Methodological Approaches to Study the Genetics of Dementia and Cognitive Function

Fan Liu

The work in this thesis was performed at the Genetic Epidemiology Unit, Department of Epidemiology & Biostatistics and Clinical Genetics, Erasmus Medical Center, Rotterdam, The Netherlands. The Rotterdam Study is supported by the Erasmus Medical Center and Erasmus University Rotterdam, the Netherlands Organization for Scientific Research (NWO), the Netherlands Organization for Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry of Health, Welfare and Sports, the European Commission (DG XII), the Municipality of Rotterdam, and the Centre for Medical Systems Biology (CMSB). The ERF Study is supported by the Centre for Medical Systems Biology (CMSB), the Netherlands Organization for Scientific Research (NWO), the Dutch Kidney Foundation, the Dutch Heart Foundation, the Hersenstichting Nederland, ISOA, and the Dutch Diabetes Foundation. Special thanks to the general practitioners and pharmacists in the Ommoord district for their contributions to the Rotterdam Study and to the general practitioners in the Erasmus Rucphen Family (ERF) region for their contributions to the ERF study. Erasmus University and the Departments of Epidemiology & Biostatistics and Clinical Genetics provided financial support for the publication of this thesis.

ISBN 978-90-8559-496-3

Layout and printing: Optima Grafische Communicatie, Rotterdam, The Netherlands

Methodological Approaches to Study the Genetics of Dementia and Cognitive Function

Een methodologische benadering voor onderzoek naar dementie en cognitieve functie

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus

Prof.dr. S.W.J. Lamberts
en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op

woensdag 18 februari 2009 om 13.45 uur

door

Fan Liu
geboren te Beijing, China

PROMOTIECOMMISSIE

Promotor	:	Prof.dr. C.M. van Duijn Prof.dr. B.A. Oostra
Overige leden	:	Prof.dr. M.H. Kayser Dr. J.C. van Swieten Prof.dr. A.G. Uitterlinden
Copromotor	:	Dr. Y.S. Aulchenko

献给爱妻温蓓

CONTENTS

CHAPTERS IN THIS THESIS ARE BASED ON THE FOLLOWING PUBLICATIONS:

Chapter 2

Liu F, Elefante S, van Duijn CM, Aulchenko YS (2006) Ignoring Distant Genealogic Loops Leads to False-positives in Homozygosity Mapping. Ann Hum Genet 70:965-970

Chapter 3

Liu F, Kirichenko A, Axenovich TI, van Duijn CM, Aulchenko YS (2008b) An approach for cutting large and complex pedigrees for linkage analysis. Eur J Hum Genet 16:854-860

Chapter 4

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

Chapter 5

Liu F, Pardo LM, Schuur M, Sanchez-Juana P, Isaacs A, Sleegers K, de Koning I, Zorkoltseva IV, Axenovich TI, Witteman JCM, Janssens ACJW, van Swieten J, Aulchenko YS, Oostra BA, van Duijna CM (2008) The apolipoprotein E gene and its age-specific effects on cognitive function. Neurobiol Aging (In press)

Chapter 6

Liu F, Ikram MA, Janssens ACJW, Schuur M, de Koning I, Isaacs A, Struchalin M, Uitterlinden AG, den Dunnen JT, Bettens K, van Broeckhoven C, Sleegers K, van Swieten J, Hofman A, Oostra BA, Aulchenko YS, Breteler MMB, van Duijn CM (2008) A study of the *SORL1* gene in Alzheimer's disease and cognitive function. Journal of Alzheimer's Disease (submitted)

Chapter 7

Ikram MA, Liu F, Oostra BA, Hofman A, van Duijn CM, Breteler MMB (2008) The *GAB2* gene is associated with the risk of Alzheimer's disease. Biological Psychiatry (In press)

Chapter 1

Introduction

GENES ASSOCIATED WITH ALZHEIMER'S DISEASE

Alzheimer's disease (AD) is the main cause of dementia and one of the most burdensome conditions of later life. The prevalence of AD grows exponentially with age, starting at less than 1% at the age of 60 years, reaching as high as 33% at the age of 85 years[1]. Smoking is one of the few non-genetic risk factors known to be involved in AD. Family history of AD has long been known as a strong predictor of the disease and the heritability of the disease was estimated as high as 79%[2]. Several genetic factors involved in AD have been identified since the end of last century. These include the amyloid precursor protein gene (*APP*), the Presenilin 1 gene (*PSEN1*) and the Presenilin 2 gene (*PSEN2*), which are involved in early onset AD. Mutations in APP cause excessive cleavage by the β- and γ-secretase enzymes, instead of normal cleavage by the α-secretase enzyme. The result is an increased production of toxic β-amyloid fragments, which are converted into insoluble aggregates that form senile plaques in brain tissue. *PSEN1* and *PSEN2* are involved in the γ-secretase complex, and mutations lead to excessive cleavage by the γ-secretes enzyme, which results in increased production and accumulation of β-amyloid fragments. To date, no monogenic mutation was described for late-onset AD. The genetic factor, which is most predominant in both early and late onset AD, is the apolipoprotein E gene (*APOE*). The *APOE* gene has 3 common allelic forms, E2 (which occurs with a frequency about 8% in Europeans), E3 (about 75%), and E4 (about 15%). The E4 allele is associated with an increased risk of AD. Compared to non-carriers, the carriers of a single copy of the E4 allele have a 3-4 fold increased risk to develop AD, and the carriers of two copies of the E4 allele have a 10-12 fold increased risk to develop AD. *APOE* is involved in cholesterol transport and β-amyloid formation[3], but the exact mechanism how it promotes AD remains unclear. Recently, a genetic test for *APOE* genotype was marketed as a tool for predicting the risk of AD (http://www.labtestsonline.org).

The four described genes together explain less than a quarter of the AD prevalence, which indicates that additional genetic risk factors remain to be identified[4,5]. Several hypothesis-free genome-wide linkage analysis targeting AD loci were conducted. As reviewed online by the Alzheimer Research Forum (http://www.alzgene.org), the replicated regions from previous genome screens include: 1p36, 1q21-31, 2p23-24, 4q35, 5p13-15, 6p21, 6q15-16, 6q25-27, 9p21-22, 10q21-22, 10q25, 12p11-12, 19q13, 21q21-22, and Xp11-21[6-21]. Several genes have been suggested to explain the linkage to chromosome 9, 10, 12 and 19, but so far these genes remain to be confirmed.

Various candidate genes were reported to be associated with late onset AD. In most cases findings have not been consistently replicated[22,23]. A large meta-analysis of all genes studied so far pinpointed thirteen potential AD susceptibility genes: angiotensin I converting enzyme (*ACE*), cholinergic receptor, nicotinic, beta 2 (*CHRNB2*), cystatin C (*CST3*), estrogen receptor 1 (*ESR1*), glyceraldehyde-3-phosphate dehydrogenase, spermatogenic (*GAPDHS*), insulin-degrading enzyme (*IDE*), 5,10-methylenetetrahydrofolate reductase (*MTHFR*), nicastrin

(*NCSTN*), prion protein (*PRNP*), *PSEN1*, transferrin (*TF*), transcription factor A, mitochondrial (*TFAM*), tumor necrosis factor (*TNF*) and neuronal sortilin-related receptor (*SORL1*)[24] (Alzgene forum http://www.alzforum.org). However, the effects of these genes were found to be small with summary odds ratios ranging from 1.11-1.38 for risk alleles and 0.92-0.67 for protective alleles[24]. One genome-wide association study of AD has been conducted, which found that alleles in GRB-associated binding protein 2 gene (*GAB2*) modify AD risk in *APOE* E4 carriers[25].

SCOPE OF THE THESIS

In this thesis, we aim to find new genes involved in AD by means of genome screen and candidate gene studies. Most of the studies are based on a genetically isolated Dutch population. During the investigation we encountered several theoretical and practical challenges in particular related to studies in the genetically isolated population. In chapter 2 we studied the effect of ignoring distant consanguineous loops on false positive findings in linkage analysis. In chapter 3, we propose an effective pedigree-cutting algorithm, which can facilitate genome-wide linkage scans in large and complex pedigrees ascertained from genetically isolated populations. In chapter 4, we searched for genes involved in late-onset AD using genome-wide analyses and high-throughput genotyping of chromosomal regions. In chapter 5, we explore the association of the *APOE* E4 allele, together on cardiovascular factors, with cognitive function. In chapter 6, we studied the *SORL1* gene, one of the latest genes implicated in AD, in relation to cognitive function and AD. Chapter 7 describes a replication study of the *GAB2* gene, which was recently reported to be associated with AD in a genome-wide association study. Finally, chapter 8 provides a general discussion of the work presented in this thesis and provides suggestions for future research on AD.

REFERENCES

1. Ferri CP, Prince M, Brayne C, Brodaty H, Fratiglioni L, Ganguli M, Hall K, Hasegawa K, Hendrie H, Huang Y, *et al.* (2005) Global prevalence of dementia: a Delphi consensus study. Lancet 366:2112-2117
2. Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S, Fiske A, Pedersen NL (2006) Role of genes and environments for explaining Alzheimer disease. Arch Gen Psychiatry 63:168-174
3. Diedrich JF, Minnigan H, Carp RI, Whitaker JN, Race R, Frey W, 2nd, Haase AT (1991) Neuropathological changes in scrapie and Alzheimer's disease are associated with increased expression of apolipoprotein E and cathepsin D in astrocytes. J Virol 65:4759-4768
4. Sleegers K, Van Duijn CM (2001) Alzheimer's Disease: Genes, Pathogenesis and Risk Prediction. Community Genet 4:197-203
5. Goedert M, Spillantini MG (2006) A century of Alzheimer's disease. Science 314:777-781
6. Kehoe P, Wavrant-De Vrieze F, Crook R, Wu WS, Holmans P, Fenton I, Spurlock G, Norton N, Williams H, Williams N, *et al.* (1999) A full genome scan for late onset Alzheimer's disease. Hum Mol Genet 8: 237-245
7. Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, Myers RH, Pericak-Vance MA, Risch N, van Duijn CM (1997) Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. *APOE* and Alzheimer Disease Meta Analysis Consortium. Jama 278:1349-1356
8. Pericak-Vance MA, Grubber J, Bailey LR, Hedges D, West S, Santoro L, Kemmerer B, Hall JL, Saunders AM, Roses AD, *et al.* (2000) Identification of novel genes in late-onset Alzheimer's disease. Exp Gerontol 35:1343-1352
9. Curtis D, North BV, Sham PC (2001) A novel method of two-locus linkage analysis applied to a genome scan for late onset Alzheimer's disease. Ann Hum Genet 65:473-481
10. Olson JM, Goddard KA, Dudek DM (2002) A second locus for very-late-onset Alzheimer disease: a genome scan reveals linkage to 20p and epistasis between 20p and the amyloid precursor protein region. Am J Hum Genet 71:154-161
11. Myers A, Holmans P, Marshall H, Kwon J, Meyer D, Ramic D, Shears S, Booth J, DeVrieze FW, Crook R, *et al.* (2000) Susceptibility locus for Alzheimer's disease on chromosome 10. Science 290:2304-2305
12. 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
13. Blacker D, Bertram L, Saunders AJ, Moscarillo TJ, Albert MS, Wiener H, Perry RT, Collins JS, Harrell LE, Go RC, *et al.* (2003) Results of a high-resolution genome screen of 437 Alzheimer's disease families. Hum Mol Genet 12:23-32
14. 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
15. Goddard KA, Olson JM, Payami H, van der Voet M, Kuivaniemi H, Tromp G (2004) Evidence of linkage and association on chromosome 20 for late-onset Alzheimer disease. Neurogenetics 5:121-128
16. Holmans P, Hamshere M, Hollingworth P, Rice F, Tunstall N, Jones S, Moore P, Wavrant DeVrieze F, Myers A, Crook R, *et al.* (2005) Genome screen for loci influencing age at onset and rate of decline in late onset Alzheimer's disease. Am J Med Genet B Neuropsychiatr Genet 135:24-32

17. 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
18. Hiltunen M, Mannermaa A, Thompson D, Easton D, Pirskanen M, Helisalmi S, Koivisto AM, Lehtovirta M, Ryynanen M, Soininen H (2001) Genome-wide linkage disequilibrium mapping of late-onset Alzheimer's disease in Finland. Neurology 57:1663-1668
19. 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
20. Wijsman EM, Daw EW, Yu CE, Payami H, Steinbart EJ, Nochlin D, Conlon EM, Bird TD, Schellenberg GD (2004) Evidence for a novel late-onset Alzheimer disease locus on chromosome 19p13.2. Am J Hum Genet 75:398-409
21. Ashley-Koch AE, Shao Y, Rimmler JB, Gaskell PC, Welsh-Bohmer KA, Jackson CE, Scott WK, Haines JL, Pericak-Vance MA (2005) An autosomal genomic screen for dementia in an extended Amish family. Neurosci Lett 379:199-204
22. Albert MS, Moss MB, Tanzi R, Jones K (2001) Preclinical prediction of AD using neuropsychological tests. J Int Neuropsychol Soc 7:631-639
23. Myers AJ, Goate AM (2001) The genetics of late-onset Alzheimer's disease. Curr Opin Neurol 14: 433-440
24. 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
25. Reiman EM, Webster JA, Myers AJ, Hardy J, Dunckley T, Zismann VL, Joshipura KD, Pearson JV, Hu-Lince D, Huentelman MJ, *et al.* (2007) *GAB2* alleles modify Alzheimer's risk in *APOE* epsilon4 carriers. Neuron 54:713-720

Chapter 2

Ignoring Distant Genealogic Loops Leads to False-positives in Homozygosity Mapping

ABSTRACT

Distant consanguineous loops are often unknown or ignored during homozygosity mapping analysis. This may potentially lead to an increased rate of false-positive linkage findings. We show that failure to take into account the distant loops may lead to seriously underestimated degree of consanguinity, especially for people from genetically isolated populations; in 6 Alzheimer's disease (AD) patients, the distant loops accounted for 57.7 % of inbreeding on average. Theoretical evaluation showed that ignoring distant loops, which account for 18-75% of inbreeding, inflates the frequency of false positive conclusions substantially in 2-point linkage analysis up to several hundred times. In multipoint linkage analysis of the 6 AD patients, a chromosome-wide "empirical" significance of 5% corresponded to a true false positive rate of 11.1%. We show that converting multiple loops to a hypothetical loop capturing all inbreeding may be a convenient solution to avoid false positive results. When extended genealogic data are not available, a hypothetical loop may still be constructed based on genomic data.

INTRODUCTION

Homozygosity mapping is a highly effective method to map disease loci. It exploits the fact that autosomal DNA regions adjacent to the mutation causing a recessive phenotype are likely to be homozygous by descent in patients of consanguineous marriages[1]. This method is very powerful and linkage can be detected with a small sample: three offspring from independent first-cousin marriages are sufficient to obtain a LOD score of 3.6. Using homozygosity mapping, various genes have been identified successfully[2,3]. In founder populations, levels of inbreeding may be high and people are often related through multiple lines of descent. The presence of multiple consanguineous loops poses computational challenges. In spite of recent advances in computational efficiency[4], exact multipoint calculations are limited to pedigrees of a few dozen of members. Approximate methods, such as those based on Markov-chain Monte Carlo algorithms, have proven very useful in resolving this problem[5]. However, in large pedigrees containing multiple loops, these methods also fall short, especially in the context of genome screens. This is one of the reasons why the analysis is often performed using the shortest loop only. Second, genealogic data may be limited to a few generations even for isolate populations. This is especially true for outbred populations, where low levels of unobserved consanguinity may also be common[6]. It has been suggested that ignoring the existence of distant loops may inflate false-positive rates[7]. For model free sib-pair design, adjustment of the expected IBD vector for the inbreeding provides a solution[8]. In this study, we aimed to (1) investigate the extent of inbreeding that can be explained by distant consanguineous loops in a genetically isolated population, (2) quantify the effect of ignoring distant loops on false positive rate in model-based homozygosity mapping, and (3) explore the potential for reducing the false positive rate by converting multiple loops to a single hypothetical loop capturing all inbreeding.

METHODS AND RESULTS

Population

The study was performed within the framework of the Genetic Research in Isolated Populations (GRIP) program. The program is set in a genetically isolated community of approximately 20,000 inhabitants located in the Southwest of The Netherlands. Between 150 and 400 individuals founded the population in the middle of the 18th century. The GRIP genealogical database currently contains information on more than 70,000 individuals from this population. We studied the genealogy of 6 Alzheimer's disease patients who participated in the program. As with the other participants examined, they, and their relatives, have provided informed consent before inclusion into the study. The scientific protocol of GRIP was approved by the Medical Ethics Committee of the Erasmus MC.

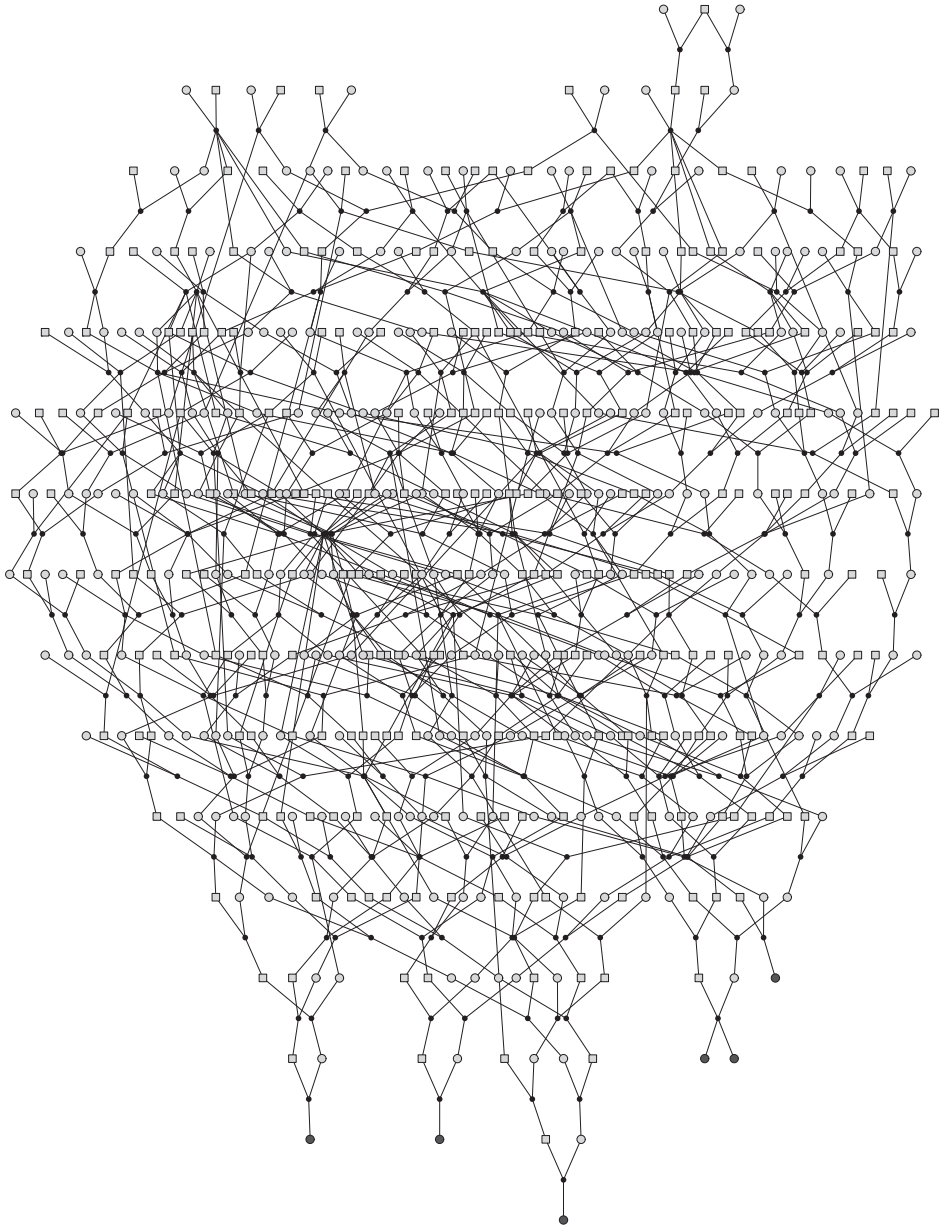


Figure 1. The pedigree of 6 Alzheimer's disease patients (black dots) coming from a genetically isolated population.

