidate gene-driven linkage analysis (Bertram et al., 2000; Blacker et al., 2003; Ertekin-Taner et al., 2000; Farrer et al., 2003; Kehoe et al., 1999; Myers et al., 2000).

Grupe and colleagues (Grupe et al., 2006) performed a screen of 1412 gene-based SNPs in 667 genes under the previously identified linkage peaks spanning 80 Mb of DNA on chromosome 10 using multiple case-control data sets. 69 SNPs reached significance ($p<0.05$) in the exploratory set. Although 5 out of 69 significant SNPs were replicated in at least one of the validation sample sets at $p<0.05$, only one SNP, rs498055 in LOC439999 (a homologue of RPS3A), was consistently significantly associated with Alzheimer disease, being significant in four out of six case-control replication sets (allelic $p=0.0001$ for a meta-analysis of all six samples, OR=1.3~1.4). Their conclusion was that variants in this RPS3A homologue, described above, were associated with late-onset

Alzheimer disease and that variants in this gene, or other functional variants close to it were likely to play a role in the pathogenesis of AD.

To independently evaluate this interesting candidate SNP, we designed an experiment using four independent data sets, three family-based and one case-control.

Materials and methods

Study populations

Four independent data sets were used in this study: (1) Families from the National Institute of Mental Health repository (NIMH) consisting of 1104 subjects from 349 pedigrees including 331 multiplex families, 488 affected sibling pairs, and 582 discordant sibling pairs. (2) Families from the Collaborative Alzheimer Project (CAP), which includes Duke and Vanderbilt Universities and University of California at Los Angeles, consisting of 755 subjects from 189 pedigrees including 84 multiplex families, 184 affected sibling pairs, and 370 discordant sibling pairs. (3) Families from the NCRAD repository at Indiana University (NCRAD) consisting of 470 subjects from 136 pedigrees including 100 multiplex families, 165 affected sibling pairs, and 216 discordant sibling pairs. (4) Sporadic cases and controls ascertained by Duke and Vanderbilt University consisting of 587 cases and 1101 controls that had substantial clinical information.

Our NIMH and NCRAD data sets overlapped with one of the data sets used by Grupe et al. They generated a case-control series by selecting one case per family from the genetic linkage sample (which included NIMH and NCRAD samples, (Myers et al., 2002)) and matching each of them to an equal number of white, nondemented controls

collected in St. Louis. But our family-based datasets used unaffected sibpairs and the family-based association was tested.

All the cases were late-onset AD (LOAD) patients (minimum age at onset (AAO) ≥ 60 years). AD was diagnosed according to the NINCDS-ADRDA criteria (McKhann et al., 1984b). Controls had no obvious signs of cognitive or neurological impairment when enrolled in the study as determined by personal interview by clinical personnel. All the age-matched controls had results within the normal range in the Mini-Mental State Exam (MMSE; ≥26) or Modified Mini-Mental State Exam (3MS; ≥85). All participants were Caucasian Americans.

Age at onset was recorded as that age at which the first symptoms were noted by the participant or a family member. Mean AAO ± SD (standard deviation) in affected individuals in the family-based sample was 72.9±6.4 years. The mean age-at examination (AAE) ± SD was 80.1±7.1 and 69.9±11.2 in affected and unaffected individuals, respectively. The control group's mean AAE ± SD was 72.0±6.3 years, while the case group's mean AAE ± SD was 76.5±6.53 years.

Genotyping

We designed specific probes to detect rs498055 using Taqman technology (Assay-by-Design) and genotyping was done on the four data sets. Systematic genotyping errors were minimized by use of a system of quality control (QC) checks with duplicated samples (Rimmler et al., 1998). The deviations from Hardy-Weinberg equilibrium (HWE) were assessed using Haploview (Barrett et al., 2005).

Family-based data set analysis

Family based association analysis was conducted using the pedigree disequilibrium test (PDT, for allelic effect) and Genotype-PDT (GenoPDT, for genotypic effect) for single-locus tests to assess association between genotypes and risk of AD in the family data (Martin et al., 2000b; Martin et al., 2002). Two-point heterogeneity LOD scores (HLODs) were computed using FASTLINK and HOMOG (Ott, 1999). Because the mode of inheritance for AD is unknown, affected-only parametric analyses were performed using both autosomal dominant and autosomal recessive models with disease allele frequencies of 0.001 and 0.20, respectively, to model the susceptibility allele. We applied Ordered Subset Analysis (OSA) (Hauser et al., 2004) to our linkage analysis for the combined data set to see if a specific subset of families might be linked. Families were ranked based on ApoE LOD score at theta=0 (low to high or high to low) and family-specific APOE-4 allele weights (low to high or high to low). OSA analysis was also performed using mean age of onset as covariate (low to high or high to low) to test for a subset generating a significantly increased LOD score relative to the overall sample.

Case-control data set analysis

Case-control association analyses for single alleles and for genotypes were conducted using logistic regression (SAS Institute, Cary, N.C., USA, version 8.1). Statistical significance was declared at $\alpha=0.05$. Gender, age, and ApoE status were used as covariates in the model to adjust for confounders.

Results

The reliability of duplicate genotyping across plates was >99% and the average genotyping efficiency was 95%. Genotypes for affecteds and unaffecteds did not violate Hardy-Weinberg Equilibrium.

Family based association was conducted using the pedigree disequilibrium test (PDT) for single-marker tests (Martin et al., 2000b) and genotype-PDT (GenoPDT). None of the three family samples nor the combined samples showed association in either allelic or genotypic association tests at $p<0.05$ (Table 7-1).

Two-point heterogeneity LOD scores (HLODs) were computed using FASTLINK and HOMOG (Ott, 1999). However, none of the data sets gave LOD scores >1.0 (Table 7-1).

Table 7-1 Association and linkage results for family-based sample

data set	Association analysis p value		Linkage analysis 2-point HLOD		HWE p value	
	PDT	Geno-PDT	Dominant	Recessive	Affected	Unaffected
NIMH	0.495	0.791	0.11	0.07	0.897	0.495
CAP	0.957	0.452	0.00	0.02	0.710	0.348
NCRAD	0.289	0.623	0.00	0.00	0.807	0.333
Combined	0.356	0.660	0.00	0.00	0.998	0.617

NIMH = National Institute of Mental Health repository; CAP = Collaborative Alzheimer Project; NCRAD = NCRAD repository at Indiana University; HWE = Hardy-Weinberg Equilibrium

In the Ordered Subset Analysis, there is no evidence indicating linkage no matter what covariates were applied. The highest LOD score was 0.82 (using mean age-of-onset as a covariate to order families from low to high, 5% of the families, the p value for the increase of the LOD score was 0.06).

Case-control association analyses were conducted for both single alleles and genotypes. The allelic association (OR=0.95) was not significant with p=0.52, 95% CI of 0.81-1.11. Analysis of genotypic association did not show significant result either (Table 7-2).

Table 7-2 Allelic and genotypic association for case-control samples

association	case	control	total	Freq.	test	OR	95% CI	P
allelic								
G	450	893	1343	0.49	G vs. T	0.95	0.81-1.11	0.52
T	486	917	1403	0.51				
genotypic								
GG	109	225	334	0.24	GG vs. TT	0.90	0.66-1.24	0.53
GT	232	443	675	0.49	GT vs. TT	0.98	0.75-1.28	0.87
TT	124	237	364	0.27				

The chi-square test for trend was p=0.80, χ^2=0.44. To adjust for potential confounders, we included gender and age at examination (AAE), and ApoE status (ApoE 4-positive and ApoE 4-negative) as covariates in the regression analysis. However, none of these tests revealed any evidence of association between this SNP and AD (all p>0.53). It is worth noting that we see an under-representation of G allele of rs498055 in our data set, which is in opposite direction of Grupe's finding but consistent with another recent report (Bertram et al., 2006).

Discussion

We have conducted a comprehensive analysis of the association between rs498055 and Alzheimer disease risk in both case-control and family-based samples. Our study has the advantage of having both a large set of unrelated cases and controls and three well-characterized family samples. In one of the family samples (NIMH), a large proportion of subjects overlapped with one of the data sets used in the initial study identifying linkage to this region (Myers et al., 2002). This makes it suitable to detect true effect under linkage peak. Grupe et al. also used this data set, but study design was different from ours. They selected one case per family and matched the cases with an unrelated controls collected in St.Louis. Although it is not easy to calculate the exact overlap between our data set and that of Grupe et al., there were more discordant sibpairs in our data set and we used the family-based association test and eliminated the possible population substructure in the data set.

The discrepancy of our results with the earlier report could be explained by several reasons. First, the LD patterns around the SNP rs498055 may be different among different sample populations. However, we found no evidence of a difference between our data sets and the HapMap data.

Second, our power could be low. Although the frequency of the risk allele of rs498055 is high (46.5% in control population), its estimated effect size is small (OR=1.26, 95%CI (1.02-1.54), and the initial report failed to find replication in two small case-control series (183 cases/127 controls; 160 cases/106 controls) (Grupe et al., 2006). They claimed that the power to detect the effect in these two small samples was low (40% and 36% respectively). In their recent response to a paper which failed to replicate this

SNP (Bertram et al., 2006), they estimated that 360 cases and 360 controls were needed to achieve 80% power in replication study to detect this small risk effect (Grupe et al., 2006). Our case-control data set has 587 cases and 1101 controls. Using the QUANTO program (http://hydra.usc.edu/gxe), we explored our statistical power to detect allele frequency differences between cases and controls. With one-sided test, we have 94% power to detect the disease-associated variation given the odds ratio of 1.26 (95% CI 1.02-1.54) and the probability of exposure of 0.46 in controls under the additive model. The power is 64% and 61% under dominant and recessive model, respectively.

Third, there could be a difference in the underlying samples. However, both cases and controls in our data sets are well characterized using standard criteria similar to that of the original report. Furthermore, all of our AD patients were late-onset AD with age-of-onset greater than 60. Fourth, our negative results may be due to chance alone. However, our lack of replication in any of four data sets argues against this.

Our results are consistent with the Bertram et al. findings that used family-based association tests to examine the effect of this SNP in two family data sets. Thus we conclude that RS498055 is not associated with an increased risk of LOAD.

CHAPTER VIII

GENOMIC CONVERGENCE TO IDENTIFY CANDIDATE GENES FOR ALZHEIMER DISEASE

Introduction

A promising solution to the growing data flood from the latest genomic technologies (e.g. Genome-wide association) is genomic convergence. Genomic convergence is a multifactor approach used in genetic research that combines different data analysis results from different data types to identify and prioritize susceptibility genes for a complex disease (Hauser et al., 2003b). Examples of such data types include genetic linkage data, association data, gene expression data, and biological function.

Our previous linkage and candidate gene studies suggested a more extensive heterogeneous genetic background on chromosome 10 for AD than we previously appreciated. Rare and common polymorphisms in multiple genes can be involved together with non-genetic factors in this complex disease. Therefore, combining different sources of data is important to selecting and prioritizing candidate genes.

In the present study, gene expression levels in the brain tissue from AD patients and controls were compared using the Serial Analysis of Gene Expression (SAGE) method (Y-J Li, personal communication). We used this Alzheimer disease gene expression study by SAGE (Li et al., 2006) as one of the important sources of data, combined with previously identified linkage data, to select candidate genes that modify the risk for AD.

SAGE is a powerful method for profiling transcripts expressed in a given tissue (Velculescu et al., 1995) and provides expression data for thousands of gene products without *a priori* knowledge of their function. In this technique, mRNAs are isolated and transcribed into cDNAs that are cleaved into short fragments called "tags" by restriction enzymes. The tags are extracted and linked together into long concatamers and sequenced. The output of SAGE is a list of short sequence tags and the number of times it is observed, which reflects the abundance of the respective mRNA level in the tissue.

Our collaborators at Duke University Center for Human Genetics (CHG) performed the SAGE analysis. We chose brain samples from AD cases (ApoE 4/4, ApoE 3/4 and ApoE 3/3 genotypes) and controls (ApoE 3/3 genotypes) to establish SAGE libraries to analyze the differential gene expression between AD cases and controls. Then we converged our linkage screen results with the gene expression data to identify genes that are under the linkage peak and differentially expressed between patients and controls. The LD structure of each genomic convergence genes was assessed.

Materials and methods

Human brain samples

Human brain tissues were collected in the Kathleen Price Bryan Brain Bank, at the Duke University Alzheimer Disease Research Center, and the Brain Bank of the Center for Human Genetics (CHG), Duke University Medical Center, following a rapid autopsy protocol (Hulette et al., 1997). The hippocampus is heavily involved in the neuropathology of AD and was selected for both neuropathological and molecular

analysis (Markesbery, 1997). In this study, the hippocampus was dissected at the time of autopsy and post-mortem delay times ranged from 1:10 to 4:15 hours (Xu et al., 2005b). Two brain tissue samples for each of the three ApoE genotypes for AD cases (ApoE 4/4, ApoE 3/4, and ApoE 3/3) and controls (ApoE 3/3) were used. The pathological diagnosis of AD was established according to CERAD criteria (Mirra et al., 1991), and the degree of AD pathological changes was staged according to Braak (Braak and Braak, 1991). The AD patients used in this study had pathological changes at the Braak and Braak (B&B) stage IV and V (B&B stage IV and V), and the controls were cognitively and pathologically normal with B&B stage I.

Construction of SAGE libraries

RNA was isolated using a standard protocol (Invitrogen) and SAGE libraries were constructed as described by Velculescu et al (Velculescu et al., 1995). Six SAGE libraries were generated. Four of those were based on the short tag (14 bp, AD ApoE 4/4, ApoE 3/4, ApoE 3/3 and control ApoE 3/3). Two libraries were based on the long tag (21-22 bp, AD ApoE 3/3 and control ApoE 3/3). Briefly, cDNA was reverse transcribed from mRNA and cut into short fragments (called "tags") by *Nla*III and *Bsm*FI restriction enzymes (New England Biolab, Beverly, MA). The tags were then concatamerized into long chain and transformed in competent Electromax DH10B cells (Invitrogen). Individual SAGE library clones were selected and PCR amplified using 96-well format Qiagen Real minipreps, and sequenced with ABI 3700 capillary sequencer using BigDye chemistry (Li et al., 2006).

SAGE data analysis

The number of times a tag observed in tissue was extracted from each library and compared between the AD and control samples using eSAGE software to form a compared ShortSAGE database. Chi-square and Fisher exact tests, as previously described (Hauser et al., 2003b), were used to test differences in expression levels between AD and control for each tag in each compared SAGE data set. The SAGE tags were compared to the UniGene to interpret the SAGE results.

Gene selection

SAGE revealed 222 unique tags that were differentially expressed at $p<0.05$ among different libraries. A False Discovery Rate (FDR) was applied to the SAGE data to correct for multiple comparisons. However, there were 77 unique tags showing significance above the FDR threshold. So we set $p<10^{-10}$ as the cutoff to select genes whose expression levels were highly significantly different between AD cases and controls. Then previous linkage study results, shown in Table 8-1, were converged with gene expression data so that only differentially expressed genes that were also under linkage peaks in previous linkage screens were selected for the next step association analysis.

Table 8-1 Summary of linkage study results on chromosome 10

Chr. Location	Linkage region (Mb)	Peak LOD score	Dataset	Group
10p	5-25	3.00	50ca/50co	*Zubenko et.al*
10q	45-75	4.10	451 ASPs	*Myers, et. al*
10q	40-60	2.69	922 ASPs	*Current study*
10q	80-100	1.70	922 ASPs	*Current study*
10q	90-120	3.80	435 ASPs	*Bertram et.al*

SNP selection

When we selected SNPs as genetic markers, several criteria were applied. At first, we flanked the gene by 10kb on each side to include potential regulatory regions. Only tagSNPs were selected with the restriction that the linkage disequilibrium r^2 between selected SNPs should be less than 0.7. The web-based program, SNPselector (Xu et al., 2005a), was used to automatically search available SNPs through the Ensembl and UCSC databases. It prioritizes these SNPs on their tagging for linkage disequilibrium based on the LD bin algorithm in LD select (Carlson et al., 2004) using HapMap data, SNP allele frequencies and source, function, regulatory potential and repeat status. Because we used the Illumina Goldengate genotyping method, we only selected SNPs with Illumina SNPscore greater than 0.6 to ensure the genotyping quality. The Illumina SNPscore reflects the ability to design a successful Oligo Pool Assay. The possible reasons for design failure of a SNP include melting temperature (T_m) exceeding assay limit, SNP in duplicated/repetitive region, the presence of another SNP within 60 bps, tri- or quad-allelic SNPs and insertion/deletion SNPs (www.illumina.com).

Minor allele frequency (MAF) is another important criterion as higher MAF SNPs are more informative, but we also want to include more coding SNPs, and these tend to have low MAF. Therefore, MAF greater than 20% was used as the cutoff for selecting the intronic SNPs. SNPs in conserved regions of eight-genome alignment (Human/chimp/mouse/rat/dog/chicken/fugu/zebra_fish, based on UCSC database), conserved transcription factor binding sites (TFBSs) in the human/mouse/rat alignment, CpG island, and microRNA genes were selected if the MAFs were greater than 10%. Coding SNPs were selected when the MAF was greater than 1%. With these criteria,

there were still many SNPs to choose from. Therefore, SNPs were selected at the 1 SNP/10 kb spacing. The gene TNFRSF6 showed significant results in linkage, candidate gene studies and SAGE gene expression analysis (Bertram et al., 2000; Feuk et al., 2000). Therefore, more weight was given to it by genotyping one SNP per five kilobases.

Genotyping

Selected 667 SNPs on chromosome 10 were genotyped as part of the 1536 SNPs multiplex panel using Illumina GoldenGate Oligo Pool Assay (OPA) on an Illumina BeadStation 500 GX (Illumina Inc., San Diego, CA). The GoldenGate assay uses an allele-specific primer extension (ASPE) gapped ligation assay, in which query oligos annealed to the solid-phase genomic DNA were extended and ligated to form an amplifiable substrate for a subsequent universal primer polymerase chain reaction (PCR). The amplified substrates were then hybridized and detected by a universal array. DNA samples from cases and controls were randomly sorted and duplicated samples were implemented across plates for genotyping quality control.

Hardy-Weinberg Equilibrium (HWE) and linkage disequilibrium test

We tested all SNPs for consistency with HWE in Haploview program (Barrett et al., 2005). Haplotype blocks were assessed on AD cases using the default algorithm of 95% confidence intervals on D’ (Gabriel et al., 2002) in Haploview. A block was created if 95% of informative (e.g. non-inconclusive) comparisons are in “strong LD”. TagSNPs were selected at different linkage disequilibrium levels using the Tagger option

incorporated in Haploview for the future haplotypic association test and gene-gene interaction test.

Results

Genomic convergence identified genes

SAGE analysis revealed 41 unigenes on chromosome 10 that were differentially expressed between AD cases and controls with $p<10^{-10}$. 28 of those genes were under previously identified linkage peaks (Bertram et al., 2000; Myers et al., 2002; Pericak-Vance et al., 2000; Zubenko et al., 1998). The genomic size of the genes ranged from 10 kb to 1.3 Mb. The number of exons ranged from 2 to 38. According to the size of the gene, we genotyped 2~124 SNPs in the genes to cover the genes at an average spacing of 1 SNP/10kb. Table 8-2 shows the gene description, size and number of genotyped SNPs.

Table 8-2 Genomic convergence identified genes on chromosome 10

Gene	Description	Position*	Size (bp)	Exons	Genotyped snps
ACTR1A	ARP1 actin-related protein 1 homolog A, centractin alpha (yeast)	104,228,976	23,482	11	3
BA108L7.2	sideroflexin 3 (SFXN3)	102,780,981	10,007	12	5
CACNB2	Calcium channel, voltage-dependent, beta 2 subunit	18,469,672	400,372	14	41
CAMK1D	Calcium/calmodulin-dependent protein kinase ID	12,431,487	486,064	11	39
CNNM2	Cyclin M2	104,668,061	160,273	8	15
CUL2	Cullin 2	35,337,485	82,091	22	8
CWF19L1	CWF19-like 1, cell cycle control (S. pombe)	101,982,045	35,382	14	3
EIF4EBP2	Eukaryotic translation initiation factor 4E binding protein 2	71,833,928	19,747	3	15
EPC1	Enhancer of polycomb homolog 1 (Drosophila)	32,596,731	110,954	15	11
HELLS	Helicase, lymphoid-specific	96,295,537	68,115	22	6
HNRPH3	heterogeneous nuclear ribonucleoprotein H3	69,760,937	12,017	4	2
IFIT3	Interferon-induced protein with tetratricopeptide repeats 3	91,077,733	12,534	2	2
MAWBP	MAWD binding protein	69,712,423	50,260	10	7
NT5C2	5'-nucleotidase, cytosolic II	104,837,764	105,281	17	9
PDZK7	PDZ domain-containing protein 7	102,757,576	23,304	15	2
PNLIP	Pancreatic lipase	118,295,433	21,924	13	5
PNLIPRP1	Pancreatic lipase-related protein 1	118,339,925	18,752	13	3
PPP3CB	Protein phosphatase 3 catalytic subunit, beta isoform	74,866,192	59,573	14	7
PRKG1	Protein kinase, cGMP-dependent, type I	52,421,124	1,306,992	18	124
PTPLA	Protein tyrosine phosphatase-like, member a	17,671,964	27,418	7	2
RPS24	Ribosomal protein S24	79,463,608	21,087	7	5
SORBS1	Sorbin and SH3 domain containing 1	97,061,518	249,643	31	64
SORCS1	Sortilin-related VPS10 domain containing receptor 1	108,323,411	590,863	27	54
SVIL	Supervillin	29,786,273	279,443	38	97
TCF7L2	transcription factor 7-like 2	114,700,201	217,226	8	38
TNFRSF6	Tumor necrosis factor receptor superfamily member 6	90,740,394	25,128	9	26
TXNL2	Thioredoxin-like 2	131,824,655	43,975	12	3
USP54	ubiquitin specific protease 54	74,927,302	19,979	5	1

* Gene location was based on NCBI build 35

Quality control of genotyped SNPs

667 SNPs in 28 genes from chromosome 10 convergence region were genotyped using the Illumina GoldenGate genotyping assay. Other 869 SNPs were in the genes on other chromosomes identified using genomic convergence by collaborators at Duke. Quality-control was performed using test of HWE and Illumina assay quality control criteria. In the GoldenGate assay, there were built in controls that measure allele-specific extension, PCR uniformity, extension gap, gender, first hybridization controls, second hybridization controls, and contamination detection controls. These control oligos (with the exception of second hybridization controls) are designed to human specific genomic DNA sequences. Each oligonucleotide (bead type) represents a specific SNP locus. Since each bead type is present an average of 30 times on each array, each Goldengate genotyping call is thus the result of the mean intensity of all 30 replicates. The consistency among duplicates also serves as a quality control criterion.

Table 8-3 shows the distribution of the quality control results of the 667 SNPs in the AD control group. 70 SNPs were dropped from the analysis because of monomorphism, out of HWE, or assay problems. Specifically, 58 SNPs failed Illumina quality control criteria, 10 SNPs were monomorphic and 2 SNPs were out of HWE in controls at $p<10^{-5}$ (Bonferroni correction for 667 markers at p=0.05). If we apply a False Discovery Rate (q=5%) to correct for multiple tests, these two SNPs were also excluded. In total, there were 597 SNPs that remained in the final analyses (Table 8-3).

Table 8-3 Quality control of 667 genotyped SNPs in the AD control group

	Count	Analyzed
Genotyped	667	
Failed Illumina_QC	58	
MAF=0	10	
Hardy-Weinberg Equilibrium (HWE) $p<10^{-5}$	2	597
0.001 < HWEp < 0.01	13	
0.01 < HWEp < 0.05	28	
0.001 < MAF < 0.005	1	

Figure 8-1 shows the distribution of the genotyped SNPs. 80% are intronic SNPs, 17% are functional SNPs (exon, promoter, UTR, exon/intron boundary, et al.) and the other 3% are from the 5'-upstream and 3'-downstream regions.

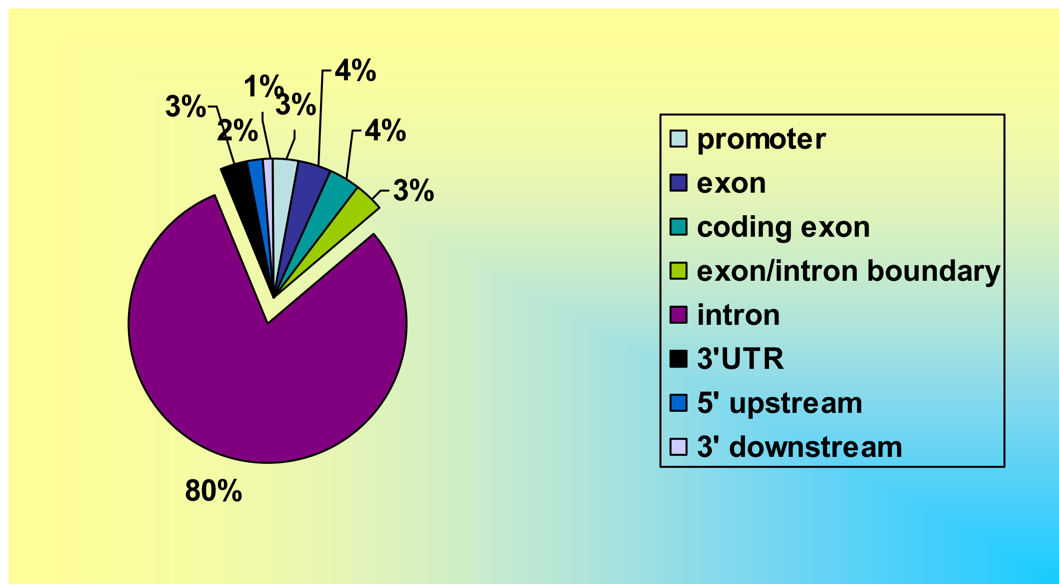


Figure 8-1 Functional classification of genotyped SNPs

Linkage Disequilibrium patterns of the 28 genes in the overall data set

The 597 SNPs were distributed across 28 genes. Linkage disequilibrium within genes was tested and Table 8-4 shows the number of blocks and tag SNPs in each gene. Genes with fewer than 4 genotyped SNPs were not analyzed for haplotype blocks. Tag SNPs were evaluated at four different linkage disequilibrium level (r^2>0.8, 0.7, 0.64, and 0.5).

Table 8-4 Haplotype blocks and tag SNPs in each gene

gene	Size (bp)	# exons	# genotyped SNPs	# blocks	# tag SNPs			
					r^2>0.8	r^2>0.7	r^2>0.64	r^2>0.5
USP54	19,979	5	1	-	-	-	-	-
HNRPH3	12,017	4	2	-	-	-	-	-
IFIT3	12,534	2	2	-	-	-	-	-
PDZK7	23,304	15	2	-	-	-	-	-
PTPLA	27,418	7	2	-	-	-	-	-
ACTR1A	23,482	11	3	-	-	-	-	-
CWF19L1	35,382	14	3	-	-	-	-	-
PNLIPRP1	18,752	13	3	-	-	-	-	-
TXNL2	43,975	12	3	-	-	-	-	-
BA108L7.2	10,007	12	5	1	5	4	4	4
PNLIP	21,924	13	5	1	2	2	2	2
RPS24	21,087	7	5	1	2	2	2	2
HELLS	68,115	22	6	1	1	1	1	1
MAWBP	50,260	10	7	1	6	5	5	4
PPP3CB	59,573	14	7	1	5	5	4	4
CUL2	82,091	22	8	1	3	3	3	3
NT5C2	105,281	17	9	1	3	2	2	2
EPC1	110,954	15	11	1	4	4	4	4
CNNM2	160,273	8	15	1	4	4	3	2
EIF4EBP2	19,747	3	15	3	11	11	9	7
TNFRSF6	25,128	9	26	2	10	9	9	8
TCF7L2	217,226	8	38	10	29	27	26	25
CAMK1D	486,064	11	39	9	36	35	35	30
CACNB2	400,372	14	41	8	37	36	36	31
SORCS1	590,863	27	54	14	38	34	30	20
SORBS1	249,643	31	64	14	48	40	37	32
SVIL	279,443	38	97	17	54	47	45	37
PRKG1	1,306,992	18	124	30	90	79	76	64

Example of haplotype blocks in genes

1. Nine genes were genotyped by 1-3 SNPs, so the haplotype block structure was not assessed. These genes are USP54, HNRPH3, IFIT3, PDZK7, PTPLA, ACTR1A, ACTR1A, CWF19L1, and TXNL2.

2. Ten genes have only 1 haplotype block. An example of such gene (HELLS) is given in Figure 8-2. The other genes are BA108L7.2, PNLIP, RPS24, MAWBP, PPP3CB, CUL2, NT5C2, EPC1, and CNNM2.

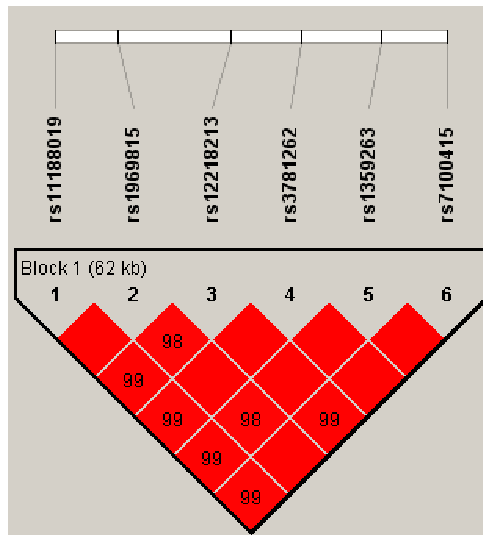


Figure 8-2 Example of the linkage disequilibrium structure of genes with one haplotype block (eg. HELLS)

3. 26 SNPs were genotyped in TNFRSF6 and there are 2 haplotype blocks in the gene (Figure 8-3).

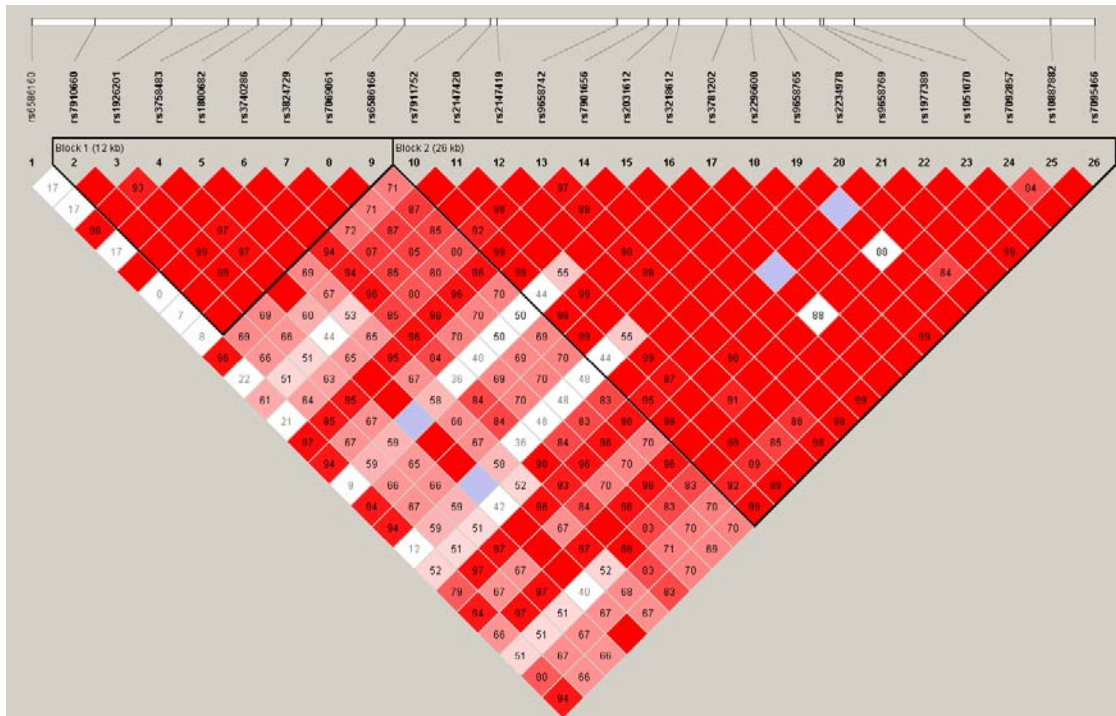


Figure 8-3 Linkage disequilibrium structure of TNFRSF6 gene

4. 15 SNPs were genotyped in EIF4EBP2 and there are 3 haplotype blocks in the gene, Figure 8-4.

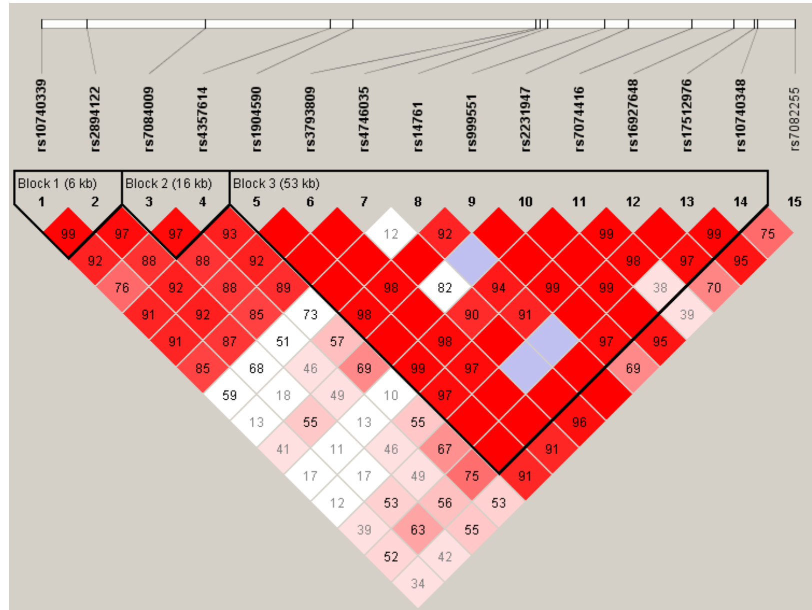


Figure 8-4 Linkage disequilibrium structure of EIF4EBP2 gene

5. 7 other genes have more than 8 haplotype blocks. Figure 8-5 shows an example (SORCS1). Other genes are TCF7L2, CAMK1D, CACNB2, SORBS1, SVIL, and PRKG1.