Contribution of distant consanguineous loops to inbreeding

Figure 1 shows the 6 Alzheimer's disease (AD) patients and their 593 ancestors in 16 generations. These patients were selected for homozygosity mapping because they are offspring of 2nd to 3rd cousin marriages. A complex, multi-looped structure however existed for each patient. To enumerate all consanguineous loops for each patient, we developed FCN program for characterizing the structure of extremely large pedigrees (ftp://mga.bionet.nsc.nu/fcn). The program exploits a recursive Depth-First-Search algorithm (Figure 2). Table 1 describes the distribution of consanguineous loops for the 6 AD patients. The total number of consanguineous loops ranged from 197 (patient 6) to 677 (patient 2) with a mean of 388. The mean length of consanguineous loops was 18.8 meioses with a standard deviation of 2.9. Although

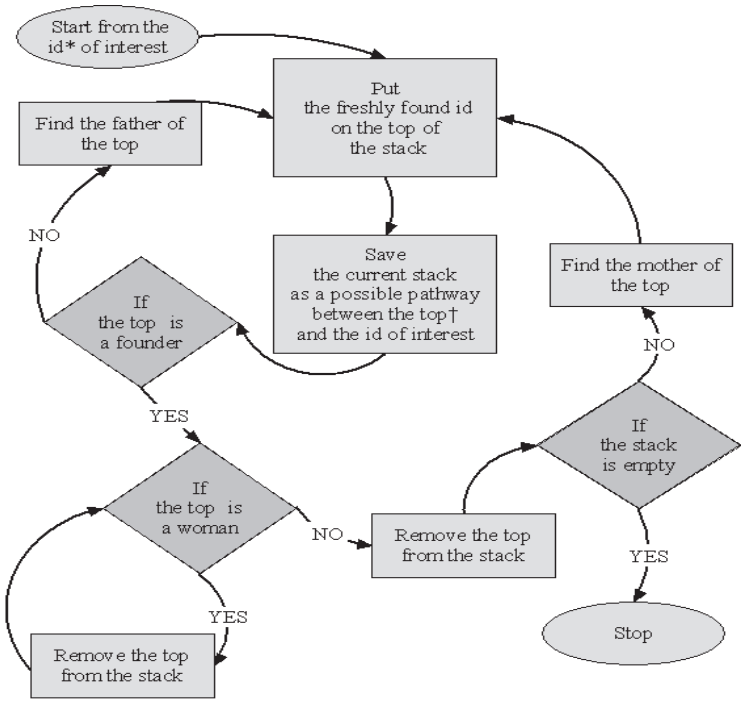


Figure 2. A Depth-First-Search (DFS) algorithm is an approach to traverse a tree structure and/or a graph. The traversing sequence is illustrated in the figure. During traversing, a dynamic stack is employed to temporally store all individuals in the current route. When DFS goes up from an individual to her/his parent, the parent is placed on the top of the stack and when it goes back, that parent (the individual on the top of the stack) is removed. Using this stack register, all of the individuals in the current route are placed in order. When the stack becomes empty, DFS has traversed all the possible pathways that start and end with the same individual. Consanguineous loops for the studied individual are those which are disjointed. This algorithm can be easily generalized to find pair-wise connections.

Table 1. Distribution of consanguineous loops for 6 Alzheimer's disease patients from a genetically isolated population with known genealogy

Patient	N^*	Shortest†	Longest†	Mean ± SD	F_t‡	F_u§	F_t/F_u
1	324	9	25	19.2 ± 3.3	0.0276	0.0078	3.5
2	677	8	25	19.5 ± 2.8	0.0294	0.0156	1.9
3	577	8	29	19.7 ± 3.2	0.0283	0.0156	1.8
4&5	277	10	24	18.7 ± 2.8	0.0137	0.0039	3.5
6	197	9	22	17.1 ± 2.7	0.0164	0.0078	2.1
Average	388.2	9.0	24.8	18.8 ± 2.9	0.0215	0.0091	2.4

* Total number of consanguineous loops.

† Number of meioses in the shortest and longest loops.

‡ True inbreeding coefficient.

§ Inbreeding coefficient taking into account only the shortest loops.

the shortest loops explains a large part of the degree of consanguinity, a major proportion of the 'true' inbreeding was contributed by multiple distant loops. The distant consanguineous loops accounted for 57.7% of inbreeding.

Inflation of false positive rate in 2-point analysis

Next, we studied the effect of ignoring distant loops on false positive rate of homozygosity mapping. In a 2-point analysis, we evaluated the false positive rate as a function of inbreeding coefficient used in the analysis, F_u, the true inbreeding coefficient F_t, the disease allele frequency q, and the marker alleles frequencies m (equal frequency assumed for all alleles).

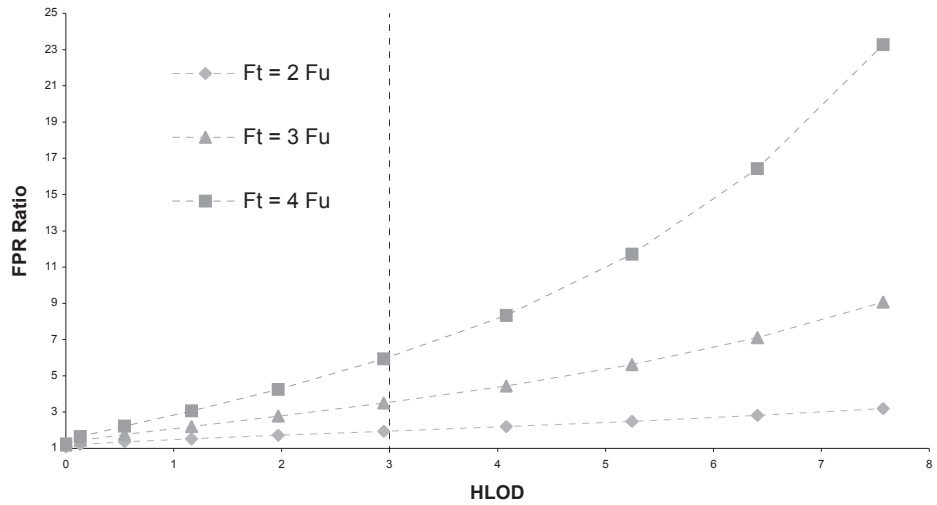


Figure 3. The inflation of false positive rate (FPR) when inbreeding coefficient is 2, 3, and 4 times underestimated in 2-point homozygosity mapping. FPR ratio was computed by using the FPR when the inbreeding values were underestimated divided by the FPR when the inbreeding values were estimated correctly. Model: $q = 0.01$, $m = 0.1$, $F_t = 1/64$. Note that because the resultant HLOD had a discrete distribution, the first HLOD ≥3 appears at HLOD = 4.1.

The heterogeneity LOD (HLOD) score at absolute linkage was derived. The false positive rate was defined as the probability of yielding a HLOD equal to or greater than a pre-specified threshold when the studied locus was not linked to the disease allele. In other words, this is the type I error of the HLOD test. The mathematical evaluation is given in appendix. Using 10 hypothetical patients with identical inbreeding values of 1/64, which corresponds to a 2nd cousin marriage, we evaluated various models in which true inbreeding values were actually 2, 3, or 4 times higher than that used in the analysis. Figure 3 shows a typical effect of ignoring distant loops on the false positive rate. When all loops were taken into account (so that the inbreeding values used in the analysis were true), the false positive rate was close to that theoretically expected under the asymptotic theory. When the true inbreeding values were higher than those used in analysis, the false positive rate increased. For example, at HLOD ≥ 3.0, the false positive rate increased 2.2, 4.3, and 8.0 times when the true inbreeding was two, three, and four times higher than the inbreeding accounted for in the analysis (Figure 3). A more extensive evaluation of different scenarios with respect of the baseline inbreeding, disease gene and marker allele frequencies, showed that ignoring distant loops, which accounted for 18-75% of inbreeding, always inflated the frequency of false positive conclusions. Under certain scenarios, the false positive rate may increase up to several hundred times. For example, when $q = 0.01$, $m = 0.01$, $F_t = 1/16$, and the true inbreeding is underestimated by a factor of 4, the false positive rate increased 1016 times when the threshold of HLOD ≥ 3.0 was applied. The false positive rate also depended on the disease and marker allele frequencies. In general, when the disease allele becomes common ($q > 0.01$), homozygosity mapping is conservative and thus is prone to produce false negatives. Once q is rare ($q \leq 0.01$) and inbreeding is adjusted correctly, the resulting false positive rates are very close to the theoretically expected ones. The effect of variation of the marker allele frequency is small and the asymptotic theory works well over a wide range of conditions (Figure 4).

Inflation of false positive rate in multipoint analysis

We next performed computer simulations to assess the false positive rate in multipoint analysis. An unlinked (to the disease locus) chromosome with a length of 100 cM with 21 equally spaced markers was simulated using the genedrop program of the MORGAN package (http://www.stat.washington.edu/thompson/Genepi/MORGAN/Morgan.shtml). For each marker, 10 alleles with equal frequency were assumed. A homozygosity mapping analyses was subsequently performed assuming $m = 0.1$, $q = 0.01$, and complete penetrance. The null distribution of the false positive rate was derived empirically by finding how many times the resultant HLOD was equal to or greater than a pre-defined threshold. The simulations were repeated 10,000 times. We used the genealogic data of the 6 AD patients (Figure 1) in the simulations. To derive the 5%, 1%, 0.5%, and 0.1% chromosome-wide significance thresholds, we repeatedly simulated and analyzed markers using the same genealogic structure (the shortest loops only). The false positive rate when only the shortest loops were used in the

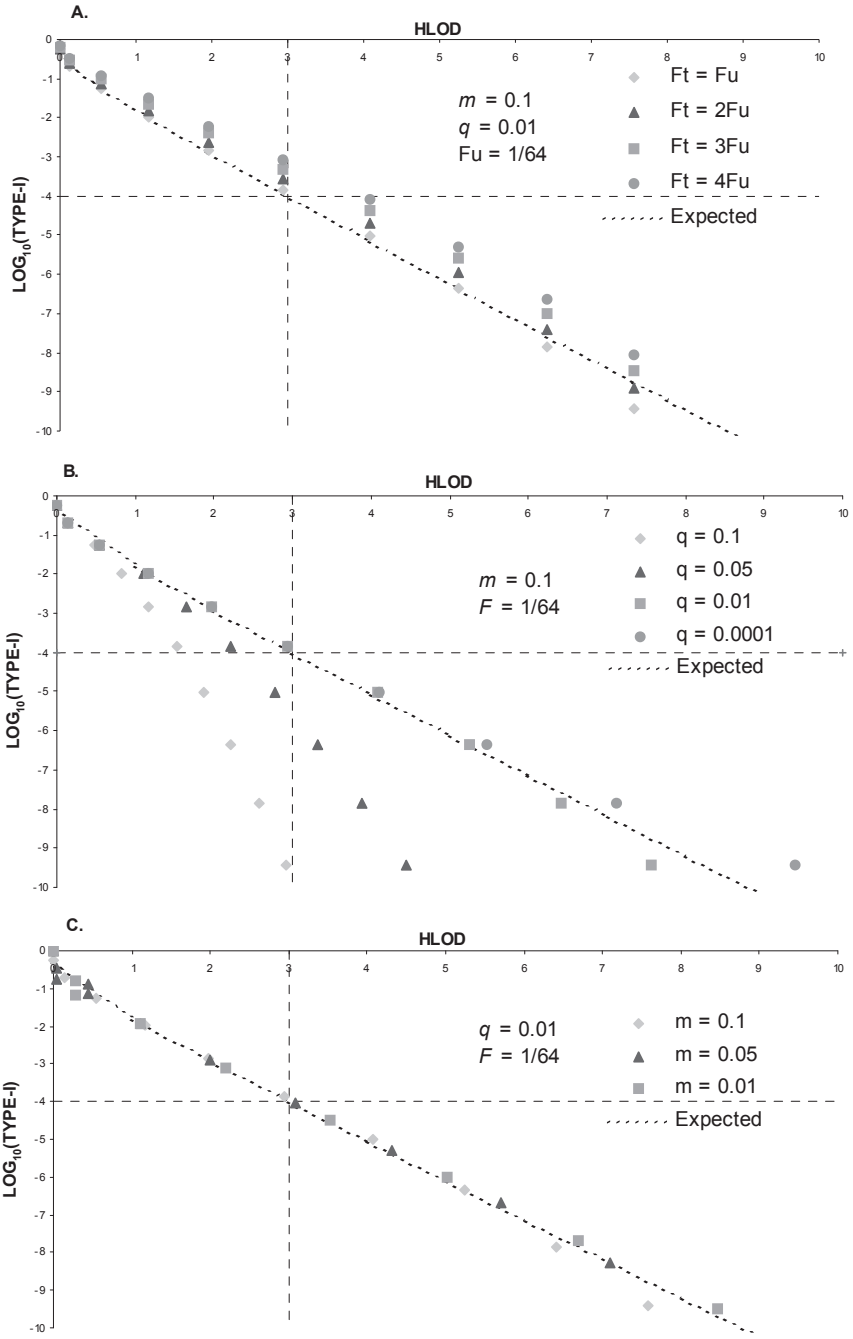


Figure 4. The effect of underestimation of inbreeding and variation of disease allele and marker allele frequency on false positive rate of 2-point homozygosity mapping using 10 hypothetical patients. The expected type-I error curve is derived from the asymptotic theory: the maximum LOD score multiplied by 2(ln10) follows a one-sided chi-square distribution with one degree of freedom.

Table 2. Effect of not taking into account all loops on false positive rate (FPR) in multipoint analysis

Expected FPR (Threshold)*	Model 1†	Model 2‡
0.05 (1.339)	0.1112	0.0341
0.01 (2.032)	0.0292	0.0067
0.005 (2.409)	0.0151	0.0022
0.001 (3.281)	0.0033	0.0002
0.0001 (4.259)	0.0005	0.0000

* The chromosome-wide significance thresholds were derived empirically by repeatedly simulating and analyzing the markers using the same genealogic structure (shortest loops only).
† Marker simulation using all loops, subsequent homozygosity mapping analysis using only the shortest loop.
‡ Marker simulation using all loops, subsequent homozygosity mapping analysis using the hypothetical loop.

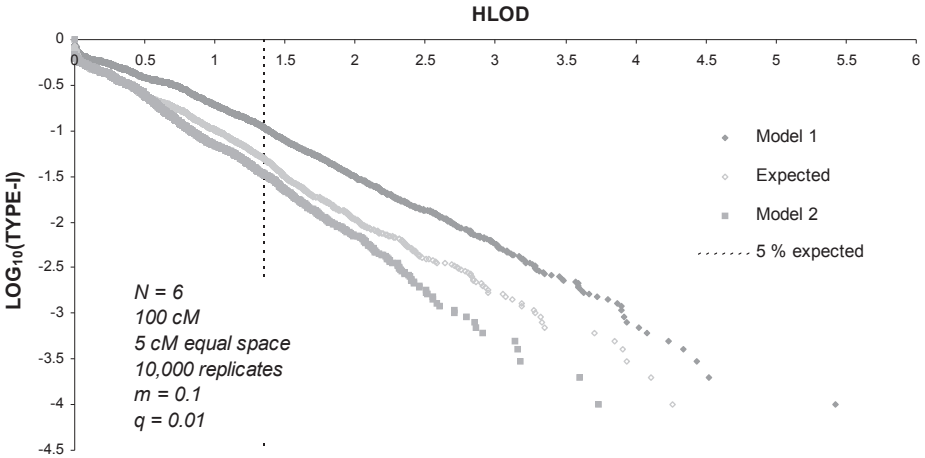


Figure 5. The effect of ignoring distant loops on false positive rate of multi-point homozygosity mapping. The expected false positive rate is derived empirically by repeatedly simulating and analyzing the markers using the shortest loops only. Model 1: maker simulation using all loops, subsequent homozygosity mapping analysis using only the shortest loops. Model 2: marker simulation using all loops, subsequent homozygosity mapping analysis using only the hypothetical loops.

analysis, in the presence of multiple distant loops contributing to inbreeding, was found by repeatedly simulating the marker inheritance in the complex, multi-loop structured pedigree, followed by a statistical analysis using the shortest loops only (model 1). Table 2 shows that when multiple loops contribute to inbreeding, using only the shortest ones inflates the false positive rate of a multipoint analysis substantially. For example, at the threshold of the expected false positive rate of 5%, taking into account only the shortest loops yielded an 11.1% false positive rate (Figure 5).

A conservative solution to avoid false positives

Thus, both two and multipoint homozygosity mapping is prone to false positives when distant loops are ignored. Ideally, one should analyze the complete pedigree to avoid the inflation of false positive rate, but, when dealing with extremely large pedigrees, this is not possible due to the computational complexity of the task. One may achieve the goal however, by simplifying the pedigree structure while keeping the inbreeding values close to, if not the same as, the true inbreeding values. For this, we suggest creating a single hypothetical loop for each patient. The number of meioses *n* in the hypothetical loop is derived using the patient's true inbreeding coefficient:

$$n = floor\,(\log_{10} F_t / \log_{10} 0.5) + 2$$

where *floor* means rounding a decimal downward to an integer. Because *n* is forced to be an integer, which is smaller than an exact conversion from the inbreeding value, the inbreeding coefficient is over-estimated, and thus, the resultant test is expected to be conservative. To access the false positive rate, we repeatedly dropped the assigned markers down through the complex multi-loop structured pedigree, and then performed a homozygosity mapping analysis using the hypothetical loops only (model 2). The results indicate that, indeed, by utilizing this approach, one is able to avoid false positives. For example, at the expected false positive rate of 5%, using the hypothetical loops yielded a 3.4% false positive rate (Table 2, Figure 5). In this study, we rounded the numbers of meioses in hypothetical loops downward to integers, which led to a conservative test. Rounding off the resultant decimals to the nearest integers may give less conservative solution.

DISCUSSION

In summary, we show that the degree of consanguinity may be seriously underestimated when only the shortest loops are known or used in the analysis, especially for people from a genetically isolated population. Although the contribution of each distant loop to inbreeding may be very small, hundreds, and even thousands, of such distant loops may exist, and they together may contribute substantially to the inbreeding. We quantified the effect of underestimation of inbreeding on the false positive rate, and showed that the frequency of false positive conclusions may be seriously inflated. To overcome this problem, we propose constructing hypothetical loops based on patients' true inbreeding values as a convenient, although not perfect, solution. The true inbreeding values can be easily computed exactly for any given pedigree regardless of size. In the absence of extended genealogy data, they may still be reliably estimated from genomic data[9].

Though use of hypothetical loops may prevent false positives, it is not a perfect solution. The classical relationship coefficients, a representation based on distance which is the "proportion of genome shared"[10], may give the same distance in a case when there is a single short connection between two people and also when there are several, and even hundreds or thousands, of longer connections. The actual genetic consequence of the sharing, however, will be quite different: people, connected via a single short path share, on average, longer genomic regions when compared to people connected via multiple long paths as a result of a increased number of recombinations. Underestimating the probability of the recombination events may reduce the power to detect linkage[11]. We developed software, which catalogs all consanguineous loops and connections for a set of related individuals, as well as calculates loop-specific and ancestor-specific inbreeding and kinship coefficients. Knowing all the consanguineous loops and the number of meioses in each loop, combined with the information on marker distances, may help computing the probability of recombination.

APPENDIX

We assume that the recombination fraction between the studied marker and the disease locus is 0.

Consider *n* independent nuclear families, each consisting of one patient with an observed genotype at the marker locus. Since the genotype G_i of the i^{th} patient could be either homozygous or heterozygous, there will be in total 2^n possible marker genotype configurations. If the HLOD of each configuration and the probability of observing such configurations are known, one can derive the null distribution of HLOD.

The likelihood ratio of linkage and non-linkage, for patient *i*, is:

$$LR_i(G_i) = \frac{L_i(G_i \mid \theta = 0)}{L_i(G_i \mid \theta = 0.5)} = \begin{cases} \dfrac{Fu_i + mq(1-Fu_i)}{[Fu_i + m(1-Fu_i)][Fu_i + q(1-Fu_i)]} & G_i = homozygosous \\ \dfrac{q}{q(1-Fu_i) + Fu_i} & G_i = hetrozygous \end{cases}.$$

Consider that a proportion (α) of the families is linked, whereas $1-\alpha$ unlinked. The likelihood ratio for the current configuration is:

$$LR = \prod_{i=1}^{n} (\alpha LR_i(G_i) + 1 - \alpha),$$

where *LR* can be maximized with respect to α, yielding MLEs $\hat{\alpha}$.

The HLOD for the current configuration is defined by a base 10 logarithmic transformation of the maximum likelihood ratio,

$$HLOD = \log_{10}(max\ LR),$$

and the probability of the current configuration is the product of the probability of observing each patient's genotype (homozygous or heterozygous) at the marker locus:

$$P(G_i) = \begin{cases} Ft_i + m(1 - Ft_i) & G_i = homozygous \\ 1 - Ft_i - m(1 - Ft_i) & G_i = heterozygous \end{cases},$$

which is not dependant on the inbreeding used in the analysis. Exhausting HLOD and probability for all configurations yields the null distribution of HLOD. The false positive rate for a given HLOD threshold is the cumulative probability of the corresponding configuration.

REFERENCES

1. Lander ES, Botstein D (1987) Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. Science 236:1567-1570
2. van Duijn CM, Dekker MC, Bonifati V, Galjaard RJ, Houwing-Duistermaat JJ, Snijders PJ, Testers L, Breedveld GJ, Horstink M, Sandkuijl LA, *et al.* (2001) Park7, a novel locus for autosomal recessive early-onset parkinsonism, on chromosome 1p36. Am J Hum Genet 69:629-634
3. Fukushima K, Ueki Y, Smith RJ (2000) Sensorineural hearing impairment, non-syndromic: DFNB5, 6, 7. Homozygosity mapping to localize genes causing autosomal recessive non-syndromic hearing loss. Adv Otorhinolaryngol 56:152-157
4. 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
5. Sobel E, Lange K (1996) Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. Am J Hum Genet 58:1323-1337
6. Broman KW, Weber JL (1999) Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. Am J Hum Genet 65:1493-1500
7. Miano MG, Jacobson SG, Carothers A, Hanson I, Teague P, Lovell J, Cideciyan AV, Haider N, Stone EM, Sheffield VC, *et al.* (2000) Pitfalls in homozygosity mapping. Am J Hum Genet 67:1348-1351
8. Leutenegger AL, Genin E, Thompson EA, Clerget-Darpoux F (2002) Impact of parental relationships in maximum lod score affected sib-pair method. Genet Epidemiol 23:413-425
9. Leutenegger AL, Prum B, Genin E, Verny C, Lemainque A, Clerget-Darpoux F, Thompson EA (2003) Estimation of the inbreeding coefficient through use of genomic data. Am J Hum Genet 73:516-523
10. Guo SW (1995) Proportion of genome shared identical by descent by relatives: concept, computation, and applications. Am J Hum Genet 56:1468-1476
11. Dyer TD, Blangero J, Williams JT, Goring HH, Mahaney MC (2001) The effect of pedigree complexity on quantitative trait linkage analysis. Genet Epidemiol 21 Suppl 1:S236-243

Chapter 3

An Approach for Cutting Large and Complex Pedigrees for Linkage Analysis

ABSTRACT

Utilizing large pedigrees in linkage analysis is a computationally challenging task. The pedigree size limits applicability of the Lander-Green-Kruglyak algorithm for linkage analysis. A common solution is to split large pedigrees into smaller computable sub-units. We present a pedigree-splitting method that, within a user supplied bit-size limit, identifies sub-pedigrees having the maximal number of subjects of interest (e.g. patients) who share a common ancestor. We compare our method with the maximum clique partitioning method using a large and complex human pedigree consisting of 50 patients with Alzheimer's disease ascertained from genetically isolated Dutch population. We show that under a bit-size limit our method can assign more patients to sub-pedigrees than the clique partitioning method, particularly when splitting deep pedigrees where the subjects of interest are scattered in recent generations and are relatively distantly related via multiple genealogic connections. Our pedigree-splitting algorithm and associated software can facilitate genome-wide linkage scans searching for rare mutations in large pedigrees coming from genetically isolated populations. The software package PedCut implementing our approach is available at http://mga.bionet.nsc.ru/soft/index.html.

INTRODUCTION

Parametric linkage analysis searching for a unique founder mutation is a powerful tool to identify rare genetic variants with large effects. Genetically isolated populations provide extended pedigrees for linkage analysis as evidenced by numerous successes for complex traits, including type 2 diabetes[1] and Alzheimer's disease[2]. In such populations, pedigrees can be reconstructed based on genealogical records resulting in deep pedigrees that include a large number of multiple lines of descent. However, utilizing such large pedigrees in linkage analysis is computationally challenging. For exact multipoint linkage analysis, several software packages implementing Lander-Green-Kruglyak algorithm[3,4], such as Genehunter[3], Merlin[5], and Allegro[6], are frequently used. The computational complexity of this algorithm increases linearly with the number of markers, but exponentially with the bit-size of the pedigree. The bit-size is defined as the twice of the number of individuals with parents presented in the pedigree minus the number of pedigree founders[3]. As long as the pedigree bit-size is small, the Lander-Green-Kruglyak algorithm can analyse a large number of markers. In modern implementations[5], the time to compute multipoint LOD score using the Lander-Green-Kruglyak algorithm also depends on the fraction of pedigree members with missing genotypic information, which may be large for deep pedigrees because phenotypes and genotypes are usually unknown from the upper generations. Therefore, exact multipoint calculations are limited to pedigrees of several dozens of bits. Programs using Markov-chain Monte Carlo (MCMC) algorithms, such as Simwalk2[7], Loki[8], and Morgan[9] can analyse larger pedigrees but still fall short in the context of large pedigrees with hundreds of loops, especially for genome-screens with large number of densely spaced markers[10,11].

A common solution to reduce the computational burden is to split a large pedigree into smaller, and thus computable, sub-units. Analyses of Hutterite pedigrees have revealed that a substantial amount of linkage information may be lost when truncating the pedigree for linkage analysis based on recent generations[12]. Manual splitting by an expert is only possible for relatively small pedigrees. For cutting large pedigrees, a semi-automatic method has been proposed based on factor analysis[13]. This method relies partly on the expert decisions and often yields sub-pedigrees that are still too complex for the Lander-Green-Kruglyak algorithm based linkage analysis. Falchi and colleagues [13]suggested a pedigree splitting method based on graph theory maximum-cliques partitioning algorithm. This algorithm, although automatic, requires a number of parameters to be pre-specified, e.g. a maximum number of generations, range for the measure of relatedness used to group individuals, and the range of the number of subjects of interest in a sub-pedigree. The optimization of these parameters largely depends on examining different sets of resultant sub-pedigrees. Furthermore, the available software implementing this algorithm does not guarantee that all of the resultant sub-pedigrees fall within specific bit-size limit and thus can be efficiently analyzed by parametric linkage analysis using the Lander-Green-Kruglyak algorithm.

In this work, we develop a fast automatic algorithm for splitting large pedigrees based on user-specified maximum bit-size restriction. The algorithm specifically aims to split deep pedigrees where patients are relatively remotely related through genealogic connections and to produce an optimal set of sub-pedigrees for parametric linkage analysis of binary traits under rare dominant mutation model.

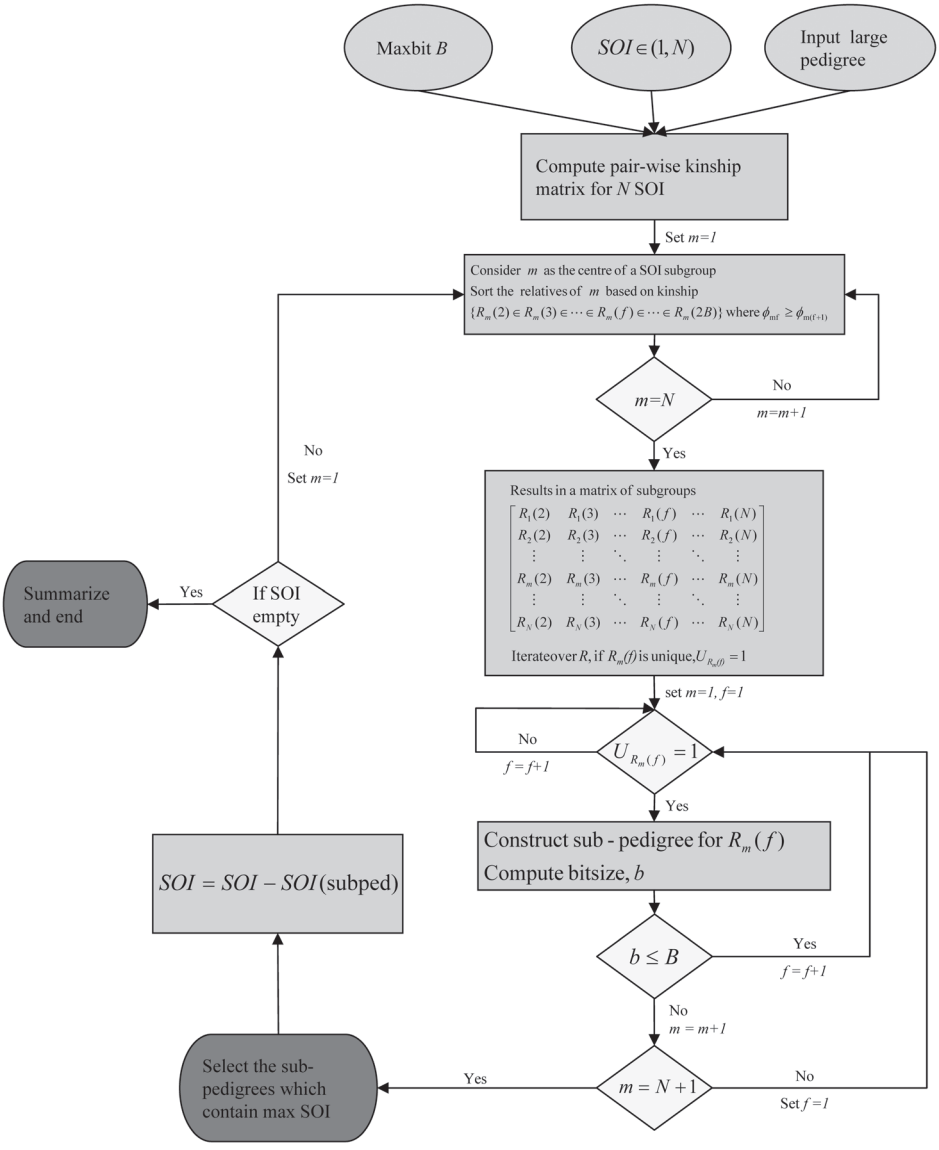


Figure 1. Flowchart of the pedigree splitting algorithm

RESULTS

Pedigree splitting algorithm

In pedigree splitting we focus on the family members who have known genotypes and/or phenotypes. These people are denoted as "subjects of interest" (SOI). We assume that some SOI share the genetic variant explaining their phenotype identical by descent from their common ancestor(s). The aim of our heuristic algorithm is to split a large pedigree into sub-pedigrees containing a maximal number of SOI who are related to a common ancestor and the bit sizes of the resultant pedigrees should be smaller than or equal to that specified by the user. This aim can be theoretically achieved by evaluating all SOI subgroups in term of bit-sizes of pedigrees relating them to a common ancestor. The number of subgroups to evaluate, however, grows exponentially with the number of SOI studied and becomes prohibitively large with more than 20-30 SOI. The number of subgroups to evaluate therefore need to be reduced.

In our algorithm the kinship coefficient, φ_{ij}, is used to measure the degree of relatedness between SOI. The kinship coefficient is defined as the expected probability that two alleles randomly sampled from a pair of relatives *i* and *j* are copies of the same ancestral allele (identical by descent). For example the kinship coefficient is 1/4 if *i* and *j* are first-degree relatives (e.g. siblings) and 1/8 if they are second-degree relatives (e.g. uncle-niece). In our work, the coefficients were calculated using a modified version of the PEDIG software developed by Didier Boichard[14].

Figure 1 is a flowchart of the algorithm. In the first step, the algorithm constructs a matrix of subgroups that are sorted based on kinship. Given a group of SOI of size *N*, consider individual $m \in (1, N)$. The set of relatives of *m* is sorted in decreasing order according to their kinship to *m*, so as *m* is the first element of this set. Let us call this set as R_m. Let $R_m(f)$ be a set of the first *f* elements of R_m, $f \in (2, N)$. When *f=2*, $R_m(2)$ includes *m* and his closest relative. When *f=3*, the next closest relative of *m is* included into $R_m(2)$, so that $R_m(1) \in R_m(2) \in \ldots \in R_m(N)$. Iterating the central individual *m* over all SOI gives an *N* by *N-1* matrix,

$$\begin{bmatrix} R_1(2) & R_1(3) & \cdots & R_1(f) & \cdots & R_1(N) \\ R_2(2) & R_2(3) & \cdots & R_2(f) & \cdots & R_2(N) \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ R_m(2) & R_m(3) & \cdots & R_m(f) & \cdots & R_m(N) \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ R_N(2) & R_N(3) & \cdots & R_N(f) & \cdots & R_N(N) \end{bmatrix},$$

where each row represents the sub-groups of SOI derived from the same central individual but having different sizes (2 to *N*). Each column represents the sub-groups of the same size, which are derived from each individual, who is considered as the centre of the corresponding subgroup. Identify unique groups of relatives by iterating over matrix R and give $R_m(f)$ a binary index of 1 when the content of $R_m(f)$ is seen for the first time, otherwise assign zero.

At the second step of our algorithm, the genealogies connecting the members of identified unique subgroups (as identified by index of 1) are constructed using the PedHunter program developed by Agarwala and colleagues[15]. For each subgroup, a sub-pedigree linking the maximal number of subgroup members to the most recent common ancestor is reconstructed and the bit-size of every sub-pedigree is computed. For each row R_m, the construction of sub-pedigrees starts at $R_m(2)$ and stops when the bit-size of $R_m(f)$ violates the maximal bit-size limit. There is no need to construct sub-pedigrees for $R_m(f)$ to $R_m(N)$ because the bit-size of $R_m(f+1)$ is always greater than that of $R_m(f)$. At this stage, all constructed sub-pedigrees satisfy the bit-size limit and contain a unique configuration of SOI. The sub-pedigree connecting the largest number of SOI , as heuristically the most interesting pedigree, is selected as the first sub-pedigree. When several non-overlapping sub-pedigrees are eligible all of them are selected. When several partly overlapping sub-pedigrees are eligible, the one with the smallest bit-size is selected. If still several sub-pedigrees are eligible, the one with the highest average kinship among SOI is selected. When, additionally, the average kinships are the same, a random sub-pedigree is selected and the alternative selections are saved in a log-file. From our experience, the latter scenario is very rare for reasonably large bit-sizes and complex pedigrees with multiple lines of descent. Within the search space, this exhaustive algorithm guarantees that the selected sub-pedigree has the maximal number of SOI in respect to the bit-size restriction *B*.

At the third step, the algorithm removes the SOI belonging to the identified pedigree(s) from further consideration and is recursively applied, starting with the Step 1, to the remaining SOI, until no further SOI can be assigned to a sub-pedigree. The described algorithm is implemented in a software package, PedCut, which is available at http://mga.bionet.nsc.ru/soft/index.html.

Splitting a large pedigree with PedCut and Greffa

We tested properties of our program PedCut using a large and complex pedigree and compared it with the Greffa program developed by Falchi and colleagues[13]. This program implements the maximum clique-partitioning algorithm. The pedigree comprises a part of the genealogy used in a genome-wide screen for late onset Alzheimer's disease (AD) in genetically isolated Dutch population[2]. The original pedigree contains 103 AD patients and 4645 family members. For demonstration purpose, we used a fraction of the original pedigree that contains 50 randomly selected AD patients and their 2460 ancestors spanning over 18 generations (Table 1 and Figure 2). These 50 AD patients, who are considered as SOI, are scattered in the most recent generations. Except 5 sibling-pairs and 2 aunt-niece pairs, all SOI are related distantly via multiple genealogic connections (table 1). Here a genealogic connection is defined as a unique genealogic pathway via which an allele identical by descent can be transmitted. For example a pair of siblings have 2 unique lines connecting them, one from the mother and another from the father whereas a pair of half sibs have only 1 line connecting them via the shared parent.

We used maximum bit-size restriction of 18, 24, and 30 bits to test PedCut and Greffa. In order to make the results from Greffa comparable with our results, we tried a number of combinations of the parameters required by Greffa and reported the results from the optimal set of parameters, which give sub-pedigrees with the maximum number of SOI within 18, 24 or 30 bits.

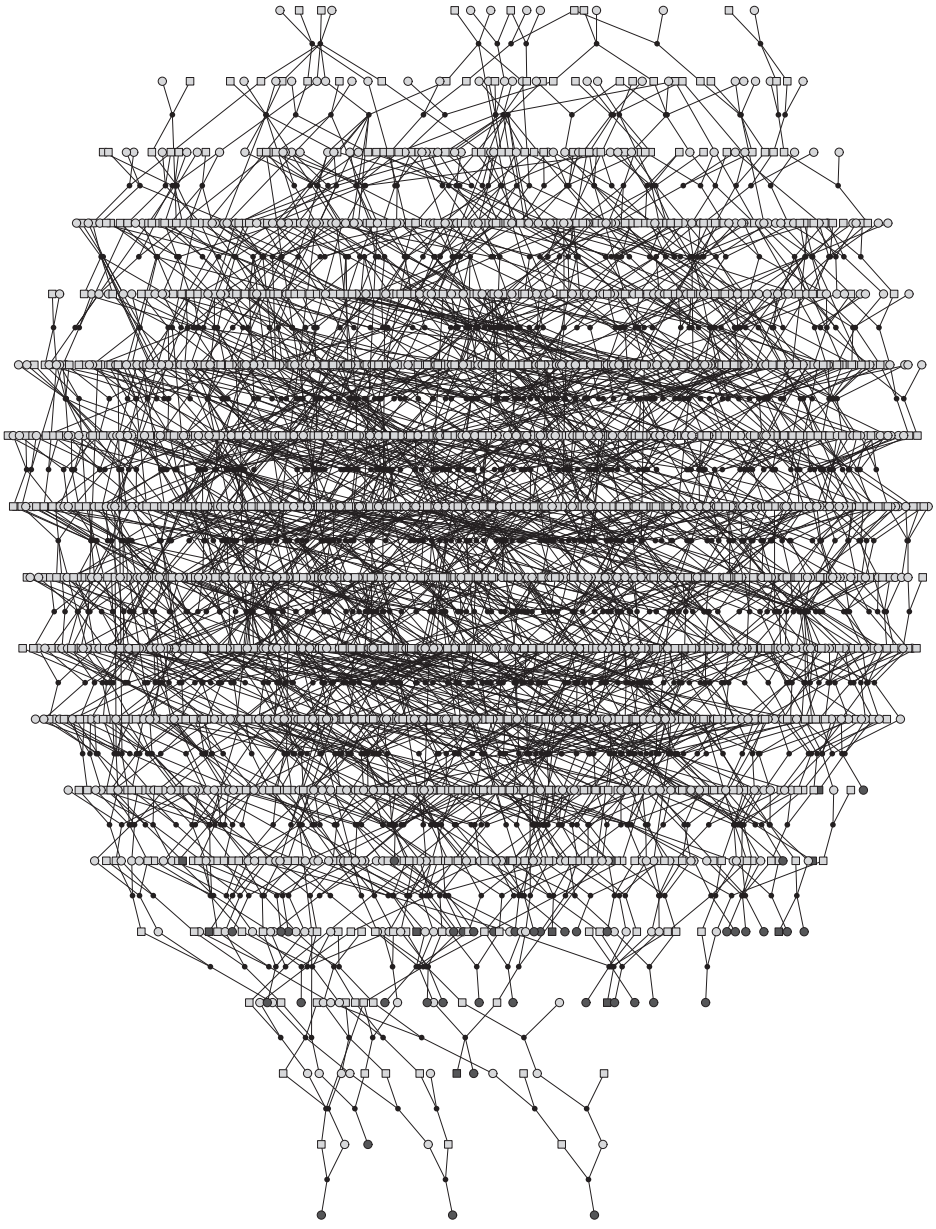


Figure 2. The initial pedigree consisting of 50 Alzheimer patients.

Table 1. Genealogic characteristics of the pedigree including 50 Alzheimer's Disease patients

Characteristic	Value ± SD (min-max)
Number of Alzheimer's Disease patients	50
Number of generations	18
Number of individuals	2510
Pedigree bit-size	2839
Average number of genealogic connections between a pair of patients	378.5 (0-2673)
Average number of meioses separating a pair of patients	18.9 ± 3.1 (2-22)
Number of SOI pairs with non-zero kinship	1214
Sum of pair-wise kinship coefficient	8.74598
Mean kinship coefficient	0.0072 ± 0.0198

Table 2. Result of cutting a pedigree of 50 AD patients using PedCut and Greffa

	Greffa			PedCut		
	18 bits	24 bits	30 bits	18 bits	24 bits	30 bits
Number of resultant sub- pedigrees	13	9	15	17	14	12
Pedigree bits (min-max)	9.4 (2-18)	9.2 (4-23)	15.1 (6-30)	12.0 (6-18)	16.1 (6-24)	23.2 (15-29)
Pedigree size (min-max)	16.2 (4-27)	15.1 (8-28)	23.5 (12-45)	20.1 (11-33)	25.2 (12-41)	37.3 (24-51)
Average number of SOI per sub-pedigree (min-max)	2.9 (2-5)	2.4 (2-4)	2.5 (2-3)	2.9 (2-6)	3.5 (2-6)	4.2 (3-7)
Average number of generations of sub-pedigrees (min-max)	4.4 (2-5)	4.2 (3-5)	5.2 (4-6)	4.6 (2-5)	5.7 (4-9)	6.1 (4-9)
Number of SOI who could not be assigned to any sub-pedigree	13	28	12	1	1	0
Number of SOI pairs with non-zero kinship	35	16	36	57	74	95
Sum of kinship derived from sub-pedigrees	1.836	0.777	1.030	2.173	2.133	2.236
Mean kinship derived from sub-pedigrees	0.052	0.049	0.029	0.038	0.029	0.024
Parameters used for the Greffa program[a]						
Minimum clique size	2	2	2			
Maximum clique size	5	6	3			
Maximum number of generations	7	7	9			
Minimum pair-wise kinship	0.01	0.001	0.0067			

[a]An optimal set of parameters was chosen for Greffa that gives sub-pedigrees with the maximum number of SOI within 18, 24, or 30 bits

Table 2 shows that PedCut can assign more patients into sub-pedigrees compared to Greffa in splitting this pedigree. Under any bit-size limit investigated, PedCut could successfully assign most SOI to sub-pedigrees whereas Greffa left a considerable number of SOI unassigned. For example, Greffa left 12 (24%) to 28 (56%) SOI unassigned, and thus completely lost for linkage analysis. On the other hand, PedCut left at maximum 1 (2%) SOI unassigned. The only person who could not be assigned to a pedigree by PedCut at bit-size of 18 and 24 is connected to a closest SOI through 10 meioses; this person was assigned when bit-size limit was set to 30.