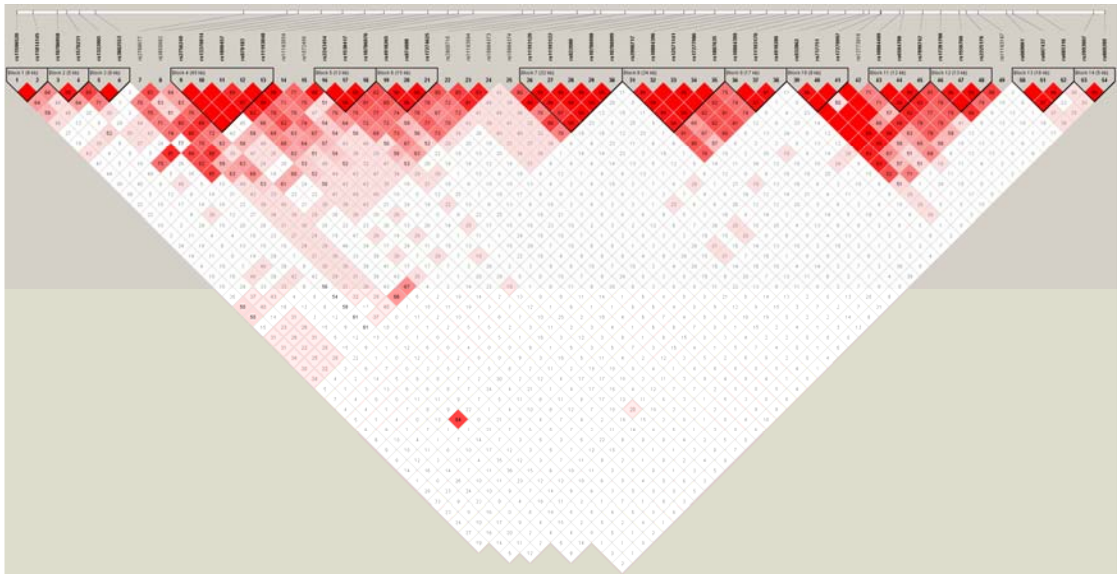


Figure 8-5 Example of linkage disequilibrium structure of genes with more than eight haplotype blocks (eg. SORCS1)

Discussion

While both linkage and expression analyses are powerful methods when applied individually, the number of possible genes they present as candidates for AD or any complex disorder remains extremely large. Thus, focusing and prioritizing effort on jointly identified candidate genes is a key to success using these techniques (Hauser et al., 2003b). By converging these two sources of data, Hauser and colleagues successfully identified candidate genes for Parkinson's disease. Similarly, Li *et al*. (Li et al., 2003) identified the glutathione S-transferase omega-1 (GSTO1) as a gene that modifies the

age-at-onset of Alzheimer disease and Parkinson's disease by converging linkage and Serial Analysis of Gene Expression (SAGE) data.

The general goal of the technique is similar to a DNA microarray approach. Unlike microarray technology, which is limited to a finite number of known gene sequences arrayed on a chip, SAGE can in theory detect all genes expressed in a tissue sample. In addition, SAGE provides more quantitative information than microarray analysis. Because the mRNA sequences do not need to be known *a priori*, genes or gene variants that are not known can be discovered. SAGE is a sequence-based sampling technique and as such observations are not based on hybridization, which gives more qualitative, analog values. However, due to the difficulty of library preparation, sequencing cost, time and labor, the total number of SAGE libraries produced for a study is generally small. Large-scale studies do not typically use SAGE.

We compared six SAGE libraries, and evaluated the gene expression level between AD cases and controls. 222 unique tags on chromosome 10 were differentially expressed at nominal $p<0.05$. Seventy-seven of those tags representing seventy-seven unigenes were significantly different between AD cases and controls after FDR correction for multiple testing. Forty-one of these genes were highly significantly different at $p<10^{-10}$. Interestingly, when we compared the direction of gene expression between AD and controls, we found that all but one gene were up-regulated in AD the ApoE 4/4 and ApoE 3/4 samples compared to the ApoE 3/3 samples. It is possible that the ApoE ε4 allele elevates the expression of other genes. The only down-regulated gene was BA108L7.2 encoding sideroflexin 3. The difference for this gene was found when we compared the long tag SAGE library for AD to control ApoE 3/3 samples.

In the genomic convergence approach, it was suggested that whether a gene was over- or under-expressed in the disease tissue was not important (Li et al., 2003). Rather, importance is placed on whether a gene has any significant difference in expression levels, suggesting it has the potential to be involved in the disease process. Genes demonstrating expression changes between the cases and controls may be involved in the primary or secondary pathways of the disease. One of the great challenges with expression analyses is the sheer number of genes demonstrating differences and the complex interactions that lead to these differences. In our approach, 28 candidate genes with significant expression differences between AD cases and controls and within previously determined linkage regions have an increased likelihood to alter the risk for AD. These 'converged' genes were then tested for their association with AD.

These twenty-eight genes varied in genomic size, number of exons and hyplotype blocks. The smallest gene was BA108L7.2 that was 10kb and 12 exons. The biggest gene was PRKG1 (protein kinase, cGMP-dependent, type 1) that was 1.3 Mb with 18 exons. To cover this gene at a 1SNP/10kb resolution, 124 SNPs were genotyped. The haplotype structures were assessed based on our genotyping data and the haplotype blocks varied by genes. We did not evaluate nine genes in which there were less than four genotyped SNPs (USP54, HNRPH3, IFIT3, PDZK7, PTPLA, ACTR1A, ACTR1A, CWF19L1, and TXNL2). Ten genes (BA108L7.2, HELLS, PNLIP, RPS24, MAWBP, PPP3CB, CUL2, NT5C2, EPC1, and CNNM2) had one single haplotype block across the genotyped SNPs with different number of TagSNPs based on different LD levels. The biggest gene, PRKG1, had 30 haplotype blocks based on 124 genotyped SNPs. There were 64 TagSNPs even at r^2>0.5 level.

The LD structures of the genes were similar to the HapMap data. The number of LD blocks for the small genes were same between our data and HapMap data. The level of LD between the two data sources was also close. Because the numbers of the genotyped SNPs across medium- to large-size genes were different between our data (a 1SNP/10kb spacing was used) and HapMap data (large amount of SNPs were genotyped), the number of LD blocks and LD levels were different. For example, 1113 SNPs were genotyped in SORCS1 gene (590.9 kb) and 25 blocks were defined in HapMap data. In our data set, we genotyped 54 SNPs and defined 14 haplotype blocks. This might be due to the high LD between HapMap SNPs. When we selected SNPs, we applied an LD bin algorithm based on HapMap data, in order to capture more LD blocks using fewer SNPs. Another potential explanation for the difference of the haplotype blocks between our data and HapMap data is that we used genotyping data from AD cases to generate the LD structure. It is likely that the haplotype structures we observed here were Alzheimer disease haplotypes. Our data could provide useful information for genotyping or haplotyping of these genes in Caucasian AD population.

CHAPTER IX

ASSOCIATION ANALYSIS OF GENOMIC CONVERGENCE GENES FOR ALZHEIMER DISEASE

Introduction

We applied the genomic convergence approach to identify twenty-eight genes that were under linkage peaks and also significantly differentially expressed between AD cases and controls using a SAGE analysis, as described in chapter VIII. These genes were genotyped and examined for association with AD. We used genetic association analysis as our approach because it overcomes the limited power of linkage analysis to detect moderate effects in small regions (Cardon and Bell, 2001; Risch and Merikangas, 1996; Risch, 2000; Tabor et al., 2002).

The Linkage analysis approach for mapping genes aims to identify a genetic marker that is physically connected on the same chromosome with disease locus at a distance that is measured at less than a 50% recombination rate. A marker is linked with the disease locus when it fails to be transmitted to offspring independently from the disease locus. Due to the late onset of the Alzheimer disease (so that is usually impossible to genotype parents), genotyping affected sibpairs is an appropriate design to map the genes altering the risk of Alzheimer disease. If the two affecteds in the sibpair have inherited the same region of the chromosome from each parent, the region is more likely to be involved in the disease than if each sibling inherits different regions. However, there is one significant limitation of linkage mapping approach—the resolution is low. The linked region could be 10-30 megabases long in complex diseases and there

are generally a number of locational candidate genes in the area. Another limitation is that the linkage analysis approach is most powerful to detect regions involved in monogenic 'Mendelian' and highly-penetrant (generally rare due to negative selection) diseases (Jimenez-Sanchez et al., 2001; Pritchard, 2001; Reich and Lander, 2001). However, linkage analysis of common, multifactorial diseases has been limited by the lack of clear genetic segregation of any DNA variants in multigenerational family material, and by the modest contribution to disease made by individual genetic variants (Altmuller et al., 2001; Cardon and Bell, 2001).

Association analysis, or linkage disequilibrium mapping, is the other approach for mapping genes that either cause or increase susceptibility to human disease. It is a nonparametric approach to look for a significantly increased or decreased frequency of a marker allele, genotype, or haplotype with a disease trait over what would be expected by chance if there were no association between marker and phenotype. Therefore the resolution is much higher (Collins, 1995; Lander and Schork, 1994; Risch and Merikangas, 1996). With the Human Genome Project, HapMap project and the high-throughput technology to genotype hundreds of thousands of alleles in parallel, association studies play a critical role in the analysis of complex traits. However, association analysis, using population based case-control data, is sensitive to population substructure that family-based studies are not as sensitive to. In addition, many different mutations in a gene might lead to a disease. Allelic heterogeneity might dilute the effect of each mutation by the presence of the others in a population-based association studies. However, to detect low-penetrance alleles in complex disease, association studies, especially the genome-wide association studies supported by the high-throughput

techniques are becoming more widespread. In this study, both allelic and genotypic association tests were applied for single-locus analysis to look for a significantly increased or decreased frequency of a SNP allele with disease status.

Most of the genome falls into segments of strong LD, within which variants are strongly correlated with each other, and most regions of chromosomes carry one of only a few possible haplotypes (Daly et al., 2001; Gabriel et al., 2002; Patil et al., 2001). Although the disease-predisposing alleles are typically unmeasured, we can use correlated marker genotypes as surrogates, which is the indirect gene mapping method. In this method, haplotypes may be better markers than single SNPs because haplotypes provide greater coverage than any single SNP. Statistical analysis based on haplotypes may be more efficient than separate analysis of individual markers in both simulation (Morris and Kaplan, 2002) and empirical studies (Drysdale et al., 2000; Martin et al., 2000a). To exploit the potential greater power of haplotype analysis, we also analyzed the haplotypic effect of the markers in AD in addition to the single marker association.

In addition, we did data simulation based on ninety genetic models before all the association analyses to evaluate the power of our current data set that contained 508 cases and 556 controls.

Materials and methods

Study populations

Case-control data set

The case-control data set had 1064 individuals (506 cases and 558 controls, Table 9-1) with substantial clinical information. The samples were collected by the Center for Human Genetics Research (CHGR) at Vanderbilt University and the Center for Human Genetics (CHG) at Duke University. The criteria were same as previously described (Liang et al. 2007). The cases were all late onset Alzheimer disease (minimum age at onset (AAO) $\geq$ 60 years), with average age-of-onset $\pm$ standard deviation of 72.5±6.3 years (Table 9-1). The average age-of-exam $\pm$ standard deviation was 75.9±12.1 and 74.3±5.9 years in cases and controls, respectively. 61.5% of the cases were females in cases and 60.4% of the controls were females. All controls were ascertained in the same catchment area as cases. Following informed consent, blood samples were collected from each individual. Genomic DNA was extracted from whole blood by use of the Puregene system (Gentra Systems, Minneapolis, MN). All samples were coded and stored at 4°C until used.

Table 9-1 Alzheimer disease case control data set

		Case	Control
Sample size (N)		506	558
Gender	Female	311 (61.5%)	337 (60.4%)
	Male	195 (38.5%)	221 (39.6%)
Age-of-Onset	All	72.5±6.3	N/A
	Female	72.8±6.6	N/A
	Male	72.1±5.9	N/A
Age-of-Exam	All	75.9±12.1	74.3±5.9
	Female	75.8±14.1	73.5±5.7
	Male	76.2±8.1	75.5±6.0

Alzheimer disease is a heterogeneous complex disease. To increase the homogeneity of the data set, the overall data set was subsetted according to four variables: Gender, ApoE status, Age-of-Onset and Diagnosis criteria. The distribution is shown in Table 9-2.

Table 9-2 Overall and subsets of study populations for Alzheimer disease

		Case	Control	Total
Gender	Female	311	337	648
	Male	195	221	416
ApoE	Positive (≥1 ApoE ε4 allele)	323	131	454
	Negative (no ApoE ε4 allele)	152	415	567
Age-of-onset (Age-of-exam in ctrl)	60-75	348	335	683
	≥76	158	223	381
Diagnosis criteria	Definite & probable AD	310	558	N/A
	Possible AD	196		N/A

Family-based data set

We also used independent family-based data sets as the validation sets to confirm any significant association identified in the case-control data set. All individuals included in this study were Caucasian late-onset AD (LOAD) patients (minimum age at onset (AAO) ≥ 60 years). Written consent was obtained from all participants in

agreement with protocols approved by the institutional review board at each contributing center. AD was diagnosed according to the NINCDS-ADRDA criteria (McKhann et al., 1984b). All controls had results within the normal range in the Mini-Mental State Exam (MMSE) or Modified Mini-Mental State Exam (3MS). Samples were ascertained by the following centers: the NCRAD repository at Indiana University (NCRAD); the Collaborative Alzheimer Project (CAP), including Duke and Vanderbilt Universities and University of California at Los Angeles; and the National Institute of Mental Health repository (NIMH). These family-based data sets were same as the data sets used in the previous studies. Table 9-3 shows the distribution of these three family-based data sets. Age at onset was recorded as that age at which the first symptoms were noted by the participant or a family member. If the affected individuals were early in the disease process, we included their report of age-at-onset as part of the determination. If the disease was more advanced, we only used information as collected from multiple family members (such as, spouse and children).

Table 9-3 Family-based data sets for Alzheimer disease

Family	Combined	NIMH	NCRAD	CAP
Total pedigrees	730	352	154	224
Affected individuals	1521	807	315	390
Unaffected individuals	974	331	162	481
Discordant Sib Pairs (DSP)	1337	629	269	439
Independent Discordant Sib Pairs	674	283	129	262
Pedigrees with at least one DSP	406	165	75	166

NIMH: the National Institute of Mental Health repository; NCRAD: the National Cell Repository for Alzheimer Disease at Indiana University; CAP: the Collaborative Alzheimer Project

Power study

Genetic models

We set up genetic models in which two disease loci underlie the phenotype, so that the models reflect the heterogeneity and complexity of complex disease better than single locus models. We simulated disease alleles as either minor alleles or major alleles, with allele frequencies for the disease locus (A) being 0.1 or 0.2 or 0.6. Allele frequencies for the disease locus (B) were 0.3 or 0.4 or 0.6. The combinations of the two disease loci in the models were A/B=0.1/0.4; 0.2/0.3; and 0.6/0.6. The 0.1/0.4 combination represented the cases in which one underlying disease allele was rare, the other was common. The 0.2/0.3 combination represented two common disease alleles and the 0.6/0.6 combination represented the case in which the two disease alleles were major alleles. The effect sizes of two disease loci were controlled by joint odds ratio (OR) of the two loci which varied from 1.1 to 2.0. Three inheritance models (dominant, recessive and additive) were applied to the simulation. Table 9-4 summarizes the parameters for the ninety genetic models. The prevalence of the disease was set as 10% to mimic the prevalence based on the age distribution of our current Alzheimer case-control data set.

Table 9-4 Genetic models

model	Inheritance Model	A allele freq	B allele freq	prevalence	OR
1~10	Dominant	0.1	0.4	0.1	1.1~2.0
11~20	Dominant	0.2	0.3	0.1	1.1~2.0
21~30	Dominant	0.6	0.6	0.1	1.1~2.0
31~40	Recessive	0.1	0.4	0.1	1.1~2.0
41~50	Recessive	0.2	0.3	0.1	1.1~2.0
51~60	Recessive	0.6	0.6	0.1	1.1~2.0
61~70	Additive	0.1	0.4	0.1	1.1~2.0
71~80	Additive	0.2	0.3	0.1	1.1~2.0
81~90	Additive	0.6	0.6	0.1	1.1~2.0

Penetrance calculation

Penetrance was calculated assuming that the prevalence of a disease was the sum of the number of people under exposure (e.g. have the risk allele(s)) who developed the disease and the number of people having the disease but not under the exposure. It is shown as the formula below.

$$\text{Prevalence} = \text{risk in population} * \text{geno.freq}_{\text{non-exposed}} + \text{penetrance} * \text{geno.freq}_{\text{exposed}}$$

Table 9-5 shows the genotype frequencies for each of the nine genotypes for 2-disease locus model. Letters A and B stand for the two risk alleles; letters a and b stand for the two non-risk alleles. The letter ‘q’ is the allele frequency of the disease allele.

q_1 = allele frequency of allele A; p_1 = allele frequency of allele a;

q_2 = allele frequency of allele B; p_2 = allele frequency of allele b.

Table 9-5 Genotype frequency table for 2-disease loci models

	AA	Aa	aa
BB	$q_1^2*q_2^2$	$2p_1q_1*q_2^2$	$p_1^2*q_2^2$
Bb	$q_1^2*2p_2q_2$	$2p_1q_1*2p_2q_2$	$p_1^2*2p_2q_2$
bb	$q_1^2*p_2^2$	$p_2^2*2p_1q_1$	$p_1^2*p_2^2$

In Table 9-6, x is the penetrance of the risk allele. The penetrance for each genotype (value in each cell) varies according to different inheritance models. The variable 'a' is the risk of getting Alzheimer disease in the general population without either one of the two disease loci. It is the baseline value of getting disease due to non-genetic risk factors.

Table 9-6 Penetrances for three inheritance models

Dominant model:

	AA	Aa	aa
BB	2x+a	2x+a	x+a
Bb	2x+a	2x+a	x+a
bb	x+a	x+a	a

Recessive model:

	AA	Aa	aa
BB	2x+a	x+a	x+a
Bb	x+a	a	a
bb	x+a	a	a

Additive model:

	AA	Aa	aa
BB	2x+a	1.5x+a	x+a
Bb	1.5x+a	x+a	x/2+a
bb	x+a	x/2+a	a

Given the notation of A and B being the disease alleles, the non-exposed and exposed genotypes are the following according to different inheritance models. In the dominant model, the non-exposed genotype is aabb; the exposed genotypes are AABB,

AABb, AAbb, AaBB, AaBb, Aabb, aaBB, and aaBb. In the recessive model, the non-exposed genotypes are AaBb, Aabb, aaBb, and aabb; the exposed genotypes are AABB, AABb, AAbb, AaBB, and aaBB. In the additive model, the non-exposed and exposed genotypes are same as dominant model with different penetrance for each genotype.

We set up the genetic models by controlling joint Odds Ratio (OR) as the joint effect size of the two disease alleles. The following is an example of joint OR of 2.0 for two disease alleles:

We assumed that we have ascertained 1,000 cases and 1,000 controls from the population. Given a certain combination of two disease allele frequencies, the numbers of people for each of the nine possible genotypes are known. However, the penetrances for different genotypes are not same. These can be calculated according to different inheritance models (Table 9-6). With this logic in mind, we applied a reverse process to calculate at which penetrance level, certain OR (e.g. 2.0) can be given under three inheritance models. The penetrance was then used in the data simulation using genomeSIM program. Tables 9-7, 9-8 and 9-9 show the joint and marginal ORs at different allele frequencies in three inheritance models.

Table 9-7 Marginal ORs at different joint ORs and disease allele frequencies in the dominant model

Allele frequency	Joint OR									
	2.00	1.90	1.80	1.70	1.60	1.50	1.40	1.30	1.20	1.10
A=0.1	1.59	1.55	1.50	1.46	1.40	1.35	1.29	1.23	1.16	1.08
B=0.4	1.75	1.68	1.61	1.54	1.47	1.40	1.32	1.25	1.17	1.08
A=0.2	1.59	1.54	1.50	1.45	1.39	1.34	1.28	1.21	1.15	1.08
B=0.3	1.63	1.58	1.53	1.47	1.41	1.35	1.29	1.22	1.15	1.08
A=0.6	1.40	1.38	1.34	1.31	1.28	1.24	1.20	1.15	1.11	1.06
B=0.6	1.40	1.38	1.34	1.31	1.28	1.24	1.20	1.15	1.11	1.06

Table 9-8 Marginal ORs at different joint ORs and disease allele frequencies in the recessive model

Allele frequency	Joint OR									
	2.00	1.90	1.80	1.70	1.60	1.50	1.40	1.30	1.20	1.10
A=0.1	1.89	1.81	1.72	1.64	1.55	1.47	1.38	1.29	1.19	1.10
B=0.4	1.98	1.88	1.79	1.69	1.59	1.49	1.39	1.30	1.20	1.10
A=0.2	1.91	1.82	1.74	1.65	1.56	1.47	1.38	1.29	1.19	1.10
B=0.3	1.94	1.85	1.76	1.67	1.57	1.48	1.38	1.29	1.19	1.10
A=0.6	1.66	1.60	1.55	1.49	1.43	1.36	1.30	1.23	1.16	1.08
B=0.6	1.66	1.60	1.55	1.49	1.43	1.36	1.30	1.23	1.16	1.08

Table 9-9 Marginal ORs at different joint ORs and disease allele frequencies in the additive model

Allele frequency	Joint OR									
	2.00	1.90	1.80	1.70	1.60	1.50	1.40	1.30	1.20	1.10
A=0.1	1.51	1.47	1.43	1.39	1.35	1.30	1.25	1.20	1.14	1.07
B=0.4	1.79	1.72	1.64	1.57	1.49	1.42	1.34	1.26	1.17	1.09
A=0.2	1.57	1.52	1.48	1.43	1.38	1.32	1.27	1.21	1.14	1.07
B=0.3	1.65	1.60	1.54	1.49	1.42	1.36	1.30	1.23	1.15	1.08
A=0.6	1.40	1.38	1.34	1.31	1.28	1.24	1.20	1.15	1.11	1.06
B=0.6	1.40	1.38	1.34	1.31	1.28	1.24	1.20	1.15	1.11	1.06

Table 9-10 summarizes the genetic models used in the power study and the criteria we normally use in the association study.

Table 9-10 Parameters for genetic models

Sample Size	500 cases, 500 controls
Markers	600
Replicates	1000
Models	2 disease loci (A/B): 0.1/0.4,0.2/0.3, 0.6/0.6
	Dominant, recessive, additive inheritance models
	Prevalence = 10%
	OR=1.1~2.0
Calculation	0.1= risk in pop. * $\text{geno.freq}_{\text{non-exposed}}$ + penetrance * $\text{geno.freq}_{\text{exposed}}$
Test	χ^2 (p = 0.05)
1-stage power	Pick up locus A, B, A or B, A and B

Data simulation

genomeSIM (Dudek et al., 2006) is a data simulation package for the simulation of large-scale genomic data in population based case-control samples. It allows for single SNP and gene-gene interaction models to be associated with disease risk. genomeSIM utilizes two different methods to generate data sets. An initial population can be generated on the basis of allele frequencies of the SNPs and then further generations are created by crossing the members of successive generations. The simulator assigns affection status only after a specified number of generations. Alternatively, the simulator can construct a case-control data set by generating individuals as above, assigning

affection status, and selecting cases and controls until the data set is complete. I used the second function of the genomeSIM to generate my data set. genomeSIM uses a penetrance table to assign the affection status of individuals. To determine status, the simulator determines the genotype of the individual at the disease SNPs. The simulation then determines the penetrance for that genotype and generates a random number to determine if this individual is affected. The penetrance table represents the disease model being simulated.

600 markers including the two disease loci were simulated assuming they were independent and there was no genotyping error. 1000 replicates for each genetic model were simulated and a χ^2 test for trend was used to compare genotype frequencies between cases and controls at every marker to evaluate the power of picking up either one of the two disease loci based on p value = 0.05.

Statistical methods

Association analysis in case-control data set

In the case-control data set, allelic association for single SNP was tested using the χ^2 test in Haploview (Barrett et al., 2005). Genotypic association was assessed by a 2 x 3 contingency table likelihood ratio test in SAS program (SAS Institute, 2003). Haplotype analysis for each gene was conducted using haplo.stats (Schaid et al., 2002). To save calculation time and satisfy computer memory limit, small genes (with less than 8 genotyped SNPs in the gene) were analyzed using individuals with less than 50% missing genotypes. Genes of medium size (with 8 to 26 genotyped SNPs) were analyzed using only individuals who had complete genotypes on all analyzed loci in that gene. There are

seven big genes that were genotyped for 38 to 124 SNPs. They were analyzed on individuals without any missing genotypes and only tagSNPs with $r^2<0.5$ were used. In addition, the tagSNPs were separated into groups of 10 with 2 SNPs overlapping across the groups. We estimated haplotype frequencies and tested association of each haplotype with a frequency of at least 1% in our case-control data set with age and gender adjusted score statistics. Haplotype logistic regression was modeled using a GLM algorithm including age and gender as covariates. The most frequent haplotype occurring in the similar percentage of cases and controls was selected as baseline haplotype. To evaluate the association of subsets of alleles from the full haplotype, a sliding window of three SNPs was used and the global score statistics were calculated.

Conditional analyses were conducted by dividing the case-control data set by age, gender and ApoE status. Then the difference in allele frequency between cases and controls was tested by a χ^2 test or Fisher's exact test when applicable (e.g. cell size <5). We also applied logistic regression to estimate the effects of the SNPs showing the most significant results across different analyses after controlling for age, gender and ApoE status. Nominal P values are reported and FDR was used to correct multiple testing. A p value of 0.05 was used as criterion for significant result. Haploview and SAS (SAS institute, Cary, NC) programs were used for the analyses.

Association analysis in family data set

Family based association analysis was used to follow up one of the SNPs showing strongest association in female subset. The allelic association analyses were conducted using the association in the presence of linkage (APL) analysis (Martin et al., 2003b) and Pedigree disequilibrium test (PDT) (Martin et al., 2000b). APL and PDT each have

distinct advantages. APL can correctly infer missing parental genotypes in regions of linkage by estimating identity-by-descent (IBD) parameters and is a more powerful test under those conditions than PDT. However, APL only uses nuclear families for the association test. PDT can use data from extended pedigrees and gains power substantially when extended pedigree data are available. Genotype-PDT (GenoPDT) is an extension of PDT and was used to assess association between genotypes and risk of AD in the family data (Martin et al., 2003a).

Results

Power study

Figure 9-1 shows the results of the power study based on ninety genetic models with different inheritance models, diseases allele frequencies and odds ratios. We define the power as the probability of selecting at least one of the two disease loci at $p<0.05$. The power to detect two disease allele frequencies at 0.1 and 0.4 was very close to that of 0.2 and 0.3 in additive and dominant model (Figure 9-1 A and B). The power to detect major disease alleles (frequency of 0.6) was low in additive and dominant model (Figure 9-1 A and B). However, the power of detecting major alleles was the highest in the recessive model (Figure 9-1 C). In the additive and dominant models, the dataset had more than 80% power to detect disease alleles with odds ratio greater than 1.5 with two disease allele frequencies of 0.1/0.4 or 0.2/0.3. In the recessive model, the power was 62% in detecting one of the two disease alleles (allele frequencies of 0.1 and 0.4) at joint odds ratio of 1.5. If the OR=1.7, the power was 86%. When the two disease alleles had

frequencies of 0.2 and 0.3, the dataset had 82% power to detect one of the disease alleles at OR=1.8. The power was 81% to detect one of the major disease alleles (frequencies of 0.6) in the recessive model.

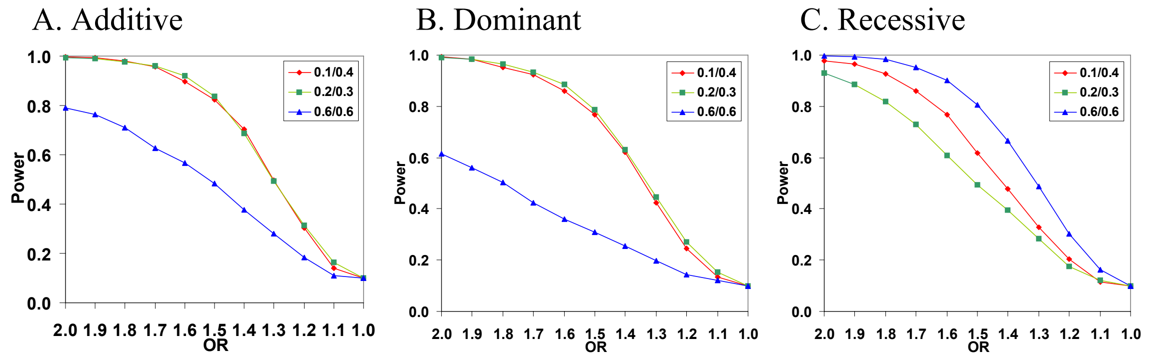


Figure 9-1 Power of the data set with 500 cases and 500 controls at p<0.05
A: additive model B: dominant model C: recessive model. Lines in each plot represent different allele frequencies for the two disease alleles.

Allelic association analysis in the overall data set and subsets

The allelic association of 597 genotyped SNPs was analyzed on the overall data set. Twenty-four SNPs in eight genes (PTPLA, SORCS1, PRKG1, SVIL, ACTR1A, BA108L7.2, CAMK1D, and PDZK7) showed association with AD at nominal p<0.05 (Figure 9-2). After we applied FDR (q=20%) based on the number of genotyped SNPs in each gene to correct for multiple testing, two SNPs in two genes remained significant. The first SNP is rs10508533 PTPLA with p=0.0022; the second is rs17277986 in SORCS1 gene with p=0.0025, Figure 9-2.

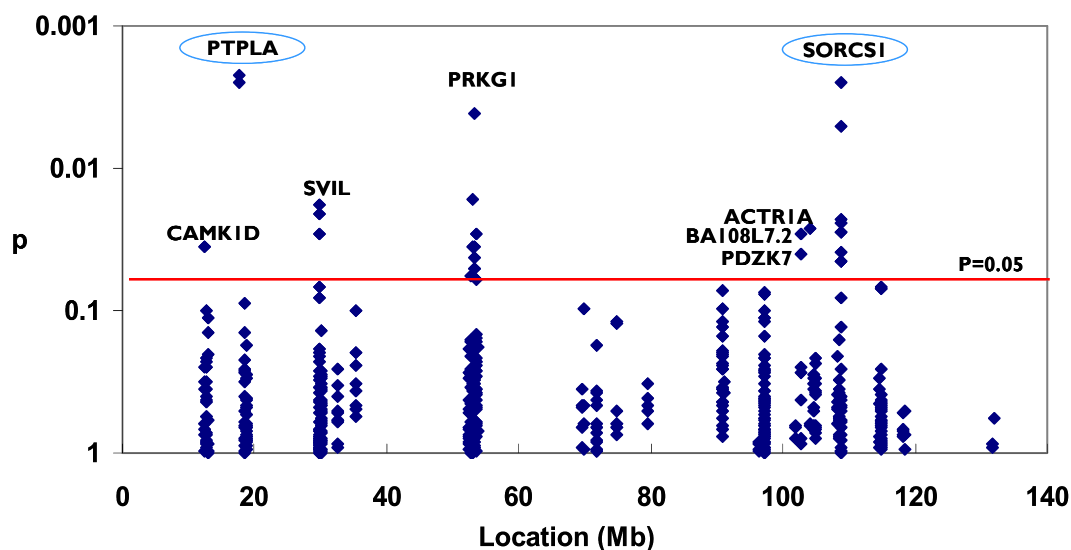


Figure 9-2 Allelic association test in the overall data set

PTPLA and SORCS1 are associated with AD after FDR correction for multiple testing

To increase the homogeneity of the data set, I analyzed the data on eight different subsets. Age is the major risk factor for Alzheimer disease as both the prevalence and incidence of AD double ever five years after age 60. To look for genetic factors that might predispose in different age groups, we subsetted our data into two groups, Age-of-onset between 60 and 75 and Age-of-onset greater than 76. ApoE is the only known genetic risk factor for AD and has very strong effect (OR=5~8). We analyzed association in the ApoE 4-positive subset in which all the cases had at least one ApoE ε4 allele and in the ApoE 4-negative subset in which none of the cases had ApoE ε4 allele to identify genes underlying Alzheimer disease susceptibility independent of ApoE locus. Alzheimer disease is more common in females than in males. Therefore the female subset and male subset were analyzed individually. Due to the uncertainty of the

diagnosis of AD, we separated the AD cases into two groups, definite plus probable Alzheimer disease and possible Alzheimer disease.

Not surprisingly, there were additional SNPs showing significant results in each data set. Five out of eight associated genes in the overall data set showed association in three or more subsets (Table 9-11). These five genes were PTPLA, SORCS1, SVIL, PRKG1 and PDZK7. The female subset had the most significant result (Table 9-11): p=0.00002 for a SNP (rs172777986) in SORCS1. Close to rs17277986, four tag SNPs were also significant at $p<0.05$.