Also, the average number of SOI per sub-pedigree from PedCut was larger (2.9, 3.5, and 4.2 for 18, 24, and 30 bits) than that from Greffa (2.9, 2.4, and 2.5). Furthermore, the maximum number of SOI from Pedcut (6, 6, and 7 for 18, 24, and 30 bits) is larger than that from Greffa (5, 4, and 3). Finally, compared to Greffa , PedCut produced more uniformly sized sub-pedigrees, as evidenced by a narrower range of bit and pedigree-sizes (Table 2).

With an Intel Pentium 4 2.4 GHz CPU, PedCut could split the pedigree in about 3 minutes. The resultant sub-pedigrees from PedCut using 24 bits as the pedigree limit are depicted in figure 3, where the 14 sub-pedigrees are ordered in the same sequence as PedCut identified them. All 5 sib-pairs were captured by the first 2 sub-pedigrees and the 2 aunt-niece pairs were captured by sub-pedigrees 3 and 4. The pedigrees 5 and 6 also contain multiple SOI who are relatively closely related to each other. The sub-pedigree 7 contains 2 SOI who are double-first cousins and the sub-pedigree 10 contains 2 SOI who are second cousins. The sub-pedigrees 8, 9, 11 and 12 contain multiple distantly related SOI. The pedigree 13 and 14 contains only 2 distantly related SOI each.

Power comparison

We compared the power to detect linkage to a rare fully penetrant variant using sub-pedigrees derived in previous section. Using these pedigrees, we simulated genetic data conditional on the phenotypes observed, using method of Boehnke[16], implemented in software package SIMLINK version 4.12. We assumed a genetically homogenous binary trait, determined by two underlying alleles (D being causal and d being normal allele). The penetrance of DD was fixed at 1.0 and penetrance of dd was fixed at 0.0. The penetrance for Dd was set at 1.0 (dominant model), 0.75, 0.5, 0.25, or 0.0 (recessive model). The frequency of D allele was set in such manner, that locus-specific population attributable fraction was 0.0001. One marker locus having 5 alleles with equal frequency of 0.2 was modeled. One thousand replicas were simulated and analyzed under each scenario concerning Dd penetrance. The distance between the trait and marker loci was set to zero. Consequent linkage analysis was done using correct model.

For 18-bits pedigrees, results of power calculation are shown in Figure 4.The sub-pedigrees derived from PedCut consistently showed higher power compared to the ones from Greffa, across all models analysed. This is most likely due to the fact that Greffa left a considerable amount of patients un-assigned to any sub-pedigree. Of interest, the sub-pedigrees derived from PedCut showed a clear trend of increase in power when the underlying genetic model was becoming more dominant. The same trend exists for the pedigrees from Greffa but in much less degree. Similar results were obtained for 24- and 30-bit pedigrees as well.

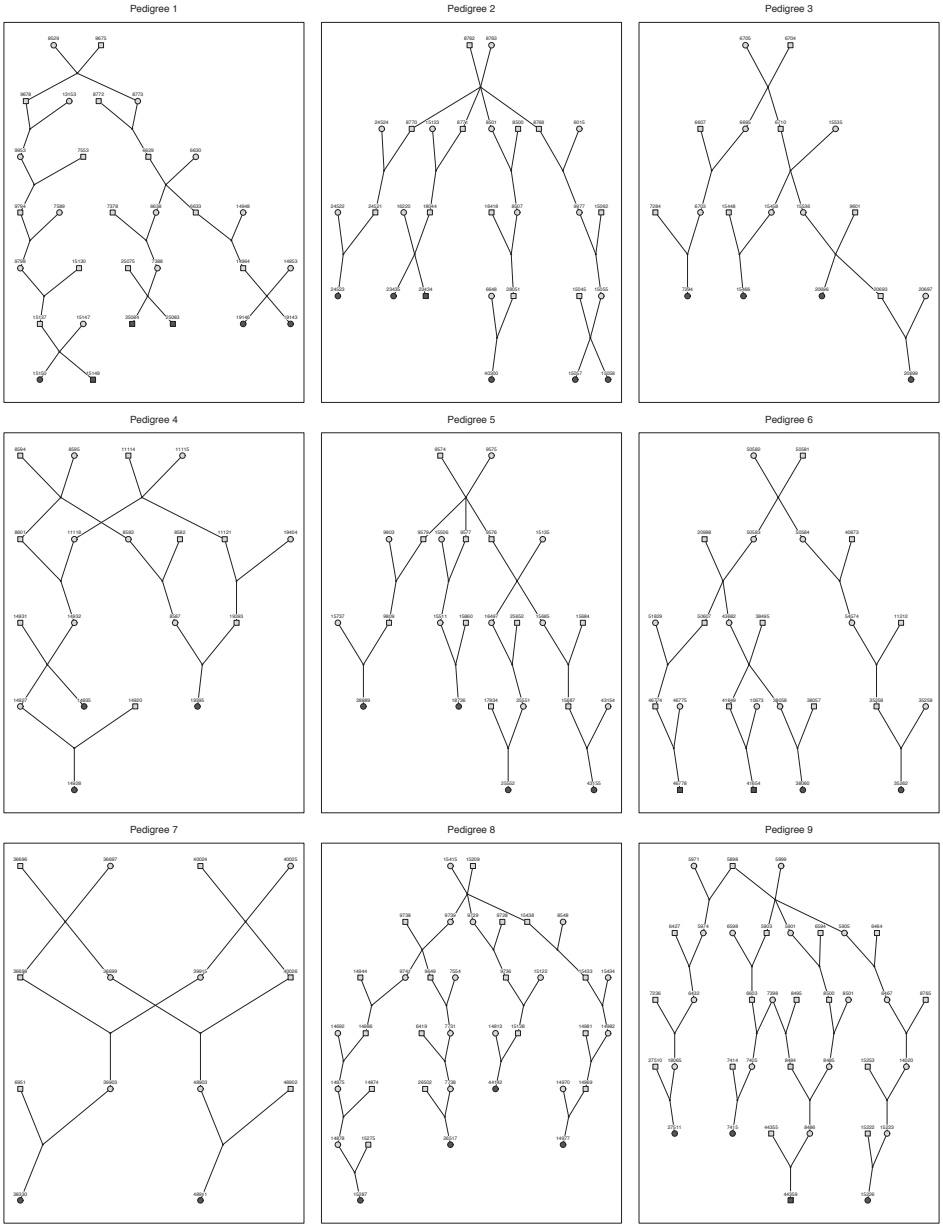


Figure 3. The resultant sub-pedigrees from PedCut using 24 bits as the pedigree size limit

DISCUSSION

In this work we have developed an algorithm that recursively groups the subjects of interest (SOI, e.g. genotyped patients) into sub-groups that fall within a certain bit-size limit and include the maximum number of SOI who share a common ancestor. With the algorithm we

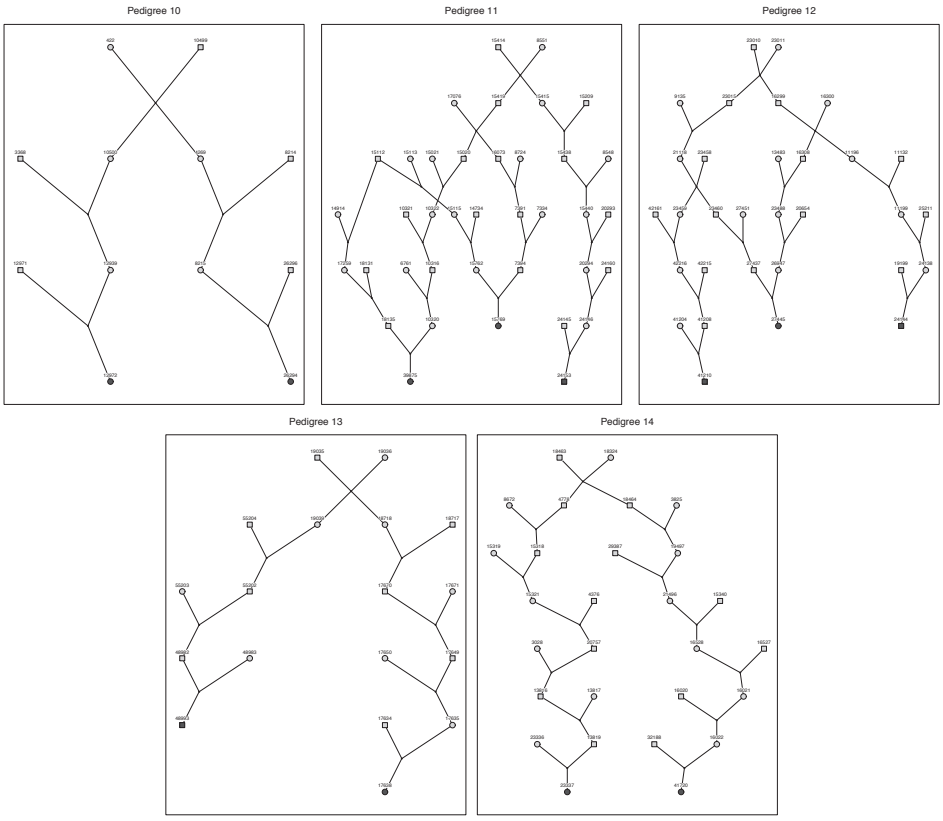


Figure 3. (continued)

exploit the basic rationale of linkage analysis, i.e. that affected relatives are likely to share the disease-causal allele identical by descent from a common ancestor. Fast grouping is achieved by prioritizing relatives using kinship. Our algorithm guarantees that the derived sub-pedigrees can be directly and efficiently analyzed by software implementing the Lander-Green-Kruglyak algorithm. Further, it is fully automated.

Any method for sub-pedigree identification involves breaking relationships between SOI. Breaking relationships may increase false positive linkage signals[17,18] or reduce the power of linkage analysis[12]. Bias is especially pronounced when close relationships are broken. In our program we include a default option to preserve close relationships between SOI, keeping all second cousin or closer relationships between SOI. In general, type 1 error for split pedigree should be worked out by gene-dropping using complete pedigree, and re-analysis using split pedigrees[2]. Exhaustive search algorithm guarantees (within the search space) that the selected sub-pedigree has the maximal number of SOI in respect to the bit-size restriction, if no equally eligible sub-pedigrees are available. The latter situation, though possible, is in our experience unlikely in deep pedigrees coming from genetically isolated populations, where mating is mostly random, and multiples genealogic connections are observed between SOI.

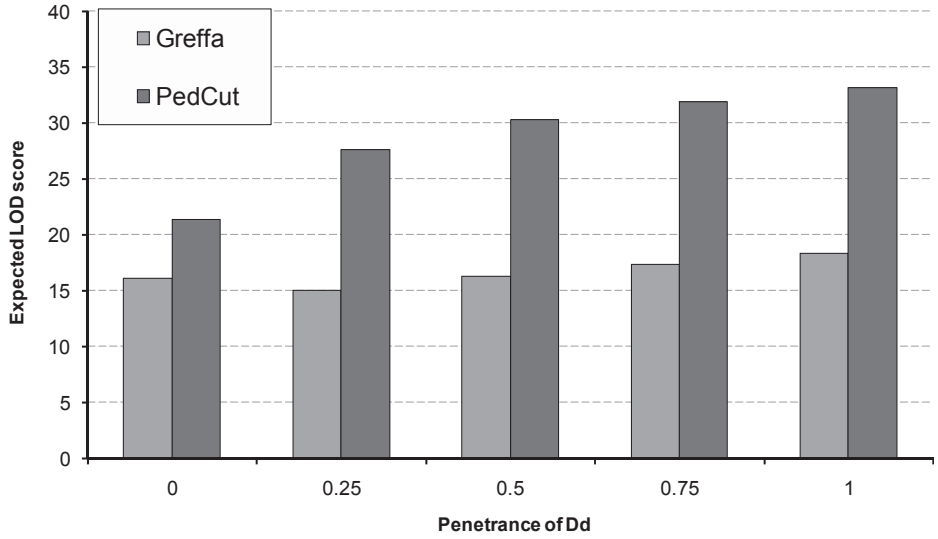


Figure 4. Expected LOD score for pedigrees derived using PedCut and Greffa with a bit-size limit of 18

We have previously applied PedCut to split a pedigree including 4,645 people from whose 112 were Alzheimer's Disease patients[2]. In that study, we could assign 103 patients to 35 pedigrees using bit-size limit of 35 in about 11 minutes. Empirical threshold for 5% genome-wide significance, established using simulations in complete pedigree, was estimated to be 3.64. Regions of significant linkage were genotyped using dense single-nucleotide polymorphisms (SNPs) marker map in an independent cohort. Significant associations were observed between some of these regions and cognitive function. This proves applicability of our approach in real studies.

In summary, we developed a heuristic algorithm that is suitable to split large and complex pedigrees coming from genetically isolated populations. Our algorithm and associated software (PedCut, http://mga.bionet.nsc.ru/soft/index.html) can facilitate fast genome-wide linkage search for rare mutations.

REFERENCES

1. Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, Helgason A, Stefansson H, Emilsson V, Helgadottir A, *et al.* (2006) Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat Genet 38:320-323
2. Gonzalez-Zuloeta Ladd AM, Liu F, Houben MP, Arias Vasquez A, Siemes C, Janssens AC, Coebergh JW, Hofman A, Janssen JA, Stricker BH, *et al.* (2007) IGF-1 CA repeat variant and breast cancer risk in postmenopausal women. Eur J Cancer 43:1718-1722
3. Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347-1363
4. Lander ES, Botstein D (1987) Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. Science 236:1567-1570
5. 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
6. Gudbjartsson DF, Thorvaldsson T, Kong A, Gunnarsson G, Ingolfsdottir A (2005) Allegro version 2. Nat Genet 37:1015-1016
7. Sobel E, Lange K (1996) Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. Am J Hum Genet 58:1323-1337
8. Heath SC (1997) Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. Am J Hum Genet 61:748-760
9. Sung YJ, Thompson EA, Wijsman EM (2007) MCMC-based linkage analysis for complex traits on general pedigrees: multipoint analysis with a two-locus model and a polygenic component. Genet Epidemiol 31:103-114
10. Service S, Molina J, Deyoung J, Jawaheer D, Aldana I, Vu T, Bejarano J, Fournier E, Ramirez M, Mathews CA, *et al.* (2006) Results of a SNP genome screen in a large Costa Rican pedigree segregating for severe bipolar disorder. Am J Med Genet B Neuropsychiatr Genet 141:367-373
11. Sieh W, Basu S, Fu AQ, Rothstein JH, Scheet PA, Stewart WC, Sung YJ, Thompson EA, Wijsman EM (2005) Comparison of marker types and map assumptions using Markov chain Monte Carlo-based linkage analysis of COGA data. BMC Genet 6 Suppl 1:S11
12. Dyer TD, Blangero J, Williams JT, Goring HH, Mahaney MC (2001) The effect of pedigree complexity on quantitative trait linkage analysis. Genet Epidemiol 21 Suppl 1:S236-243
13. Ciullo M, Bellenguez C, Colonna V, Nutile T, Calabria A, Pacente R, Iovino G, Trimarco B, Bourgain C, Persico MG (2006) New susceptibility locus for hypertension on chromosome 8q by efficient pedigree-breaking in an Italian isolate. Hum Mol Genet 15:1735-1743
14. Boichard D (2002) PEDIG: a FORTRAN package for pedigree analysis studied for large populations. Proceeding of the 7th World Congress of Genet Appl Livest Prod, Montpellier, France 28-13
15. Agarwala R, Biesecker LG, Hopkins KA, Francomano CA, Schaffer AA (1998) Software for constructing and verifying pedigrees within large genealogies and an application to the Old Order Amish of Lancaster County. Genome Res 8:211-221
16. Boehnke M (1986) Estimating the power of a proposed linkage study: a practical computer simulation approach. Am J Hum Genet 39:513-527
17. Liu F, Elefante S, van Duijn CM, Aulchenko YS (2006) Ignoring Distant Genealogic Loops Leads to False-positives in Homozygosity Mapping. Ann Hum Genet 70:965-970
18. Miano MG, Jacobson SG, Carothers A, Hanson I, Teague P, Lovell J, Cideciyan AV, Haider N, Stone EM, Sheffield VC, *et al.* (2000) Pitfalls in homozygosity mapping. Am J Hum Genet 67:1348-1351

Chapter 4

A Genomewide Screen for Late-onset Alzheimer Disease in a Genetically Isolated Dutch Population

ABSTRACT

Alzheimer's disease (AD) is the most common cause of dementia. We conducted a genome screen in 103 late onset AD patients that were ascertained as part of the Genetic Research in Isolated Populations (GRIP) program that is embedded in a recently isolated population from the Southwest of the Netherlands. All patients and their 170 closely related relatives were genotyped using 402 microsatellite markers. Extensive genealogy was collected, resulting in an extremely large and complex pedigree including 4,645 members. The pedigree was split into 35 sub-pedigrees in order to reduce the computational burden of linkage analysis. Simulations aiming to evaluate the effect of pedigree splitting on false positive probabilities showed that a LOD score of 3.64 corresponds to 5% genome-wide type-I error. Multipoint analysis revealed four significant and one suggestive linkage peaks. The strongest evidence of linkage was found for chromosome 1q21 (HLOD = 5.20, at marker D1S498). About 30 cM upstream of this locus, we found another peak at 1q25 (HLOD = 4.0 at D1S218). These two loci are in a previously established linkage region. We also confirmed the AD locus at 10q22-24 (HLOD = 4.15, at marker D10S185). There was significant evidence of linkage of AD to chromosomes 3q22-24 (HLOD = 4.44 at marker D3S1569). For chromosome 11q24-25 there was suggestive evidence of linkage (HLOD = 3.29, at marker D11S1320). We next tested for association between cognitive function and 4173 single nucleotide polymorphisms (SNPs) in the linked regions in an independent sample consisting of 197 individuals from the GRIP region. After adjusting for multiple testing we were able to detect significant associations for cognitive function in four out of five AD-linked regions, including the new region on chromosome 3q22-24 and regions1q25, 10q22-24, and 11q25. Using cognitive function as an endophenotype of AD, our study indicates the *RGSL2*, *RALGPS2*, and *C1orf49* genes at 1q25. Our analysis on chromosome 10q22-24 points to *HTR7 [MIM 182137]*, *MPHOSPH1*, and *CYP2C* cluster. This is the first genome-wide screen that showed significant linkage to chromosome 3q23 markers. For this region our analysis identified *NMNAT3 [MIM 608702]* and *CLSTN2* genes. Our findings confirm linkage to chromosome 11q25. We could not confirm *SORL1 [MIM 602005]*, instead, our analysis points to *OPCML [MIM 600632]* and *HNT [MIM 607938]* genes.

INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative disorder that accounts for the vast majority of dementia. The population prevalence of the disease rises steeply with age from below 2% at 65 years to above 35% after the age of 90 years[1,2]. Family history is an important risk factor for AD and in a large number of families the disease segregates as an autosomal dominant trait. The heritability for AD was recently estimated to be 79%[3]. Several dominant mutations have been identified including mutations in presenilin 1 (*PSEN1 [MIM 104311]*)[4], presenilin 2 (*PSEN2 [MIM 600759]*)[4,5], and amyloid precursor protein (*APP [MIM 104760]*) genes[6]. A common polymorphism (ε4) in the gene encoding apolipoprotein E (*APOE [MIM 107741]*) increases susceptibility to both early and late onset AD[7,8]. These four genes together explain less than a quarter of the disease prevalence, indicating additional genetic risk factors remain to be identified[9,10]. In addition to *APOE*, various candidate genes were reported to be associated with late onset AD. In most cases findings have not been consistently replicated[11,12]. A large meta-analysis of all genes studied so far pinpointed thirteen potential AD susceptibility genes: angiotensin I converting enzyme (*ACE [MIM 106180]*), cholinergic receptor, nicotinic, beta 2 (*CHRNB2 [MIM 118507]*), cystatin C (*CST3 [MIM 604312]*), estrogen receptor 1 (*ESR1 [MIM 133430]*), glyceraldehyde-3-phosphate dehydrogenase, spermatogenic (*GAPDHS [MIM 609169]*), insulin-degrading enzyme (*IDE [MIM 146680]*), 5,10-methylenetetrahydrofolate reductase (*MTHFR [MIM 607093]*), nicastrin (*NCSTN [MIM 605254]*), prion protein (*PRNP [MIM 176640]*), *PSEN1*, transferrin (*TF [MIM 190000]*), transcription factor A, mitochondrial (*TFAM [MIM 600438]*) and tumor necrosis factor (*TNF [MIM 191160]*)[13]. Furthermore, genome screens targeting AD loci have been conducted. As reviewed online by the Alzheimer Research Forum (http://www.alzgene.org), the replicated regions from previous genome screens include: 1p36, 1q21-31, 2p23-24, 4q35, 5p13-15, 6p21, 6q15-16, 6q25-27, 9p21-22, 10q21-22, 10q25, 12p11-12, 19q13, 21q21-22, and Xp11-21[8,14-28]. Several genes have been suggested to explain the linkage to chromosome 9, 10, 12 and 19, but so far also these genes remain to be confirmed. Finally, there is evidence for linkage to chromosome 11[21], which was explained recently by *SORL1 [MIM 602005]*[29].

Each of the established loci for AD (*APP*, *PSEN1*, *PSEN2* and *APOE*) has been initially localized by linkage analyses. However, pedigrees suitable to localize genes have become scarce particularly for late onset forms of AD. Genetically isolated populations provide opportunities for linkage analysis. Using genealogical records, extended pedigrees can be constructed. Furthermore, the complexity of disease may be reduced in terms of number of genes involved, in particular for rare Mendelian forms[30,31]. Linkage analysis of complex traits has been used successfully in Iceland for complex diseases such as type 2 diabetes and stroke[32,33], while for AD genome screens have been conducted successfully in Caribbean Hispanics[34]. We have followed this approach in a genetically isolated community from the Southwest of the Netherlands as part of the Genetic Research in Isolated Populations (GRIP) program[35]. A total

of 103 late onset AD patients were ascertained and connected into a large pedigree based on genealogical records. In this paper, we present a genome-wide screen in these families. The linkage analysis was followed by an association study of cognitive function in a series of 197 unrelated and non-demented people from the GRIP region who were extensively characterized by a cognitive battery. In order to further investigate the evidence for linkage, the regions identified in the linkage study were characterized with a dense panel of single nucleotide polymorphisms (SNPs). Decline in cognitive function, particularly mild cognitive impairment, is an early predictor of AD[36-38] and the heritability of cognitive function is as high as 56% suggesting cognition is a valuable endophenotype[39-41]. Further, memory function was found to be an endophenotype for families multiply affected with AD[42].

METHODS

Population & Genealogy

This study was performed within the framework of the previously described Genetic Research in Isolated Populations (GRIP) program[9,31,43]. The Medical Ethics Committee of the Erasmus MC approved study protocol. The GRIP population is a genetically isolated community in the southwest of The Netherlands. Less than 400 individuals were present in the region in the middle of the 18th century. Considerable population growth occurred in the period 1850-1900, as was the case in many European populations. An estimated 20,000 descendants of the population are now scattered over eight adjacent communities. There was minimal immigration. The genealogical database currently contains information on 107,091 people spanning 23 generations. Residents in the GRIP area are generally related via multiple lines of descent and inbred via multiple consanguineous loops[43].

Patients with AD were traced through general practitioners, neurologists and nursing-home physicians. Data relevant for the diagnosis of AD were collected by a research physician and the diagnosis of AD was verified by two independent neurologists according to the NINCDS ADRDA criteria[44]. Data on the presence of AD, parkinsonism, essential tremor and dementia were collected for first, second and third-degree relatives by means of a family-history questionnaire. First-degree relatives also underwent a brief neurological examination. All patients and their relatives who were invited to participate in this study provided informed consent. A total of 112 probable late onset AD patients (age of onset ≥ 65 years, mean age of onset = 75 ± 5.3 years) and 170 unaffected first-degree relatives (mean age = 63.5 ± 13.1 years, range 40 to 102 years) were ascertained.

Tracing the genealogy of the 112 probable late onset AD patients, we were able to connect 103 patients to a single pedigree containing 4,645 individuals in 18 generations as depicted in Figure 1. The other 9 patients were singletons and, therefore, were not included in the linkage analysis. This large pedigree showed multiple, distant lines of descent and consanguineous

loops (Table 1). The average kinship coefficient among patients was 0.0018. This value is in between a 3rd cousin once removed and a 4th cousin. Utilizing such a pedigree in linkage

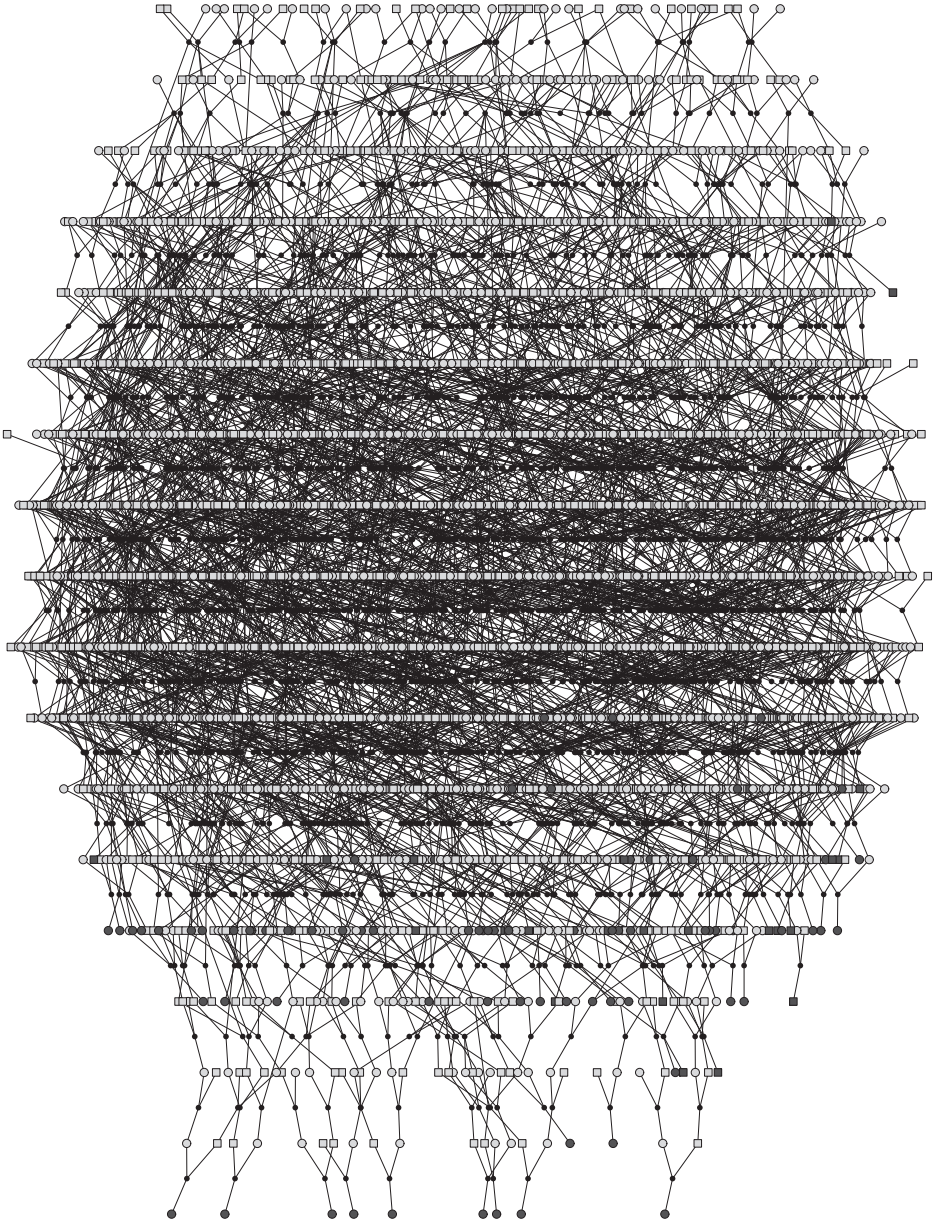


Figure 1: The entire pedigree contains 103 late onset AD patients and 4,645 relatives. Men are represented with squares and women with circles. Black dots represent marriage nodes. Affected individuals are represented with the dark blue color. Unknown affection status is represented with yellow. For simplicity, unaffected relatives of the patients are not shown.

Table 1. Genealogic characteristics of 103 late onset AD patients and their relatives

	Value ± SD	(Range or %)
The complete genealogy		
Family size	4645	
Number of generations	18	
Average number of consanguineous loops per patient	71.7	(0-677)
Average number of meioses in a consanguineous loop	9.9 ± 1.2	(0-29)
Mean inbreeding coefficient × 100	0.39 ± 0.73	(0-3.2)
Average number of lines of descent between a pair of patients	141.7	(0-2673)
Average number of meioses separating a pair of patients	17.1 ± 1.6	(0-34)
Mean kinship coefficient ×100	0.18 ± 1.06	(0-26.4)
After clustering patients into sub-pedigrees		
Number of sub-pedigrees	35	
Number of founders	564	0.46
Number of females	630	0.51
Mean pedigree size	29.6	(18-75)
Mean number of generations	7.5	(6-10)
Mean number of genotyped individuals per pedigree	7.8	(2-14)
Mean number of patients per pedigree	2.9	(2-6)

analysis is computationally impossible. A common approach to reduce the computational complexity is to split the large pedigree into smaller and computable units. For this purpose we used a kinship clustering method that is similar to the maximal cliques partitioning method proposed by Falchi et al[45], adding a restriction that the resulting sub-pedigrees should have no more than 35 bits, where the bit-size is computed as (2 × number of founders – number of non-founders). Our software for splitting large pedigrees, PedCut, is freely available at http://mga.bionet.nsc.ru/soft.

We further studied a series of 197 individuals who were not closely related (≥ 5 generations) and not related to the AD patients. The average age of these people was 31.2 ± 6.4 years, with 51% being female. These individuals were characterized by an extensive cognitive battery[46]. In brief, the selection of tests included the 15-word test, the colour word card of the Stroop Colour Word test (Stroop), the part B of the Trail making test (TMTB) and the verbal fluency test. These tests were selected to target early cognitive problems related to AD. From the 15-word test we derived 3 scores for further analysis, i.e. learning (or working memory), delayed recall and recognition. The verbal fluency consists of two sub-domains: semantic fluency and phonological fluency. The performance of each individual on each test was scored quantitatively. Power calculation showed that this sample has 80% power to detect a SNP explaining 4% of phenotypic variance with alpha of 0.05.

Genotyping

For all patients and their 170 first-degree relatives, DNA was extracted from peripheral leucocytes following a standard protocol[47]. Mutations in the *APP*, *PSEN1* and *PSEN2* genes were

previously excluded[35]. The *APOE* genotype was determined in all DNA samples using TaqMan allelic discrimination technology on an ABI Prism 7900HT Sequence Detection System with SDS V 2.1 (Applied Biosystems, Foster City, CA). Patients and their first-degree relatives underwent a full genome-screen in two sequential experiments. Both screens were conducted using the same set of micro-satellite markers, evenly spaced by approximately 10 cM (ABI Prism Linkage Mapping Set MD-10 Versions 2 and 2.5, Applied Biosystems, Foster City, CA, USA). Polymerase chain reactions (PCR) were performed according to the manufacturer's specified conditions. PCR products were separately pooled and analyzed on ABI377 and ABI3100 automated sequencers (Applied Biosystems). Because the genome scan had been performed with different sequencing devices, the genetic data had to be merged. The genotypic data was pooled using Pool_STR-1.1, based on the allele lengths and allele frequencies observed in each group[48]. Two independent technicians read the results from the sequencers and a third reader resolved the discordant results. Only the markers with a discordance proportion less than five percent were selected for further analysis (N=402). Genotyping errors leading to Mendelian inconsistencies were detected using PedCheck[49]. Unlikely double recombination events were detected using Merlin[20]. Definitive genotyping errors and unlikely genotypes were rechecked using the data from the laboratory. Regions linked to late onset AD were later fine typed by placing 45 additional microsatellite markers in between those from the initial set at a distance of one to five cM apart.

SNPs in the linkage regions were selected from the 250K Nsp array of the GeneChip® Human Mapping 500K Array Set (Affymetrix). Genomic DNA was extracted from whole blood samples drawn at the baseline examination, utilizing the salting out method[47]. The 250K Nsp array from Affymetrix was utilized to determine genotypes. The chips were run and analyzed according to the manufacturer's protocols. A total of 4173 SNPs were selected for association test based on the following criteria: (1) position within the regions which show significant or suggestive evidence for linkage after fine mapping, (2) minor allele frequency ≥ 2.5%, (3) P-value for Hardy-Weinberg equilibrium test ≥ 0.01, and (4) call rate ≥ 95%.

Statistical analysis

In linkage analysis, we assumed a dominant model of inheritance with age dependent penetrance. Seven liability classes were defined based on age (years): <65, 65-69, 70-74, 75-79, 80-84, 85-90, and >90. For each age group *j*, age dependent population prevalence, P_j, was obtained from the Rotterdam Study[1]. The disease gene penetrance, f_j, of the j^{th} age group can be estimated:

$$f_j = \frac{PAF \times P_j}{q^2 + 2q(1-q)}$$

where PAF stands for the population-attributable fraction, the proportion of the population prevalence that can be explained by the studied gene (10% assumed), and *q* is the disease allele frequency (1% assumed). The estimated penetrance for each defined age group is

Table 2. Age dependent liability classes and penetrances

Liability Class	Age (years)	Population prevalence	Penetrances	Nbr patients	Nbr unaffected
1	< 65	< 0.02	0	0	129
2	65-69	0.02	0.09	4	6
3	70-74	0.05	0.23	22	11
4	75-79	0.09	0.46	32	14
5	80-84	0.23	0.99	30	8
6	85-89	0.35	0.99	24	1
7	>=90	>0.35	0.99	0	1

shown in table 2. Marker allele frequencies were estimated based on 144 chromosomes from unaffected elderly GRIP population members. For small pedigrees (bits ≤ 18), we used the exact calculation of multi-locus likelihood, Lander-Green algorithm implemented in GENEHUNTER 2.0[50]. For larger pedigrees, we used Markov chain Monte Carlo estimation methods implemented in SIMWALK 2.91. Overall LOD scores and Heterogeneity LOD (HLOD) scores were computed by combining results per family using standard formulas below.

HLOD = $\log_{10}$ (max *LR*), where maxLR is maximized with respect to *α*, the proportion of the linked families, yielding Maximum Likelihood Estimate $\hat{\alpha}$,

$$\max LR = \prod_{i=1}^{n} (\hat{\alpha} LR_i + 1 - \hat{\alpha}) \quad .$$

Haplotypes were reconstructed based on the genotypes of patients, spouses of patients and their offspring using MERLIN package. Haplotypes are shown only for the linked families with the highest LOD scores. These families are further expanded in order to depict the haplotype sharing of other patients who are relatively closely related to the patients in the families with high LOD scores, who were assigned to different families in the pedigree splitting procedure.

Breaking pedigrees may increase the possibility of spurious linkage findings[51]. Therefore, we estimated the threshold for statistical significance using simulations. To evaluate genome-wide type-I error, we simulated our genome scan 100 times. We used the complete pedigree including all 4,645 members for marker simulation. Unlinked markers were dropped in the complete pedigree. Number of markers, intermarker distances and marker allele frequencies were simulated according to the typed marker set. We performed linkage analysis using the splitted sub-pedigrees. Disease allele frequency, liability classes, genetic model and penetrances were the same as we used later in the actual linkage analysis. Genotypes of untyped individuals were set to missing. For each genome-screen the highest HLOD score was stored. Cumulative density function of the obtained 100 maximum HLOD scores approximates the distribution of the genome-wide type-I error rates. Our simulations showed that a HLOD score of 3.64 corresponds to a genome-wide type-I error rate of 5% and HLOD of 2.11 corresponds to a genome-wide type-I error of 50% (Table 3).

Table 3. LOD scores and corresponding genome-wide type-I error rates based on 100 simulations

LOD	HLOD	Type-I error
4.08	4.08	1%
3.64	3.64	5%
3.24	3.24	10%
2.11	2.11	50%

We used linear regression to test for association between a single SNP with a single cognitive trait. According to Affymetrix annotation, SNP genotypes were coded as 0=AA, 1=AB and 2=BB where A corresponds the allele in lower alphabetical order and B for the other allele. Thus, in case a C→T change where T is the minor allele C is coded as the A allele and T as the B allele, whereas a T→C change where C is the minor allele, A also denotes the C allele and B denotes the T allele. We adjusted for age, sex, intelligence, and highest education in the model. Considering that a causal SNP (or a SNP in linkage disequilibrium with the causal SNP) is likely to be associated with multiple cognitive domains, we used the Fisher product method for combining the findings of all cognitive tests[52]. Because the SNPs are in linkage disequilibrium and cognitive traits are also correlated, we used a permutation method to evaluate significance level for each SNP empirically (500,000 replicates). To break the associations between the markers and traits while keeping the correlations between traits and the LD pattern between markers, we permuted the vectors of individuals' traits (scores of cognitive tests and covariates) between individuals, without replacement. For each permutation, we tested for association between SNPs in each region and cognitive traits, and derived corresponding Fisher products. The cumulative density function of all Fisher products for each region empirically approximates the regional-wide type-I error rate. Therefore, the empirical P-value for each SNP can be defined as the probability of observing an equal or smaller Fisher product by chance regional-widely.

RESULTS

For linkage analysis, we clustered all patients and 170 first-degree relatives into 35 sub-pedigrees (Figure 2). During pedigree splitting, distant ancestors who have no phenotypic and genotypic information and do not contribute to linkage information were discarded. The resultant sub-pedigrees contained a total of 1,227 individuals. The characteristics of the sub-pedigrees are shown in table 1.

Multipoint LOD and HLOD score plots for the initial scan are shown in the figure 3. A total of eight regions showed suggestive linkage (LOD or HLOD>2.11), of which the chromosome 1 region exceeded the threshold of 3.64 (LOD=4.1 at marker D1S484). These 8 regions were fine typed with 45 additional markers and include: chromosome 1 (14 additional markers), chromosome 3 (10 additional markers), chromosome 5 (2 additional markers), chromosome

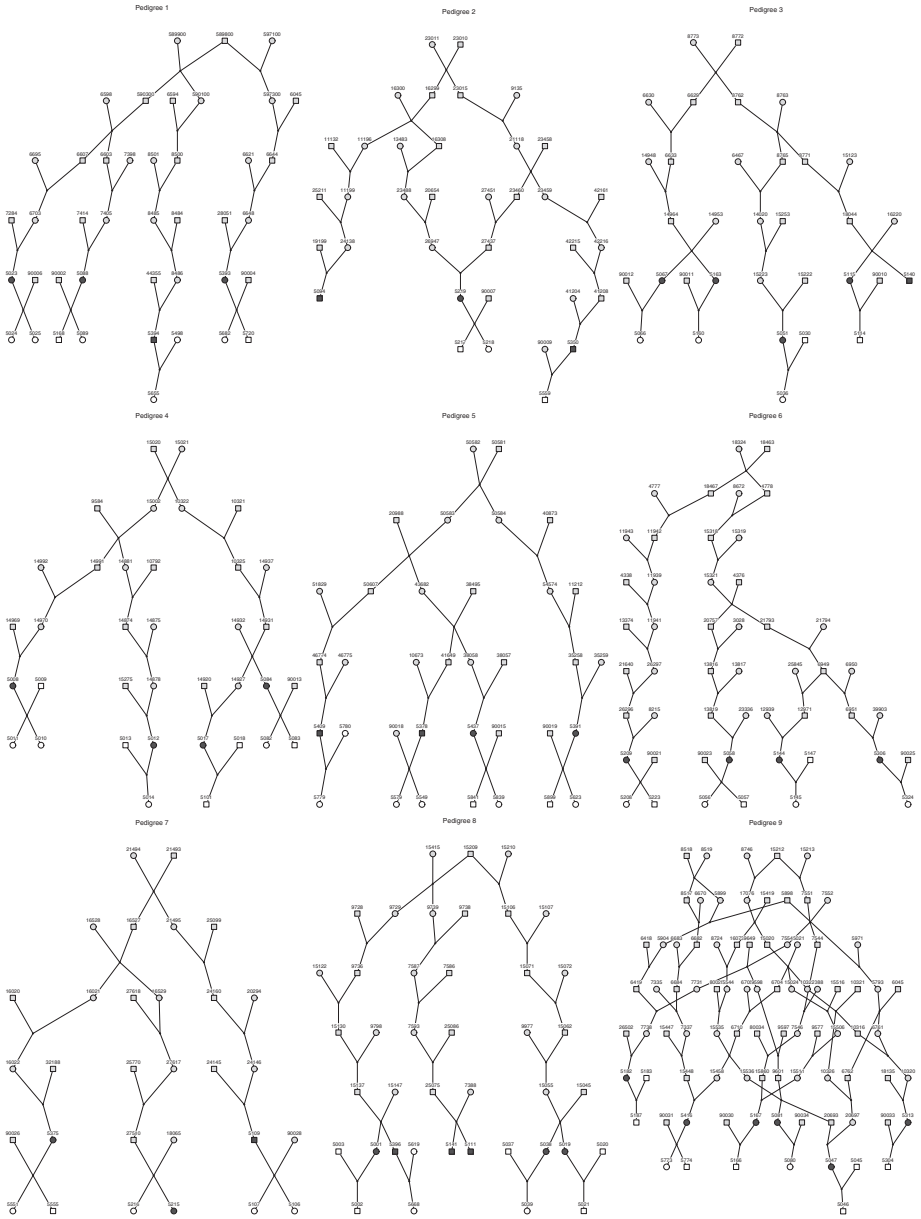


Figure 2. The 35 sub-pedigrees obtained by applying a kinship-partitioning algorithm to the entire pedigree. Affected individuals are shown with blue squares and circles, unaffected first-degree relatives of the patients were added to the sub-pedigrees, and are represented with white squares and circles and ancestors from the patients and relatives are shown with yellow squares and circles.