Table 9-11 The allelic association results (nominal P-values) in subsets for the SNPs showing significant association (p<0.05) in the overall data set

Name	location	gene	function	overall	aao6075	aao76plus	apoe+	apoe-	female	male	defprobAD	possAD
rs10508533	17689894	PTPLA	Intron	0.0022	0.0202	0.0571	0.2215	0.0835	0.0199	0.0474	0.0023	0.0931
rs17277986	108748715	SORCS1	Intron1	0.0025	0.0484	0.0178	0.0073	0.0412	0.00002	0.6617	0.0034	0.0751
rs4453117	17666099	PTPLA	3' UTR	0.0025	0.0255	0.0403	0.0021	0.1825	0.0125	0.0882	0.0012	0.1607
rs6480499	53263984	PRKG1	Intron	0.0041	0.002	0.4276	0.0325	0.1385	0.0012	0.5973	0.0058	0.0794
rs12571141	108738659	SORCS1	Intron1	0.0051	0.0697	0.0282	0.0143	0.0629	0.00005	0.5906	0.0079	0.0859
rs16921605	53125867	PRKG1	Exon	0.0164	0.0452	0.1953	0.4001	0.058	0.0216	0.3326	0.0388	0.0917
rs10826650	29854289	SVIL	Intron	0.0181	0.0243	0.234	0.0559	0.2062	0.0133	0.4908	0.0967	0.0237
rs3780849	29844250	SVIL	Intron	0.0207	0.0423	0.1834	0.0391	0.2352	0.043	0.2403	0.0932	0.0322
rs2900717	108721676	SORCS1	Intron	0.0227	0.1752	0.0376	0.0263	0.0948	0.0008	0.5817	0.0043	0.6258
rs4918280	108780747	SORCS1	Intron	0.0239	0.0656	0.2274	0.3215	0.0103	0.0018	0.7999	0.0041	0.6718
rs17709610	104240267	ACTR1A	Intron (boundary)	0.0263	0.1152	0.1094	0.439	0.0769	0.1281	0.0952	0.028	0.2134
rs17279997	108799513	SORCS1	Intron	0.0278	0.009	0.8353	0.5525	0.1031	0.0937	0.1547	0.0359	0.1833
rs10884399	108762924	SORCS1	Intron	0.0281	0.0481	0.3128	0.2698	0.0225	0.002	0.757	0.0084	0.5317
rs1961742	53440631	PRKG1	Intron	0.0287	0.0678	0.19	0.8253	0.0276	0.0735	0.1966	0.0166	0.3484
rs807051	102792559	BA108L7.2	3' UTR	0.0291	0.0627	0.1702	0.4537	0.1143	0.0679	0.2259	0.1648	0.0253
rs2479687	29943066	SVIL	Intron	0.0293	0.2035	0.0244	0.6622	0.1936	0.2032	0.0542	0.0415	0.1703
rs4415704	53231618	PRKG1	Intron	0.0355	0.0349	0.4359	0.0276	0.7619	0.0159	0.7201	0.0873	0.0894
rs4451621	12471372	CAMK1D	Intron	0.0356	0.0579	0.4086	0.0572	0.6932	0.4252	0.0175	0.0193	0.3962
rs11594325	53152739	PRKG1	Intron	0.0357	0.0111	0.9497	0.069	0.2973	0.0074	0.9739	0.014	0.4923
rs6584788	108811850	SORCS1	Intron	0.0385	0.0517	0.3944	0.4538	0.2455	0.0611	0.3319	0.0378	0.2522
rs3740496	102773134	PDZK7	Intron (boundary)	0.0404	0.2477	0.0455	0.6924	0.0194	0.0054	0.7721	0.0762	0.1095
rs6415910	53405760	PRKG1	Intron	0.0426	0.043	0.4000	0.2198	0.2335	0.1471	0.1528	0.0289	0.3578
rs717751	108793300	SORCS1	Intron	0.045	0.2944	0.0687	0.1051	0.4857	0.5104	0.0179	0.1253	0.073
rs10884409	108809774	SORCS1	Intron	0.0452	0.0768	0.3328	0.3748	0.2996	0.0874	0.2837	0.0466	0.2567

The SORCS1 SNP rs17277986 remained significant after FDR correction for multiple testing. SNP rs6480499 in PRKG1 is also significant in the female subset after FDR correction (Figure 9-3).

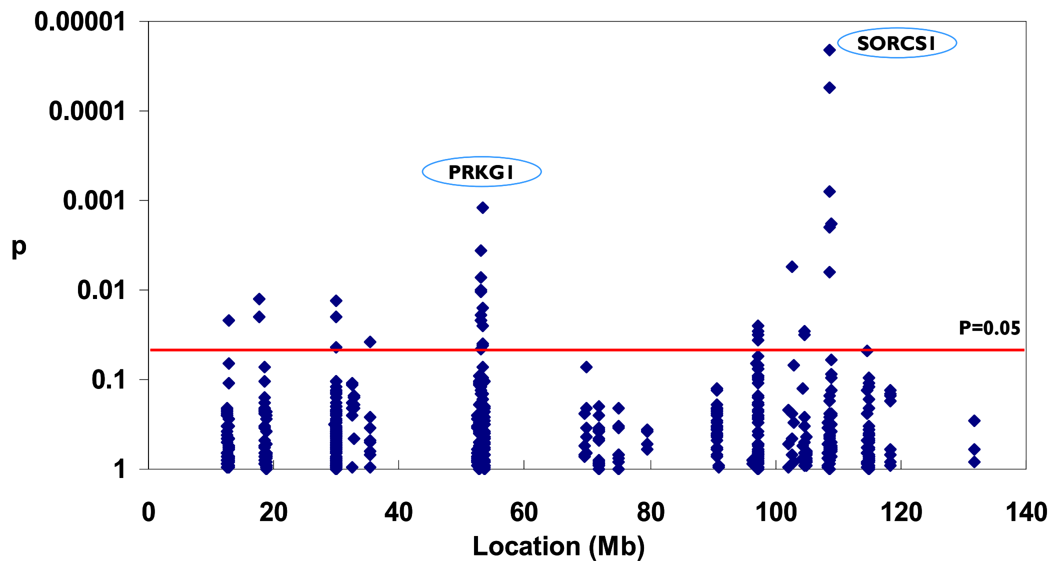


Figure 9-3 Allelic association analysis in the female subset
SORCS1 and PRKG1 showed association after FDR correction for multiple testing

Genotypic association analysis in the overall data set and subsets

Genotypic association analysis of 597 SNPs was performed on not only the overall dataset but also the eight subsets. Twenty-seven SNPs in twelve genes (CACNB2, PTPLA, SORCS1, SVIL, SORBS1, PRKG1, CAMK1D, CNNM2, PDZK7, TCF7L2, TNFRSF6, and NT5C2) showed association with AD at nominal $p<0.05$. Figure 9-4 shows the genotypic association in overall data set. SNPs in CACNB2 and PTPLA were significant after FDR correction (q=0.2) for multiple testing (p value of 0.0030 and 0.0036 for rs1277738 in CACNB2 and rs10508533 in PTPLA, respectively). Table 9-12 shows genotypic association in subsets for SNPs having genotypic association with AD at $p<0.05$ in the overall data set. Similar to allelic association, there were additional SNPs

showing significant results in each subset. The female subset had the most significant results. The SNP rs17277986 in SORCS1 showed genotypic association after FDR correction (Figure 9-5) in this subset (p=0.0002).

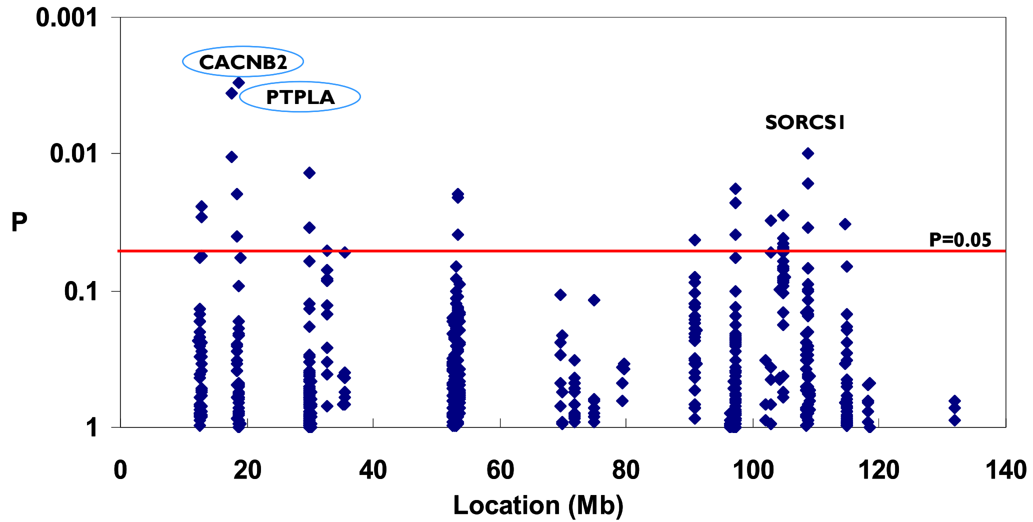


Figure 9-4 Genotypic association analysis in the overall data set
CACNB2 and PTPLA showed association after FDR correction for multiple testing

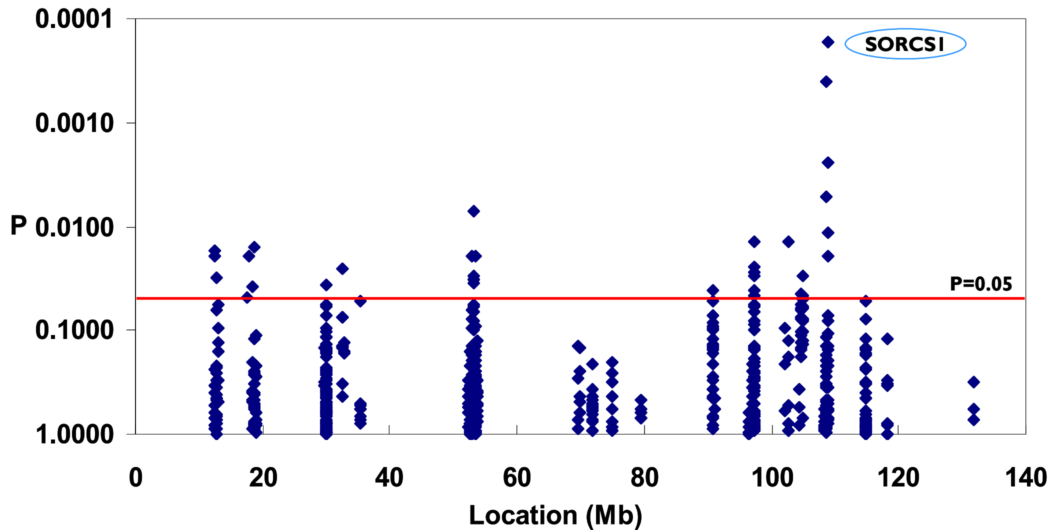


Figure 9-5 Genotypic association analysis in the female subset
SORCS1 showed association after FDR correction for multiple testing

Table 9-12 The genotypic association in subsets for the SNPs showing significant association (p<0.05) in overall data set

Name	location	gene	function	overall	aao6075	aao76plus	ApoE 4+	ApoE 4-	female	male	defprobAD	possAD
rs1277738	18,614,317	CACNB2	Intron	0.0030	0.0450	0.0853	0.3220	0.0182	0.0158	0.1254	0.0834	0.0019
rs10508533	17,689,894	PTPLA	Intron	0.0036	0.0386	0.0848	0.0982	0.1258	0.0191	0.1239	0.0057	0.1189
rs17277986	108,748,715	SORCS1	Intron	0.0100	0.1509	0.0534	0.0248	0.1411	0.0002	0.6625	0.0087	0.2043
rs4453117	17,666,099	PTPLA	3' UTR	0.0104	0.0861	0.1113	0.0074	0.3148	0.0478	0.1738	0.0065	0.2431
rs1247419	29,948,133	SVIL	Intron	0.0136	0.0318	0.4135	0.1976	0.0135	0.0584	0.1811	0.0233	0.0803
rs12571141	108,738,659	SORCS1	Intron	0.0162	0.1773	0.0909	0.0425	0.1996	0.0004	0.4887	0.0164	0.2340
rs7085647	97,290,625	SORBS1	Intron	0.0178	0.0128	0.8267	0.4235	0.0249	0.0138	0.5966	0.0219	0.1248
rs1961742	53,440,631	PRKG1	Intron	0.0194	0.0944	0.1734	0.3148	0.0935	0.1614	0.0343	0.0134	0.3543
rs7090118	18,499,287	CACNB2	Intron	0.0199	0.0688	0.0650	0.1958	0.4992	0.2014	0.0109	0.1468	0.0246
rs6480499	53,263,984	PRKG1	Intron	0.0205	0.0144	0.6523	0.0126	0.1597	0.0071	0.8824	0.0250	0.2601
rs500470	97,318,685	SORBS1	Promoter	0.0229	0.0055	0.7650	0.0918	0.8158	0.0589	0.2449	0.0400	0.1257
rs17580343	12,721,118	CAMK1D	Intron	0.0242	0.1129	0.1174	0.3580	0.1018	0.6641	0.0070	0.2421	0.0129
rs1046411	104,827,805	CNNM2	Exon	0.0283	0.0321	0.2795	0.2390	0.1069	0.0299	0.2612	0.0068	0.2970
rs4750253	12,818,977	CAMK1D	Intron	0.0287	0.0236	0.7315	0.4986	0.0432	0.0308	0.1156	0.0680	0.1326
rs3740496	102,773,134	PDZK7	boundary	0.0310	0.0553	0.1077	0.2467	0.0518	0.0140	0.4321	0.0457	0.1888
rs2094405	114,705,678	TCF7L2	Intron	0.0330	0.0021	0.8824	0.1759	0.7720	0.0520	0.1876	0.0769	0.1257
rs3780849	29,844,250	SVIL	Intron	0.0341	0.0705	0.3843	0.1522	0.2926	0.1158	0.2454	0.1442	0.0643
rs4918280	108,780,747	SORCS1	Intron	0.0344	0.0739	0.4558	0.6125	0.0320	0.0025	0.9533	0.0035	0.8687
rs10826650	29,854,289	SVIL	Intron	0.0345	0.0636	0.3742	0.1904	0.2499	0.0369	0.5335	0.1415	0.0597
rs2252101	53,428,892	PRKG1	Intron	0.0390	0.1407	0.2273	0.1761	0.5187	0.1602	0.2082	0.0109	0.5945
rs1326935	97,250,237	SORBS1	Intron	0.0392	0.0020	0.7169	0.3272	0.0507	0.1007	0.3641	0.0927	0.1347
rs2482100	18,521,085	CACNB2	Intron	0.0402	0.1145	0.1732	0.2063	0.6905	0.0387	0.6468	0.3447	0.0137
rs11191506	104,758,623	CNNM2	Intron	0.0410	0.2013	0.0463	0.3753	0.0316	0.1013	0.0395	0.0346	0.1593
rs3781202	90,759,443	TNFRSF6	Intron	0.0421	0.1497	0.1096	0.2101	0.4982	0.0723	0.4626	0.0085	0.7879
rs7914558	104,765,897	CNNM2	Intron	0.0454	0.0871	0.3402	0.2439	0.0462	0.0650	0.5697	0.0789	0.1126
rs11191547	104,833,137	NT5C2	3' UTR	0.0485	0.0376	0.4294	0.3624	0.1250	0.0460	0.3796	0.0240	0.3218
rs746293	104,887,243	NT5C2	Intron	0.0493	0.2011	0.1174	0.1775	0.0648	0.0611	0.6135	0.0840	0.0732

Effect sizes of the most significant SNPs in two genes

SORCS1 gene

The allelic Odds Ratio of having the risk allele vs. not having risk allele was 1.34 (p =0.003, Table 9-13) for SNP rs17277986 in the SORCS1 gene. By looking at the genotypic effect, the odds ratio of getting AD for homozygosity of the risk allele vs. homozygosity of the non-risk allele was 1.93, with p=0.005. In the female subset, the effect was stronger (Table 9-14). The odds ratio of having the risk allele vs. not having risk allele was 1.70 and the p value was very significant (p<0.0001). The odds ratio of getting AD for homozygosity of the risk allele vs. homozygosity of the non-risk allele was 3.02, with p=0.001. The odds ratio for heterozygote was not significant (Table 9-13 and 9-14). After adjusting the odds ratio by age, gender and ApoE status, the odds ratios remained significant (Table 9-15).

Table 9-13 Effect size of SORCS1 (rs17277986) in the overall data set

association	case	control	total	Freq.	test	OR	95% CI	P
allelic								
C	713	718	1431	0.71	C vs. T	1.34	***1.11-1.63***	***0.003***
T	253	342	595	0.29				
genotypic								
CC	263	248	511	0.50	CC vs. TT	1.93	***1.22-3.05***	***0.005***
CT	187	222	409	0.40	CT vs. TT	1.53	0.96-2.44	0.074
TT	33	60	93	0.10				

Table 9-14 Effect size of SORCS1 (rs17277986) in the female subset

association	case	control	total	Freq.	test	OR	95% CI	P
allelic								
C	448	417	865	0.71	C vs. T	1.70	***1.33-2.19***	***<0.0001***
T	138	219	357	0.29				
genotypic								
CC	171	138	309	0.51	CC vs. TT	3.02	***1.62-5.63***	***0.001***
CT	106	141	247	0.40	CT vs. TT	1.83	0.97-3.45	0.061
TT	16	39	55	0.09				

Table 9-15 Effect size of SORCS1 (rs17277986) adjusted by covariates

	Crude			Adjusted by age, gender, ApoE status		
	OR	**95% CI**	**p**	**OR**	**95% CI**	**p**
ALL						
allelic effect						
Dominant-having risk allele vs. no risk allele	1.74	1.12-2.71	***0.014***	1.84	1.11-3.06	***0.019***
Recessive-2 risk alleles vs. 0~1 risk allele	1.36	1.06-1.74	***0.015***	1.49	1.12-1.98	***0.006***
Additive-linear effect (0, 1, 2 risk allele)	1.33	1.10-1.62	***0.003***	1.42	1.14-1.78	***0.002***
genotypic effect						
2 risk alleles vs. 0 risk allele	1.93	1.22-3.05	***0.005***	2.12	1.25-3.60	***0.005***
1 risk allele vs. 0 risk allele	1.53	0.96-2.44	0.074	1.54	0.90-2.63	0.116
FEMALE_ONLY						
allelic effect						
Dominant-having risk allele vs. no risk allele	2.42	1.32-4.43	***0.004***	2.85	1.40-5.80	***0.004***
Recessive-2 risk alleles vs. 0~1 risk allele	1.83	1.33-2.52	***0.002***	1.96	1.35-2.83	***<0.0001***
Additive-linear effect (0, 1, 2 risk allele)	1.70	1.32-2.19	***<0.0001***	1.83	1.36-2.45	***<0.0001***
genotypic effect						
2 risk alleles vs. 0 risk allele	3.02	1.62-5.63	**0.001**	3.66	1.75-7.63	***0.001***
1 risk allele vs. 0 risk allele	1.83	0.97-3.45	0.061	2.11	1.00-4.44	***0.049***

PTPLA gene

The allelic and genotypic effect sizes of this gene are presented in Table 9-16. Both of the two genotyped SNPs were significant. For SNP rs10508533, the allelic odds ratio for having the risk allele (allele 'A') vs. not having the risk allele was 1.34 (p=0.003). By looking at the genotypic effect, the odds ratio of getting AD for risk allele homozygotes vs. non-risk allele homozygotes of was 2.29 (p=0.001). Heterozygotes with the risk allele had an OR of 1.91 (p=0.014) compared to the homozygotes with the non-risk allele (Table 9-16). The odds ratios were also significant after the adjustment for age, gender and ApoE status (Table 9-17).

For SNP rs4453117, the allelic odds ratio for having the risk allele (allele 'C') vs. not having the risk allele was 1.31 (p=0.003). By looking at the genotypic effect, the odds ratio of getting AD for homozygotes of risk allele vs. homozygotes of non-risk allele was 1.69 (p=0.004). Heterozygotes with the risk allele had an OR of 1.36 (p=0.031) compared to the homozygotes with the non-risk allele (Table 9-16). The odds ratios were also significant after the adjustment for age, gender and ApoE status (Table 9-17). The linkage disequilibrium between the two SNPs (rs10508533 and rs4453117) had $r^2 = 0.22$ (Figure 9-7).

Table 9-16 Effect size of PTPLA in the overall data set

Association	Case	Control	Total	Freq.	Test	OR	95% CI	P
rs10508533								
allelic								
A	732	757	1489	0.72	A vs. C	1.34	1.10-1.62	***0.003***
C	248	343	591	0.28				
genotypic								
AA	266	261	527	0.51	AA vs. CC	2.29	1.38-3.82	***0.001***
AC	200	235	435	0.42	AC vs. CC	1.91	1.14-3.21	***0.014***
CC	24	54	78	0.07				
rs4453117								
allelic								
C	453	424	877	0.43	C vs. A	1.31	1.10-1.57	***0.003***
A	511	628	1139	0.57				
genotypic								
CC	107	89	196	0.19	CC vs. AA	1.69	1.18-2.41	***0.004***
AC	239	246	485	0.48	AC vs. AA	1.36	1.03-1.81	***0.031***
AA	136	191	327	0.33				

Table 9-17 Effect size of PTPLA adjusted by covariates

	Crude			Adjusted by age, gender, ApoE status		
	OR	**95% CI**	**p**	**OR**	**95% CI**	**p**
rs10508533						
allelic effect						
Dominant-having risk allele vs. no risk allele	2.11	1.29-3.47	***0.003***	2.29	1.31-4.03	***0.004***
Recessive-2 risk alleles vs. 0~1 risk allele	1.32	1.03-1.68	***0.028***	1.18	0.89-1.56	0.250
Additive-linear effect (0, 1, 2 risk allele)	1.35	1.11-1.64	***0.003***	1.28	1.02-1.60	***0.032***
genotypic effect						
2 risk alleles vs. 0 risk allele	2.29	1.38-3.82	***0.001***	2.34	1.31-4.18	***0.004***
1 risk allele vs. 0 risk allele	1.91	1.14-3.21	***0.014***	2.24	1.24-4.02	***0.007***
rs4453117						
allelic effect						
Dominant-having risk allele vs. no risk allele	1.45	1.11-1.89	***0.006***	1.53	1.13-2.08	***0.006***
Recessive-2 risk alleles vs. 0~1 risk allele	1.40	1.02-1.92	***0.035***	1.49	1.05-2.13	***0.027***
Additive-linear effect (0, 1, 2 risk allele)	1.37	1.09-1.56	***0.003***	1.37	1.12-1.67	***0.002***
genotypic effect						
2 risk alleles vs. 0 risk allele	1.69	1.18-2.41	***0.004***	1.85	1.23-2.78	***0.003***
1 risk allele vs. 0 risk allele	1.36	1.03-1.81	***0.031***	1.42	1.03-1.97	***0.035***

Haplotypic association analysis

Haplotypic association analysis for each gene was done on the overall data set. Global p values for score statistics are shown in Table 9-18. The haplotype in PTPLA had the most significant p value of 0.009. A haplotype containing ten tagSNPs in SVIL gene was marginally associated with AD (p=0.049). Since a SNP in SORCS1 was strongly associated with AD in the female subset, it was also analyzed for haplotypic association in the female subset only. The global p value was 0.147, which was not significant (Table 9-18). However, the three 3-SNP sliding window haplotypes were highly significant (smallest p=0.0005, Figure 9-6). All these three haplotypes contained rs17277986, which showed the most significant single SNP allelic association in the same data set.

Table 9-18 Haplotypic association in the overall data set

Gene	# SNPs	Tag SNPs only	Global P value
USP54	1	N/A	N/A
HNRPH3	2	N/A	0.16249
IFIT3	2	N/A	0.52978
PDZK7	2	N/A	0.09629
PTPLA	2	N/A	0.00928
ACTR1A	3	N/A	0.22319
CWF19L1	3	N/A	0.88097
PNLIPRP1	3	N/A	0.91184
TXNL2	3	N/A	0.79160
BA108L7.2	5	N/A	0.17506
PNLIP	5	N/A	0.85250
RPS24	5	N/A	0.67050
HELLS	6	N/A	0.39395
MAWBP	7	N/A	0.81463
PPP3CB	7	N/A	0.70719
CUL2	8	N/A	0.33924
NT5C2	9	N/A	0.47201
EPC1	11	N/A	0.79317
CNNM2	15	N/A	0.66829
EIF4EBP2	15	N/A	0.55133
TNFRSF6	26	N/A	0.29154
TCF7L2	38	25	0.13144
CAMK1D	39	30	0.38370
CACNB2	41	31	0.14616
SORCS1*	54	20	0.14659
SORBS1	64	32	0.14904
SVIL	97	37	0.04925
PRKG1	124	64	0.30425

* Analyzed in female only.

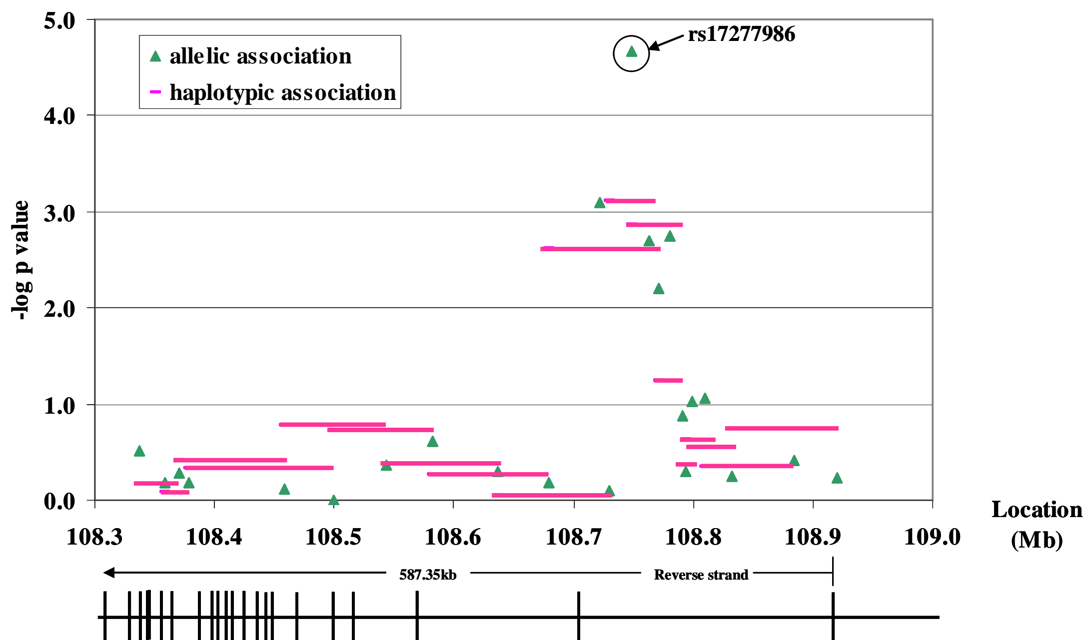


Figure 9-6 Haplotype analysis in SORCS1 in the female subset

20 tagSNPs were analyzed in SORCS1. Bars show the haplotypes in the 3-SNP sliding window; triangles show the single marker allelic association –log p values. The single marker allelic association results for three tagSNPs around the SNP rs17277986 with $r^2<0.8$ are also shown in the figure. The gene structure is shown at the bottom.

Analytical convergence of the genetic association tests

Across all the association analyses including allelic, genotypic and haplotypic association tests, several genes showed association. After we applied FDR to correct for multiple testing for allelic and genotypic association tests, only five genes showed association in at least one of the analyses. SNPs in PTPLA were significant in the overall data set across all three association tests. SNPs in SORCS1 had a significant allelic effect in the overall data set, and it was highly significant in the female subset across all three association tests. Although SNPs in CACNB2 did not show allelic association in the overall data set, SNP rs1277738 in CACNB2 had the strongest genotypic effect in overall

data set, and showed genotypic association at nominal p<0.05 in four subsets (Age-of-onset between 60 and 75, ApoE-negative, female and possible AD). SNPs in SVIL were marginally significant in the haplotypic association test in the overall data set.

Follow up of SNPs in SORCS1 and PTPLA genes

SORCS1 gene

SNPs in the SORCS1 gene had the strongest allelic effect in the female subset and were also significant in the overall data set. To confirm the effect of this gene, we genotyped the most significant SNP (rs17277986) in our validation data sets. When we analyzed female and male individuals together, none of the datasets showed significant results at p<0.05. Two data sets (CAP and NIMH) had family-based association at p<0.1 using PDT test. When we analyzed females only, two sets were significant at p<0.05 (Table 9-19). The CAP data set showed association in PDT with p=0.03 and the NCRAD data set showed association in APL at p=0.01. The results in family data sets were similar to the population based case-controls data set, but not significant after correction for multiple comparisons.

Table 9-19 Family-based association tests for rs17277986 in SORCS1 in validation data sets

	Female & Male				Female_only			
	All	**CAP**	**NIMH**	**NCRAD**	**All**	**CAP**	**NIMH**	**NCRAD**
APL	0.189	0.696	0.301	0.189	0.057	0.526	0.309	***0.010***
Sum_PDT	0.862	0.085	0.099	0.742	0.904	***0.030***	0.254	0.668
Geno_PDT	0.606	0.269	0.052	0.402	0.798	0.103	0.182	0.562

All: combined data set; CAP: the Collaborative Alzheimer Project; NCRAD: the NCRAD repository at Indiana University; NIMH: the National Institute of Mental Health repository

PTPLA gene

Two genotyped SNPs in PTPLA were significant across all the association analyses. This gene is small, spanning only 27 kb. We followed the results up by genotyping seven more SNPs to give this gene a better coverage. The linkage disequilibrium between all genotyped SNPs in cases is shown in Figure 9-7.

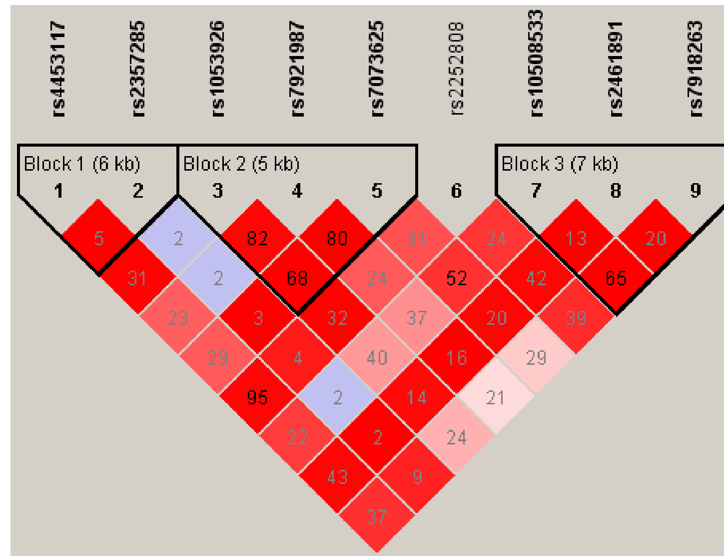


Figure 9-7 Linkage disequilibrium between all nine genotyped SNPs in PTPLA
r^2 is shown in the figure.

Most of the SNPs were significant in allelic and genotypic association test (Table 9-20). The overall haplotypic association test considering all SNPs in the gene was significant with global score statistic p value = 0.03. The 3-SNP sliding window haplotypic association test was significant for all haplotypes (Figure 9-8).

Table 9-20 Allelic and genotypic association of all genotyped SNPs in PTPLA in the overall data set

	SNP	location	function	MAF	Allelic Asso.	Genotypic Asso.
1	**rs4453117**	17666100	3'UTR	0.42	***0.0011***	***0.0039***
2	**rs2357285**	17672568	Intron	0.06	0.4504	0.7087
3	**rs1053926**	17676315	Exon 6, ns	0.30	***0.0141***	***0.0432***
4	**rs7921987**	17680752	Intron	0.32	0.0688	0.1316
5	**rs7073625**	17682131	Intron	0.37	***0.0223***	0.0857
6	**rs2252808**	17685820	boundary	0.44	***0.0017***	***0.0035***
7	**rs10508533**	17689895	Intron	0.29	***0.0029***	***0.0082***
8	**rs2461891**	17695987	Intron	0.25	***0.0035***	***0.0102***
9	**rs7918263**	17697158	Intron	0.38	***0.0007***	***0.0014***

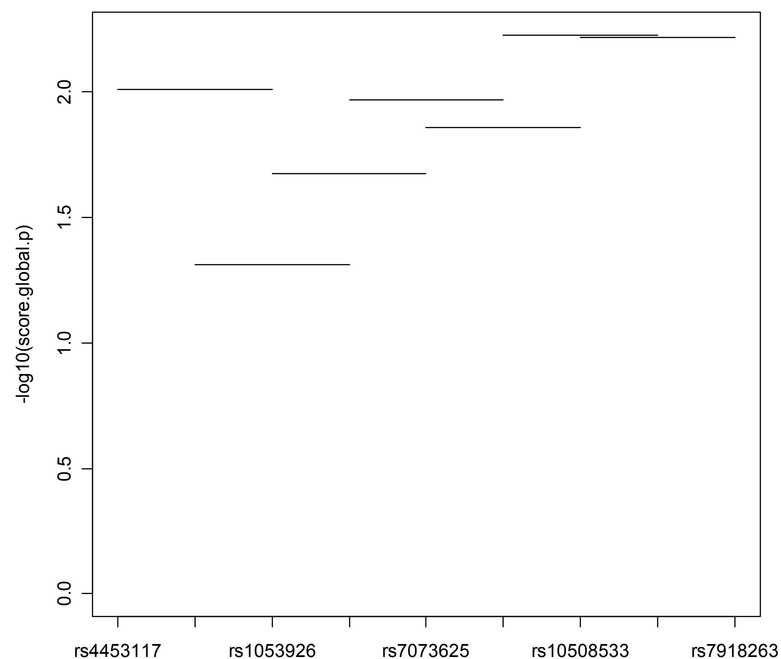


Figure 9-8 Haplotypic association analysis using 3-SNP sliding window in PTPLA
All haplotypes are significantly associated with AD at $p<0.05$

Discussion

Numerous genetic variants in many functional candidate genes have been associated with AD. However, none of the associations has been consistently replicated except the ApoE locus (Bertram et al., 2007). The inconsistencies may be due to false positive results, genetic heterogeneity, lack of power due to small sample size, or population stratification. Our group performed candidate gene studies to test VR22, LRRTM3, PLAU, IDE, TNFRSF6, CDC2 and LOC439999, which were suggested to be associated with AD (Bertram et al., 2000; Ertekin-Taner et al., 2003; Feuk et al., 2000; Finckh et al., 2003; Grupe et al., 2006; Johansson et al., 2003; Johansson et al., 2005; Martin et al., 2005), combined with an unbiased dense linkage screen on chromosome 10. However, our data suggested much more extensive locus heterogeneity than we previously suspected. With this in mind, we designed the current study to converge all available data (linkage, candidate gene, and gene expression data) to identify susceptibility genes altering AD risk.