6 (5 additional; markers), chromosome 7 (3 additional markers), chromosome 10 (6 additional markers), chromosome 11 (2 additional markers), and chromosome 18 (3 additional markers).

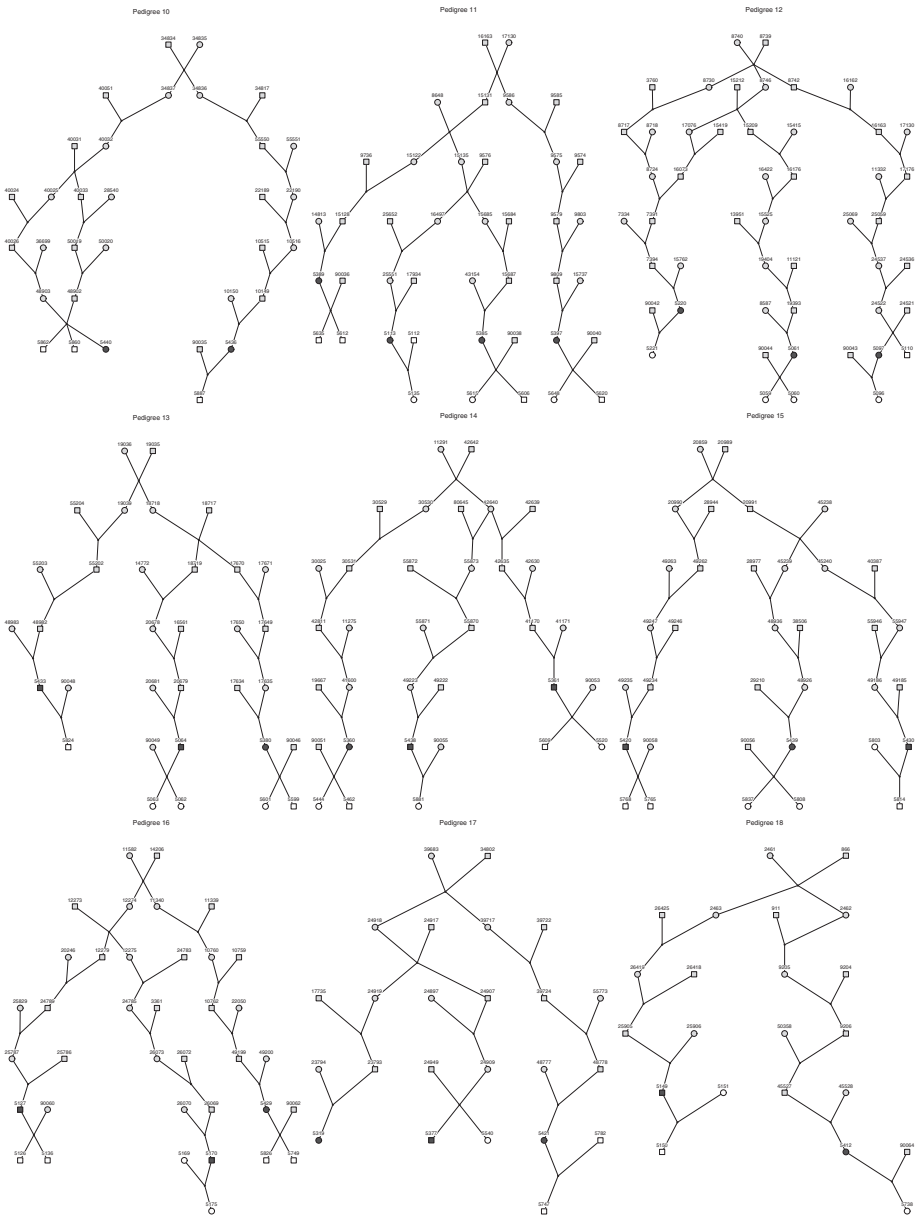


Figure 2. (continued)

Table 4 shows the regions for which the evidence for linkage remained significant or suggestive (LOD>2.11) after fine mapping. AD remained linked to three known regions, two of them on chromosomes 1 (figure 4A) and one on chromosome 10 (figure 4C). The maximum HLOD at 1q21 was 5.2 at D1S498. This is the highest peak over the genome. The maximum HLOD at 1q25 was 4.0 at D1S218 and 4.2 at D10S185. REFERENCE for the previously identified

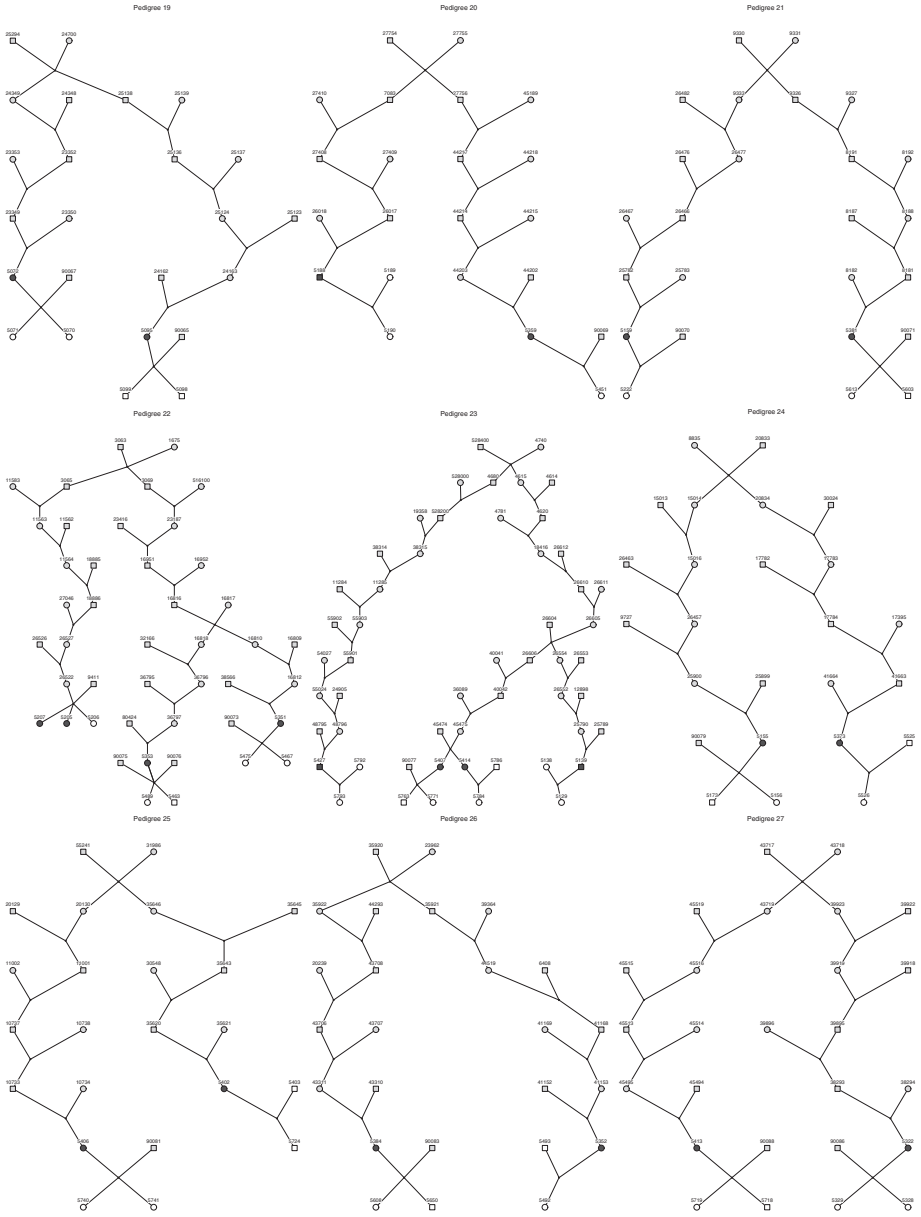


Figure 2. (continued)

regions can be found in table 4. In addition to the known regions, we found genome-wide significant evidence for linkage of AD to a region on chromosome 3 that spanned 18 cM from marker D3S3514 to D3S3626, and reached a maximum HLOD of 4.4 at marker D3S1569 (figure 4B). This is the second highest peak over the genome. In table 4 we also included chromosome 11, in which recently a new gene (*SORL1*) responsible for AD was reported[29].

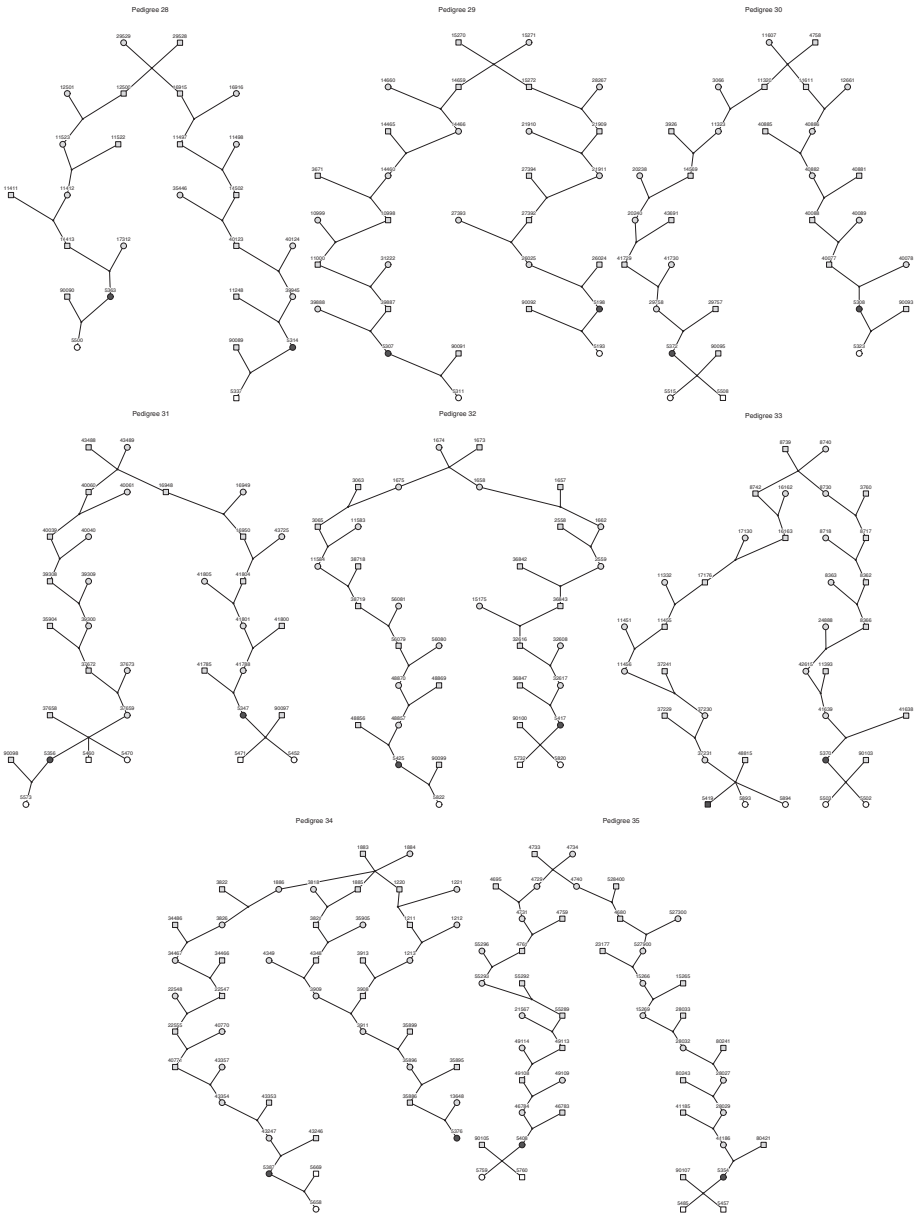


Figure 2. (continued)

There is suggestive evidence for linkage of AD to chromosome 11 (HLOD = 3.3 at D11S1320, figure 4D), which overlaps with a region reported earlier. On chromosome 11 the HLOD at the position of *SORL1* gene (118cM) is 1.1.

Haplotype analysis showed that the two linkage peaks on chromosome 1q21 and 1q25 are explained by different haplotypes segregating in different families. On chromosome 1q21,

Table 4. Regions with genome-wide empirically significant (in bold) or suggestive linkage after fine mapping

Chromosome	Marker	Position (cM)	LOD	HLOD	α	Previously identified regions[1]
1A	D1S498	164	**5.1**	**5.2**	0.9	A[24], D[25], F[53], G[21]
	D1S305	167	**4.5**	**4.5**	1.0	
1B	D1S218	201	2.6	**4.0**	0.6	A[24], D[25], F[53], G[21]
	D1S366	208	2.7	3.5	0.6	
3	D3S1549	151	2.8	3.6	0.6	B[58]
	D3S1569	158	**4.3**	**4.4**	0.8	
10	D10S1686	105	**3.7**	**3.7**	1.0	C[17], E[18], F[53], G[21], H[26], I[23]
	D10S185	116	**4.2**	**4.2**	1.0	
11*	D11S4151	127	0.3	2.8	0.4	G[21]
	D11S4131	138	1.3	3.1	0.5	
	D11S1320	142	1.6	3.3	0.6	
	D11S968	148	0.3	2.0	0.5	

[1]Overlaps with regions reported with suggestive linkage or significant association in previous genome screens, including: A (Zubenko et al. 1998), B (Poduslo et al. 1999), C (Curtis et al. 2001), D (Hiltunen et al. 2001), E (Olson et al. 2002), F (Myers et al. 2002), G (Blacker et al. 2003), H (Farrer et al. 2003), and I (Holmans et al. 2005). Note that B screened for only 2 chromosomes.
*Chromosome included to confirm a recent report of the *SORL1* gene (Rogaeva et al. 2007).

we identified a 15 cM region shared by 4 patients in family 1 and 6 other closely related patients, who were assigned to different pedigrees for computational reasons in the process of pedigree splitting (figure 5A). The 21 cM haplotype of 1q25 segregates in family 3 (4 patients) and is shared by four other closely related patients who were assigned to different sub-pedigrees (figure 5B). Six patients from family 9, and 6 closely related patients carry the haplotype of chromosome 3q23 (18cM) as shown in figure 5C. The linkage of AD to the region on chromosome 10 was based on moderate contributions from multiple families with different haplotypes. There is not a single haplotype segregating in this region (data not shown). For chromosome 11q24, which showed suggestive linkage, we observed a single haplotype (3.4 cM) shared by 4 patients from family 4, and 2 additional closely related patients (figure 5D).

Next we tested for association between cognitive function and a set of 4173 SNPs within regions 1q21, 1q25, 3q23, 10q22-24 and 11q25 using an independent sample consisting of 197 individuals from the GRIP population (table 5). All of the linked regions except 1q21 contain at least one SNP showing significant association using an empirical p-value of 0.05 (table 6). Statistically the most significant SNP is rs7071717 at 10q23, both for the nominal p-value in single test (P=0.000005 for Stroop test) and for the empirical Fisher product (P=0.002), which combines the results of cognitive tests and adjust for multiple testing. This SNP together with rs17129662 and rs11185978 is in a range of 80 kb and shows evidence of association with the Stroop test, TMTB, semantic (except rs17129662) and phonological fluency, all of which are sub domains of executive function. These 3 SNPs are 2 to 80 kb downstream of the *MPHOSPH1* gene coding M phase phosphoprotein 1and about 760kb upstream of the

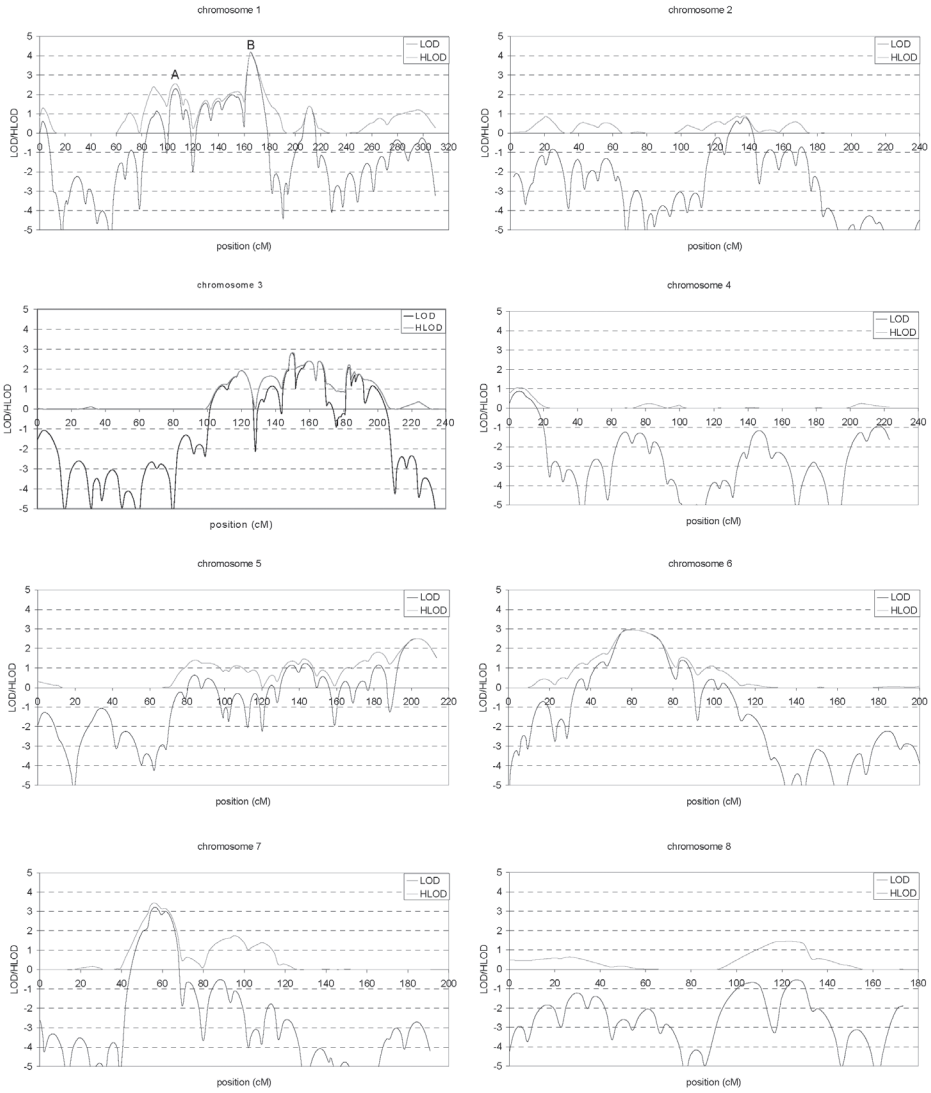


Figure 3. Multipoint LOD and HLOD scores for each autosome in the genome screen of late onset Alzheimer's disease. Marker locations are given in Kosambi centi-Morgans.

5-hydroxytryptamine receptor 7 (*HTR7 [MIM 182137]*) gene. Another SNP rs4110517 at 10q23 showed association with semantic (P=0.00003) and phonological (P=0.04) fluency (empirical P=0.02). This SNP is 37.6 kb downstream to the *CYP2C19 [MIM 124020]* gene and 48.1kb upstream to the *CYP2C9 [MIM 601130]* gene. At 1q25, the SNP rs2584820 was associated with Stroop test (P=0.0001) and phonological fluency (P=0.03), with an empirical P value of 0.04. This SNP is in intron 4 of the Regulator of G-protein Signaling Like 2 (*RGSL2*) gene. Two other SNPs in this region showed association to the TMTB test (P=0.0003, empirical P=0.04).