Several genes showed association with AD in either the overall data set or the various subsets. To correct for multiple testing, we applied the False Discovery Rate (FDR) to access potential associations. In the overall data set, PTPLA showed allelic, genotypic and haplotypic association. SORCS1 showed allelic association, CACNB2 showed genotypic association, and SVIL showed haplotypic association. In the age-of-onset 60-75 subset, SORBS1 was the only gene that showed association (genotypic) after the FDR correction for multiple testing. TNFRSF6 showed both allelic and genotypic associations in the age-of-onset 76 plus subset; it also showed allelic association in the possible AD subset. PTPLA showed both allelic and genotypic associations in the ApoE-

positive subset and definite and probable AD subset. SORCS1 showed allelic, genotypic and haplotypic associations in female subset and genotypic association in the definite and probable AD subset. None of the genes showed association in male subset after correction for multiple comparisons. Table 9-21 summarizes the significant associations in the overall and subsets.

Table 9-21 Associated genes in the overall and subsets after the FDR correction for multiple comparisons (q=0.2)

	All	AOO 60-75	AOO 76+	ApoE+	ApoE-	Female	Male	Definite /probAD	PossAD
PTPLA	Ya,Yg, Yh	N	N	Ya, Yg	N	N	N	Ya, Yg	N
SORCS1	Ya	N	N	N	N	Ya,Yg, Yh	N	N	N
TNFRSF6	N	N	Ya,Yg	N	N	N	N	N	Ya
PRKG1	N	N	N	N	N	Ya	N	N	N
CACNB2	Yg	N	N	N	N	N	N	N	Yg
SORBS1	N	Yg	N	N	N	N	N	N	N
CWF19L1	N	N	N	N	Yg	N	N	N	N
CNNM2	N	N	N	N	N	N	N	Yg	N
SVIL	Yh	N	N	N	N	N	N	N	N

*Ya=significant allelic association; Yg=significant genotypic association; Yh= significant haplotypic association; N=not significant.

Given the assumption that genes showing positive association in multiple tests have higher likelihood of being important in the disease process, we performed an analytical convergence on the results from allelic, genotypic and haplotypic association tests. The most consistently associated SNPs across the three analyses are rs10508533 in PTPLA in the overall data set (allelic association p=0.0022) and rs17277986 in SORCS1 in the female subset (allelic association p=0.00002).

PTPLA is protein tyrosin phosphatase-like, member A. Little is known about its function, but it may be involved in the protein phosphorylation process that is important for the function of some proteins in the AD pathology. For example, phosphorylation of

glycogen synthase kinase 3β (GSK-3β), a binding partner of β-catenin, is important in regulating the function of β-catenin which binds to transcription factors and starts the cell cycle (Caricasole et al., 2003). PTPLA may also affect the phosphorylation of Tau which forms neurofibrillary tangles, a hallmark of AD. Among all nine genotyped SNPs in PTPLA, seven showed allelic association and six showed genotypic association. Three of those six SNPs are functional SNPs. rs4453117 is in 3'-UTR with allelic p=0.001 and genotypic p=0.004. rs1053926 is a non-synonymous SNP located in exon 6 with allelic p=0.014 and genotypic p=0.043. rs2252808 is located in the intron/exon boundary and has allelic p=0.002 and genotypic p=0.004. This gene is not in the same LD block with any nearby genes based on HapMap data. Therefore, the three functional SNPs could have potential roles in the biological function of PTPLA gene or are in LD with a susceptible SNP affecting PTPLA function, but are unlikely to be in LD with susceptibility SNPs in another gene.

SORCS1 encodes sortilin-related VPS10p domain containing type 1 receptor. It contains a leucine-rich domain and mediates intracellular sorting and trafficking functions. It is highly expressed in the brain (Hermey et al., 1999) and neuronal activity can differentially affect its expression (Hermey et al., 2004). SORCS1 is a substrate of γ-secretase, and γ-secretase cuts amyloid precursor protein (APP) and generates amyloid β peptide (Aβ), one of the hallmarks of Alzheimer disease. rs17277986 in SORCS1 was significant in the overall, age-of-onset greater than 76 year old, ApoE 4-positive, and definite/probAD data sets, with p=0.0025, 0.0178, 0.0073, and 0.0034 respectively. It was most significant in the female subset (p=0.00002). It survived not only an FDR multiple testing correction, but also a more conservative Bonferroni correction. Three

nearby tagSNPs ($r^2<0.8$) are also significant in the female subset (p=0.0008, 0.002, 0.0064, 0.0018 for rs2900717, rs10884399, rs11193170 and rs4918280, respectively). Another SNP (rs12571141) had a p=0.00006 in this subset, but it was in high LD with rs17277986 ($r^2=0.97$), suggesting it was measuring the same effect. Haplotypes containing the most significant SNP (rs17277986) were all significant with the smallest p value = 0.0005.

All the interesting SNPs are in intron 1 of SORCS1. This gene is 590kb and intron 1 itself is 207kb and the largest intron in the gene. Most of the regions in intron 1 are conserved (>70% identity) over the 100bp calculation window among mammalians (Figure 9-9). Thus, this region may be important as a functional element in the gene.

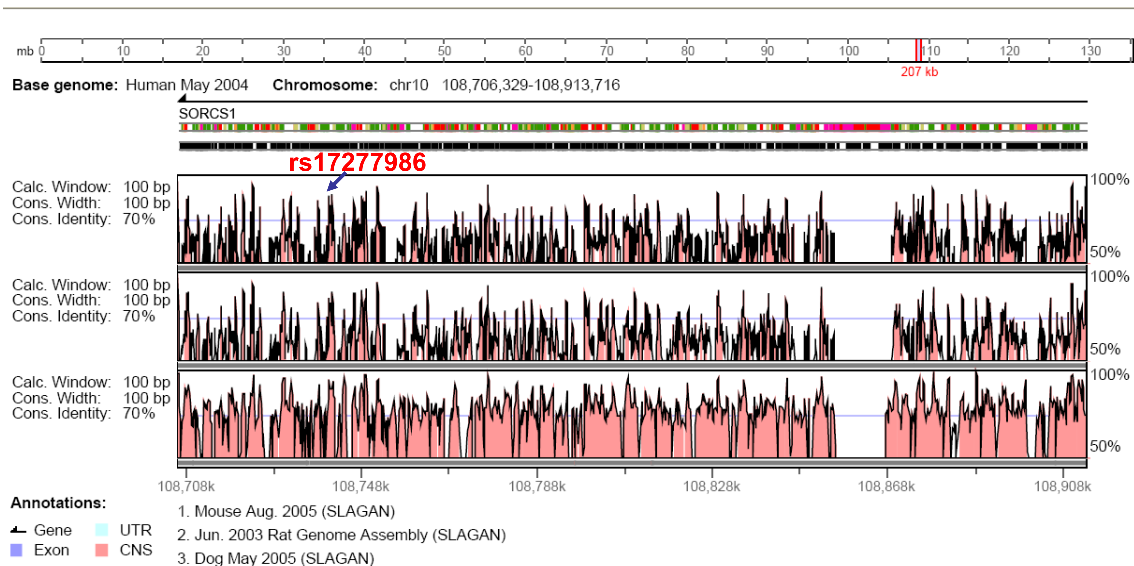


Figure 9-9 Genomic DNA comparisons of intron 1 in SORCS1 between Human and other Mammals

Output from VISTA analysis of the conserved region among mammals is shown. The first panel is mouse genome, the second is rat genome and the third is dog genome. The whole region is the intron 1 (207 kb) in SORCS1gene. Conserved region (>70% identity) is highlighted by pink shading. SNP rs17277986 with the most significant result in the association tests is labeled in the figure.

Grupe et al. (Grupe et al., 2006) found an association between SNP rs600879 in SORCS1 and Alzheimer disease with a p=0.0043 in their combined data set. SNP rs600879 (108,913,108 bp) is 608 bp away from the first intron/exon boundary. However, it was not consistently replicated in all of their data sets. It was significant in two of their datasets (p=0.017 and 0.040). However, the sample size of the homozygotes with the minor allele was small (between 4 and 9 samples for cases and controls) because the minor allele frequency of this SNP was 11.2% and 8.5% for cases and controls, respectively. The SNP was not significant in the other two data sets.

In our study, SNP rs17277986 (164 kb away from the SNP rs600879 that showed association in the Grupe et al. study) in SORCS1 was significantly associated with AD in females in the case-control and independent family-based data sets. There are four explanations for such findings.

First, it could be due to the estrogen effect. We checked the transcription factor binding sites within 1000 bp of rs17277986 on each side of the SNP using Transcription Elements Search System (TESS) (http://www.cbil.upenn.edu/tess). There were three regions showing a high likelihood of binding to the estrogen receptor (ER). One of those three regions was only 32 bp from the SNP rs17277986. Thus the genomic fragment including this SNP might have higher binding affinity to ER transcription factors or higher enhancer activity, suggesting a potential role of estrogen in AD. Studies on the estrogen-replacement therapy in Alzheimer disease suggested that estrogen may provide some protection against memory loss and lower the risk of developing AD (Burns and Murphy, 1996; Tang et al., 1996; Wickelgren, 2003). However, this intronic SNP was 165kb away from the start codon, suggesting its role in transcription regulation was likely

to be low. On the other hand, this SNP was only 42 kb away from exon 2. There is possibility that it is in LD with a SNP close to the intron/exon boundary that may affect mRNA splicing. The DNA sequence containing the SNP rs17277986 is a binding site of NF-κB (nuclear factor κB) that is important during AD pathogenesis. Studies of postmortem brain tissue from AD patients hade revealed increased NF-κB activity in cells involved in the neurodegenerative process (Baldwin, Jr., 1996). SNP rs17277986 in SORCS1 may play a role in the interaction or translocation of the NF-κB in the pathophysiology of AD. Second, this variant may be in or near intronic regulatory sequences that might govern cell type-specific or tissue-specific expression of SORCS1. Third, genomic imprinting may also play a role for the significant difference of the alleles between AD cases and controls, where differential gene expression depends upon whether the inheritance is through the mother or father (Hall, 1990). Methylation has been proposed as a mechanism of imprinting (Holliday, 1989) and is supported by the findings of the increased number of unmethylated sites in AD patients in comparison to controls (Payao et al., 1998). Fourth, there may be interaction between mitochondrial DNA mutations and this autosomal locus such that a particular mitochondrial genotype is required for individuals carrying this autosomal risk variant to express disease. It was suggested that mitochondria are involved in apoptosis and there is evidence of oxidative damage to mitochondrial DNA in AD cases (Mecocci et al., 1994; Green and Reed, 1998). Although there is not evidence that any of these possible explanations are in fact correct, they provide plausible areas for further research.

Taken together, our genomic convergence study suggests that genetic variations in PTPLA and SORCS1 may be associated and have modest effect to the risk of AD.

The replication of the effect of these genes in different study populations and search for susceptible variants and functional studies of these genes are necessary to get a better understanding of the roles of the genes in Alzheimer disease.

CHAPTER X

GENE-GENE INTERACTION AMONG GENOMIC CONVERGENCE GENES IN ALZHEIMER DISEASE

Introduction

Common diseases are likely to have complex etiology and Alzheimer disease is not an exception. Multiple genes with multifaceted interactions among genes and other risk factors undoubtedly play roles in the disease process. However, the inconsistent results from numerous studies indicate that finding the underlying genes is difficult. This may be due to several reasons. First of all, both genetic heterogeneity (allelic and locus heterogeneity) and phenotypic heterogeneity make the process of finding genes involved in the common complex diseases slow. For example, hundreds of candidate genes have been associated with Alzheimer disease, but all except ApoE have failed to be consistently replicated (Pericak-Vance and Haines, 1999). Among the possible reasons for this failure are false positive results due to population stratification and true differences in genetic etiology between study populations (Hirschhorn et al., 2002). Secondly, gene-gene interaction may play an important role in the disease etiology. There is clear and convincing evidence that gene-gene interactions are not only possible but also probably ubiquitous (Moore, 2003; Tong et al., 2004). Thus, it is crucial that complex genetic data sets be properly interrogated for possible underlying interactions (Thornton-Wells et al., 2004). Thirdly, traditional statistical methods of genetic analysis are suitable for detecting main effect in simple, Mendelian disease, such as Huntington disease, but were not designed and indeed fail in the face of more complex architectures.

Statistical and computational methodology has been slow to address the complexity of common diseases. There have been a few attempts to address the issue, including the development of non-parametric tools (e.g. the transmission disequilibrium test, TDT, ordered subset analysis, OSA) to find homogeneous subset using continuous or ordinal covariates, cluster analysis, latent class analysis and factor analysis to produce clusters with high intraclass similarity and/or interclass similarity based on genetic background (Hauser et al., 1998; Hauser et al., 2004; Ott and Hoh, 2003; Spielman et al., 1993; Hoh et al., 2001; Slonim, 2002; Mountain and Cavalli-Sforza, 1997; Grigull et al., 2001). Thornton-Wells et al. detailed a complete list of the tools of dissecting heterogeneity (Thornton-Wells et al., 2004).

However, traditional statistical methods to detect multilocus gene-gene interaction are limited due to the curse of dimensionality (Bellman, 1961). Regression analysis is one of the traditional approaches that is still widely used today. Logistic regression directly models the relationship of genetic (SNPs) and other risk factors (age, gender, et al.) to binary disease status (cases or controls). However, when the distribution of data across numerous combinations of factors becomes sparse, the parameter estimates become unreasonably biased, particularly when the ratio of sample size to independent variables is below ten to one (Concato et al., 1993; Moore and Williams, 2002; Peduzzi et al., 1996). In addition, logistic regression requires significant main effects to detect interactions between factors. However, this is not always the case in a complex disease, in which disease loci may have relatively small main effects but more substantial interactive effects.

Multifactor Dimensionality Reduction (MDR) is a computational data reduction method that addresses several limitations of traditional methods (Ritchie et al., 2001). MDR has been successful at finding gene-gene interactions in both simulated data (Hahn et al., 2003; Hahn and Moore, 2004; Ritchie et al., 2003) and real data (Cho et al., 2004; Ritchie et al., 2001; Ashley-Koch et al., 2006; Williams et al., 2004; Ma et al., 2005; Qin et al., 2005; Tsai et al., 2004). MDR reduces the dimensionality of multilocus information by pooling high-risk genotype combinations into one group and low-risk combinations into another group. Then the new one-dimensional multilocus genotype variable is evaluated for its ability to classify and predict disease status using cross-validation and permutation testing to identify optimal models.

Since MDR is a nonparametric, genetic model-free method, no hypothesis concerning any statistical parameter and no genetic inheritance model is assumed (Ritchie et al., 2001). MDR improves the ability to identify the high-order gene-gene interactions with the use of relatively small sample sizes. Another strength of the MDR method is its ability to detect significant interactions in the absence of main effects. The potential lack of main effects in complex disease makes MDR an attractive method to investigate the susceptibility of the disease.

To dissect the heterogeneity of Alzheimer disease, we applied a genomic convergence approach to identify genes under linkage peaks that are differentially expressed between AD cases and controls (Liang X et al. in preparation, as described in chapter VIII and IX; Sliffer M et al. in preparation). In the present study, with the appreciation of the importance of epistasis or gene-gene interaction in the complex disease, we applied MDR to the 1536 SNPs in the genomic convergence candidate genes

that were genotyped on a population data set (506 cases and 558 controls) to identify potential gene-gene interactions involved in Alzheimer disease.

Materials and methods

Study population

The data set had 1064 individuals (506 cases and 558 controls, Table 9-1) with substantial clinical information. The samples were collected by the Center for Human Genetics Research (CHGR) at Vanderbilt University and the Center for Human Genetics (CHG) at Duke University. The criteria were same as previously described in chapter IX (Liang et al. 2007). The cases are all late onset Alzheimer disease (minimum age at onset (AAO) $\geq$ 60 years), with average age-of-onset of 71.7 years. The average age-of-exam is 75.2 and 74.3 years in cases and controls, respectively. 61.5% of the cases are females in cases and 60.4% of the controls are females. All controls were ascertained in the same catchment area as cases. Following informed consent, blood samples were collected from each individual. Genomic DNA was extracted from whole blood by use of the Puregene system (Gentra Systems, Minneapolis, MN). All samples were coded and stored at 4°C until used.

Genotyping

Genes and SNPs were selected as described in chapter VIII (Liang X, et al. in preparation, Sliffer M et al. in preparation). Briefly, genes were selected from the convergence of linkage, association and expression data. A gene was declared

convergent if significant effects were found in at least three independent experimental methods (e.g. genetic linkage, genetic association, and gene expression). 869 SNPs in 19 candidate genes were selected at approximately 1 SNP per 2.5 kb. For the genes on chromosome 10, if they were under previously identified linkage peaks from linkage screens and they were differentially expressed between AD cases and controls at $p<10^{-10}$, the genes were selected. 667 SNPs in 28 candidate genes on chromosome 10 were selected at 1 SNP per 10kb. These 1536 SNPs were customized and multiplexed into one OPA (Oligo Pooled Assay) using Illumina Goldengate genotyping method. Detailed SNP selection and quality control criteria were described in previous chapter.

Statistical analysis

All SNPs were tested for the deviation from Hardy-Weinberg Equilibrium (HWE) in controls using the Haploview program (Barrett and Cardon, 2006). The SNPs were excluded if HWE $p<0.001$. TagSNPs were selected using $r^2>0.5$ as a threshold to exclude any SNPs that were in high LD by the Tagger option in Haploview.

Potential gene-gene interactions were identified using MDR. MDR performs an exhaustive search of all possible single-locus through k-locus interactions to evaluate all possible high/low risk models of disease. Briefly, an exhaustive list of n combinations are generated from the pool of all independent variables in step one. In step two, for $k = 1$ to M, the combinations are represented in k-dimensional space, and the number of cases and controls are counted in each multifactor cell. In step three, the ratio of cases to controls is calculated within each cell. In step four, each multifactor cell in the k-dimensional space is labeled as high risk (HR) if the ratio of affected individuals to

unaffected individuals exceeds a threshold (T = 1), and low risk (LR) if the threshold is not exceeded. In step five, the training accuracy is calculated. Training accuracy for each model is calculated based on the number of individuals within the model that are actually cases in genotype combinations classified as high risk and the number of individuals that are controls in the genotype combinations classified as low risk. In step six, this is repeated for each multifactor combination. In step seven, the model with the best training accuracy is selected and evaluated in the test set. In step eight, the testing accuracy of the model is estimated. Steps 1 through 8 are repeated for each possible cross-validation interval (e.g. 5 or 10 fold cross validation intervals). In step nine, a permutation test is conducted to determine the statistical significance of the model(s). These steps are illustrated in Figure 10-1 adapted from Ritchie et al. (Ritchie and Motsinger, 2005)

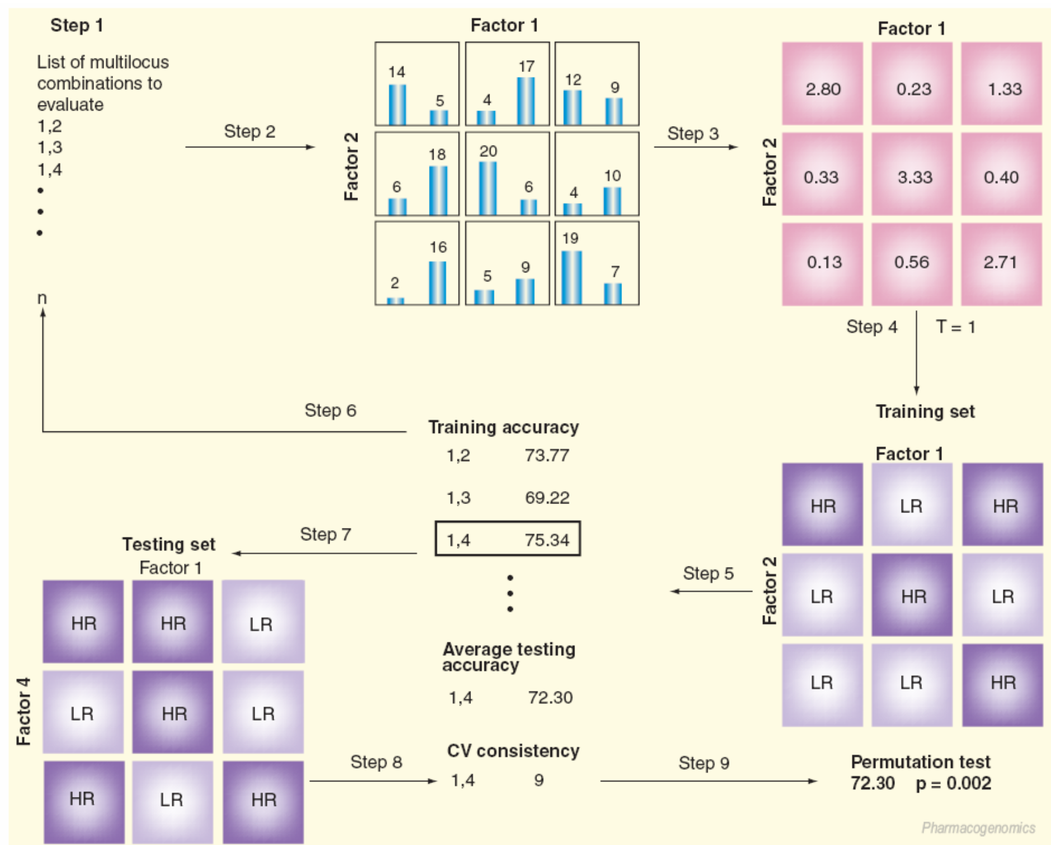


Figure 10-1 Summary of the general steps to implement the MDR method
Bars represent hypothetical distributions of cases (left) and controls (right) with each multifactor combination. Dark-shaded cells represent HR genotype combinations, while light-shaded cells represent LR genotype combinations. CV: Cross-validation; HR: High risk; LR: Low risk; MDR: Multifactor dimensionality reduction.

The statistical significance of the final best model is determined through permutation testing. Permutation testing involves creating 1000 permuted data sets by randomizing the disease status labels. The entire MDR procedure is repeated for each permuted data set, generating a distribution of 1000 prediction errors and cross-validation consistencies that could be expected by chance alone. The significance of the final model is determined by comparing the prediction error and cross validation consistency of the final model with the distribution. A p value is extracted for the model by its theoretical

location in the permutation distribution (Motsinger et al., 2006). In considering the possible unbalanced ratio of cases to controls in the data set, a balanced accuracy approach was used to evaluate the performance of the models (Velez et al., 2007).

In the present study, 1, 2, and 3 locus combinations were run for chromosome 10 only and cross-chromosome (all genomic convergence genes) models. Logistic regression (SAS program) was used to verify the interaction between genes in the identified model.

Results

Interaction among genes on chromosome 10

Overall data set

597 out of 667 genotyped SNPs on chromosome 10 passed the quality control process. There is a possibility that MDR method may pick up models containing different SNPs in different cross-validation intervals when there are SNPs in high linkage disequilibrium (representing the same effect) and results in the low cross validation consistency. Therefore, only SNPs in low LD were considered in the model. 300 tagSNPs were selected to represent the 597 SNPs at an $r^2>0.5$ threshold to test for gene-gene interaction in MDR.

Table 10-1 shows the results of the MDR analysis conducted for genes on chromosome 10. The single best model was found between two SNPs in CACNB2. This 2-locus model involves rs1277738 and rs10741083. The average prediction accuracy was 51.43% and the cross-validation consistency was 40%, using 5-fold cross-validation.

Because other models outperformed this model in other cross validation intervals, the p value was not significant in the permutation test (p=0.374). However, this model is in the top 10 models in 3 out 5 of the cross-validation intervals, being the top model 67% of the time.

We ran the whole MDR analysis 10 times using 10 different random seeds when splitting the data set. The best 2-locus model was identified as the best model in more than two cross-validation intervals with all the different splits of the data. Thus, the way to split the data set for training and test sets does not affect the selection of this model.

When we eliminate noise by using the "FORCELOCI" function in MDR and let the program do the permutation only on the two SNPs in the best 2-locus model, the permutation p value was significant (p=0.003). The result was confirmed in logistic regression by putting the genotypes of these two SNPs and the interaction between them in the model. The interaction was significant with p=0.004 in the model. Table 10-1 shows that SNP rs1277738 in CACNB2 was consistently identified across 1-, 2-, 3-locus models in MDR. It showed a main effect (best 1-locus model) in the MDR analysis. This SNP also had strongest genotypic effect in single marker association test (p=0.003).

Table 10-1 Summary of MDR results for genes on chromosome 10

# locus	Best model	Accuracy (%)	Cross validation	P
1-locus	CACNB2(rs1277738)	50.35	40%	0.636
2-locus	CACNB2(rs1277738) CACNB2(rs10741083)	51.43	40%	0.374
3-locus	CACNB2(rs1277738) CACNB2(rs2489214) CAMK1D(rs4750255)	49.96	20%	0.710

The low cross validation consistencies in the models could be due to locus heterogeneity or small effect of the genes. Several models may perform almost equally in different cross validation intervals. Looking at only the top model might not be sufficient to identify important genes in this complex disease. Therefore, we considered the frequency of each gene in the top ten MDR models ranked by prediction accuracy. Table 10-2 shows three most frequently identified genes in the top 10 MDR models. CACNB2 was in the top 10 models among all cross-validation intervals in the 1-, 3-locus models, and in 80% of cross-validations for 2-locus model. In more than 60% of cross validation intervals, PTPLA was in the top 10 MDR models. SORCS1 was in top 10 models among 40% of cross-validation intervals.

Female subset

Our previous study showed a strong effect of SORCS1 in a female subset. We performed MDR to find potential gene-gene interactions between this gene and other genes. Table 10-3 shows the results of MDR analysis in female subset. SNP rs17277986 in SORCS1 was identified as the single best model with prediction accuracy of 54.74% and 80% cross-validation consistency (p=0.098). This SNP was the one showing the strongest single marker allelic effect in the female data set. There isn't a good 2-locus model because none of the 2-locus models had prediction accuracy greater than 50%. The 3-locus model contains rs17277986 in SORCS1 and two SNPs (rs1409207 and rs2482100) in CACNB2 with 52.49% prediction accuracy and 40% cross-validation consistency. Although the permutation p was not significant at p<0.05, the SORCS1 gene was always in the top 10 MDR models among all cross-validation intervals for 1-,

2- locus models. CACNB2 was in the top 10 1-locus models among all cross-validation intervals and 40% of the cross-validation intervals for 2-, 3-locus models (Table 10-2).

Table 10-2 The frequency of 3 genes in the top 10 MDR models from 5 cross-validations

	1-locus model	2-locus model	3-locus model
Overall			
CACNB2	100%	80%	100%
PTPLA	80%	60%	60%
SORCS1	40%	40%	40%
Female			
CACNB2	100%	40%	40%
PTPLA	20%	40%	40%
SORCS1	100%	100%	80%

Table 10-3 Summary of MDR results in the female subset

# locus	Best model	Accuracy (%)	Cross validation	P
1-locus	SORCS1(rs17277986)	54.74	80%	0.098
2-locus	SORCS1(rs11193054) CACNB2(rs1277738)	48.67	20%	0.884
3-locus	SORCS1(RS17277986) CACNB2(rs1409207) CACNB2(rs2482100)	52.49	40%	0.297

Interaction among all genomic convergence genes in AD

To eliminate the possibility of getting low cross-validation consistency due to the presence of SNPs in high LD (as stated in the previous section) in the MDR method, 576 tagSNPs were selected at an $r^2 < 0.5$ level to represent the 1536 genotyped SNPs in all the genomic convergence genes. Table 10-4 shows best interaction models for all

genomic convergence genes in Alzheimer disease. We observed a statistically significant effect of ApoC1 (rs4420638 in 3'-UTR) with an average prediction accuracy of 69.04% ($p<0.001$). In addition, a statistically significant two-locus interaction between ApoC1 (rs4420638) and PTPLA (rs10508533) was detected, which predicts disease status correctly 67.98% of the time ($p<0.001$). A statistically significant 3-locus model was identified and involved ApoC1 (rs4420638), PRKG1 (rs10740413) and CUL2 (rs16935840) with prediction accuracy of 68.27%. However, the prediction accuracy and cross-validation consistency for the 2-, 3-locus models are worse than that in the one-locus model, suggesting the strongest main effect of ApoC1. Because there are "0"s in some of the contingency tables for 2-, 3-locus models, logistic regression is not appropriate for testing gene-gene interaction for the SNPs in the identified MDR models.

Table 10-4 Summary of MDR results among genomic convergence genes in AD

# locus	Best model	Accuracy (%)	Cross validation	P
1-locus	APOC1(rs4420638)	69.04	100%	<0.001
2-locus	APOC1(rs4420638) PTPLA(rs10508533)	67.98	40%	<0.001
3-locus	APOC1(rs4420638) PRKG1(rs10740413) CUL2(rs16935840)	68.27	40%	<0.001

Discussion

Alzheimer disease is a complex disease that is likely to be the result of many genetic and non-genetic factors. One of the biggest challenges in complex, common diseases, such as Alzheimer disease, is the identification of the relationship between genetic variations and disease risk. Epistasis, or gene-gene interaction, is crucial in detecting polymorphisms associated with an increased risk of disease (Moore, 2003; Thornton-Wells et al., 2004; Sing et al., 2004). In the present study, we applied MDR to investigate the potential associations between AD susceptibility and candidate genes coverged from linkage, gene expression and association studies.

MDR identified two SNPs (rs1277738 and rs10741083) in CACNB2 as having an interaction effect among genes on chromosome 10. These two SNPs were both intronic SNPs, one was in intron 2 and the other was in intron 5. They were 216 kb apart and not in linkage disequilibrium with each other (r^2=0.002). One SNP, rs1277738, had a very strong genotypic association (p=0.003) with AD and it was identified as having a main effect in MDR analysis. Another SNP, rs10741083 did not show single marker association with AD. However, the interaction effect between this SNP and the main effect SNP had higher prediction accuracy (Table 10-1), suggesting that the interaction effect was not absolutely driven by rs1277738, which had the main effect.

CACNB2 encodes a voltage-dependent calcium channel beta 2 subunit. The beta 2 subunit works as a complex with other subunit (e.g. alpha unit) and the protein complexes play pivotal roles in signal transduction and homeostasis processes (Opatowsky et al., 2003). The beta 2 subunit modulates calcium channel activity and enables trafficking by both chaperoning the voltage-dependent calcium channel complex

to the membrane and modulating gating (Yamaguchi et al., 1998; Brice et al., 1997; Bichet et al., 2000). It permits the flow of Ca^{2+} ions through the cellular membrane as a function of membrane potential. The polymorphisms in the introns may affect the binding site that binds alpha unit. If the interaction between the alpha and beta subunit is affected, the function of the whole protein complex is subsequently affected. In addition, CACNB2 is involved in MAPK signaling pathway. A downstream protein, c-Jun N-terminal Kinase interacting protein 1 (JIP1), in this pathway is suggested to be involved in AD by regulating APP neuronal transportation and controlling APP processing (Matsuda et al., 2003).

Interpreting results is a challenge in exploring gene-gene interactions. MDR analysis on genes on chromosome 10 had low cross-validation consistency on the best model (two SNPs in CACNB2), although the interaction is significant in the logistic regression analysis when noise was eliminated from the data set. It may be the case that multiple models are all good predictors of disease risk. If a gene is always in the top MDR models, it also suggests that the gene may have an effect. We discovered three genes (CACNB2, PTPLA, and SORCS1) that showed consistent evidence of association with AD in our previous studies. MDR identified CACNB2 as the single best model, but not PTPLA or SORCS1. By considering the frequencies of specific models in the top ten MDR models, we were able to find that these three genes were in the top ten MDR models most of the time. This is another way to identify important effects in the disease pathology. The data suggest that each gene has a modest effect and that an extensive epitasis or gene-gene interaction is underlying the disease etiology.

When we tested potential gene-gene interactions among all genomic convergence genes for Alzheimer disease, MDR identified a statistically significant main effect and 2-, 3-locus models ($p<0.001$). All the models contained ApoC1, which is a surrogate for the known effect of the ApoE locus ($r^2=0.74$, distance between two SNPs is 11kb). Although the permutation p values for 2-, 3-locus models were significant ($p<0.001$), the cross-validation consistency and prediction accuracy of each model were worse than that in the one-locus model. This may suggest that the interaction is mainly driven by the main effect from ApoC1, a surrogate of ApoE.

Due to the limits of the sample size, there are empty cells in the contingency tables for different genotype combinations for 2-, 3- locus interactions. It would be beneficial to conduct high-order gene-gene interaction analysis using larger sample size of replication data set.

CHAPTER XI

CONCLUSIONS AND FUTURE DIRECTIONS

Summary and conclusions

Knowledge of direct Mendelian genetic causes of Alzheimer disease and genes that predispose to late-onset Alzheimer disease has led to greater understanding of the pathophysiology of the neurodegenerative process.

In the present study, we located a major linkage peak in the region between 40 to 60 Mb and a minor peak between 80 to 100 Mb by linkage analysis using a large family-based data set. Five previously reported candidate genes also showed association in our study population in the candidate gene study. After we compared gene expression levels between AD cases and controls, we converged linkage and gene expression data from Serial Analysis of Gene Expression (SAGE) and identified and examined 28 candidate genes on chromosome 10. We investigated the potential associations between candidate genes on chromosome 10 and the risk of Alzheimer disease (AD) using a clinically well-defined case-control data set to uncover multiple possible associations with AD.

To our knowledge, the present study is the first comprehensive investigation of the allelic, genotypic and haplotypic association together with an investigation of potential gene-gene interactions using a set of genes and SNPs identified through the genomic convergence of linkage, candidate gene and gene expression studies.

Several approaches were used to control for false-positive results. First, genes were selected by genomic convergence, so that selected genes had a substantial *a priori*

likelihood of the involvement in AD pathology. Second, we applied an analytical convergence approach to interpret our findings based on the assumption that SNPs showing significant results in multiple tests have higher likelihood of being important in the disease process. We looked for the convergence of results across several methods, rather than relying on results from a single analytic tool.

The most promising findings in this study were the significant associations found in PTPLA and SORCS1. The associations were consistent among all association analysis (allelic, genotypic and haplotypic association) and these two genes were also seen in the gene-gene effect in MDR analysis. CACNB2 had the strongest genotypic effect and was the best model in gene-gene interaction test in MDR (Figure 11-1).

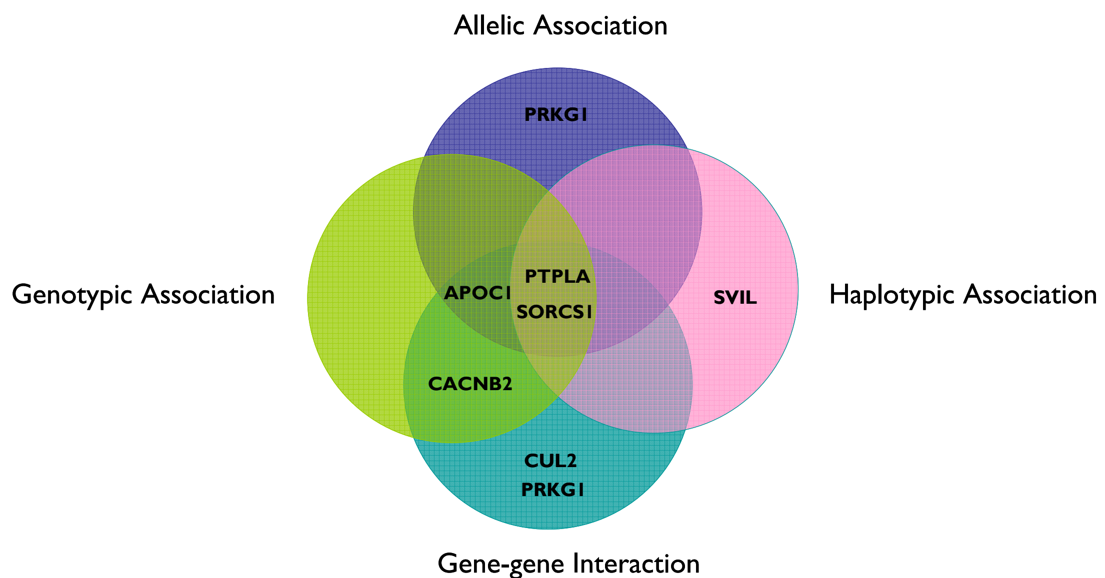


Figure 11-1 Analytical convergence in Alzheimer disease
Allelic, genotypic, haplotypic association and gene-gene interaction analysis are shown in the figure.

Figure 11-2 shows the potential biological relevance of these three genes in AD. In Alzheimer disease, the amyloid precursor protein (APP) is cut to generate Amyloid-β.

Amyloid-β accumulates and aggregates to form an oligomer and cause apoptosis which can cause AD (Hardy and Higgins, 1992). The SORLA gene has been associated with AD and it inhibits the generation of Amyloid-β (Rogaeva et al., 2007). SORCS1 is a homolog of SORLA and might have a similar function as SORLA to inhibit the generation of Aβ. CACNB2 is a calcium channel protein. It affects the calcium level that could cause mitochondrial damage and then induce apoptosis (Finlin et al., 2006). PTPLA is a phosphatase, but relatively little is known about its biological function. Based on its phosphatase function, we might conjecture that it is involved in the phosphorylation of the Tau protein; phosphorylated Tau forms neurofibrillary tangles, which is one of the hallmarks of AD. PTPLA might also be involved in the phosphorylation of GSK-3β protein which binds β-catenin. β-catenin is associated with elevated Aβ 42 (Prager et al., 2007; Shim et al., 2007). β-catenin can also bind to transcription factor and induce unscheduled cell cycle, which cause AD (Figure 11-2).

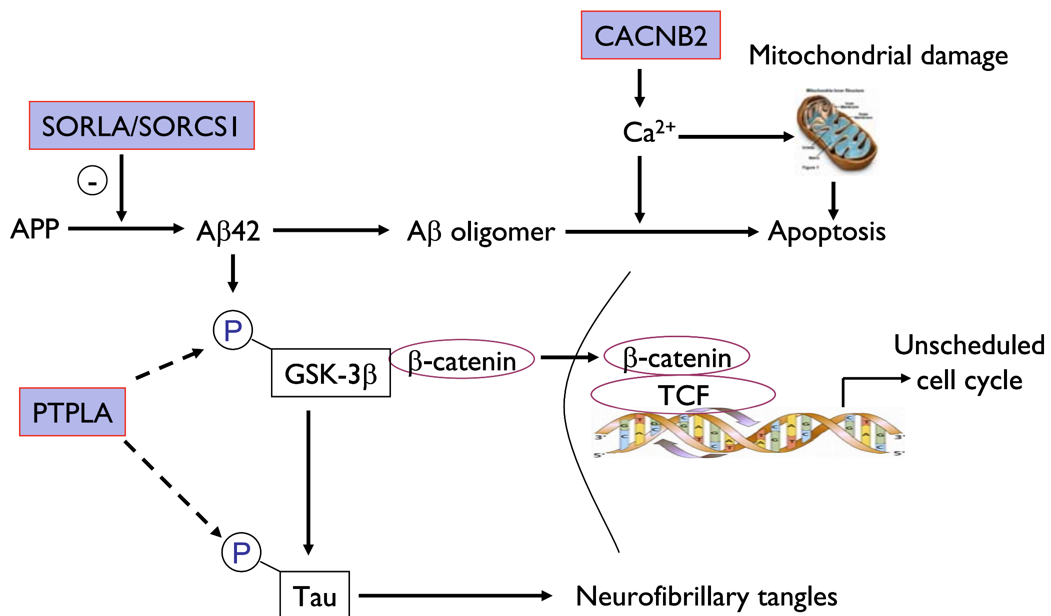


Figure 11-2 Hypothesized pathways involved in Alzheimer Disease

In conclusion, this study suggests that genetic variations in PTPLA, SORCS1 and CACNB2 genes might alter the risk for Alzheimer disease by affecting multiple pathways. Furthermore, these findings support the hypothesis that the etiology of Alzheimer disease is complex and that complex interactions may contribute to Alzheimer disease risk.

Future directions

In the future, larger sample sizes will be necessary to increase power and to develop independent confirmation datasets. To dissect the complex etiology of heterogeneous common human diseases (e.g. Alzheimer disease), finding the most homogeneous subset is a critical factor. With the detailed clinical information, larger sample sizes will give us larger subsets and offset the power issues present with current subsets of the data. The rich phenotypic information of the data set could help us effectively analyze the rapidly generated genotypes.

Due to the genetic (locus and allelic) heterogeneity and the phenotypic heterogeneity underlying the complex diseases and the false positive results, replication studies are very important to confirm a finding. The replication of the associated SNPs in the suggested genes in independent datasets will help us to confirm the associations found in the current data set. Genotyping more SNPs in the interesting genes is also a possible follow-up study, which is made possible by the availability of the large number of SNPs in publicly available databases (e.g. HapMap, Perlegen, dbSNP).

Over the past twenty years, technology is improving at a great rate. As genotyping throughput increases and costs per genotyping decrease dramatically, genome-wide association studies are underway and have the potential to detect genes

with small to moderate effect without *a priori* knowledge about the gene function or chromosome location.

Genetic variation in the human genome takes many forms, ranging from large, chromosome anomalies to single-nucleotide changes. Recently, multiple studies have discovered an abundance of submicroscopic copy number variation (CNV) of DNA segments ranging from kilobases (kb) to megabases (Mb) in size (Iafrate et al., 2004; Sebat et al., 2004; Sharp et al., 2005). CNVs, including deletions, insertions, duplications and complex multi-site variants (Fredman D, 2004), influence gene expression, phenotypic variation and adaptation by disrupting genes and altering gene dosage, and can cause disease or confer risk to complex disease traits. A recent study reported that duplication of the APP locus on chromosome 21 causes early-onset Alzheimer disease with cerebral amyloid angiopathy (Rovelet-Lecrux et al., 2006). Studies assessing genome-wide CNVs in Alzheimer disease using large-scale technology could help us to understand the underlying mechanisms of the gene functions in the disease.

Epigenetic variations may have a great affect on particular genotypes as they relate to disease risk. Epigenetic variation may be a necessary and specific target for environmental influences. Alzheimer disease (AD) is among the few diseases that may display high homocysteine (HCY) and low B12 and folate in blood. DNA methylations in elderly are consistently lower than in young and mid-aged people (Scarpa et al., 2006). These observations have raised the suspicion that Aβ overproduction and accumulation, which may be the cause of the disease, could be due to the loss of epigenetic control in the expression of the genes involved in APP processing. Investigation of the unknown

and unexamined epigenetic differences may make it clearer in studying gene-environment interactions in AD.

Additional technological advances should be made to handle the large amount of data being generated. Advances in genotyping technology have far outpaced those in statistical and computational methods for analyzing genetic data. Thornton-Wells et al. suggested that a haplotype in LRRTM3 might be a good stratifier to subset the data set using the clustering analysis. Both the clustering method and the haplotype in LRRTM3 could be used to subset the data set and test association in each of the clusters. Applications of the new methods (e.g. subsetting data set, detecting gene-gene, and gene-environment interactions) are promising to find genetic variations altering disease risk.

It is a very plausible hypothesis that genes in a common biological pathway may interact to yield disease; however, the complexity of such a pathway is likely greater than our current understanding. Our present study suggested that multiple genes in multiple pathways are involved in the Alzheimer disease. Strategies to tease apart both locus and allelic heterogeneity must be developed and will likely play a key role in examination of complex genetic disease.

In addition, further experiments to locate the susceptible variants in the associated genes and assess the biological relevance of the genes using knock-down/knock-out cell lines and transgenic animal models would give a better understanding of the pathophysiology of AD.

APPENDIX

Complete linkage and candidate genes studies results

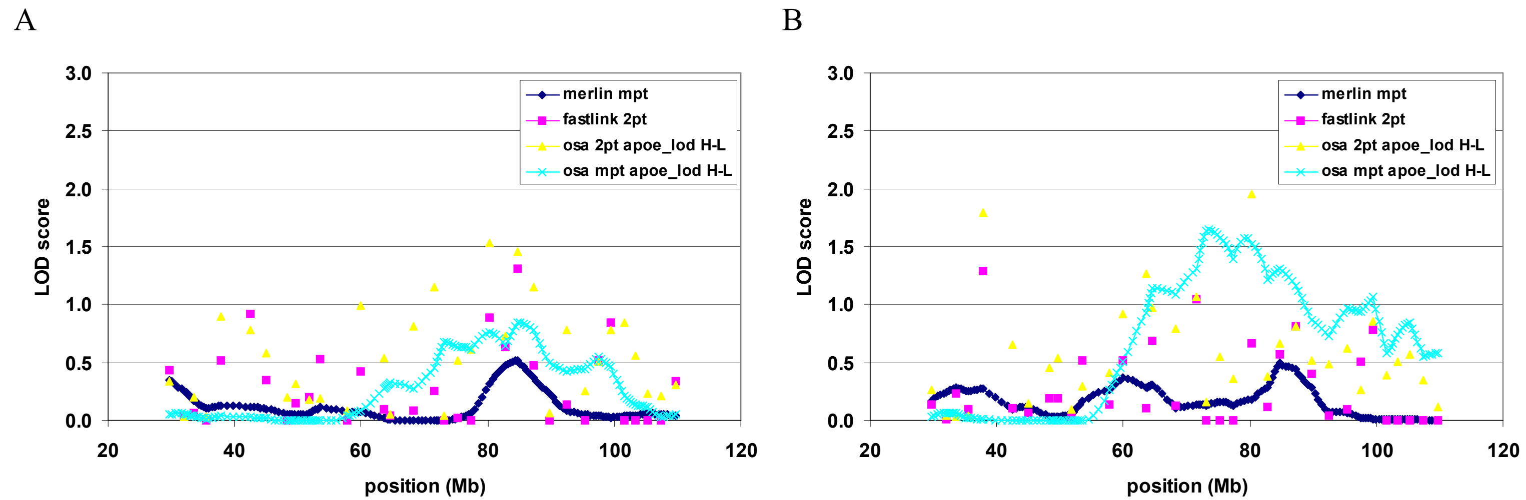
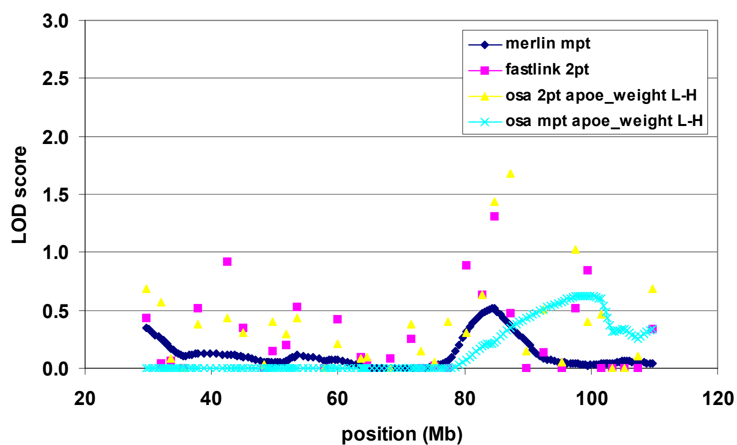


Figure A-1 Two point and multipoint linkage analyses using the ApoE LOD score as a covariate to order families from high to low in the overall and autopsy subset
A. Overall data set B. Autopsy-confirmed subset

A



B

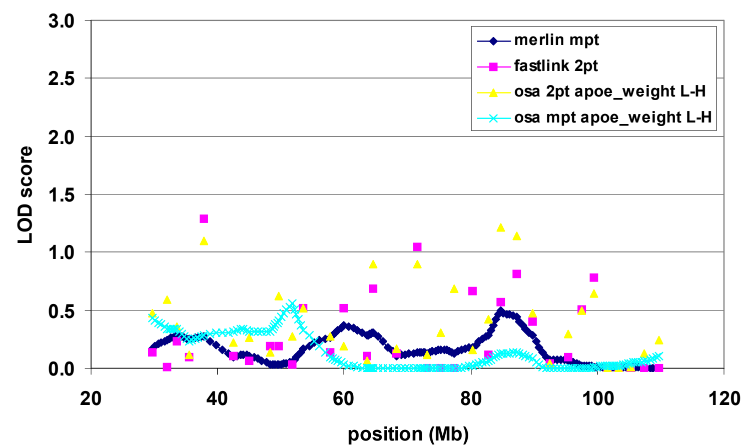
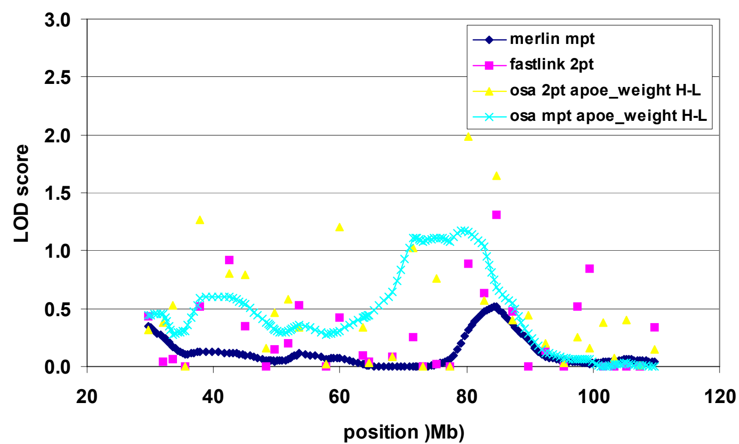


Figure A-2 Two point and multipoint linkage analyses using the ApoE weight as a covariate to order families from low to high in the overall and autopsy subset
A. Overall data set B. Autopsy-confirmed subset

A



B

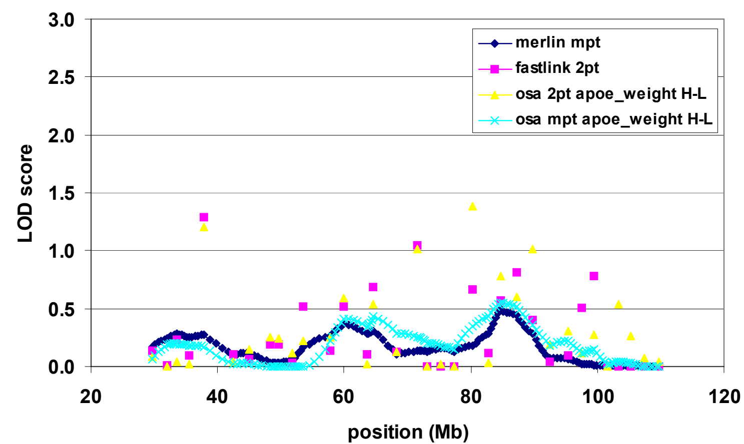


Figure A-3 Two point and multipoint linkage analyses using the ApoE weight as a covariate to order families from high to low in the overall and autopsy subset

A. Overall data set B. Autopsy-confirmed subset

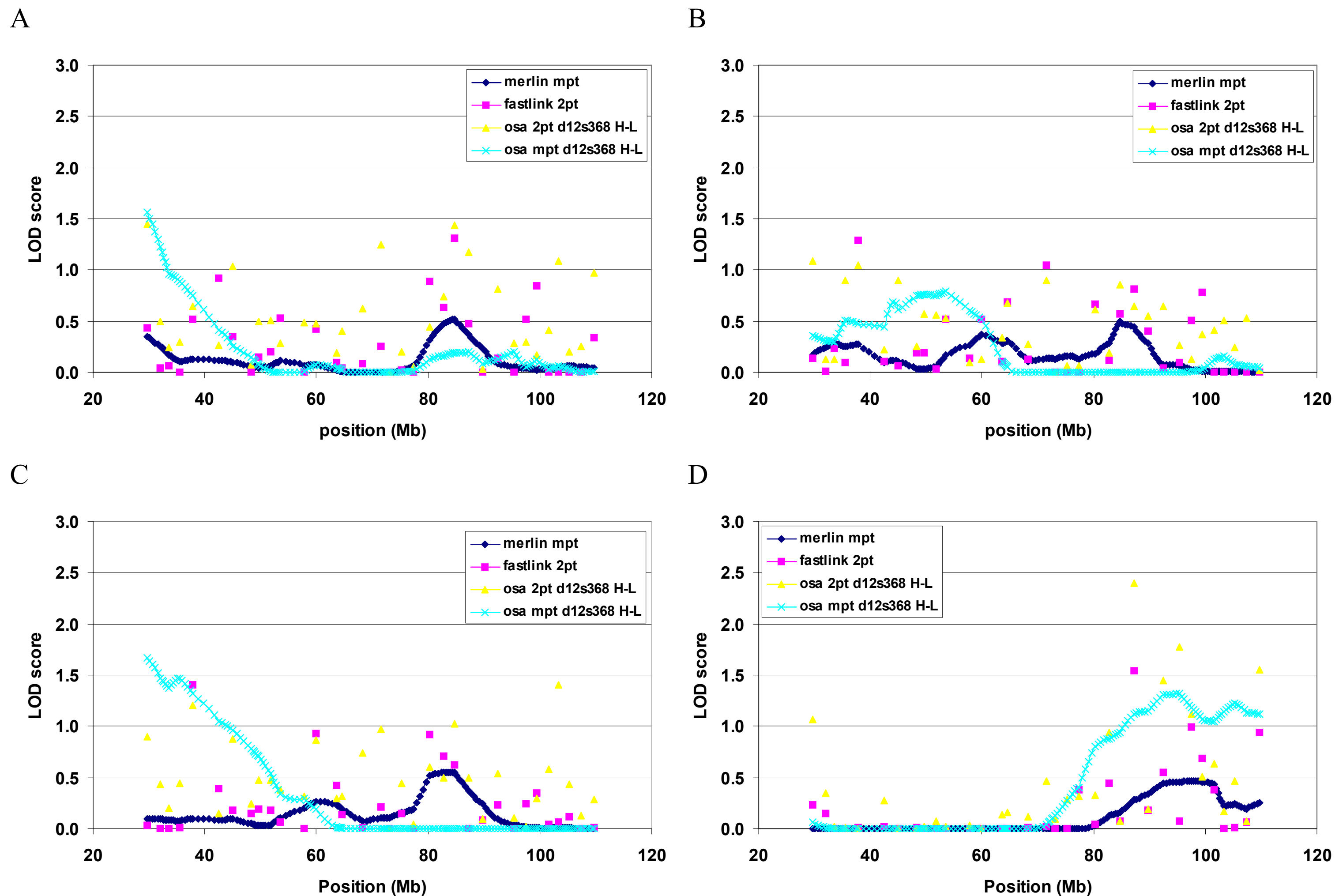


Figure A-4 Two point and multipoint linkage analyses using the D12S368 LOD score as a covariate to order families from high to low in the overall and three subsets

A. Overall data set B. Autopsy-confirmed subset C. ApoE ε4-positive subset D. ApoE ε4-negative subset

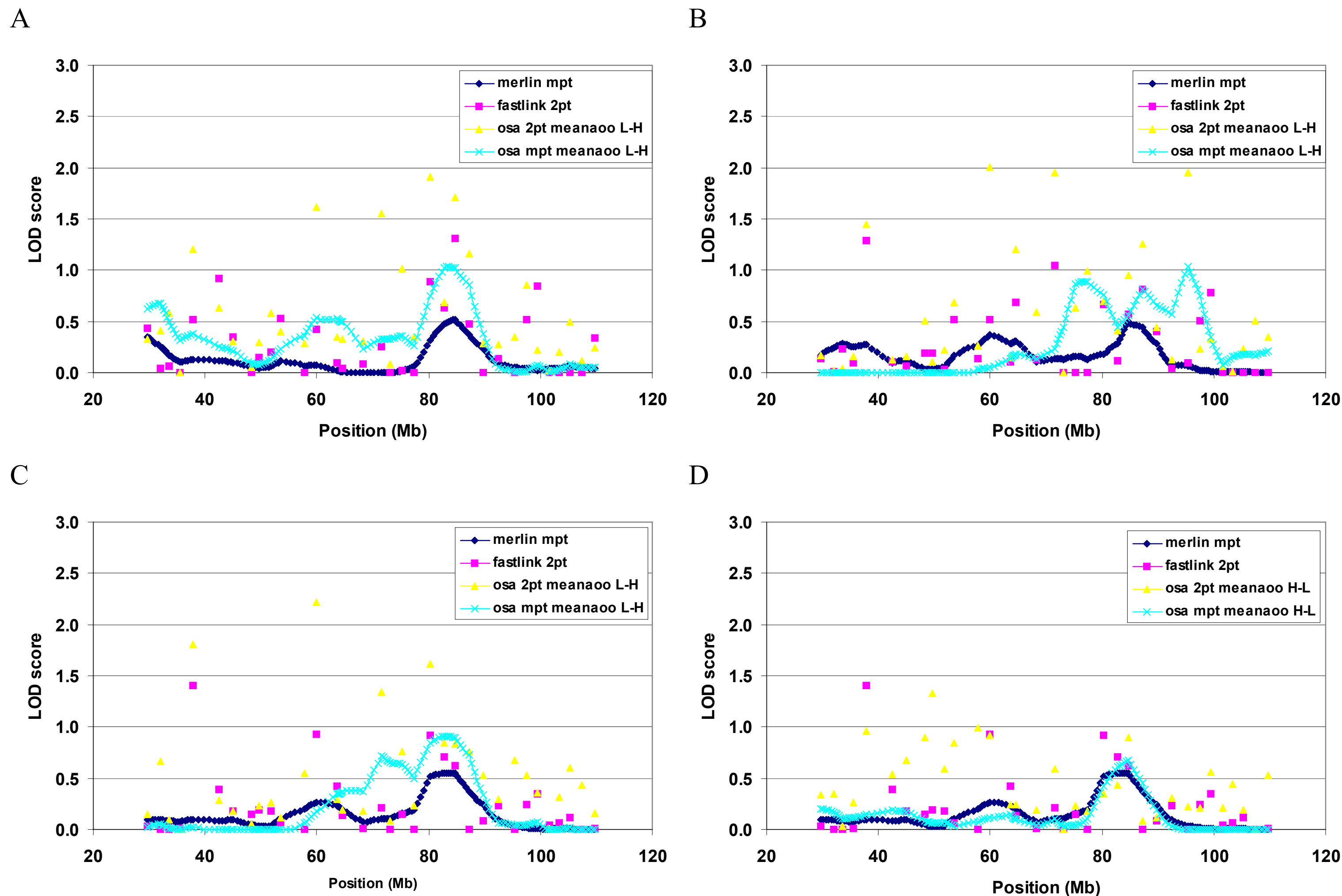


Figure A-5 Two point and multipoint linkage analyses using the Mean Age-of-onset (aoo) as a covariate to order families from low to high in the overall and three subsets

A. Overall data set B. Autopsy-confirmed subset C. ApoE ε4-positive subset D. ApoE ε4-negative subset

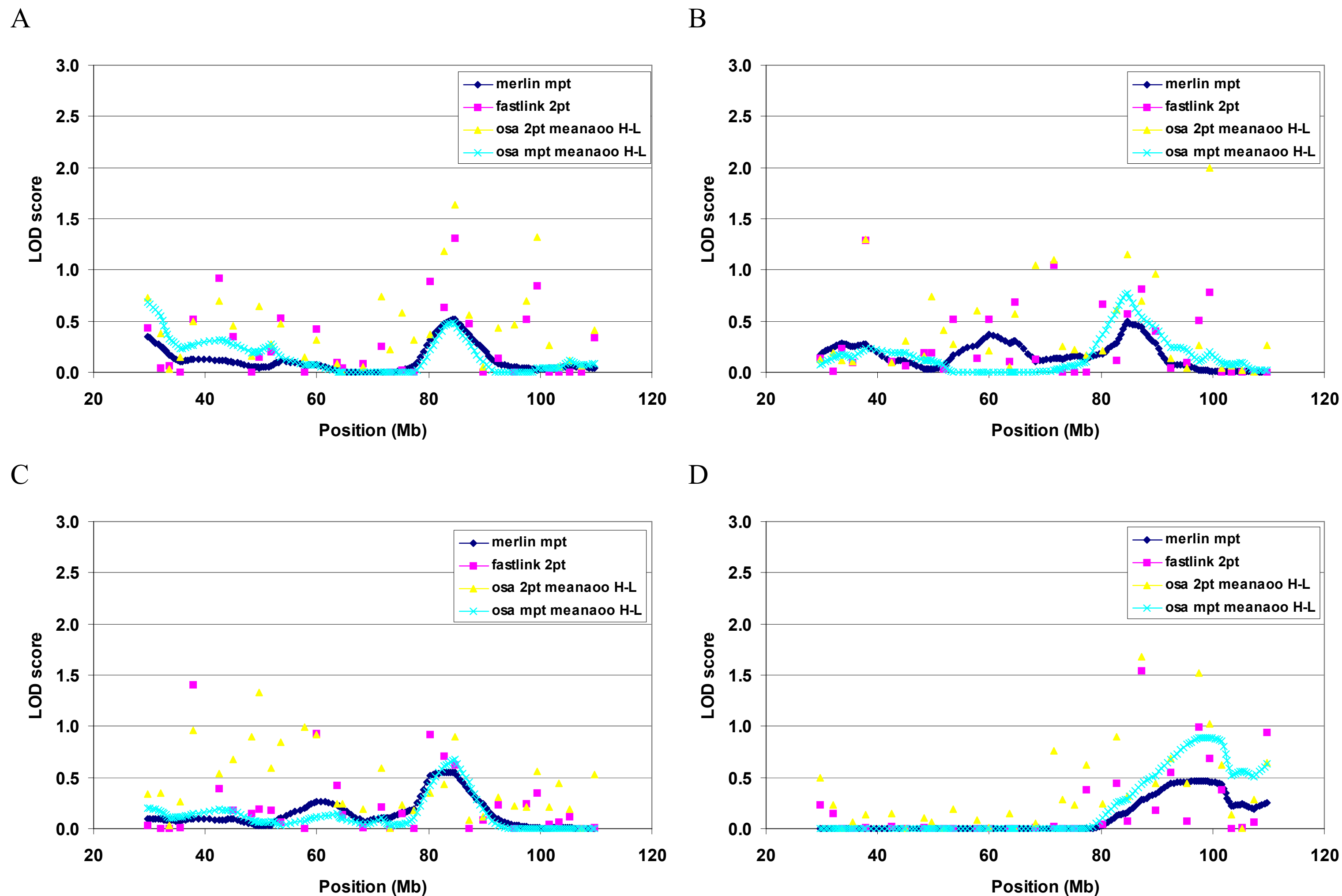


Figure A-6 Two point and multipoint linkage analyses using the Mean Age-of-onset as a covariate to order families from high to low in the overall and three subsets

A. Overall data set B. Autopsy-confirmed subset C. ApoE ε4-positive subset D. ApoE ε4-negative subset

Table A-1 Allele distribution of SNPs in five candidate genes in case-control data set

SNP	GENE	Allele	Total	Cases	Controls	Odds ratio (95% CI)	p-value
RS1786927	**VR22**	100	974	405	569	1.08	0.3784
		300	1394	605	789	(0.91,1.27)	
RS2126750	**VR22**	10	519	279	240	1.13	0.2713
		40	861	489	372	(0.91,1.41)	
RS7911820	**VR22**	100	864	382	482	0.91	0.2638
		200	1562	654	908	(0.77,1.07)	
RS12357560	**VR22**	100	1834	783	1051	1.02	0.8016
		300	552	239	313	(0.85,1.24)	
RS7070570	**VR22**	1	1215	662	553	1.18	0.1332
		2	447	262	185	(0.95,1.47)	
RS7074454	**VR22**	1	917	503	414	0.92	0.4792
		2	501	265	236	(0.74,1.15)	
RS10822719	**VR22**	200	484	193	291	1.15	0.1653
		400	1972	855	1117	(0.94,1.41)	
RS6480140	**VR22**	1	554	292	262	1.12	0.3156
		2	866	480	386	(0.90,1.38)	
RS922347	**VR22**	100	1464	621	843	1.01	0.9389
		300	876	373	503	(0.85,1.19)	
RS4463744	**VR22**	100	1793	766	1027	1.00	0.9440
		400	681	292	389	(0.84,1.20)	
RS2441718	**VR22**	100	1650	728	922	**0.83**	**0.0369**
		300	714	282	432	**(0.69,0.99)**	
RS2939947	**VR22**	100	1279	538	741	0.95	0.1942
		300	105	494	611	(1.11,1.31)	
RS2456737	**VR22**	100	2091	908	1183	**0.76**	**0.0168**
		300	375	138	237	**(0.60,0.95)**	
RS997224	**VR22**	1	268	160	108	0.95	0.7102

		2	890	520	370	(0.72,1.25)	
RS746606	**VR22**	200	1579	660	919	1.07	0.4430
		400	871	378	493	(0.90,1.26)	
RS7909676	**VR22**	100	1206	524	682	0.94	0.4941
		200	1132	476	656	(0.80,1.11)	
RS11593235	**VR22**	200	880	388	492	0.97	0.7588
		400	1386	602	784	(0.82,1.15)	
RS10997591	**VR22**	200	1478	644	834	1.04	0.6898
		400	792	352	440	(0.87,1.23)	
RS7903421	**VR22**	100	2320	997	1323	0.80	0.2089
		300	146	55	91	(0.57,1.13)	
RS3096244	**VR22**	200	862	363	499	1.13	0.1617
		400	1412	637	775	(0.95,1.34)	
RS1001016	**LRRTM3**	200	316	127	189	1.12	0.3350
		300	2108	905	1203	(0.88,1.42)	
RS12769870	**LRRTM3**	100	1260	506	754	**1.18**	**0.0463**
		300	1062	470	592	**(1.00,1.40)**	
RS1925583	**LRRTM3**	300	1333	558	775	1.10	0.2577
		400	1101	486	615	(0.93,1.29)	
RS2394314	**LRRTM3**	100	1112	490	622	0.91	0.2478
		300	1332	556	776	(0.77,1.07)	
RS10762122	**LRRTM3**	200	746	312	434	1.05	0.6014
		400	1690	726	964	(0.88,1.25)	
RS942780	**LRRTM3**	200	472	196	276	1.06	0.5835
		400	1962	842	1120	(0.86,1.30)	
RS1925617	**LRRTM3**	300	1084	452	632	1.09	0.3003
		400	1352	592	760	(0.93,1.28)	
RS1925622	**LRRTM3**	100	1392	600	792	0.98	0.7585
		300	1078	458	620	(0.83,1.15)	
RS1925632	**LRRTM3**	100	1082	441	641	1.12	0.1800