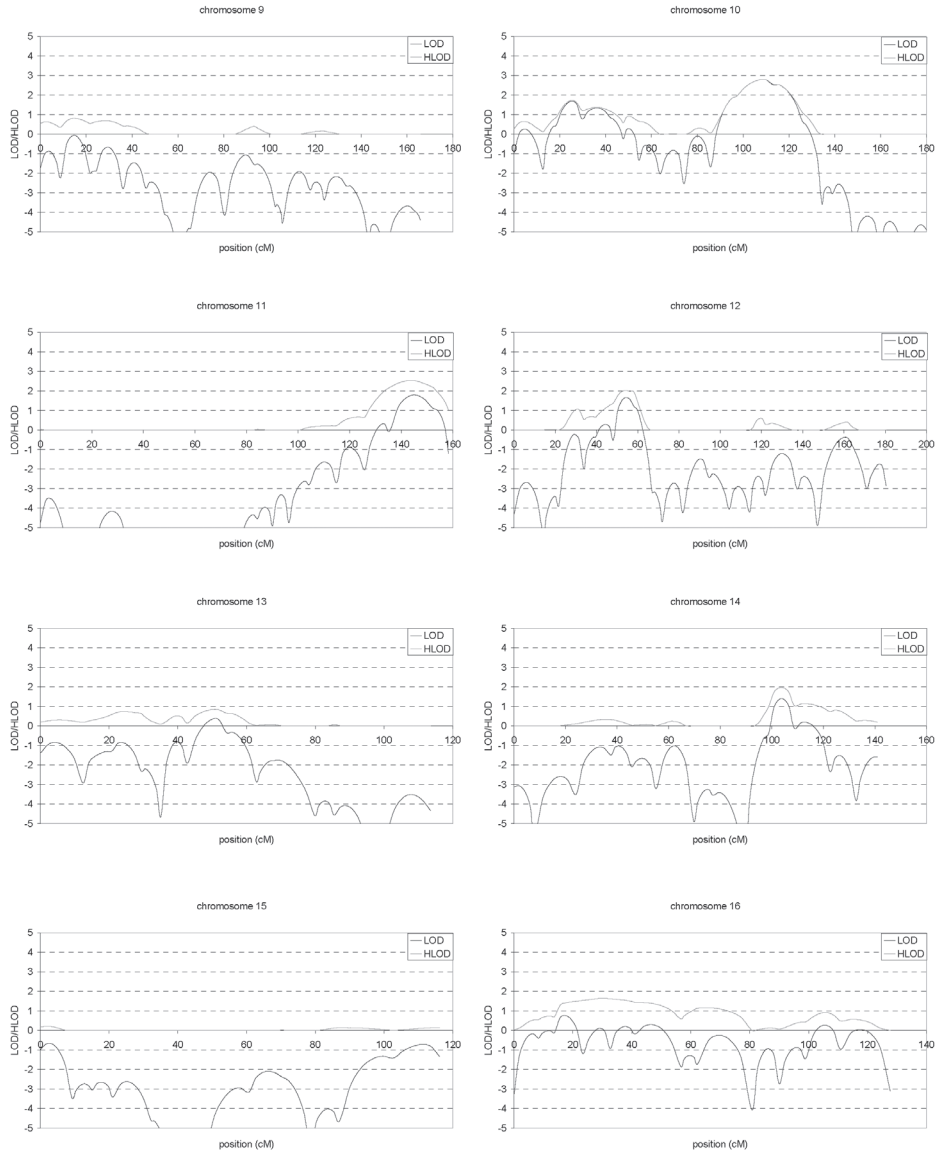


Figure 3. (continued)

They are 4kb downstream to the *C1orf49* gene and 149kb upstream to the Ral GEF with PH domain and SH3 binding motif 2 (*RALGPS2*) gene. At 3q23 the SNP rs952797 was associated with the Stroop test (P=0.0001), Block test (P=0.0002) and learning (P=0.06). When evaluating all tests simultaneously, the association was significant (empirical P=0.04). This SNP is 126 kb downstream of the gene encoding nicotinamide nucleotide adenylyltransferase 3 (*NMNAT3 [MIM 608702]*) and 131kb upstream of the gene encoding calsyntein 2 (*CLSTN2*). SNP rs11223225(C→T) at 11q25 showed a consistent allelic effect across key cognitive do-

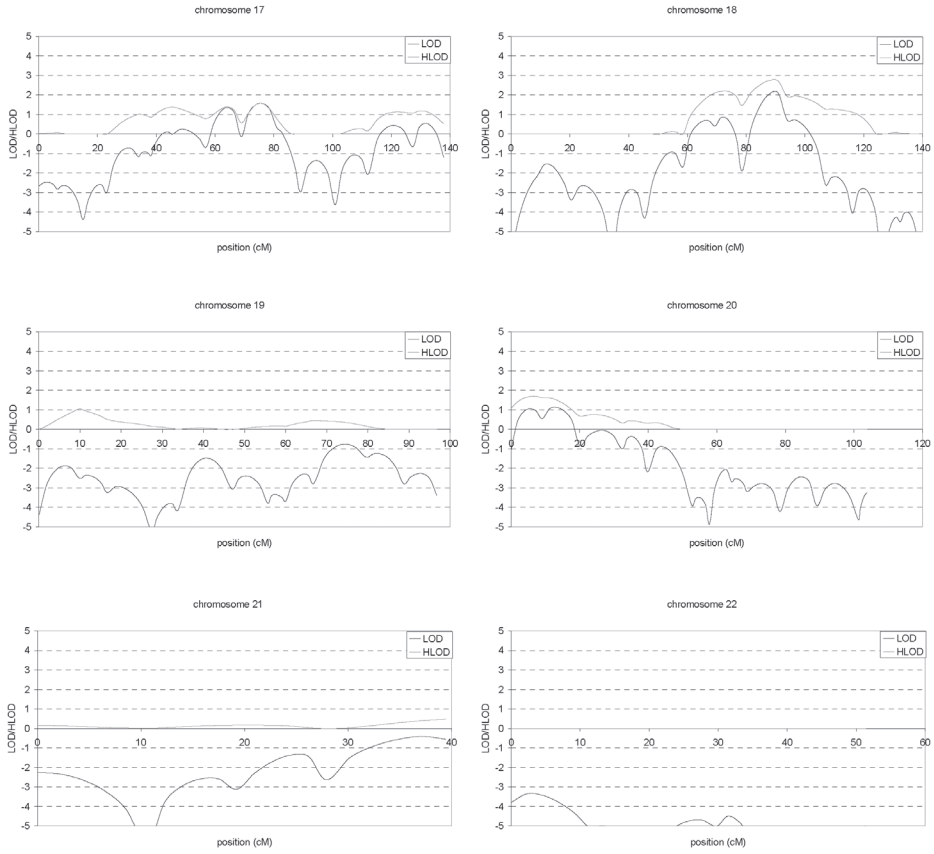


Figure 3. (continued)

mains for AD including learning, delayed recall and concept shifting (Stroop and TMTB test), where the minor allele of this SNP is associated with poorer performance on delayed recall (P=0.0004), learning (P=0.03), the Stroop test (P=0.02), TMTB test (P=0.09), and the Block test (P=0.07). When combining the effect of various tests, the overall empirical P-value is 0.03. This SNP is in intron 1 of the gene encoding opioid binding protein/cell adhesion molecule-like (*OPCML [MIM 600632]*). Four close SNPs rs1629316, rs1547897, rs1122931, and rs11222932 at 11q25 were associated with TMTB and phonological fluency. These SNPs are in intron 1 of the gene encoding neurotrimin (*HNT [MIM 607938]*). The *OPCML* and *HNT* genes are less than 80 kb apart.

DISCUSSION

This study confirms earlier findings suggesting linkage of AD to a wide region that spans chromosome 1q21-31[24,25,53]. The 1q21 region yielded the most significant evidence of linkage

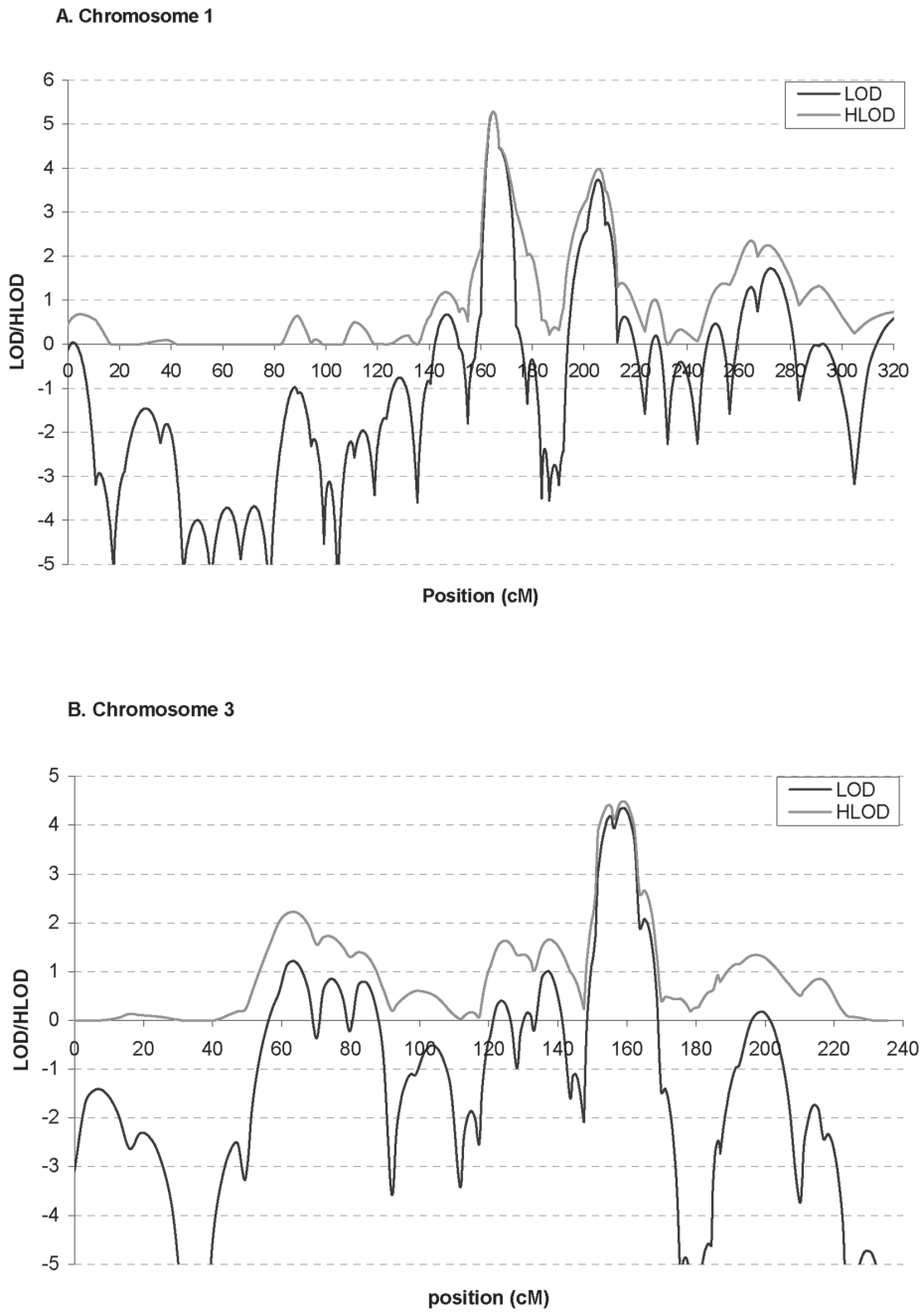
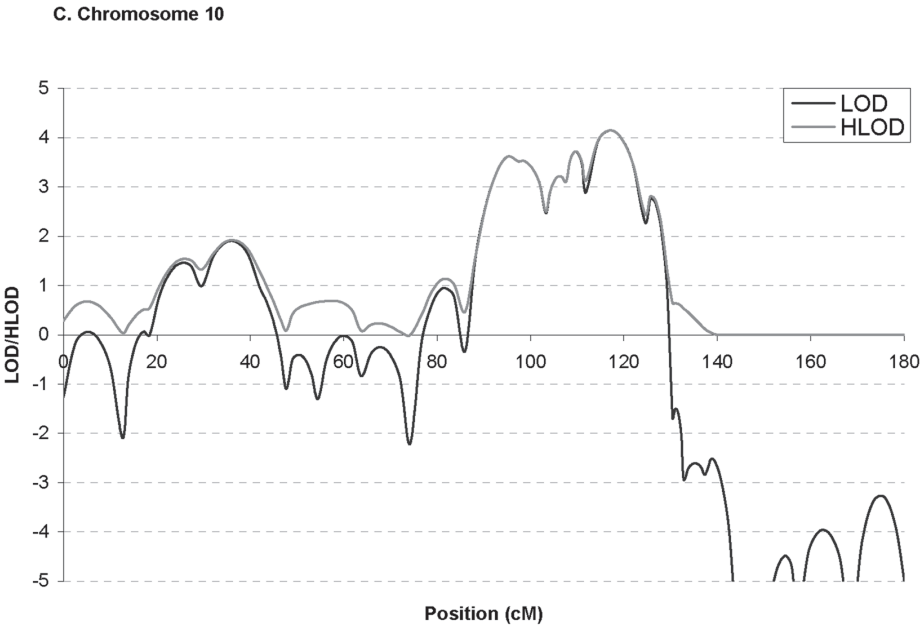


Figure 4. Multipoint LOD (blue) and HLOD (pink) scores for chromosomes 1, 3, 10 and 11 in the genome screen of late onset Alzheimer's disease after fine typing. Marker locations are given in Kosambi centi-Morgans.



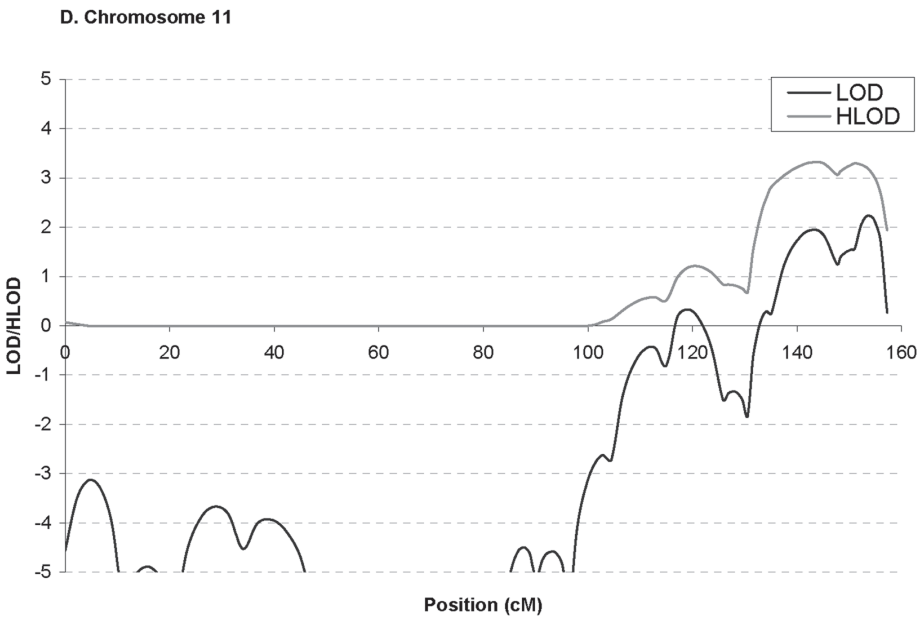
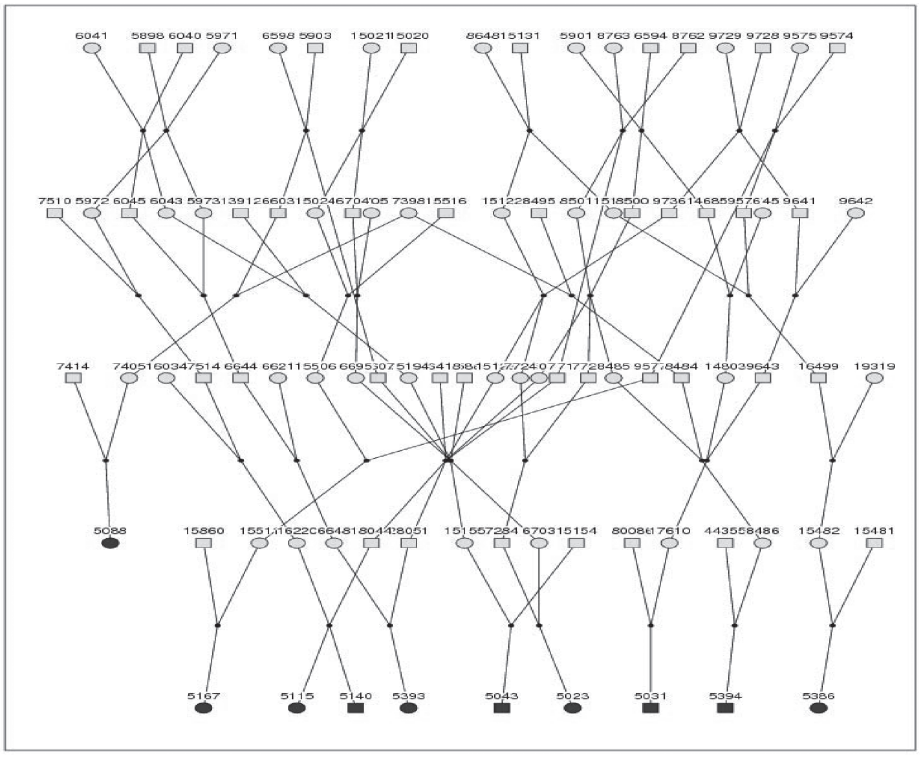


Figure 4. (continued)

A. Haplotype of chromosome 1q21

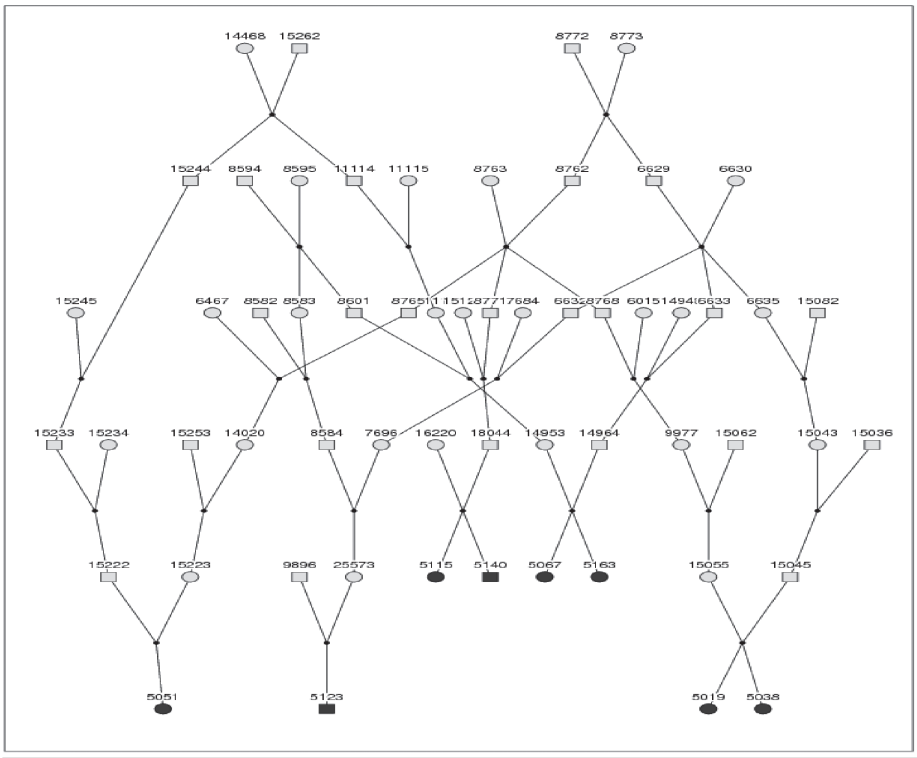


Markers	5088*		5167		5115		5140		5393*		5043		5023*		5031		5394*		5386	
D1S2746	(1)	(3)	2	1	7	1	7	1	1	3	1	8	7	6	1	1	7	7	7	3
D1S252	2	3	1	4	1	1	1	1	1	2	1	4	1	1	1	1	1	4	1	6
D1S514	(6)	(4)	6	1	6	4	6	7	6	5	6	5	6	4	6	1	6	5	6	3
D1S498	1	9	1	4	1	10	1	6	1	5	1	4	1	2	1	6	1	4	3	9
D1S2635	3	1	3	1	3	3	3	6	3	6	3	4	(3)	(6)	3	3	(3)	(6)	(3)	(1)
D1S484	3	3	3	2	3	3	3	3	3	1	3	1	3	2	3	1	1	3	1	2

Figure 5. Haplotypes of chromosomes 1, 3 and 11 segregating with AD families from the GRIP population.

over the genome in our study (HLOD = 5.2). This region was not replicated when testing for association with cognition in a series of 197 distantly related subjects. Although we cannot exclude the possibility of a false positive finding, given the strength of the linkage signal and previous evidence, it is more likely that there is a rare mutation in a major gene in this region, which could not be identified by association analysis in a small sample. This region

B. Haplotype of chromosome 1q25

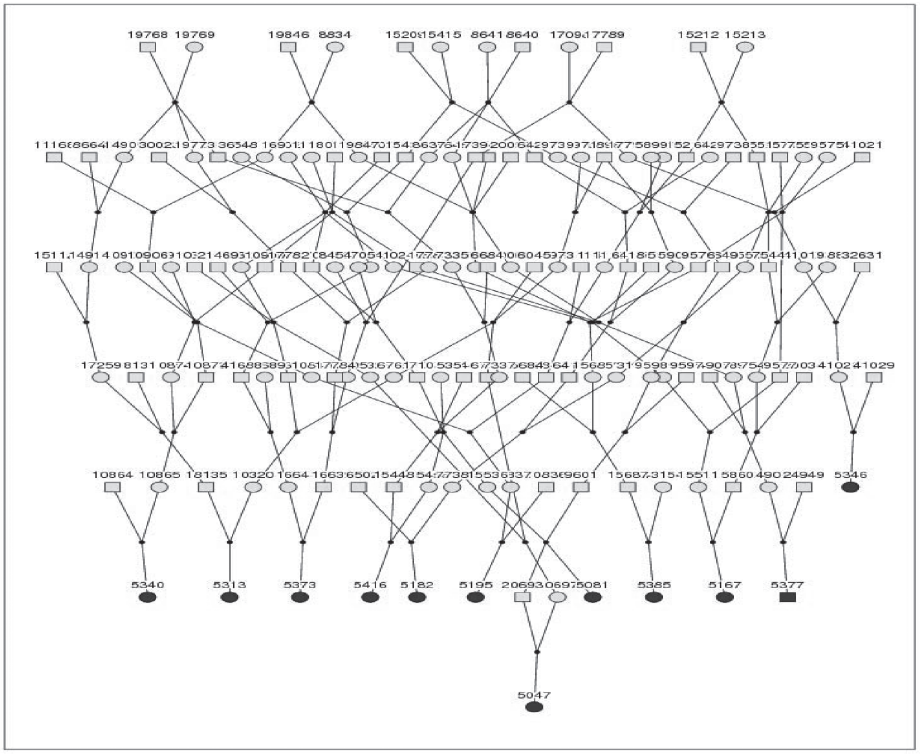


Markers	5051*		5123		5115*		5140*		5067*		5163*		5019		5038	
D1S2799	3	8	3	4	3	1	4	1	7	1	7	1	1	2	1	2
D1S218	1	1	1	5	5	1	5	1	1	1	1	1	1	2	1	2
D1S466	5	1	1	3	3	1	(1)	(1)	1	1	1	1	1	1	1	1
D1S238	1	2	1	7	2	2	7	2	4	1	4	1	2	1	2	1

Figure 5. (continued)

contains the *NCSTN* gene, which binds presenilin and is required for γ-secretase activity and Aβ generation[54]. Mutations in this gene have been found to be related to early onset AD and we have reported association in a sub-group of patients with familial early onset AD, particularly in those who lack the *APOE* E4 allele[55]. We have sequenced all the exons and splice sites of this gene in 6 patients, but have not found variants. Another obvious candidate gene in this region is the gene encoding C-reactive protein (*CRP [MIM 123260]*), which acts as a scavenger for chromatin released by dead cells during the acute inflammatory process[56].