		200	1346	585	761	(0.95,1.31)	
RS1952060	**LRRTM3**	200	1146	494	652	0.96	0.6217
		400	1320	556	764	(0.82,1.13)	
RS2147886	**LRRTM3**	200	1367	594	773	0.90	0.1814
		400	1089	444	645	(0.76,1.05)	
RS2251000	**LRRTM3**	100	1093	444	649	1.12	0.1610
		300	1335	580	755	(0.95,1.32)	
RS2764807	**LRRTM3**	200	1185	515	670	0.93	0.3834
		400	1287	537	750	(0.79,1.09)	
RS10762136	**LRRTM3**	100	1252	502	750	1.09	0.3192
		200	1092	460	632	(0.92,1.28)	
RS1916341	**PLAU**	300	1011	433	578	0.97	0.7359
		400	1445	609	836	(0.83,1.14)	
RS2227564	**PLAU**	200	1915	802	1113	1.17	0.1115
		400	529	242	287	(0.96,1.42)	
RS2227566	**PLAU**	200	1005	429	576	1.00	0.9872
		400	1449	619	830	(0.85,1.18)	
RS2227568	**PLAU**	200	2023	880	1143	0.85	0.1746
		400	363	144	219	(0.68,1.07)	
RS4065	**PLAU**	200	966	423	543	0.95	0.5288
		400	1466	623	843	(0.81,1.12)	
RS1800682	**TNFRSF6**	100	1281	548	733	1.02	0.8097
		300	1165	504	661	(0.87,1.20)	
RS1324551	**TNFRSF6**	200	1198	505	693	1.05	0.5501
		400	1248	541	707	(0.89,1.23)	
RS2031612	**TNFRSF6**	200	973	387	586	**1.22**	**0.0173**
		400	1481	661	820	**(1.04,1.44)**	
RS2296600	**TNFRSF6**	200	1472	655	817	0.83	**0.0220**
		300	972	387	585	**(0.70,0.97)**	
RS2251101	**IDE**	200	648	287	361	0.94	0.4966

		400	1622	693	929	(0.78,1.13)	
RS2251101	**IDE**	200	693	299	394	0.98	0.7927
		400	1741	741	1000	(0.82,1.17)	
RS7076966	**IDE**	200	1292	553	739	0.96	0.6005
		400	1152	481	671	(0.82,1.13)	
RS4646954	**IDE**	100	180	77	103	1.01	0.9690
		300	2262	971	1291	(0.74,1.37)	
RS3758505	**IDE**	100	2236	958	1278	0.96	0.7758
		200	182	76	106	(0.70,1.30)	
RS7099761	**IDE**	200	1200	501	699	1.02	0.7976
		400	1228	519	709	(0.87,1.20)	
RS1544210	**IDE**	100	1052	430	622	0.96	0.6041
		300	1196	476	720	(0.81,1.13)	

Table A-2 Family-based association test in overall data set

SNP	GENE	sum PDT	geno PDT	Fbat	HWE Affecteds	HWE Normals
RS1786927	VR22	0.925	0.962	0.321	0.3356	1.0000
RS2126750	VR22	0.265	0.588	0.465	0.5256	1.0000
RS7911820	VR22	**0.028**	0.123	0.076	0.3146	0.9175
RS12357560	VR22	0.106	0.072	0.189	0.6584	0.4100
RS7070570	VR22	0.262	0.282	0.499	0.4441	0.3825
RS7074454	VR22	**0.009**	0.066	0.070	0.2769	0.2859
RS10822719	VR22	0.582	0.662	0.800	0.7175	0.0463
RS6480140	VR22	0.601	0.134	0.818	0.3297	1.0000
RS922347	VR22	0.476	0.665	0.806	0.9428	0.9206
RS4463744	VR22	0.537	0.380	0.782	1.0000	0.7081
RS2441718	VR22	0.588	0.875	0.472	0.1200	0.2806
RS2939947	VR22	0.249	0.329	0.493	0.6334	0.8409
RS2456737	VR22	0.583	0.844	0.500	0.1678	0.5453
RS997225	VR22	0.582	0.587	0.165	**0.0472**	0.8634
RS4746606	VR22	0.703	0.680	0.314	0.3906	0.4656
RS7909676	VR22	0.551	**0.043**	0.754	0.1147	0.6150
RS11593235	VR22	0.465	0.204	0.351	0.1722	0.7516
RS10997591	VR22	0.680	0.053	0.889	0.8566	0.5778
RS7903421	VR22	0.148	0.247	0.319	0.2394	0.5953
RS3096244	VR22	0.474	0.535	0.598	0.4484	0.3738
RS1001016	LRRTM3	0.727	0.120	0.258	0.4191	0.0403
RS12769870	LRRTM3	0.271	0.467	0.105	0.0025	0.9244
RS1925583	LRRTM3	0.531	0.835	0.479	0.0187	0.4987
RS2394314	LRRTM3	0.351	0.588	0.324	0.0881	0.5022
RS10762122	LRRTM3	0.908	0.106	0.942	0.0725	0.9016
RS942780	LRRTM3	0.099	0.294	0.125	0.3316	0.6563
RS1925617	LRRTM3	0.559	0.002	0.086	0.6300	0.1438
RS1925622	LRRTM3	0.267	0.009	0.043	0.5259	0.2359
RS1925632	LRRTM3	0.356	0.142	0.105	0.2131	1.0000
RS1952060	LRRTM3	0.756	0.000	0.414	0.0228	0.4997
RS2147886	LRRTM3	0.584	0.189	0.211	0.1928	0.7609
RS2251000	LRRTM3	0.359	0.126	0.133	0.1875	0.9197
RS2764807	LRRTM3	0.353	0.492	0.382	0.3969	0.9197
RS10762136	LRRTM3	0.479	0.292	0.599	0.2765	1.0000
RS1916341	PLAU	0.062	0.136	0.051	0.9369	0.1559
RS2227564	PLAU	0.635	0.727	0.675	0.2669	0.9009
RS2227566	PLAU	0.040	0.128	0.022	0.8775	0.2688
RS2227568	PLAU	0.020	0.016	0.084	0.0291	1.0000
RS4065	PLAU	0.174	0.308	0.026	0.6188	0.2813

RS1800682	TNFRSF6	0.873	0.984	0.458	0.8134	0.8478
RS1324551	TNFRSF6	0.924	0.795	0.283	0.6625	0.6359
RS2031612	TNFRSF6	0.867	0.951	0.900	0.4691	1.0000
RS2296600	TNFRSF6	0.910	0.959	0.852	0.4663	1.0000
RS2251101	IDE	0.391	0.350	0.880	0.4875	0.5391
RS2251101/ IDE-7 (2)	IDE	0.483	0.301	0.914	0.8600	1.0000
RS7076966	IDE	0.678	0.867	0.882	1.0000	1.0000
RS4646954	IDE	1.000	0.958	0.417	0.6615	0.7463
RS3758505	IDE	0.955	0.998	0.405	0.6134	0.7428
RS7099761	IDE	0.421	0.764	0.558	1.0000	0.8388
RS1544210	IDE	0.658	0.546	0.491	0.8906	0.6772

Table A-3 Family-based association test in Duke subset (N=147)

SNP	GENE	sum PDT	geno PDT	Fbat	HWE Affecteds	HWE Normals
RS1786927	VR22	0.625	0.498	0.825	0.3169	1.0000
RS2126750	VR22	0.382	0.058	0.987	0.2956	0.1556
RS7911820	VR22	0.215	0.339	0.707	0.4884	0.5372
RS12357560	VR22	0.245	0.109	0.327	0.0775	0.1325
RS7070570	VR22	0.125	**0.034**	0.254	**0.0075**	0.7897
RS7074454	VR22	0.104	0.272	0.345	1.0000	**0.0084**
RS10822719	VR22	0.838	0.836	0.374	0.3269	0.0713
RS6480140	VR22	0.637	0.841	0.559	0.6863	0.8247
RS922347	VR22	0.229	0.452	0.423	0.7169	**0.0356**
RS4463744	VR22	0.899	0.805	0.411	0.3488	0.6375
RS2441718	VR22	**0.026**	0.112	0.096	1.0000	0.4766
RS2939947	VR22	0.099	0.313	0.173	0.2625	0.8534
RS2456737	VR22	0.231	0.389	0.243	1.0000	0.7013
RS997225	VR22	0.821	0.659	0.656	**0.0194**	0.4875
RS4746606	VR22	0.760	0.952	0.503	0.5856	0.1669
RS7909676	VR22	0.298	0.527	0.489	0.8575	1.0000
RS11593235	VR22	0.245	0.306	0.846	0.2553	0.8447
RS10997591	VR22	0.653	0.310	0.977	0.7053	0.8322
RS7903421	VR22	0.439	0.281	**	1.0000	0.0969
RS3096244	VR22	0.510	0.579	0.729	0.2481	0.6769
RS1001016	LRRTM3	0.514	0.745	0.670	1.0000	0.3065
RS12769870	LRRTM3	0.744	0.761	0.276	0.4320	0.8447
RS1925583	LRRTM3	0.534	0.469	0.664	0.4153	1.0000
RS2394314	LRRTM3	0.459	0.524	0.577	0.7116	1.0000
RS10762122	LRRTM3	1.000	0.090	0.733	0.7125	0.8250
RS942780	LRRTM3	0.139	0.246	0.028	0.8641	1.0000
RS1925617	LRRTM3	0.947	0.086	0.762	0.3434	0.0553
RS1925622	LRRTM3	1.000	0.089	0.780	0.2613	0.0378
RS1925632	LRRTM3	0.817	0.225	0.419	0.7156	0.4537
RS1952060	LRRTM3	1.000	0.007	0.241	0.7078	0.2775
RS2147886	LRRTM3	0.764	0.266	0.325	0.0303	0.6919
RS2251000	LRRTM3	0.614	0.202	0.501	0.0188	0.5763
RS2764807	LRRTM3	0.336	0.239	0.388	0.1003	0.8650
RS10762136	LRRTM3	0.851	0.132	0.168	0.0753	0.8547
RS1916341	PLAU	0.604	0.560	0.769	0.3709	0.4463
RS2227564	PLAU	0.797	0.597	0.947	0.1516	0.3500
RS2227566	PLAU	0.895	0.744	0.755	0.2869	0.7178
RS2227568	PLAU	0.560	0.665	0.676	0.2850	0.2634
RS4065	PLAU	1.000	1.000	0.467	0.5788	0.4553
RS1800682	TNFRSF6	0.682	0.027	0.649	0.7328	0.1425

RS1324551	TNFRSF6	0.799	0.189	0.294	0.1625	0.4575
RS2031612	TNFRSF6	1.000	0.247	0.848	0.3516	0.6984
RS2296600	TNFRSF6	0.823	0.262	0.622	0.4378	0.6903
RS2251101	IDE	0.371	0.661	0.579	0.3981	0.7972
RS2251101/ IDE-7(2)	IDE	0.371	0.672	0.576	0.6763	0.6147
RS7076966	IDE	0.192	0.549	0.474	0.4772	0.5709
RS4646954	IDE	0.912	0.443	0.816	0.2400	1.0000
RS3758505	IDE	0.646	0.548	0.886	0.1881	1.0000
RS7099761	IDE	0.475	0.387	0.669	0.3775	0.3606
RS1544210	IDE	0.245	0.455	0.233	0.4675	0.8438

**Less than 2 major alleles

Table A-4 Family-based association test in Vanderbilt subset (N=113)

SNP	GENE	sum PDT	geno PDT	Fbat	HWE Affecteds	HWE Normals
RS1786927	VR22	0.736	0.562	0.502	0.3169	0.1888
RS2126750	VR22	*	*	*	*	*
RS7911820	VR22	0.726	0.715	0.743	0.2031	0.4184
RS12357560	VR22	0.540	0.823	0.563	1.0000	0.7628
RS7070570	VR22	0.366	0.658	0.513	1.0000	0.5672
RS7074454	VR22	*	*	*	*	*
RS10822719	VR22	0.800	0.956	0.628	1.0000	1.0000
RS6480140	VR22	*	*	*	*	*
RS922347	VR22	0.424	0.468	0.976	1.0000	0.3128
RS4463744	VR22	0.827	0.895	0.950	0.5806	0.2356
RS2441718	VR22	0.206	0.324	0.204	0.7656	0.7872
RS2939947	VR22	0.257	0.355	0.511	0.4891	0.1122
RS2456737	VR22	0.777	0.337	0.953	0.6781	0.6900
RS997225	VR22	*	*	*	*	*
RS4746606	VR22	0.322	0.533	0.390	0.8078	1.0000
RS7909676	VR22	**0.033**	0.093	**0.018**	1.0000	0.7906
RS11593235	VR22	1.000	0.152	0.975	1.0000	1.0000
RS10997591	VR22	0.782	0.471	0.969	0.4344	0.1656
RS7903421	VR22	0.876	0.876	0.760	1.0000	1.0000
RS3096244	VR22	0.922	0.735	0.393	0.7794	1.0000
RS1001016	LRRTM3	0.475	0.654	0.417	1.0000	0.5681
RS12769870	LRRTM3	0.323	0.198	0.616	0.4794	0.6109
RS1925583	LRRTM3	0.230	0.383	0.521	0.8175	0.4291
RS2394314	LRRTM3	0.235	0.455	0.483	1.0000	0.7934
RS10762122	LRRTM3	0.931	0.629	0.570	0.7563	0.5378
RS942780	LRRTM3	0.317	0.190	0.462	0.7578	0.4481
RS1925617	LRRTM3	0.753	0.589	0.922	0.3472	0.4278
RS1925622	LRRTM3	0.929	0.828	0.707	0.2153	0.5966
RS1925632	LRRTM3	0.157	0.244	0.295	0.4447	0.4463
RS1952060	LRRTM3	0.905	0.899	0.866	1.0000	1.0000
RS2147886	LRRTM3	0.317	0.355	0.351	0.2356	0.8000
RS2251000	LRRTM3	0.204	0.182	0.268	0.1422	0.6041
RS2764807	LRRTM3	0.251	0.143	0.298	1.0000	0.8097
RS10762136	LRRTM3	0.480	0.510	0.675	0.6347	0.8053
RS1916341	PLAU	0.923	0.435	0.445	1.0000	0.2994
RS2227564	PLAU	0.377	0.420	0.768	0.3925	0.7572
RS2227566	PLAU	1.000	0.519	0.480	0.8194	0.3488
RS2227568	PLAU	0.615	0.317	0.287	0.0275	0.6769
RS4065	PLAU	0.576	0.843	0.681	0.6225	1.0000
RS1800682	TNFRSF6	0.388	0.426	0.568	0.6359	0.0897

RS1324551	TNFRSF6	0.710	0.160	0.740	0.8044	0.0425
RS2031612	TNFRSF6	0.467	0.294	0.398	0.6290	0.1222
RS2296600	TNFRSF6	0.789	0.576	0.591	0.4656	0.1300
RS2251101	IDE	0.074	0.237	0.354	0.7578	0.4616
RS2251101/ IDE-7 (2)	IDE	0.134	0.405	0.895	0.5569	0.2719
RS7076966	IDE	0.324	0.101	0.959	0.3406	0.4369
RS4646954	IDE	0.732	0.268	0.891	1.0000	0.5688
RS3758505	IDE	0.602	0.196	0.752	1.0000	0.5622
RS7099761	IDE	0.174	0.065	0.668	1.0000	0.0438
RS1544210	IDE	0.294	0.587	0.727	1.0000	0.6047

* No genotype information on these markers for VAN individuals

Table A-5 Family-based association test in NIMH subset (N=352)

SNP	GENE	sum PDT	geno PDT	Fbat	HWE Affecteds	HWE Normals
RS1786927	VR22	0.629	0.833	0.369	0.1950	0.5350
RS2126750	VR22	0.938	0.423	0.836	0.4753	0.0588
RS7911820	VR22	0.090	0.270	0.097	0.4938	0.8638
RS12357560	VR22	0.158	0.166	0.197	0.6684	0.5213
RS7070570	VR22	0.406	0.607	0.851	1.0000	0.5534
RS7074454	VR22	0.193	0.438	0.330	0.7400	0.8625
RS10822719	VR22	0.637	0.896	0.487	0.6059	0.2672
RS6480140	VR22	0.273	0.127	0.625	0.6591	1.0000
RS922347	VR22	0.974	0.551	0.491	0.9034	1.0000
RS4463744	VR22	0.746	0.665	0.651	0.5091	0.3616
RS2441718	VR22	0.255	0.348	0.130	**0.0272**	0.7000
RS2939947	VR22	0.381	0.469	0.419	0.8253	0.3466
RS2456737	VR22	0.325	0.473	0.203	0.3247	0.4631
RS997224	VR22	0.498	0.708	0.324	0.2603	0.8253
RS4746606	VR22	0.820	0.629	0.793	0.1731	0.4844
RS7909676	VR22	0.315	**0.002**	0.838	0.1325	0.0559
RS11593235	VR22	0.858	0.216	0.640	0.5734	1.0000
RS10997591	VR22	0.126	0.283	0.111	0.3409	0.6047
RS7903421	VR22	0.134	0.134	0.149	0.6284	1.0000
RS3096244	VR22	**0.047**	0.218	**0.035**	0.3556	0.6122
RS1001016	LRRTM3	0.601	0.373	0.869	0.3266	0.0547
RS12769870	LRRTM3	0.203	0.077	0.237	**0.0156**	0.4384
RS1925583	LRRTM3	0.507	0.234	0.526	**0.0463**	0.1569
RS2394314	LRRTM3	0.360	0.139	0.392	**0.0831**	0.2153
RS10762122	LRRTM3	0.965	0.328	0.425	**0.0375**	0.2697
RS942780	LRRTM3	0.655	0.272	0.726	0.3734	1.0000
RS1925617	LRRTM3	0.683	0.228	0.188	0.5072	0.7553
RS1925622	LRRTM3	0.365	0.468	0.106	0.5784	0.2803
RS1925632	LRRTM3	0.238	0.488	**0.044**	1.0000	0.7628
RS1952060	LRRTM3	0.774	**0.030**	0.213	0.3900	0.8759
RS2147886	LRRTM3	0.645	0.841	0.137	0.8244	0.6244
RS2251000	LRRTM3	0.477	0.772	0.070	0.8250	0.6444
RS2764807	LRRTM3	0.624	0.880	0.405	0.5944	0.7525
RS10762136	LRRTM3	0.375	0.688	0.178	0.9222	0.5431
RS1916341	PLAU	0.101	0.185	0.161	0.6484	0.5384
RS2227564	PLAU	0.905	0.883	0.810	0.7559	0.1569
RS2227566	PLAU	0.053	0.145	0.064	0.5063	0.7644
RS2227568	PLAU	0.024	0.008	0.036	0.8375	0.8066
RS4065	PLAU	0.133	0.268	0.159	0.2691	0.7534

RS1800682	TNFRSF6	0.626	0.210	0.628	0.5191	0.7578
RS1324551	TNFRSF6	0.572	0.169	0.581	0.7463	0.2803
RS2031612	TNFRSF6	0.682	0.529	0.377	0.4040	0.6184
RS2296600	TNFRSF6	0.631	0.618	0.360	0.3675	0.7563
RS2251101	IDE	0.447	0.217	0.943	0.7981	0.5913
RS2251101/ IDE-7 (2)	IDE	0.721	0.220	0.868	0.7216	0.7200
RS7076966	IDE	0.646	0.557	0.998	0.3797	0.2141
RS4646954	IDE	0.747	0.907	0.742	0.7159	0.5241
RS3758505	IDE	0.508	0.754	0.581	0.6938	0.5263
RS7099761	IDE	0.275	0.214	0.477	0.5113	0.1688
RS1544210	IDE	0.500	0.377	0.512	1.0000	0.6359

Table A-6 Family-based association test in IU subset (N=154)

SNP	GENE	sum PDT	geno PDT	Fbat	HWE Affecteds	HWE Normals
RS1786927	VR22	0.786	0.812	0.859	0.1231	1.0000
RS2126750	VR22	0.172	0.337	0.263	0.4600	0.3000
RS7911820	VR22	0.210	0.556	0.405	0.2181	1.0000
RS12357560	VR22	0.574	0.792	0.710	0.1225	1.0000
RS7070570	VR22	0.805	0.957	0.591	0.5363	0.7969
RS7074454	VR22	0.074	0.266	0.221	0.2816	0.7856
RS10822719	VR22	0.657	0.261	0.635	0.4059	0.7588
RS6480140	VR22	1.000	0.484	0.585	0.3863	0.7897
RS922347	VR22	0.864	0.748	0.812	0.4875	0.2541
RS4463744	VR22	0.459	0.534	0.309	1.0000	0.7728
RS2441718	VR22	**0.047**	0.164	0.317	0.7016	0.5556
RS2939947	VR22	0.390	0.564	0.119	0.3225	1.0000
RS2456737	VR22	0.071	0.052	0.059	0.1788	1.0000
RS997225	VR22	0.923	0.246	0.349	0.8063	0.7141
RS4746606	VR22	0.347	0.613	0.090	0.6022	0.1888
RS7909676	VR22	0.813	0.713	0.874	0.6238	0.3641
RS11593235	VR22	0.797	0.757	0.308	0.3263	0.8159
RS10997591	VR22	0.182	0.184	0.086	0.6006	0.1525
RS7903421	VR22	0.639	0.687	0.689	0.6050	1.0000
RS3096244	VR22	0.414	0.168	0.160	0.2734	0.6056
RS1001016	LRRTM3	0.756	0.363	0.040	0.6888	0.2919
RS12769870	LRRTM3	0.481	0.811	0.339	0.0850	1.0000
RS1925583	LRRTM3	0.603	0.872	0.541	0.1984	0.8016
RS2394314	LRRTM3	0.526	0.826	0.437	0.4163	1.0000
RS10762122	LRRTM3	0.699	0.901	0.906	1.0000	0.0519
RS942780	LRRTM3	0.644	0.920	0.391	0.4241	0.7209
RS1925617	LRRTM3	0.501	0.052	0.087	0.3000	0.3503
RS1925622	LRRTM3	0.397	0.030	0.081	0.5003	0.2469
RS1925632	LRRTM3	0.724	0.815	0.264	1.0000	0.2163
RS1952060	LRRTM3	0.854	0.178	0.215	0.1934	0.8200
RS2147886	LRRTM3	0.848	0.735	0.253	1.0000	0.4606
RS2251000	LRRTM3	0.923	0.728	0.406	0.7525	0.3188
RS2764807	LRRTM3	0.880	0.691	0.392	0.7303	0.8131
RS10762136	LRRTM3	0.805	0.692	0.608	0.4231	0.4697
RS1916341	PLAU	0.265	0.425	0.170	0.8713	0.8066
RS2227564	PLAU	0.941	0.750	0.382	0.3816	0.7781
RS2227566	PLAU	0.162	0.419	0.148	1.0000	0.6165
RS2227568	PLAU	0.518	0.658	0.614	0.1181	1.0000
RS4065	PLAU	0.185	0.381	0.070	1.0000	0.4431
RS1800682	TNFRSF6	0.763	0.343	0.166	0.6191	0.0078

RS1324551	TNFRSF6	0.368	0.212	0.061	0.7390	0.0075
RS2031612	TNFRSF6	0.368	0.213	0.049	0.5081	0.6197
RS2296600	TNFRSF6	0.496	0.266	0.089	0.7350	0.8053
RS2251101	IDE	0.458	0.816	0.712	0.1534	0.1231
RS2251101 IDE-7 (2)	IDE	0.293	0.663	0.650	0.3513	0.2394
RS7076966	IDE	0.858	0.683	0.712	0.7242	0.8019
RS4646954	IDE	0.508	0.747	0.417	0.1919	0.5172
RS3758505	IDE	0.423	0.651	0.436	0.1731	0.4656
RS7099761	IDE	0.898	0.613	0.870	0.8641	0.8163
RS1544210	IDE	0.428	0.682	0.379	0.7347	0.4669

Table A-7 Two point linkage analysis of five candidate genes in overall data set (N=566)

SNP	GENE	Dominant Model Heterogeneity LOD Score	Recessive Model Heterogeneity LOD Score	ASPEX MLS
RS1786927	VR22	0.00	0.00	0.00
RS2126750	VR22	0.31	0.21	0.26
RS7911820	VR22	0.24	0.12	0.06
RS12357560	VR22	0.00	0.00	0.00
RS7070570	VR22	**1.30**	0.59	0.53
RS7074454	VR22	0.81	0.58	0.29
RS10822719	VR22	0.00	0.00	0.00
RS6480140	VR22	0.00	0.00	0.00
RS922347	VR22	0.07	0.00	0.00
RS4463744	VR22	0.17	0.44	0.29
RS2441718	VR22	**1.02**	**1.01**	0.72
RS2939947	VR22	0.95	0.81	0.63
RS2456737	VR22	0.56	0.92	0.78
RS997225	VR22	0.48	0.17	0.04
RS4746606	VR22	0.07	0.00	0.07
RS7909676	VR22	0.73	0.58	0.26
RS11593235	VR22	0.09	0.07	0.07
RS1099751	VR22	0.50	0.40	0.32
RS7903421	VR22	0.00	0.00	0.00
RS3096244	VR22	0.00	0.00	0.01
RS1001016	LRRTM3	0.00	0.00	0.00
RS12769870	LRRTM3	0.00	0.00	0.00
RS1925583	LRRTM3	0.05	0.01	0.04
RS2394314	LRRTM3	0.05	0.02	0.02
RS10762122	LRRTM3	0.00	0.00	0.01
RS942780	LRRTM3	0.02	0.07	0.03
RS1925617	LRRTM3	0.60	0.74	0.59
RS1925622	LRRTM3	0.54	0.56	0.64
RS1925632	LRRTM3	0.34	0.25	0.29
RS1952060	LRRTM3	0.14	0.12	0.13
RS2147886	LRRTM3	0.12	0.07	0.06
RS2251000	LRRTM3	0.24	0.10	0.12
RS2764807	LRRTM3	0.10	0.06	0.07
RS10762136	LRRTM3	0.51	0.41	0.28
RS1916341	PLAU	0.00	0.00	0.00
RS2227564	PLAU	0.04	0.04	0.07
RS2227566	PLAU	0.01	0.04	0.01
RS2227568	PLAU	0.07	0.05	0.00
RS4065	PLAU	0.00	0.00	0.00

RS1800682	TNFRSF6	0.49	0.39	0.20
RS1324551	TNFRSF6	0.31	0.17	0.10
RS2031612	TNFRSF6	0.88	0.85	0.56
RS2296600	TNFRSF6	0.75	0.72	0.50
RS2251101	IDE	0.00	0.00	0.00
RS2251101/ IDE-7 (2)	IDE	0.00	0.00	0.00
RS7076966	IDE	0.28	0.11	0.08
RS4646954	IDE	0.02	0.17	0.06
RS3758505	IDE	0.11	0.27	0.07
RS7099761	IDE	0.35	0.14	0.10
RS1544210	IDE	0.00	0.00	0.00

Table A-8 OSA Two-point Linkage Analysis using APOE LOD score (at theta=0) as covariate to order families from low to high in overall data set

SNP	GENE	MAX LOD	EMP p	FAM Used	TOT FAMS	PROP FAMS
RS1786927	VR22	0.615	0.100	60	504	0.12
RS2126750	VR22	0.364	0.630	57	504	0.11
RS7911820	VR22	0.156	0.920	448	523	0.86
RS12357560	VR22	0.595	0.260	112	527	0.21
RS7070570	VR22	0.659	0.800	477	486	0.98
RS7074454	VR22	0.378	0.830	438	505	0.87
RS10822719	VR22	0.020	0.810	14	529	0.03
RS6480140	VR22	0.042	0.490	14	506	0.03
RS9223471	VR22	0.819	0.190	43	516	0.08
RS4463744	VR22	0.901	0.320	60	533	0.11
RS2441718	VR22	0.681	0.840	495	514	0.96
RS2939947	VR22	0.846	0.780	530	532	1.00
RS2456737	VR22	1.028	0.870	530	532	1.00
RS997224	VR22	0.248	0.830	428	477	0.90
RS4746606	VR22	0.317	0.480	3	514	0.01
RS7909676	VR22	0.748	0.600	497	514	0.96
RS11593235	VR22	0.098	0.930	503	523	0.96
RS10997591	VR22	0.633	0.710	401	513	0.78
RS7903421	VR22	0.146	0.570	36	523	0.07
RS3096244	VR22	0.265	0.510	67	524	0.13
RS1001016	LRRTM3	0.013	0.850	4	516	0.01
RS12769870	LRRTM3	0.509	0.250	7	507	0.01
RS1925583	LRRTM3	0.491	0.250	18	520	0.03
RS2394314	LRRTM3	0.620	0.320	9	522	0.02
RS10762122	LRRTM3	0.363	0.300	57	528	0.11
RS942780	LRRTM3	0.198	0.720	474	529	0.90
RS1925617	LRRTM3	0.551	0.890	275	521	0.53
RS1925622	LRRTM3	0.494	0.820	509	528	0.96
RS1925632	LRRTM3	0.164	0.985	528	528	1.00
RS1952060	LRRTM3	0.102	0.950	510	529	0.96
RS2147886	LRRTM3	0.027	0.970	9	533	0.02
RS2251000	LRRTM3	0.098	0.970	61	531	0.11
RS2764807	LRRTM3	0.010	0.980	11	533	0.02
RS10762136	LRRTM3	0.303	0.990	523	523	1.00
RS1916341	PLAU	0.185	0.740	34	519	0.07
RS2227564	PLAU	0.218	0.750	10	529	0.02
RS2227566	PLAU	0.196	0.840	34	506	0.07
RS2227568	PLAU	0.452	0.550	46	503	0.09
RS4065	PLAU	0.386	0.410	41	509	0.08
RS1800682	TNFRSF6	0.541	0.550	520	534	0.97

RS1324551	TNFRSF6	0.276	0.780	496	521	0.95
RS2031612	TNFRSF6	1.565	0.260	275	528	0.52
RS2296600	TNFRSF6	1.319	0.500	173	531	0.33
RS2251101	IDE	0.000	0.940	1	495	0.00
RS2251101/ IDE-7 (2)	IDE	0.000	0.945	1	520	0.00
RS7076966	IDE	0.165	0.950	518	526	0.98
RS4646954	IDE	0.176	0.910	473	532	0.89
RS3758505	IDE	0.263	0.890	464	519	0.89
RS7099761	IDE	0.153	0.890	516	524	0.98
RS1544210	IDE	0.000	0.965	1	508	0.00

Table A-9 OSA Two-point Linkage Analysis using APOE LOD score (at theta=0) as covariate to order families from high to low in overall data set

SNP	GENE	MAX LOD	EMP p	FAM Used	TOT FAMS	PROP FAMS
RS1786927	VR22	0.218	0.490	17	504	0.03
RS2126750	VR22	0.563	0.480	367	504	0.73
RS7911820	VR22	0.659	0.300	396	523	0.76
RS12357560	VR22	1.134	0.062	58	527	0.11
RS7070570	VR22	1.436	0.154	49	486	0.10
RS7074454	VR22	1.411	0.100	271	505	0.54
RS10822719	VR22	0.114	0.620	305	529	0.58
RS6480140	VR22	0.195	0.240	7	506	0.01
RS9223471	VR22	0.077	0.850	39	516	0.08
RS4463744	VR22	0.476	0.560	299	533	0.56
RS2441718	VR22	1.344	0.310	373	514	0.73
RS2939947	VR22	1.150	0.560	291	522	0.56
RS2456737	VR22	**2.303**	**0.049**	273	532	0.51
RS997224	VR22	0.661	0.490	277	478	0.58
RS4746606	VR22	0.830	0.106	36	514	0.07
RS7909676	VR22	1.653	0.083	302	517	0.58
RS11593235	VR22	0.589	0.540	399	523	0.76
RS10997591	VR22	1.151	0.350	265	513	0.52
RS7903421	VR22	0.127	0.560	97	523	0.19
RS3096244	VR22	0.488	0.320	272	524	0.52
RS1001016	LRRTM3	0.092	0.460	155	516	0.30
RS12769870	LRRTM3	0.315	0.380	162	507	0.32
RS1925583	LRRTM3	0.160	0.610	116	520	0.22
RS2394314	LRRTM3	0.300	0.440	158	522	0.30
RS10762122	LRRTM3	0.000	0.915	1	528	0.00
RS942780	LRRTM3	0.522	0.310	384	529	0.73
RS1925617	LRRTM3	0.952	0.350	49	521	0.09
RS1925622	LRRTM3	0.796	0.410	55	528	0.10
RS1925632	LRRTM3	1.229	0.100	56	528	0.11
RS1952060	LRRTM3	0.539	0.340	52	529	0.10
RS2147886	LRRTM3	0.685	0.310	55	533	0.10
RS2251000	LRRTM3	1.000	0.142	56	531	0.11
RS2764807	LRRTM3	0.977	0.164	55	533	0.10
RS10762136	LRRTM3	1.125	0.340	168	523	0.32
RS1916341	PLAU	0.224	0.630	39	519	0.08
RS2227564	PLAU	0.400	0.510	326	529	0.62
RS2227566	PLAU	0.238	0.760	46	506	0.09
RS2227568	PLAU	0.206	0.750	8	503	0.02
RS4065	PLAU	0.163	0.610	5	509	0.01
RS1800682	TNFRSF6	0.794	0.200	162	534	0.30

RS1324551	TNFRSF6	0.276	0.750	485	521	0.93
RS2031612	TNFRSF6	1.194	0.680	492	528	0.93
RS2296600	TNFRSF6	1.045	0.710	495	531	0.93
RS2251101	IDE	0.055	0.660	1	495	0.00
RS2251101/ IDE-7 (2)	IDE	0.055	0.630	1	520	0.00
RS7076966	IDE	0.504	0.460	281	526	0.53
RS4646954	IDE	0.808	0.164	295	532	0.55
RS3758505	IDE	0.846	0.125	284	519	0.55
RS7099761	IDE	0.820	0.240	161	524	0.31
RS1544210	IDE	0.420	0.210	110	508	0.22

Table A-10 OSA Two-point Linkage Analysis using APOE weight as covariate to order families from low to high in overall data set

SNP	GENE	MAX LOD	EMP p	FAM Used	TOT FAMS	PROP FAMS
RS1786927	VR22	0.000	0.835	101	504	0.20
RS2126750	VR22	0.285	0.530	187	504	0.37
RS7911820	VR22	0.118	0.755	477	523	0.91
RS12357560	VR22	0.000	0.940	108	527	0.20
RS7070570	VR22	0.607	0.710	447	486	0.92
RS7074454	VR22	0.347	0.665	464	505	0.92
RS10822719	VR22	0.000	0.810	110	529	0.21
RS6480140	VR22	0.117	0.142	102	506	0.20
RS9223471	VR22	0.000	0.915	104	516	0.20
RS4463744	VR22	0.254	0.595	482	533	0.90
RS2441718	VR22	0.661	0.930	514	514	1.00
RS2939947	VR22	0.862	0.580	418	522	0.80
RS2456737	VR22	1.018	0.845	532	532	1.00
RS997224	VR22	0.246	0.660	378	478	0.79
RS4746606	VR22	0.026	0.740	455	514	0.89
RS7909676	VR22	1.114	0.164	417	517	0.81
RS11593235	VR22	0.055	0.845	106	523	0.20
RS10997591	VR22	0.454	0.900	513	513	1.00
RS7903421	VR22	0.001	0.870	107	523	0.20
RS3096244	VR22	0.071	0.630	415	524	0.79
RS1001016	LRRTM3	0.007	0.435	103	516	0.20
RS12769870	LRRTM3	0.000	0.785	104	507	0.21
RS1925583	LRRTM3	0.027	0.690	424	520	0.82
RS2394314	LRRTM3	0.042	0.310	107	522	0.20
RS10762122	LRRTM3	0.000	0.795	106	528	0.20
RS942780	LRRTM3	**1.618**	**0.021**	115	529	0.22
RS1925617	LRRTM3	0.491	0.840	521	521	1.00
RS1925622	LRRTM3	0.440	0.810	528	528	1.00
RS1925632	LRRTM3	0.294	0.470	115	528	0.22
RS1952060	LRRTM3	0.067	0.620	468	529	0.88
RS2147886	LRRTM3	0.074	0.570	117	533	0.22
RS2251000	LRRTM3	0.143	0.460	116	531	0.22
RS2764807	LRRTM3	0.030	0.770	119	533	0.22
RS10762136	LRRTM3	0.385	0.630	114	523	0.22
RS1916341	PLAU	0.004	0.885	106	519	0.20
RS2227564	PLAU	0.008	0.930	529	529	1.00
RS2227566	PLAU	0.030	0.955	506	506	1.00
RS2227568	PLAU	0.036	0.820	210	503	0.42
RS4065	PLAU	0.000	0.880	103	509	0.20
RS1800682	TNFRSF6	0.392	0.810	534	534	1.00

RS1324551	TNFRSF6	0.148	0.700	114	521	0.22
RS2031612	TNFRSF6	1.091	0.530	220	528	0.42
RS2296600	TNFRSF6	0.992	0.440	221	531	0.42
RS2251101	IDE	0.000	0.855	30	157	0.19
RS2251101/ IDE-7 (2)	IDE	0.000	0.880	34	171	0.20
RS7076966	IDE	0.735	0.131	60	172	0.35
RS4646954	IDE	0.285	0.131	139	171	0.81
RS3758505	IDE	0.254	0.190	140	172	0.81
RS7099761	IDE	0.283	0.480	59	172	0.34
RS1544210	IDE	0.152	0.117	32	169	0.19

Table A-11 OSA Two-point Linkage Analysis using APOE weight as covariate to order families from high to low in overall data set

SNP	GENE	MAX LOD	EMP p	FAM Used	TOT FAMS	PROP FAMS
RS1786927	VR22	0.000	0.780	41	504	0.08
RS2126750	VR22	0.287	0.570	55	504	0.11
RS7911820	VR22	0.368	0.220	332	523	0.63
RS12357560	VR22	0.343	0.470	96	527	0.18
RS7070570	VR22	1.054	0.079	379	486	0.78
RS7074454	VR22	0.995	0.270	317	505	0.63
RS10822719	VR22	0.330	0.142	62	529	0.12
RS6480140	VR22	0.000	0.755	41	506	0.08
RS9223471	VR22	0.157	0.200	92	516	0.18
RS4463744	VR22	0.529	0.390	309	533	0.58
RS2441718	VR22	1.509	0.200	58	514	0.11
RS2939947	VR22	1.022	0.480	327	522	0.63
RS2456737	VR22	1.348	0.220	38	532	0.58
RS997224	VR22	0.490	0.375	383	478	0.80
RS4746606	VR22	0.057	0.490	318	514	0.62
RS7909676	VR22	0.553	0.855	517	517	1.00
RS11593235	VR22	0.803	0.155	110	523	0.21
RS10997591	VR22	**1.342**	**0.056**	324	513	0.63
RS7903421	VR22	0.849	**0.033**	106	523	0.20
RS3096244	VR22	0.203	0.520	306	524	0.58
RS1001016	LRRTM3	0.110	**0.008**	58	516	0.11
RS12769870	LRRTM3	0.059	0.121	45	507	0.09
RS1925583	LRRTM3	0.069	0.520	62	520	0.12
RS2394314	LRRTM3	0.056	0.230	62	522	0.12
RS10762122	LRRTM3	0.000	0.790	48	528	0.09
RS942780	LRRTM3	0.028	0.960	529	529	1.00
RS1925617	LRRTM3	0.575	0.375	408	521	0.78
RS1925622	LRRTM3	0.550	0.345	419	528	0.79
RS1925632	LRRTM3	0.164	0.945	528	528	1.00
RS1952060	LRRTM3	0.146	0.390	317	529	0.60
RS2147886	LRRTM3	0.177	0.480	63	533	0.12
RS2251000	LRRTM3	0.114	0.670	61	531	0.11
RS2764807	LRRTM3	0.897	0.190	62	533	0.12
RS10762136	LRRTM3	0.595	0.410	107	523	0.20
RS1916341	PLAU	0.593	**0.050**	94	519	0.18
RS2227564	PLAU	0.909	**0.018**	62	529	0.12
RS2227566	PLAU	0.734	**0.020**	93	506	0.18
RS2227568	PLAU	0.173	0.300	104	503	0.21
RS4065	PLAU	0.576	**0.011**	91	509	0.18
RS1800682	TNFRSF6	0.392	0.785	534	534	1.00

RS1324551	TNFRSF6	0.218	0.500	105	521	0.20
RS2031612	TNFRSF6	1.060	0.750	528	528	1.00
RS2296600	TNFRSF6	0.926	0.750	531	531	1.00
RS2251101	IDE	0.042	0.300	96	495	0.19
RS2251101/ IDE-7 (2)	IDE	0.021	0.420	104	520	0.20
RS7076966	IDE	0.235	0.570	110	526	0.21
RS4646954	IDE	0.129	0.955	532	532	1.00
RS3758505	IDE	0.203	0.965	519	519	1.00
RS7099761	IDE	0.404	0.310	110	524	0.21
RS1544210	IDE	0.010	0.730	60	508	0.12

Table A-12 OSA Two-point Linkage Analysis using mean age-of-onset as covariate to order families from low to high in overall data set

SNP (RS #)	GENE	MAX LOD	EMP p	FAM Used	TOT FAMS	PROP FAMS
RS1786927	VR22	0.680	0.078	25	504	0.05
RS2126750	VR22	0.474	0.570	350	504	0.69
RS7911820	VR22	1.457	0.070	185	523	0.35
RS12357560	VR22	0.109	0.710	159	527	0.30
RS7070570	VR22	0.950	0.460	364	486	0.75
RS7074454	VR22	1.378	0.076	184	505	0.36
RS10822719	VR22	0.444	0.200	186	529	0.35
RS6480140	VR22	0.063	0.575	1	506	0.00
RS9223471	VR22	0.345	0.380	157	516	0.30
RS4463744	VR22	0.287	0.680	182	533	0.34
RS2441718	VR22	**1.839**	**0.046**	220	514	0.43
RS2939947	VR22	1.132	0.370	215	522	0.41
RS2456737	VR22	2.028	0.121	246	532	0.46
RS997224	VR22	0.181	0.950	169	477	0.35
RS4746606	VR22	0.132	0.650	200	514	0.39
RS7909676	VR22	1.307	0.120	183	517	0.35
RS11593235	VR22	0.729	0.230	138	523	0.26
RS10997591	VR22	1.112	0.190	152	513	0.30
RS7903421	VR22	0.295	0.190	247	523	0.47
RS3096244	VR22	0.210	0.420	240	524	0.46
RS1001016	LRRTM3	0.014	0.800	132	516	0.26
RS12769870	LRRTM3	0.067	0.720	195	507	0.38
RS1925583	LRRTM3	0.142	0.680	4	520	0.01
RS2394314	LRRTM3	0.134	0.670	9	522	0.02
RS10762122	LRRTM3	0.004	0.870	7	528	0.01
RS942780	LRRTM3	0.197	0.500	374	529	0.71
RS1925617	LRRTM3	0.491	0.970	521	521	1.00
RS1925622	LRRTM3	0.440	0.980	528	528	1.00
RS1925632	LRRTM3	0.164	0.990	528	528	1.00
RS1952060	LRRTM3	0.086	0.940	186	529	0.35
RS2147886	LRRTM3	0.116	0.840	182	533	0.34
RS2251000	LRRTM3	0.118	0.880	189	531	0.36
RS2764807	LRRTM3	**1.123**	**0.051**	180	533	0.34
RS10762136	LRRTM3	1.511	0.112	184	523	0.35
RS1916341	PLAU	0.483	0.164	14	519	0.03
RS2227564	PLAU	0.168	0.730	181	529	0.34
RS2227566	PLAU	0.474	0.280	14	506	0.03
RS2227568	PLAU	0.582	0.330	23	503	0.05
RS4065	PLAU	0.203	0.400	13	509	0.03
RS1800682	TNFRSF6	0.650	0.620	439	534	0.82

RS1324551	TNFRSF6	0.334	0.760	475	521	0.91
RS2031612	TNFRSF6	1.798	0.173	432	528	0.82
RS2296600	TNFRSF6	1.573	0.280	435	531	0.82
RS2251101	IDE	0.015	0.705	1	495	0.00
RS2251101/ IDE-7 (2)	IDE	0.000	0.910	2	520	0.00
RS7076966	IDE	0.192	0.920	2	526	0.00
RS4646954	IDE	0.738	0.180	24	532	0.05
RS3758505	IDE	0.585	0.380	24	519	0.05
RS7099761	IDE	0.825	0.290	46	524	0.09
RS1544210	IDE	0.296	0.420	17	508	0.03

Table A-13 OSA Two-point Linkage Analysis using mean age-of-onset as covariate to order families from high to low in overall data set

SNP (RS #)	GENE	MAX LOD	EMP p	FAM Used	TOT FAMS	PROP FAMS
RS1786927	VR22	**1.019**	**0.014**	23	504	0.05
RS2126750	VR22	0.796	0.220	21	504	0.04
RS7911820	VR22	0.173	0.820	485	523	0.93
RS12357560	VR22	0.532	0.090	12	527	0.02
RS7070570	VR22	0.950	0.370	451	486	0.93
RS7074454	VR22	0.507	0.620	467	505	0.92
RS10822719	VR22	0.201	0.450	4	529	0.01
RS6480140	VR22	0.168	0.330	17	506	0.03
RS9223471	VR22	0.071	0.750	16	516	0.03
RS4463744	VR22	0.464	0.550	12	533	0.02
RS2441718	VR22	0.778	0.700	436	514	0.85
RS2939947	VR22	1.083	0.420	466	522	0.89
RS2456737	VR22	1.181	0.620	500	532	0.94
RS997224	VR22	1.176	0.200	93	477	0.19
RS4746606	VR22	0.359	0.320	12	514	0.02
RS7909676	VR22	0.691	0.700	485	517	0.94
RS11593235	VR22	0.470	0.340	7	523	0.01
RS10997591	VR22	0.573	0.790	435	513	0.85
RS7903421	VR22	0.191	0.380	6	523	0.01
RS3096244	VR22	0.059	0.770	4	524	0.01
RS1001016	LRRTM3	0.640	0.082	16	516	0.03
RS12769870	LRRTM3	0.044	0.790	4	507	0.01
RS1925583	LRRTM3	0.096	0.750	139	520	0.27
RS2394314	LRRTM3	0.129	0.600	126	522	0.24
RS10762122	LRRTM3	0.301	0.180	31	528	0.06
RS942780	LRRTM3	0.455	0.320	442	529	0.84
RS1925617	LRRTM3	1.105	0.360	126	521	0.24
RS1925622	LRRTM3	0.801	0.520	491	528	0.93
RS1925632	LRRTM3	0.850	0.129	21	528	0.04
RS1952060	LRRTM3	0.986	0.095	24	529	0.05
RS2147886	LRRTM3	0.849	**0.058**	21	533	0.04
RS2251000	LRRTM3	0.849	0.105	21	531	0.04
RS2764807	LRRTM3	0.297	0.430	4	533	0.01
RS10762136	LRRTM3	0.343	0.880	520	523	0.99
RS1916341	PLAU	0.950	0.113	23	519	0.04
RS2227564	PLAU	0.299	0.530	30	529	0.06
RS2227566	PLAU	0.686	0.140	22	506	0.04
RS2227568	PLAU	0.752	0.280	12	503	0.02
RS4065	PLAU	0.775	0.050	24	509	0.05
RS1800682	TNFRSF6	0.455	0.850	509	534	0.95

RS1324551	TNFRSF6	0.439	0.600	401	521	0.77
RS2031612	TNFRSF6	1.390	0.460	429	528	0.81
RS2296600	TNFRSF6	1.217	0.490	263	531	0.50
RS2251101	IDE	0.117	0.360	42	495	0.08
RS2251101/ IDE-7 (2)	IDE	0.201	0.230	42	520	0.08
RS7076966	IDE	0.949	0.230	239	526	0.45
RS4646954	IDE	0.551	0.460	343	532	0.64
RS3758505	IDE	0.721	0.350	155	519	0.30
RS7099761	IDE	0.825	0.290	46	524	0.09
RS1544210	IDE	0.296	0.420	17	508	0.03

REFERENCES

Abecasis GR, Cherny SS, Cookson WO, Cardon LR. 2002. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 30:97-101.

Abecasis GR, Cookson WO. 2000. GOLD--graphical overview of linkage disequilibrium. Bioinformatics 16:182-183.

Akesson HO. 1969. A population study of senile and arteriosclerotic psychoses. Hum Hered 19:546-566.

Altmuller J, Palmer LJ, Fischer G, Scherb H, Wjst M. 2001. Genomewide scans of complex human diseases: true linkage is hard to find. Am J Hum Genet 69:936-950.

Ashley-Koch AE, Mei H, Jaworski J, Ma DQ, Ritchie MD, Menold MM, DeLong GR, Abramson RK, Wright HH, Hussman JP, Cuccaro ML, Gilbert JR, Martin ER, Pericak-Vance MA. 2006. An analysis paradigm for investigating multi-locus effects in complex disease: examination of three GABA receptor subunit genes on 15q11-q13 as risk factors for autistic disorder. Ann Hum Genet 70:281-292.

Baldwin AS, Jr. 1996. The NF-kappa B and I kappa B proteins: new discoveries and insights. Annu Rev Immunol 14:649-683.

Barrett JC, Cardon LR. 2006. Evaluating coverage of genome-wide association studies. Nat Genet 38:659-662.

Barrett JC, Fry B, Maller J, Daly MJ. 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263-265.

Bassett SS, Avramopoulos D, Fallin D. 2002. Evidence for parent of origin effect in late-onset Alzheimer disease. Am J Med Genet 114:679-686.

Bassett SS, Kusevic I, Cristinzio C, Yassa MA, Avramopoulos D, Yousem DM, Fallin MD. 2005. Brain activation in offspring of AD cases corresponds to 10q linkage. Ann Neurol 58:142-146.

Baumann K, Mandelkow EM, Biernat J, Piwnica-Worms H, Mandelkow E. 1993. Abnormal Alzheimer-like phosphorylation of tau-protein by cyclin-dependent kinases cdk2 and cdk5. FEBS Lett 336:417-424.

Beach TG. 1987. The history of Alzheimer's disease: three debates. J Hist Med Allied Sci 42:327-349.

Beach TG, Kuo YM, Spiegel K, Emmerling MR, Sue LI, Kokjohn K, Roher AE. 2000. The cholinergic deficit coincides with Abeta deposition at the earliest histopathologic stages of Alzheimer disease. J Neuropathol Exp Neurol 59:308-313.

Bellman R. 1961. Adaptive control processes. Princeton: Princeton University Press.

Bennett C, Crawford F, Osborne A, Diaz P, Hoyne J, Lopez R, Roques P, Duara R, Rossor M, Mullan M. 1995. Evidence that the APOE locus influences rate of disease progression in late onset familial Alzheimer's Disease but is not causative. Am J Med Genet 60:1-6.

Berchtold NC, Cotman CW. 1998. Evolution in the conceptualization of dementia and Alzheimer's disease: Greco-Roman period to the 1960s. Neurobiol Aging 19:173-189.

Bergen AL. 1994. Heredity in dementia of the Alzheimer type. Clinical Genetics 46:144-149.

Bergen AL, Engedal K, Kringlen E. 1997. The role of heredity in late-onset Alzheimer disease and vascular dementia: A twin study. Archives of General Psychiatry 54:264-270.

Bertram L, Blacker D, Mullin K, Keeney D, Jones J, Basu S, Yhu S, McInnis MG, Go RC, Vekrellis K, Selkoe DJ, Saunders AJ, Tanzi RE. 2000. Evidence for genetic linkage of Alzheimer's disease to chromosome 10q. Science 290:2302-2303.

Bertram L, Hiltunen M, Parkinson M, Ingelsson M, Lange C, Ramasamy K, Mullin K, Menon R, Sampson AJ, Hsiao MY, Elliott KJ, Velicelebi G, Moscarillo T, Hyman BT, Wagner SL, Becker KD, Blacker D, Tanzi RE. 2005. Family-based association between Alzheimer's disease and variants in UBQLN1. N Engl J Med 352:884-894.

Bertram L, Hsiao M, Lange C, Blacker D, Tanzi RE. 2006. Single-nucleotide polymorphism rs498055 on chromosome 10q24 is not associated with Alzheimer disease in two independent family samples. Am J Hum Genet 79:180-183.

Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. 2007. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. Nat Genet 39:17-23.

Bichet D, Cornet V, Geib S, Carlier E, Volsen S, Hoshi T, Mori Y, De WM. 2000. The I-II loop of the Ca2+ channel alpha1 subunit contains an endoplasmic reticulum retention signal antagonized by the beta subunit. Neuron 25:177-190.

Blacker D, Bertram L, Saunders AJ, Moscarillo TJ, Albert MS, Wiener H, Perry RT, Collins JS, Harrell LE, Go RC, Mahoney A, Beaty T, Fallin MD, Avramopoulos D, Chase GA, Folstein MF, McInnis MG, Bassett SS, Doheny KJ, Pugh EW, Tanzi RE. 2003. Results of a high-resolution genome screen of 437 Alzheimer's disease families. Hum Mol Genet 12:23-32.

Blacker D, Wilcox MA, Laird NM, Rodes L, Horvath SM, Go RCP, Perry R, Watson B, Bassett SS, McInnis MG, Albert MS, Hyman BT, Tanzi RE. 1998. Alpha-2 macroglobulin is genetically associated with Alzheimer disease. Nature Genetics 19:357-360.

Blessed G, Tomlinson BE, Roth M. 1968. The association between quantitative measures of dementia and of senile change in the cerebral grey matter of elderly subjects. Br J Psychiatry 114:797-811.

Bonney GE. 1984. On the statistical determination of major gene mechanisms in continuous human traits: Regressive models. Am J Hum Genet 18:731-749.

Bonney GE. 1986. Regressive logistic models for familial disease and other binary traits. Biometrics 42:611-625.

Braak H, Braak E. 1991. Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol (Berl) 82:239-259.

Braak H, Braak E. 1995. Staging of Alzheimer's disease-related neurofibrillary changes. Neurobiol Aging 16:271-278.

Braak H, Braak E. 1997. Diagnostic criteria for neuropathologic assessment of Alzheimer's disease. Neurobiol Aging 18:S85-S88.

Breitner JC, Gatz M, Christian JC, Mortimer JA, McClearn GE, Heston LL, Welsh KA, Anthony JC, Folstein MF. 1993. Use of twin cohorts for research in Alzheimer's disease. Neurology 43:261-267.

Breitner JC, Murphy EA. 1992. Twin studies of Alzheimer disease: II. Some predictions under a genetic model. Am J Med Genet 44:628-634.

Breitner JC, Welsh KA, Gau BA, McDonald WM, Steffens DC, Saunders AM, Magruder KM, Helms MJ, Plassman BL, Folstein MF. 1995a. Alzheimer's disease in the National Academy of Sciences-National Research Council Registry of Aging Twin Veterans. III. Detection of cases, longitudinal results, and observations on twin concordance. Arch Neurol 52:763-771.

Breitner JC, Welsh KA, Helms MJ, Gaskell PC, Gau BA, Roses AD, Pericak-Vance MA, Saunders AM. 1995b. Delayed onset of Alzheimer's disease with nonsteroidal anti-inflammatory and histamine H2 blocking drugs. Neurobiology of Aging 16:523-530.

Brice NL, Berrow NS, Campbell V, Page KM, Brickley K, Tedder I, Dolphin AC. 1997. Importance of the different beta subunits in the membrane expression of the alpha1A and alpha2 calcium channel subunits: studies using a depolarization-sensitive alpha1A antibody. Eur J Neurosci 9:749-759.

Burns A, Murphy D. 1996. Protection against Alzheimer's disease? Lancet 348:420-421.

Cardon LR, Bell JI. 2001. Association study designs for complex diseases. Nat Rev Genet 2:91-99.

Caricasole A, Copani A, Caruso A, Caraci F, Iacovelli L, Sortino MA, Terstappen GC, Nicoletti F. 2003. The Wnt pathway, cell-cycle activation and beta-amyloid: novel therapeutic strategies in Alzheimer's disease? Trends Pharmacol Sci 24:233-238.

Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. 2004. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am J Hum Genet 74:106-120.

Castano EM, Prelli F, Wisniewski T, Golabek A, Kumar RA, Soto C, Frangione B. 1995. Fibrillogenesis in Alzheimer's disease of amyloid beta peptides and apolipoprotein E. Biochem J 306 (Pt 2):599-604.

Chase GA, Folstein MF, Breitner JC, Beaty TH, Self SG. 1983. The use of life tables and survival analysis in testing genetic hypotheses, with an application to Alzheimer's disease. Am J Epidemiol 117:590-597.

Cho YM, Ritchie MD, Moore JH, Park JY, Lee KU, Shin HD, Lee HK, Park KS. 2004. Multifactor-dimensionality reduction shows a two-locus interaction associated with Type 2 diabetes mellitus. Diabetologia 47:549-554.

Citron M. 2004. Strategies for disease modification in Alzheimer's disease. Nat Rev Neurosci 5:677-685.

Collins FS. 1995. Positional cloning moves from periditional to traditional. Nature Genetics 9:347-350.

Concato J, Feinstein AR, Holford TR. 1993. The risk of determining risk with multivariable models. Ann Intern Med 118:201-210.

Conrad DF, Jakobsson M, Coop G, Wen X, Wall JD, Rosenberg NA, Pritchard JK. 2006. A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. Nat Genet 38:1251-1260.

Cook RH, Schneck SA, Clark DB. 1981. Twins with Alzheimer's disease. Arch Neurol 38:300-301.

Cooper DN, Smith BA, Cooke HJ, Niemann S, Schmidtke J. 1985. An estimate of unique DNA sequence heterozygosity in the human genome. Hum Genet 69:201-205.

Corder EH, Saunders AM, Risch N, Strittmatter WJ, Schmechel DE, Gaskell PC, Rimmler JB, Locke PA, Conneally PM, Schmader KE, Small GW, Roses AD, Haines JL, Pericak-Vance MA. 1994. Apolipoprotein E type 2 allele decreases the risk of late onset Alzheimer disease. Nature Genetics 7:180-184.

Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, Pericak-Vance MA. 1993. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science 261:921-923.

Cupples LA, Risch N, Farrer LA, Myers RH. 1991. Estimation of morbid risk and age at onset with missing information. Am J Hum Genet 49:76-87.

Cupples LA, Terrin NC, Myers RH, D'Agostino RB. 1989. Using survival methods to estimate age-at-onset distributions for genetic diseases with an application to Huntington disease. Genet Epidemiol 6:361-371.

Curran M, Middleton D, Edwardson J, Perry R, McKeith I, Morris C, Neill D. 1997. HLA-DR antigens associated with major genetic risk for late-onset Alzheimer's disease. Neuroreport 8:1467-1469.

Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. 2001. High-resolution haplotype structure in the human genome. Nat Genet 29:229-232.

Drysdale CM, McGraw DW, Stack CB, Stephens JC, Judson RS, Nandabalan K, Arnold K, Ruano G, Liggett SB. 2000. Complex promoter and coding region beta 2-adrenergic receptor haplotypes alter receptor expression and predict in vivo responsiveness. Proc Natl Acad Sci U S A 97:10483-10488.

Dudek S, Motsinger AA, Velez D, Williams SM, Ritchie MD. 2006. Data simulation software for whole-genome association and other studies in human genetics. Pac Symp Biocomput in press.

Dupont WD, Plummer WD, Jr. 1990. Power and sample size calculations. A review and computer program. Control Clin Trials 11:116-128.

Ertekin-Taner N, Graff-Radford N, Younkin LH, Eckman C, Baker M, Adamson J, Ronald J, Blangero J, Hutton M, Younkin SG. 2000. Linkage of plasma Abeta42 to a quantitative locus on chromosome 10 in late-onset Alzheimer's disease pedigrees. Science 290:2303-2304.

Ertekin-Taner N, Ronald J, Asahara H, Younkin L, Hella M, Jain S, Gnida E, Younkin S, Fadale D, Ohyagi Y, Singleton A, Scanlin L, de AM, Petersen R, Graff-Radford N, Hutton M, Younkin S. 2003. Fine mapping of the alpha-T catenin gene to a quantitative trait locus on chromosome 10 in late-onset Alzheimer's disease pedigrees. Hum Mol Genet 12:3133-3143.

Evans DA, Funkenstein HH, Albert MS, Scherr PA, Cook NR, Chown MJ, Hebert LE, Hennekens CH, Taylor JO. 1989. Prevalence of Alzheimer's disease in a community population of older persons. Higher than previously reported [see comments]. JAMA 262:2551-2556.

Evans KC, Berger EP, Cho CG, Weisgraber KH, Lansbury PT, Jr. 1995. Apolipoprotein E is a kinetic but not a thermodynamic inhibitor of amyloid formation: implications for the pathogenesis and treatment of Alzheimer disease. Proc Natl Acad Sci U S A 92:763-767.

Ewens WJ, Spielman RS. 1995. The Transmission/Disequilibrium Test: History, Subdivision, and Admixture. Am J Hum Genet 57:455-464.

Farrer LA, Bowirrat A, Friedland RP, Waraska K, Korczyn AD, Baldwin CT. 2003. Identification of multiple loci for Alzheimer disease in a consanguineous Israeli-Arab community. Hum Mol Genet 12:415-422.

Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, Myers RH, Pericak-Vance MA, Risch N, van Duijn CM, APOE and Alzheimer Disease Meta Analysis Consortium. 1997. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. JAMA 278:1349-1356.

Farrer LA, Myers RH, Connor L, Cupples A, Growdon JH. 1991. Segregation analysis reveals evidence of a major gene for Alzheimer disease. Am J Hum Genet 48:1026-1033.

Feuk L, Prince JA, Breen G, Emahazion T, Carothers A, St Clair D, Brookes AJ. 2000. apolipoprotein-E dependent role for the FAS receptor in early onset Alzheimer's disease: finding of a positive association for a polymorphism in the TNFRSF6 gene. Hum Genet 107:391-396.

Finckh U, Van Hadeln K, Muller-Thomsen T, Alberici A, Binetti G, Hock C, Nitsch RM, Stoppe G, Reiss J, Gal A. 2003. Association of late-onset Alzheimer disease with a genotype of PLAU, the gene encoding urokinase-type plasminogen activator on chromosome 10q22.2. Neurogenetics.

Finlin BS, Correll RN, Pang C, Crump SM, Satin J, Andres DA. 2006. Analysis of the complex between Ca2+ channel beta-subunit and the Rem GTPase. J Biol Chem 281:23557-23566.

Folstein MF, Folstein SE, McHugh PR. 1975. "Mini-Mental State": a practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research 12:189-198.

Forero DA, Casadesus G, Perry G, Arboleda H. 2006. Synaptic dysfunction and oxidative stress in Alzheimer's disease: emerging mechanisms. J Cell Mol Med 10:796-805.

Francis PT, Palmer AM, Snape M, Wilcock GK. 1999. The cholinergic hypothesis of Alzheimer's disease: a review of progress. J Neurol Neurosurg Psychiatry 66:137-147.

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. 2002. The structure of haplotype blocks in the human genome. Science 296:2225-2229.

Geling A, Steiner H, Willem M, Bally-Cuif L, Haass C. 2002. A gamma-secretase inhibitor blocks Notch signaling in vivo and causes a severe neurogenic phenotype in zebrafish. EMBO Rep 3:688-694.

Gervais F, Paquette J, Morissette C, Krzywkowski P, Yu M, Azzi M, Lacombe D, Kong X, Aman A, Laurin J, Szarek WA, Tremblay P. 2007. Targeting soluble Abeta peptide with Tramiprosate for the treatment of brain amyloidosis. Neurobiol Aging 28:537-547.

Goate AM, Chartier-Harlin MC, Mullan MC, Brown J, Crawford F, Fidani L, Guiffra L, Haynes A, Irving N, James L, Mant R, Newton P, Rooke K, Roques P, Talbot C, Pericak-Vance MA, Roses AD, Williamson R, Rossor M, Owen M, Hardy J. 1991. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. Nature 33:53-56.

Goedert M, Spillantini MG. 2006. A century of Alzheimer's disease. Science 314:777-781.

Green DR, Reed JC. 1998. Mitochondria and apoptosis. Science 281:1309-1312.

Greenberg DA, Abreu P, Hodge SE. 1998. The power to detect linkage in complex disease by means of simple LOD-score analyses. Am J Hum Genet 63:870-879.

Grigull J, Alexandrova R, Paterson AD. 2001. Clustering of pedigrees using marker allele frequencies: impact on linkage analysis. Genet Epidemiol 21 Suppl 1:S61-S66.

Growden JH. 1995. Advances in the Diagnosis of Alzheimer's Disease. In: Iqbal K, Mortimer JA, Winblad B, Wisniewski H, editors. Research Advances in Alzheimer's Disease and Related Disorders. New York: John Wiley and Sons, Ltd. p 139-153.

Grupe A, Li Y, Rowland C, Nowotny P, Hinrichs AL, Smemo S, Kauwe JS, Maxwell TJ, Cherny S, Doil L, Tacey K, van LR, Myers A, Wavrant-De VF, Kaleem M, Hollingworth P, Jehu L, Foy C, Archer N, Hamilton G, Holmans P, Morris CM, Catanese J, Sninsky J, White TJ, Powell J, Hardy J, O'donovan M, Lovestone S, Jones L, Morris JC, Thal L, Owen M, Williams J, Goate A. 2006. A scan of chromosome 10 identifies a novel locus showing strong association with late-onset Alzheimer disease. Am J Hum Genet 78:78-88.

Gudbjartsson DF, Jonasson K, Frigge ML, Kong A. 2000. Allegro, a new computer program for multipoint linkage analysis. Nat Genet 25:12-13.

Guttman R, Altman RD, Nielsen NH. 1999. Alzheimer disease. Report of the Council on Scientific Affairs. Arch Fam Med 8:347-353.

Hahn LW, Moore JH. 2004. Ideal discrimination of discrete clinical endpoints using multilocus genotypes. In Silico Biol 4:183-194.

Hahn LW, Ritchie MD, Moore JH. 2003. Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions. Bioinformatics 19:376-382.

Hall JG. 1990. Genomic imprinting: review and relevance to human diseases. Am J Hum Genet 46:857-873.

Hardy JA, Higgins GA. 1992. Alzheimer's disease: the amyloid cascade hypothesis. Science 256:184-185.

Hauser ER, Watanabe RM, Duren WL, Bass MP, Langefeld CD, Boehnke M. 2004. Ordered subset analysis in genetic linkage mapping of complex traits. Genet Epidemiol 27:53-63.

Hauser ER, Watanabe RM, Duren WL, Boehnke M. 1998. Stratified Linkage Analysis of Complex Genetic Traits Using Relate Covariates.

Hauser MA, Li YJ, Takeuchi S, Walters R, Noureddine M, Maready M, Darden T, Hulette C, Martin E, Hauser E, Xu H, Schmechel D, Stenger JE, Dietrich F, Vance J. 2003b. Genomic convergence: identifying candidate genes for Parkinson's disease by combining serial analysis of gene expression and genetic linkage. Hum Mol Genet 12:671-677.

Hebert LE, Scherr PA, Bienias JL, Bennett DA, Evans DA. 2003. Alzheimer disease in the US population: prevalence estimates using the 2000 census. Arch Neurol 60:1119-1122.

Hermey G, Plath N, Hubner CA, Kuhl D, Schaller HC, Hermans-Borgmeyer I. 2004. The three sorCS genes are differentially expressed and regulated by synaptic activity. J Neurochem 88:1470-1476.

Hermey G, Riedel IB, Hampe W, Schaller HC, Hermans-Borgmeyer I. 1999. Identification and characterization of SorCS, a third member of a novel receptor family. Biochem Biophys Res Commun 266:347-351.

Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. 2002. A comprehensive review of genetic association studies. Genet Med 4:45-61.

Hoffman JM, Welsh-Bohmer KA, Hanson M, Crain B, Hulette C, Earl N, Coleman RE. 2000. FDG PET imaging in patients with pathologically verified dementia. J Nucl Med 41:1920-1928.

Hoh J, Wille A, Ott J. 2001. Trimming, weighting, and grouping SNPs in human case-control association studies. Genome Res 11:2115-2119.

Holliday R. 1989. DNA methylation and epigenetic mechanisms. Cell Biophys 15:15-20.

Horvath S, Xu X, Laird NM. 2001. The family based association test method: strategies for studying general genotype--phenotype associations. Eur J Hum Genet 9:301-306.

Hulette CM, Welsh-Bohmer KA, Crain B, Szymanski MH, Sinclaire NO, Roses AD. 1997. Rapid brain autopsy. The Joseph and Kathleen Bryan Alzheimer's Disease Research Center experience. Arch Pathol Lab Med 121:615-618.

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. 2004. Detection of large-scale variation in the human genome. Nat Genet 36:949-951.

Jarvik G, Larson EB, Goddard K, Schellenberg GD, Wijsman EM. 1996. Influence of apolipoprotein E genotype on the transmission of Alzheimer disease in a community-based sample [published erratum appears in Am J Hum Genet 1996 Mar;58(3):648]. Am J Hum Genet 58:191-200.

Jimenez-Sanchez G, Childs B, Valle D. 2001. Human disease genes. Nature 409:853-855.

Johansson A, Hampel H, Faltraco F, Buerger K, Minthon L, Bogdanovic N, Sjogren M, Zetterberg H, Forsell L, Lilius L, Wahlund LO, Rymo L, Prince JA, Blennow K. 2003. Increased frequency of a new polymorphism in the cell division cycle 2 (cdc2) gene in patients with Alzheimer's disease and frontotemporal dementia. Neurosci Lett 340:69-73.

Johansson A, Zetterberg H, Hampel H, Buerger K, Prince JA, Minthon L, Wahlund LO, Blennow K. 2005. Genetic association of CDC2 with cerebrospinal fluid tau in Alzheimer's disease. Dement Geriatr Cogn Disord 20:367-374.

Kamboh MI, Sanghera DK, Ferrell RE, DeKosky ST. 1995. APOE*4-associated Alzheimer's disease risk is modified by alpha 1-antichymotrypsin polymorphism. Nature Genetics 10:486-488.

Kang J, Lemaire HG, Unterbeck A, Salbaum JM, Masters CL, Grzeschik KH, Multhaup G, Beyreuther K, Muller-Hill B. 1987. The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor. Nature 325:733-736.

Kehoe P, Wavrant-De VF, Crook R, Wu WS, Holmans P, Fenton I, Spurlock G, Norton N, Williams H, Williams N, Lovestone S, Perez-Tur J, Hutton M, Chartier-Harlin MC, Shears S, Roehl K, Booth J, Van Voorst W, Ramic D, Williams J, Goate A, Hardy J, Owen MJ. 1999. A full genome scan for late onset Alzheimer's disease. Hum Mol Genet 8:237-245.

Khachaturian Z. 1985. Diagnosis of Alzheimer's disease. Arch Neurol 42:1097-1104.

Khachaturian ZS. 2006. Diagnosis of Alzheimer's disease: two-decades of progress. J Alzheimers Dis 9:409-415.

Khoury MJ, Beaty TH. 1994. Applications of the case-control method in genetic epidemiology. Epidemiol Rev 16:134-150.

Kruglyak L. 1997. The use of a genetic map of biallelic markers in linkage studies. Nature Genetics 17:21-24.

Lander ES, Kruglyak L. 1995. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nature Genetics 11:241-247.

Lander ES, Schork NJ. 1994. Genetic dissection of complex traits. Science 265:2037-2048.

Lautenschlager NT, Cupples LA, Rao VS, Auerbach SA, Becker R, Burke J, Chui H, Duara R, Foley EJ, Glatt SL, Green RC, Jones R, Karlinsky H, Kukull WA, Kurz A, Larson EB, Martelli K, Sadovnick AD, Volicer L, Waring SC, Growdon JH, Farrer LA. 1996. Risk of dementia among relatives of Alzheimer's disease patients in the MIRAGE study: What is in store for the oldest old? Neurology 46:641-650.

Lendon CL, Ashall F, Goate AM. 1997. Exploring the etiology of Alzheimer disease using molecular genetics. JAMA 277:825-831.

Levy-Lahad E, Wasco W, Poorkaj P, Romano DM, Oshima J, Pettingell WH, Yu CE, Jondro PD, Schmidt SD, Wang K, Crowley AC, Fu Y-H, Guenette SY, Galas D, Nemens E, Wijsman EM, Bird TD, Schellenberg GD, Tanzi RE. 1995. Candidate gene for the chromosome 1 familial Alzheimer's disease locus. Science 269:973-977.

Li WH, Sadler LA. 1991. Low nucleotide diversity in man. Genetics 129:513-523.

Li YJ, Oliveira SA, Xu P, Martin ER, Stenger JE, Scherzer CR, Hauser MA, Scott WK, Small GW, Nance MA, Watts RL, Hubble JP, Koller WC, Pahwa R, Stern MB, Hiner BC, Jankovic J, Goetz CG, Mastaglia F, Middleton LT, Roses AD, Saunders AM, Schmechel DE, Gullans SR, Haines JL, Gilbert JR, Vance JM, Pericak-Vance MA. 2003. Glutathione S-Transferase Omega 1 modifies age-at-onset of Alzheimer Disease and Parkinson Disease. Hum Mol Genet 12:3259-3267.

Li YJ, Xu P, Qin X, Schmechel DE, Hulette CM, Haines JL, Pericak-Vance MA, Gilbert JR. 2006. A comparative analysis of the information content in long and short SAGE libraries. BMC Bioinformatics 7:504.

Liang X, schnetz-boutaud N, Kenealy SJ, Jiang L, Bartlett J, Lynch B, Gaskell PC, Gwirtsman H, McFarland L, Bembe ML, Bronson P, Gilbert JR, Martin ER, Pericak-Vance MA, Haines JL. 2005. Covariate analysis of late-onset Alzheimer disease refines the chromosome 12 locus. Mol Psychiatry 11: 280-285.

Ma DQ, Whitehead PL, Menold MM, Martin ER, shley-Koch AE, Mei H, Ritchie MD, DeLong GR, Abramson RK, Wright HH, Cuccaro ML, Hussman JP, Gilbert JR, Pericak-Vance MA. 2005. Identification of significant association and gene-gene interaction of GABA receptor subunit genes in autism. Am J Hum Genet 77:377-388.

Ma J, Yee A, Brewer HB, Jr., Das S, Potter H. 1994. Amyloid-associated proteins alpha 1-antichymotrypsin and apolipoprotein E promote assembly of Alzheimer beta-protein into filaments. Nature 372:92-94.

Markesbery WR. 1997. Neuropathological criteria for the diagnosis of Alzheimer's disease. Neurobiol Aging 18:S13-S19.

Martin ER, Bass M.P., Hauser E.R. 2002. A Genotype-Based Association Test for General Pedigrees: The Geno-PDT.

Martin ER, Bass MP, Gilbert JR, Pericak-Vance MA, Hauser ER. 2003a. Genotype-based association test for general pedigrees: the genotype-PDT. Genet Epidemiol 25:203-213.

Martin ER, Bass MP, Hauser ER, Kaplan NL. 2003b. Accounting for linkage in family-based tests of association with missing parental genotypes. Am J Hum Genet 73:1016-1026.

Martin ER, Bronson PG, Li YJ, Wall N, Chung RH, Schmechel DE, Small G, Xu PT, Bartlett J, schnetz-boutaud N, Haines JL, Gilbert JR, Pericak-Vance MA. 2005. Interaction between the alpha-T catenin gene (VR22) and APOE in Alzheimer's disease. J Med Genet 42:787-792.

Martin ER, Lai EH, Gilbert JR, Rogala AR, Afshari AJ, Riley J, Finch KL, Stevens JF, Livak KJ, Slotterbeck BD, Slifer SH, Warren LL, Conneally PM, Schmechel DE, Purvis I, Pericak-Vance MA, Roses AD, Vance JM. 2000a. SNPing away at complex diseases: analysis of single-nucleotide polymorphisms around APOE in Alzheimer disease. Am J Hum Genet 67:383-394.

Martin ER, Monks SA, Warren LL, Kaplan NL. 2000b. A test for linkage and association in general pedigrees: the pedigree disequilibrium test. Am J Hum Genet 67:146-154.

Matsuda S, Matsuda Y, D'Adamio L. 2003. Amyloid beta protein precursor (AbetaPP), but not AbetaPP-like protein 2, is bridged to the kinesin light chain by the scaffold protein JNK-interacting protein 1. J Biol Chem 278:38601-38606.

Mayeux R, Lee JH, Romas SN, Mayo D, Santana V, Williamson J, Ciappa A, Rondon HZ, Estevez P, Lantigua R, Medrano M, Torres M, Stern Y, Tycko B, Knowles JA. 2002. Chromosome-12 mapping of late-onset Alzheimer disease among Caribbean Hispanics. Am J Hum Genet 70:237-243.

McGowan E, Eriksen J, Hutton M. 2006. A decade of modeling Alzheimer's disease in transgenic mice. Trends Genet 22:281-289.

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

McKhann G, Drachman G, Folstein M. 1984b. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of the department of health and human services task force on Alzheimer's disease. Neurology 34:939-944.

Mecocci P, MacGarvey U, Beal MF. 1994. Oxidative damage to mitochondrial DNA is increased in Alzheimer's disease. Ann Neurol 36:747-751.

Mirra SS, Heyman A, McKeel D, Sumi SM, Crain BJ, Brownlee LM, Vogel FS, Hughes JP, van Belle G, Berg L. 1991. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. Neurology 41:479-486.

Moore JH. 2003. The ubiquitous nature of epistasis in determining susceptibility to common human diseases. Hum Hered 56:73-82.

Moore JH, Williams SM. 2002. New strategies for identifying gene-gene interactions in hypertension. Ann Med 34:88-95.

Morris RW, Kaplan NL. 2002. On the advantage of haplotype analysis in the presence of multiple disease susceptibility alleles. Genet Epidemiol 23:221-233.

Morton NE. 1995. LODs past and present. Genetics 140:7-12.

Morton NE, Collins A. 1998. Tests and estimates of allelic association in complex inheritance. Proc Natl Acad Sci U S A 95:11389-11393.

Morton NE, MacLean CJ. 1974. Analysis of family resemblance. 3. Complex segregation of quantitative traits. Am J Hum Genet 26:489-503.

Motsinger AA, Ritchie MD, Shafer RW, Robbins GK, Morse GD, Labbe L, Wilkinson GR, Clifford DB, D'Aquila RT, Johnson VA, Pollard RB, Merigan TC, Hirsch MS, Donahue JP, Kim RB, Haas DW. 2006. Multilocus genetic interactions and response to efavirenz-containing regimens: an adult AIDS clinical trials group study. Pharmacogenet Genomics 16:837-845.

Mountain JL, Cavalli-Sforza LL. 1997. Multilocus genotypes, a tree of individuals, and human evolutionary history. Am J Hum Genet 61:705-718.

Mullan M, Crawford F, Buchanan J. 1994. Technical feasibility of genetic testing for Alzheimer's disease. Alzheimer Dis Assoc Disord 8:102-115.

Myers A, Holmans P, Marshall H, Kwon J, Meyer D, Ramic D, Shears S, Booth J, DeVrieze FW, Crook R, Hamshere M, Abraham R, Tunstall N, Rice F, Carty S, Lillystone S, Kehoe P, Rudrasingham V, Jones L, Lovestone S, Perez-Tur J, Williams J, Owen MJ, Hardy J, Goate AM. 2000. Susceptibility locus for Alzheimer's disease on chromosome 10. Science 290:2304-2305.

Myers A, Wavrant De-Vrieze F, Holmans P, Hamshere M, Crook R, Compton D, Marshall H, Meyer D, Shears S, Booth J, Ramic D, Knowles H, Morris JC, Williams N, Norton N, Abraham R, Kehoe P, Williams H, Rudrasingham V, Rice F, Giles P, Tunstall N, Jones L, Lovestone S, Williams J, Owen MJ, Hardy J, Goate A. 2002. Full genome screen for Alzheimer disease: stage II analysis. Am J Med Genet 114:235-244.

Myers AJ, Goate AM. 2001. The genetics of late-onset Alzheimer's disease. Curr Opin Neurol 14:433-440.

Nagy Z, Hindley NJ, Braak H, Braak E, Yilmazer-Hanke DM, Schultz C, Barnetson L, King EM, Jobst KA, Smith AD. 1999. The progression of Alzheimer's disease from limbic regions to the neocortex: clinical, radiological and pathological relationships. Dement Geriatr Cogn Disord 10:115-120.

Nee LE, Eldridge R, Sunderland T, Thomas CB, Katz D, Thompson KE, Weingartner H, Weiss H, Julian C, Cohen R. 1987. Dementia of the Alzheimer type: clinical and family study of 22 twin pairs. Neurology 37:359-363.

Noureddine MA, Li YJ, Van Der Walt JM, Walters R, Jewett RM, Xu H, Wang T, Walter JW, Scott BL, Hulette C, Schmechel D, Stenger JE, Dietrich F, Vance JM, Hauser MA. 2005. Genomic convergence to identify candidate genes for Parkinson disease: SAGE analysis of the substantia nigra. Mov Disord 20:1299-1309.

Olichney JM, Hansen LA, Galasko D, Saitoh T, Hofstetter CR, Katzman R, Thal LJ. 1996. The apolipoprotein E epsilon 4 allele is associated with increased neuritic plaques and cerebral amyloid angiopathy in Alzheimer's disease and Lewy body variant. Neurology 47:190-196.

Oliver C, Holland AJ. 1986. Down's syndrome and Alzheimer's disease: a review. Psychol Med 16:307-322.

Opatowsky Y, Chomsky-Hecht O, Kang MG, Campbell KP, Hirsch JA. 2003. The voltage-dependent calcium channel beta subunit contains two stable interacting domains. J Biol Chem 278:52323-52332.

Ott J. 1999. Analysis of Human Genetic Linkage. Baltimore: Johns Hopkins University Press.

Ott J, Hoh J. 2003. Set association analysis of SNP case-control and microarray data. J Comput Biol 10:569-574.

Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, Kautzer CR, Lee DH, Marjoribanks C, McDonough DP, Nguyen BT, Norris MC, Sheehan JB, Shen N, Stern D, Stokowski RP, Thomas DJ, Trulson MO, Vyas KR, Frazer KA, Fodor SP, Cox DR. 2001. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science 294:1719-1723.

Payao SL, Smith MD, Bertolucci PH. 1998. Differential chromosome sensitivity to 5-azacytidine in Alzheimer's disease. Gerontology 44:267-271.

Peduzzi P, Concato J, Kemper E, Holford T, Feinstein A. 1996. A simulation study of the number of events per variable in logistic regression. J Clin Epidemiol 49:1373-1379.

Pericak-Vance MA, Bass ML, Yamaoka LH, Gaskell PC, Scott WK, Terwedow HA, Menold MM, Conneally PM, Small GW, Saunders AM, Roses AD, Haines JL. 1998. Complete genomic screen in late-onset familial Alzheimer's disease. Neurobiol Aging 19:S39-S42.

Pericak-Vance MA, Bass MP, Yamaoka LH, Gaskell PC, Scott WK, Terwedow HA, Menold MM, Conneally PM, Small GW, Vance JM, Saunders AM, Roses AD, Haines JL. 1997. Complete genomic screen in late-onset familial Alzheimer disease: evidence for a new locus on chromosome 12. JAMA 278:1237-1241.

Pericak-Vance MA, Bebout JL, Gaskell PC, Yamaoka LH, Hung W-Y, Alberts MJ, Walker AP, Bartlett RJ, Haynes CS, Welsh KA, Earl NL, Heyman A, Clark CM, Roses AD. 1991. Linkage studies in familial Alzheimer's disease: evidence for chromosome 19 linkage. Am J Hum Genet 48:1034-1050.

Pericak-Vance MA, Grubber J, Bailey LR, Hedges D, West S, Santoro L, Kemmerer B, Hall JL, Saunders AM, Roses AD, Small GW, Scott WK, Conneally PM, Vance JM, Haines JL. 2000. Identification of novel genes in late-onset Alzheimer's disease. Exp Gerontol 35:1343-1352.

Pericak-Vance MA, Haines JL. 1995. Genetic susceptilbility to Alzheimer disease. Trends in Genetics 11:504-508.

Pericak-Vance MA, Haines JL. 1999. The Genetics of Alzheimer disease. In: King RA, Rotter JI, Motulsky AG, editors. Genetic Basis of Common Diseases. Oxford: Oxford University Press.

Pfeifer M, Boncristiano S, Bondolfi L, Stalder A, Deller T, Staufenbiel M, Mathews PM, Jucker M. 2002. Cerebral hemorrhage after passive anti-Abeta immunotherapy. Science 298:1379.

Poirier J, Delisle MC, Quirion R, Aubert I, Farlow M, Lahiri D, Hui S, Bertrand P, Nalbantoglu J, Gilfix BM, Gauthier S. 1995. Apolipoprotein E4 allele as a predictor of cholinergic deficits and treatment outcome in Alzheimer disease. Proc Natl Acad Sci U S A 92:12260-12264.

Prager K, Wang-Eckhardt L, Fluhrer R, Killick R, Barth E, Hampel H, Haass C, Walter J. 2007. A structural switch of presenilin 1 by GSK-3beta mediated phosphorylation regulates the interaction with beta -catenin and its nuclear signaling. J Biol Chem.

Pritchard JK. 2001. Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet 69:124-137.

Qin S, Zhao X, Pan Y, Liu J, Feng G, Fu J, Bao J, Zhang Z, He L. 2005. An association study of the N-methyl-D-aspartate receptor NR1 subunit gene (GRIN1) and NR2B subunit gene (GRIN2B) in schizophrenia with universal DNA microarray. Eur J Hum Genet 13:807-814.

Rall SC, Jr., Weisgraber KH, Innerarity TL, Mahley RW. 1982. Structural basis for receptor binding heterogeneity of apolipoprotein E from type III hyperlipoproteinemic subjects. Proc Natl Acad Sci U S A 79:4696-4700.

Rao VS, Cupples A, van Duijn CM, Kurz A, Green RC, Chui H, Duara R, Auerbach SA, Volicer L, Wells J, Van Broeckhoven C, Growdon JH, Haines JL, Farrer LA. 1996. Evidence for major gene inheritance of Alzheimer disease in families of patients with and without apolipoprotein E epsilon 4. Am J Hum Genet 59:664-675.

Rao VS, van Duijn CM, Connor-Lacke L, Cupples LA, Growdon JH, Farrer LA. 1994. Multiple etiologies for Alzheimer disease are revealed by segregation analysis. Am J Hum Genet 55:991-1000.

Rapoport SI, Pettigrew KD, Schapiro MB. 1991. Discordance and concordance of dementia of the Alzheimer type (DAT) in monozygotic twins indicate heritable and sporadic forms of Alzheimer's disease. Neurology 41:1549-1553.

Rebeck GW, Harr SD, Strickland DK, Hyman BT. 1995. Multiple, diverse senile plaque-associated proteins are ligands of an apolipoprotein E receptor, the alpha 2-macroglobulin receptor/low-density-lipoprotein receptor-related protein. Ann Neurol 37:211-217.

Rebeck GW, Reiter JS, Strickland DK, Hyman BT. 1993. Apolipoprotein E in sporadic Alzheimer's disease: allelic variation and receptor interactions. Neuron 11:575-580.

Reich DE, Lander ES. 2001. On the allelic spectrum of human disease. Trends Genet 17:502-510.

Renvoize EB. 1984. An HLA and family study of Alzheimer's disease. Psychol Med 14:515-520.

Rimmler J, McDowell J, Slotterbeck BD, Haynes CS, Menold MM, Rogala A, Speer MC, Gilbert J.R., Hauser E.R., Vance J.M., Pericak-Vance M.A. 1998. Development of a data coordinating center (DCC): data quality control for complex disease studies. Am J Hum Genet 63:A240.

Risch N, Merikangas K. 1996. The future of genetic studies of complex human disorders. Science 273:1516-1517.

Risch NJ. 2000. Searching for genetic determinants in the new millennium. Nature 405:847-856.

Ritchie MD, Hahn LW, Moore JH. 2003. Power of multifactor dimensionality reduction for detecting gene-gene interactions in the presence of genotyping error, missing data, phenocopy, and genetic heterogeneity. Genet Epidemiol 24:150-157.

Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, Moore JH. 2001. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. Am J Hum Genet 69:138-147.

Ritchie MD, Motsinger AA. 2005. Multifactor dimensionality reduction for detecting gene-gene and gene-environment interactions in pharmacogenomics studies. Pharmacogenomics 6:823-834.

Rocca WA, Hofman A, Brayne C, Breteler MM, Clarke M, Copeland JR, Dartigues JF, Engedal K, Hagnell O, Heeren TJ. 1991. Frequency and distribution of Alzheimer's disease in Europe: a collaborative study of 1980-1990 prevalence findings. The EURODEM- Prevalence Research Group. Ann Neurol 30:381-390.

Rocchi A, Pellegrini S, Siciliano G, Murri L. 2003. Causative and susceptibility genes for Alzheimer's disease: a review. Brain Res Bull 61:1-24.

Rogaev EI, Sherrington R, Rogaeva EA, Levesque G, Ikeda M, Liang G, Chi H, Lin C, Holman K, Tsuda T, Mar L, Sorbi S, Nacmias B, Placentini S, Amaducci L, Chumakov I, Cohen D, Lannfelt L, Fraser PE, Rommens JM, St George-Hyslop PH. 1995. Familial Alzheimer's disease in kindreds with missense mutations in a gene on chromosome 1 related to the Alzheimer's disease type 3 gene. Nature 376:775-778.

Rogaeva E, Meng Y, Lee JH, Gu Y, Kawarai T, Zou F, Katayama T, Baldwin CT, Cheng R, Hasegawa H, Chen F, Shibata N, Lunetta KL, Pardossi-Piquard R, Bohm C, Wakutani Y, Cupples LA, Cuenco KT, Green RC, Pinessi L, Rainero I, Sorbi S, Bruni A, Duara R, Friedland RP, Inzelberg R, Hampe W, Bujo H, Song YQ, Andersen OM, Willnow TE, Graff-Radford N, Petersen RC, Dickson D, Der SD, Fraser PE, Schmitt-Ulms G, Younkin S, Mayeux R, Farrer LA, St George-Hyslop P. 2007. The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. Nat Genet 39:168-177.

Rovelet-Lecrux A, Hannequin D, Raux G, Le MN, Laquerriere A, Vital A, Dumanchin C, Feuillette S, Brice A, Vercelletto M, Dubas F, Frebourg T, Campion D. 2006. APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. Nat Genet 38:24-26.

Sanan DA, Weisgraber KH, Russell SJ, Mahley RW, Huang D, Saunders A, Schmechel D, Wisniewski T, Frangione B, Roses AD, . 1994. Apolipoprotein E associates with beta amyloid peptide of Alzheimer's disease to form novel monofibrils. Isoform apoE4 associates more efficiently than apoE3. J Clin Invest 94:860-869.

SAS Institute. 2003. SAS, Version 9.1.

Scarpa S, Cavallaro RA, D'Anselmi F, Fuso A. 2006. Gene silencing through methylation: an epigenetic intervention on Alzheimer disease. J Alzheimers Dis 9:407-414.

Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. 2002. Score tests for association between traits and haplotypes when linkage phase is ambiguous. Am J Hum Genet 70:425-434.

Schoenberg BS, Kokmen E, Okazaki H. 1987. Alzheimer's disease and other dementing illnesses in a defined United States population: incidence rates and clinical features. Ann Neurol 22:724-729.

Scott WK, Saunders AM, Gaskell PC, Locke PA, Growdon JH, Farrer LA, Auerbach SA, Roses AD, Haines JL, Pericak-Vance MA. 1997. Apolipoprotein E ε2 does not increase risk of early-onset sporadic Alzheimer disease. Annals of Neurology 42:376-378.

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M. 2004. Large-scale copy number polymorphism in the human genome. Science 305:525-528.

Selkoe DJ. 2001. Clearing the brain's amyloid cobwebs. Neuron 32:177-180.

Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S, Segraves R, Oseroff VV, Albertson DG, Pinkel D, Eichler EE. 2005. Segmental duplications and copy-number variation in the human genome. Am J Hum Genet 77:78-88.

Sherrington R, Rogaev E, Liang Y, Rogaeva EA, Levesque G, Ikeda M, Chi H, Lin C, Li G, Holman K, Tsuda T, Mar L, Foncin J-F, Bruni AC, Montesi MP, Sorbi S, Rainero I, Pinessi L, Nee L, Chumakov I, Pollen D, Brookes A, Sanseau P, Polinsky RJ, Wasco W, DaSilva HAR, Haines JL, Pericak-Vance MA, Tanzi RE, Roses AD, Fraser PE, Rommens JM, St George-Hyslop PH. 1995. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. Nature 375:754-760.

Shim SB, Lim HJ, Chae KR, Kim CK, Hwang DY, Jee SW, Lee SH, Sin JS, Leem YH, Lee SH, Cho JS, Lee HH, Choi SY, Kim YK. 2007. Tau overexpression in transgenic mice induces glycogen synthase kinase 3beta and beta-catenin phosphorylation. Neuroscience.

Shoffner JM, Brown MD, Torroni A, Lott MT, Cabell MF, Mirra SS, Beal MF, Yang CC, Gearing M, Salvo R, . 1993. Mitochondrial DNA variants observed in Alzheimer disease and Parkinson disease patients. Genomics 17:171-184.

Siemers ER, Quinn JF, Kaye J, Farlow MR, Porsteinsson A, Tariot P, Zoulnouni P, Galvin JE, Holtzman DM, Knopman DS, Satterwhite J, Gonzales C, Dean RA, May PC. 2006. Effects of a gamma-secretase inhibitor in a randomized study of patients with Alzheimer disease. Neurology 66:602-604.

Sing CF, Stengard JH, Kardia SL. 2004. Dynamic relationships between the genome and exposures to environments as causes of common human diseases. World Rev Nutr Diet 93:77-91.

Slonim DK. 2002. From patterns to pathways: gene expression data analysis comes of age. Nat Genet 32 Suppl:502-508.

Small GW, Ebeling SC, Matsuyama SS, Heyman A, Reisner EG, Renvoize EB, Sulkava R. 1991. Variable association of HLA-A2 in men with early-onset Alzheimer disease. Neurobiol Aging 12:375-377.

Soininen H, Kosunen O, Helisalmi S, Mannermaa A, Paljarvi L, Talasniemi S, Ryynanen M, Riekkinen P, Sr. 1995. A severe loss of choline acetyltransferase in the frontal cortex of Alzheimer patients carrying apolipoprotein epsilon 4 allele. Neurosci Lett 187:79-82.

Spielman RS, McGinnis RE, Ewens WJ. 1993. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52:506-516.

St George-Hyslop P, Haines JL, Rogaev E, Mortilla M, Vaula G, Pericak-Vance MA, Foncin J-F, Montesi M, Bruni A, Sorbi S, Rainero I, Pinessi L, Pollen D, Polinsky R, Nee L, Kennedy J, Macciardi F, Rogaeva E, Liang Y, Alexandrova N, Lukiw W, Schlumpf K, Tanzi R, Tsuda T, Farrer L, Cantu J-M, Duara R, Amaducci L, Bergamini L, Gusella J, Roses A, McLachlan DC. 1992. Genetic evidence for a novel familial Alznheimer's disease locus on chromosome 14. Nature Genetics 2:330-334.

St George-Hyslop PH, Myers RH, Haines JL, Farrer LA, Tanzi RE, Abe K, James MF, Conneally PM, Polinsky RJ, Gusella JF. 1989. Familial Alzheimer's disease: progress and problems. Neurobiology of Aging 10:417-425.

Strittmatter WJ, Weisgraber KH, Huang DY, Dong L-M, Salvesen GS, Pericak-Vance MA, Schmechel D, Saunders AM, Goldgaber D, Roses AD. 1993. Binding of human apolipoprotein E to synthetic amyloid β peptide: isoform-specific effects and implications for late-onset Alzheimer disease. Proceedings of the National Academy of Science, USA 90:8098-8102.

Tabor HK, Risch NJ, Myers RM. 2002. Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat Rev Genet 3:391-397.

Tang MX, Jacobs D, Stern Y, Marder K, Schofield P, Gurland B, Andrews H, Mayeux R. 1996. Effect of oestrogen during menopause on risk and age at onset of Alzheimer's disease. Lancet 348:429-432.

Teng EL, Chui HC. 1987. The modified Mini-Mental State (3MS) examination. Journal of Clinical Psychiatry 48:314-318.

Thornton-Wells TA, Moore JH, Haines JL. 2004. Genetics, statistics and human disease: analytical retooling for complexity. Trends Genet 20:640-647.

Tong AH, Lesage G, Bader GD, Ding H, Xu H, Xin X, Young J, Berriz GF, Brost RL, Chang M, Chen Y, Cheng X, Chua G, Friesen H, Goldberg DS, Haynes J, Humphries C, He G, Hussein S, Ke L, Krogan N, Li Z, Levinson JN, Lu H, Menard P, Munyana C, Parsons AB, Ryan O, Tonikian R, Roberts T, Sdicu AM, Shapiro J, Sheikh B, Suter B, Wong SL, Zhang LV, Zhu H, Burd CG, Munro S, Sander C, Rine J, Greenblatt J, Peter M, Bretscher A, Bell G, Roth FP, Brown GW, Andrews B, Bussey H, Boone C. 2004. Global mapping of the yeast genetic interaction network. Science 303:808-813.

Tsai CT, Lai LP, Lin JL, Chiang FT, Hwang JJ, Ritchie MD, Moore JH, Hsu KL, Tseng CD, Liau CS, Tseng YZ. 2004. Renin-angiotensin system gene polymorphisms and atrial fibrillation. Circulation 109:1640-1646.

van Duijn CM. 1996. Epidemiology of the dementias: recent developments and new approaches. J Neurol Neurosurg Psychiatry 60:478-488.

van Duijn CM, Clayton D, Chandra V, Fratiglioni L, Graves AB, Heyman A, Jorm AF, Kokmen E, Kondo K, Mortimer JA, . 1991. Familial aggregation of Alzheimer's disease and related disorders: a collaborative re-analysis of case-control studies. EURODEM Risk Factors Research Group. Int J Epidemiol 20 Suppl 2:S13-S20.

van Duijn CM, Farrer LA, Cupples LA, Hofman A. 1993. Genetic transmission of Alzheimer's disease among families in a Dutch population based study. J Med Genet 30:640-646.

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. 1995. Serial analysis of gene expression. Science 270:484-487.

Velez DR, White BC, Motsinger AA, Bush WS, Ritchie MD, Williams SM, Moore JH. 2007. A balanced accuracy function for epistasis modeling in imbalanced datasets using multifactor dimensionality reduction. Genet Epidemiol.

Vincent I, Jicha G, Rosado M, Dickson DW. 1997. Aberrant expression of mitotic cdc2/cyclin B1 kinase in degenerating neurons of Alzheimer's disease brain. J Neurosci 17:3588-3598.

Wall JD. 2001. Insights from linked single nucleotide polymorphisms: what we can learn from linkage disequilibrium. Curr Opin Genet Dev 11:647-651.

Wall JD, Pritchard JK. 2003a. Assessing the performance of the haplotype block model of linkage disequilibrium. Am J Hum Genet 73:502-515.

Wall JD, Pritchard JK. 2003b. Haplotype blocks and linkage disequilibrium in the human genome. Nat Rev Genet 4:587-597.

Weeks DE, Lathrop GM. 1995. Polygenic disease: methods for mapping complex diesase traits.. Trends in Genetics 11:513-519.

Weiner HL, Lemere CA, Maron R, Spooner ET, Grenfell TJ, Mori C, Issazadeh S, Hancock WW, Selkoe DJ. 2000. Nasal administration of amyloid-beta peptide decreases cerebral amyloid burden in a mouse model of Alzheimer's disease. Ann Neurol 48:567-579.

Weisgraber KH, Innerarity TL, Mahley RW. 1982. Abnormal lipoprotein receptor-binding activity of the human E apoprotein due to cysteine-arginine interchange at a single site. J Biol Chem 257:2518-2521.

Wickelgren I. 2003. Estrogen research. Brain researchers try to salvage estrogen treatments. Science 302:1138-1139.

Williams SM, Haines JL, Moore JH. 2004. The use of animal models in the study of complex disease: all else is never equal or why do so many human studies fail to replicate animal findings? Bioessays 26:170-179.

Wisniewski T, Golabek A, Matsubara E, Ghiso J, Frangione B. 1993. Apolipoprotein E: binding to soluble Alzheimer beta-amyloid. Biochemical & Biophysical Research Communications 192:359-365.

Wyss-Coray T. 2006. Inflammation in Alzheimer disease: driving force, bystander or beneficial response? Nat Med 12:1005-1015.

Xu H, Gregory SG, Hauser ER, Stenger JE, Pericak-Vance MA, Vance JM, Zuchner S, Hauser MA. 2005a. SNPselector: a web tool for selecting SNPs for genetic association studies. BioInformatics 21:4181-4186.

Xu PT, Li YJ, Qin XJ, Scherzer CR, Xu H, Schmechel DE, Hulette CM, Ervin J, Gullans SR, Haines J, Pericak-Vance MA, Gilbert JR. 2005b. Differences in apolipoprotein E3/3 and E4/4 allele-specific gene expression in hippocampus in Alzheimer disease. Neurobiol Dis.

Yamaguchi H, Hara M, Strobeck M, Fukasawa K, Schwartz A, Varadi G. 1998. Multiple modulation pathways of calcium channel activity by a beta subunit. Direct evidence of beta subunit participation in membrane trafficking of the alpha1C subunit. J Biol Chem 273:19348-19356.

Zaykin D, Zhivotovsky L, Weir BS. 1995. Exact tests for association between alleles at arbitrary numbers of loci. Genetica 96:169-178.

Zubenko GS, Hughes HB, Stiffler JS, Hurtt MR, Kaplan BB. 1998. A genome survey for novel Alzheimer disease risk loci: results at 10-cM resolution. Genomics 50:121-128.