C. Haplotype of chromosome 3q23

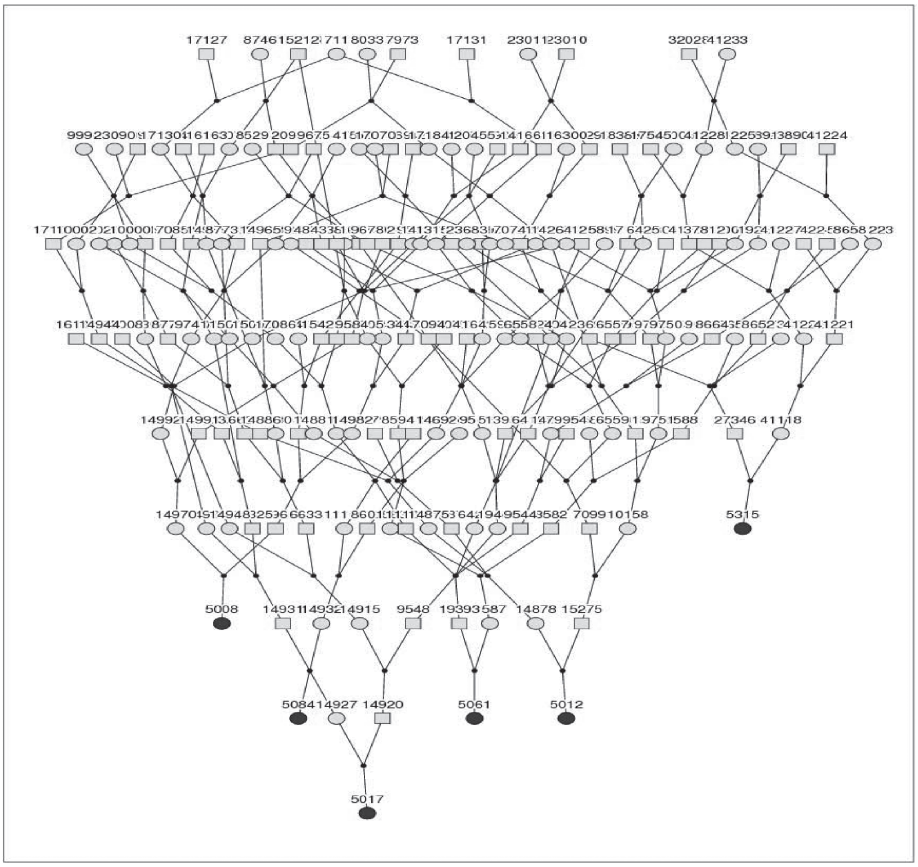


Markers	5340		5313*		5373		5416*		5182*		5195		5047*		5081*		5385		5167*		5377		5346	
D3S3514	2	3	2	9	4	4	2	2	2	5	2	1	2	2	2	2	2	2	2	1	7	2	2	1
D3S1292	2	9	2	4	5	2	2	2	2	9	2	1	2	2	2	2	2	6	2	1	2	2	2	1
D3S1549	5	2	5	3	5	5	5	2	5	3	5	3	5	2	5	5	5	5	5	8	5	1	5	4
D3S1569	3	1	3	1	3	1	3	2	3	3	3	2	2	6	3	1	3	1	3	1	3	4	3	3
D3S3626	2	8	2	1	3	2	7	3	3	1	2	2	3	3	2	2	2	2	2	8	2	1	2	2

Figure 5. (continued)

We also sequenced the exons and splice sites of this gene in 7 patients (5088, 5167, 5115, 5140, 5393, 5023,and 5394 of figure 3A) and found that all patients, except 5167 and 5393, carry the rare allele of SNPs rs1130864 (C→T) and rs1417938 (T→A). The SNP rs1130864 has been reported as a tagging SNP for a haplotype associated with higher levels of CRP[57]. We specifically tested the association of polymorphisms in *CRP* with cognitive function but failed to show any association (data not shown). As CRP is a key protein involved in inflammation, a key process in life by itself, a major mutation in *CRP* seems unlikely for late onset diseases

D. Haplotype of chromosome 11q24



Markers	5008*		5084*		5017*		5061		5012*		5315	
D11S925	1	2	8	6	8	3	3	9	8	1	11	2
D11S1328	5	5	3	5	3	5	5	5	3	1	9	3
D11S4151	5	2	2	5	2	2	2	3	2	3	2	4
D11S4131	6	3	6	7	6	3	6	3	6	7	6	3
D11S1320	1	1	1	1	1	1	1	2	1	3	1	3
D11S968	7	1	5	5	5	1	3	1	1	1	5	3

Figure 5. (continued)

suggesting another gene in the region may explain our high LOD score. In the 1q25 region, there was a second haplotype segregating. This region was confirmed in our association analysis by a SNP in intron 4 of RGSL2 gene, which may be involved in G-protein coupled

Table 5. Select SNPs in the regions linked to AD

	START				END					
Chr	marker	cM	SNP	position	marker	cM	SNP	position	Nbr SNPs	After QC
1	D1S514	152.5	rs2790308	141501392	D1S2635	165.6	rs16827466	156462773	828	585
1	D1S218	191.5	rs17838246	171261041	D1S466	198.5	rs16860717	179972352	750	584
3	D3S1549	151.5	rs7632392	139168654	D3S3626	164.3	rs10513332	149350824	954	769
10	D10S580	96.7	rs7101263	77715269	D10S205	125.4	rs12765878	105659612	2518	2006
11	D11S4131	138.6	rs1526562	131160829	D11S968	147.8	rs7936592	133620325	260	229
18	D18S474	71.3	rs9963534	47896056	D18S64	84.8	rs1622784	55603302	898	696

receptor protein signaling pathway. Also two SNPs upstream showed evidence for association. However, these SNPs were significant only for the TMTB test and are intergenic, making it more likely that *RGSL2* is the relevant gene in the region 1q25.

The second highest linkage signal was found at chromosome 3q22-24. This region was reported earlier to be linked to AD without tau pathology in a study of a small family with 4 affected relatives[58]. A significant LOD score of 4.1 between markers D3S1569 and D3S3554 was reported, where in our study D3S1569 is also the marker that gives the highest HLOD over chromosome 3. In the study of Poduslo no genome wide screen was conducted but only chromosomes 3 and 17 were screened since the disease was expected to be related to frontotemporal dementia (FTD) and the phenotype was apparently considered to be compatible with that of FTD. As we do not have pathology information of our patients, we cannot exclude that part of our patients also suffer from this atypical form of AD. However, all patients were carefully evaluated by an expert neurologist on FTD. A recent linkage-based genome-scan using Caribbean Hispanic families revealed a new locus on chromosome 3q28 with 2-point LOD score of 3.09 at marker D3S2418[34]. However, this region is about 50cM downstream from the region we identified in our study. The linked region on chromosome 3q22-24 contains various possible candidate genes including the *TF* gene, the gene encoding for butyrylcholinesterase (*BCHE [MIM 177400]*), the neprilysin gene (*MME [MIM 1205203]*) and the somatostatin gene (*SST [MIM 182450]*). We screened these genes for mutations but no variants were found. The SNP rs952797 at 3q23 was consistently associated with cognitive function in the 197 unrelated subjects from GRIP population. This SNP is 126 kb downstream of the *NMNAT3* gene encoding nicotinamide nucleotide adenylyltransferase 3 (NAD3). The coenzyme NAD and its derivatives are involved in hundreds of metabolic redox reactions and are utilized in protein ADP-ribosylation, histone deacetylation, and in some Ca(2+) signaling pathways. NMNAT is a central enzyme in NAD biosynthesis, catalyzing the condensation of nicotinamide mononucleotide (NMN) or nicotinic acid mononucleotide (NaMN) with the AMP moiety of ATP to form NAD or NaAD (Zhang *et al.*, 2003) and thus, the *NMNAT3* gene may relevant in AD. The SNP rs952797 is 131 kb upstream of the *CLSTN2* gene. It has been reported recently that SNP rs6439886, a common T→C substitution within the first intron of *CLSTN2*, was significantly associated with memory performance[59]. Our SNP rs952797 is 160 kb upstream to the reported SNP.

Table 6. SNPs significantly associated to various aspects of cognitive function in 197 unrelated individuals

Linked Region	SNP[1]	cM[2]	Physical position[3]	Delayed			Stroop	TMTB	Semantic fluency	Phonological fluency	Block	Combined test	Empirical P value[4]
				Learning	recall	Recognition							
1q25	rs17361286	193.9	176788539	0.68	0.56	0.91	0.44	**0.0003**	0.96	0.64	0.51	1.2×10^{-5}	**0.04**
	rs2493864	194.0	176811626	0.76	0.63	0.96	0.46	**0.0003**	0.99	0.62	0.49	1.7×10^{-5}	**0.04**
	rs2584820	197.6	180705709	0.20	0.19	0.76	**0.0001**	0.89	0.07	**0.03**	0.98	5.9×10^{-9}	**0.02**
3q23	rs952797	152.6	141005823	0.06	0.47	0.57	**0.05**	0.89	0.82	0.24	**0.0002**	2.8×10^{-8}	**0.04**
10q23	rs17129662	111.8	91730916	0.23	0.55	0.48	**0.00003**	**0.002**	0.10	**0.002**	0.84	7.0×10^{-13}	**0.02**
	rs11185978	111.8	91778768	0.36	0.60	0.60	**0.00002**	**0.01**	**0.05**	**0.001**	0.61	8.9×10^{-13}	**0.01**
	rs7071717	111.8	91808267	0.30	0.73	0.67	**0.000005**	**0.01**	**0.04**	**0.01**	0.90	2.9×10^{-12}	**0.002**
	rs4110517	116.8	96640318	0.10	0.09	0.94	0.18	0.24	**0.00003**	**0.04**	0.87	4.7×10^{-10}	**0.02**
11q25	rs1629316	139.7	131440484	0.07	0.21	0.41	0.94	**0.0007**	0.35	**0.05**	0.10	8.0×10^{-9}	**0.05**
	rs1547897	139.7	131451988	0.06	0.16	0.39	1.00	**0.0006**	0.29	**0.05**	0.16	4.8×10^{-9}	**0.04**
	rs11222931	139.8	131483366	0.07	0.15	0.33	0.87	**0.0007**	0.27	**0.05**	0.16	4.5×10^{-9}	**0.04**
	rs11222932	139.8	131483471	0.07	0.15	0.34	0.86	**0.0006**	0.30	**0.05**	0.18	4.6×10^{-9}	**0.04**
	rs11223225	142.7	132198420	**0.03**	**0.0004**	0.26	**0.02**	0.09	0.47	0.26	0.06	3.2×10^{-11}	**0.03**

1. All SNPs with empirical P-value smaller or equal to 0.05 are shown; 2 Centimorgan according to the Marshfield map;. 3. Physical positions according to NCBI reference map build 35 the SNPs departing from each other less than 100kb are in box; 4. The probability of observing an equal or smaller Fisher product by chance per region, based on permutation test of 500000 replicates; 5. For intergenic SNPs, only the closest neighbor genes are listed.

Chromosome 10q22-24 is the 3rd highest peak over the genome. This finding is consistent with previous findings on AD[14,19-21,23] and plasma amyloid β42 levels [60]. Our finding of linkage to late onset AD at 10q22-24 is the first replication using a data set fully independent from the NIMH sample. So far it has been difficult to identify the causal mutation(s) in this region. A series of genes have been densely genotyped and several genes have been noted to be susceptibility genes for AD, including *IDE*, *CH25H [MIM 604551]*, *PLAU [MIM 191840]*, and *LIPA [MIM 278000]*. In our linkage analysis, there was not a single haplotype segregating in the region suggesting that multiple mutations in one or multiple genes may contribute to the linkage. In our association analysis, the most significant evidence of association with cognitive function was seen for this region. Three SNPs: rs17129662, rs11185978, and rs7071717 together at 91.7 Mb showed association with multiple cognitive domains in the 197 unrelated subjects from the GRIP population. These SNPs are intergenic SNPs of known genes. All 3 SNPs are less then 1 Mb upstream to the *CH25H* gene and the *LIPA* gene and less than 3 Mb downstream to the *IDE* gene. The genes most closely flanking these SNPs are the *MPHOSPH1* gene and the *HTR7* gene, which encodes 5-hydroxytryptamine receptor 7. These two genes, however, have not been extensively investigated. Another associated SNP rs4110517 at 116.8 cM is surrounded by 4 similar genes: *CYP2C18 [MIM 601131]*, *CYP2C19*, *CYP2C9*, *CYP2C8 [MIM 601129]* in a range less than 350kb. Since no significant association was found between the *CYP2C19* gene and familial AD patients in a previous study[61], it is more likely that the SNP rs4110517 is in LD with the causal gene(s) in the region.

We also found suggestive evidence for linkage to chromosome 11q25. Blacker et al (2003) previously described this region in their study of the NIMH sample, including 437 families with AD. Recent evidence suggests the *SORL1* gene may be responsible[29]. Our linkage peak is, however, about 23 cM downstream to the *SORL1* gene. We specifically tested the association between polymorphisms flanking *SORL1* and cognitive function but failed to detect consistent associations (data not shown), suggesting that our linkage peak may be explained by other gene(s). The association for SNP rs11223225, a C→T substitution at 11q25 is one of the most promising results from our association analysis. The T allele of this SNP is consistently associated with reduced cognitive performance on multiple domains and this SNP is an intronic SNP of the *OPCML* gene, which encodes the opioid binding protein. There is evidence that the opioidergic system is affected in AD. Furthermore, performance on immediate memory and mental flexibility tasks has been suggestively linked to 11q25 in a recent genome-wide linkage study in 260 families[62]. Other 4 close SNPs at 11q25 also showed association with cognitive function. These SNPs are in intron 1 of the *HNT* gene, which encodes neurotrimin. Notably, the *OPCML* and *HNT* genes are separated in less than 80kb.

In summary, we confirmed two previously well described linkage regions for late onset AD on chromosomes 1q21-25 and 10q22-24. Using cognitive function as an endophenotype of AD, our study indicates the *RGSL2*, *RALGPS2*, and *C1orf49* genes at 1q25. Our analysis on chromosome 10q22-24 points to *HTR7*, *MPHOSPH1*, and *CYP2C* cluster. This is the first

genome-wide screen that showed significant linkage to chromosome 3q23 markers. For this region our analysis identified *NMNAT3* and *CLSTN2* genes. Our findings confirm linkage to chromosome 11q25. We could not confirm *SORL1*, instead, our analysis points to *OPCML* and *HNT* genes.

Web Resources

The program, pedcut, for breaking large pedigrees is available at: http://mga.bionet.nsc.ru/soft/index.html

The program, FCN, for characterization of large genealogy is available at: http://mga.bionet.nsc.ru/soft/index.html

The software package PedHunter facilitates creation and verification of pedigrees within large genealogies, which is available at: http://www.ncbi.nlm.nih.gov/CBBresearch/Schaffer/pedhunter.html.

The PEDIG package for relationship coefficients calculation in large pedigrees is available at: http://www-sgqa.jouy.inra.fr/diffusions.html.

The software package pedfiddler version 0.5 (JC Loredo-Osti and K Morgan) for drawing large pedigrees is available at:

http://www.medicine.mcgill.ca/statgene/software.html.

The program Simwalk2 for multipoint linkage analyses using Markov chain Monte Carlo is available at: http://watson.hgen.pitt.edu/register.

The program GENHUNTER for multipoint linkage analysis using Lander-Green algorithm is available at: http://linkage.rockefeller.edu/soft/gh.

The program Pedcheck for detecting Mendilian errors is available at: http://watson.hgen.pitt.edu/register.

The program MERLIN for detecting unlikely double-recombination events and haplotype construction is available at: http://www.sph.umich.edu/csg/abecasis/Merlin/download/

The AlzGene Database and Alzheimer Research Forum: http://www.alzgene.org.

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/.

REFERENCES

1. Hofman A, Grobbee DE, de Jong PT, van den Ouweland FA (1991) Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. Eur J Epidemiol 7:403-422
2. Rocca WA, Hofman A, Brayne C, Breteler MM, Clarke M, Copeland JR, Dartigues JF, Engedal K, Hagnell O, Heeren TJ, *et al.* (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
3. Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S, Fiske A, Pedersen NL (2006) Role of genes and environments for explaining Alzheimer disease. Arch Gen Psychiatry 63:168-174
4. Rogaev EI, Sherrington R, Rogaeva EA, Levesque G, Ikeda M, Liang Y, Chi H, Lin C, Holman K, Tsuda T, *et al.* (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
5. Levy-Lahad E, Wasco W, Poorkaj P, Romano DM, Oshima J, Pettingell WH, Yu CE, Jondro PD, Schmidt SD, Wang K, *et al.* (1995) Candidate gene for the chromosome 1 familial Alzheimer's disease locus. Science 269:973-977
6. Goate A, Chartier-Harlin MC, Mullan M, Brown J, Crawford F, Fidani L, Giuffra L, Haynes A, Irving N, James L, *et al.* (1991) Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. Nature 349:704-706
7. van Duijn CM, de Knijff P, Cruts M, Wehnert A, Havekes LM, Hofman A, Van Broeckhoven C (1994) Apolipoprotein E4 allele in a population-based study of early-onset Alzheimer's disease. Nat Genet 7:74-78
8. 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
9. Sleegers K, Van Duijn CM (2001) Alzheimer's Disease: Genes, Pathogenesis and Risk Prediction. Community Genet 4:197-203
10. Bertram L, Hiltunen M, Parkinson M, Ingelsson M, Lange C, Ramasamy K, Mullin K, Menon R, Sampson AJ, Hsiao MY, *et al.* (2005) Family-based association between Alzheimer's disease and variants in UBQLN1. N Engl J Med 352:884-894
11. Albert MS, Moss MB, Tanzi R, Jones K (2001) Preclinical prediction of AD using neuropsychological tests. J Int Neuropsychol Soc 7:631-639
12. Myers AJ, Goate AM (2001) The genetics of late-onset Alzheimer's disease. Curr Opin Neurol 14: 433-440
13. 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
14. Kehoe P, Wavrant-De Vrieze F, Crook R, Wu WS, Holmans P, Fenton I, Spurlock G, Norton N, Williams H, Williams N, *et al.* (1999) A full genome scan for late onset Alzheimer's disease. Hum Mol Genet 8: 237-245
15. Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, Myers RH, Pericak-Vance MA, Risch N, van Duijn CM (1997) Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. *APOE* and Alzheimer Disease Meta Analysis Consortium. Jama 278:1349-1356

16. Pericak-Vance MA, Grubber J, Bailey LR, Hedges D, West S, Santoro L, Kemmerer B, Hall JL, Saunders AM, Roses AD, *et al.* (2000) Identification of novel genes in late-onset Alzheimer's disease. Exp Gerontol 35:1343-1352
17. Curtis D, North BV, Sham PC (2001) A novel method of two-locus linkage analysis applied to a genome scan for late onset Alzheimer's disease. Ann Hum Genet 65:473-481
18. Olson JM, Goddard KA, Dudek DM (2002) A second locus for very-late-onset Alzheimer disease: a genome scan reveals linkage to 20p and epistasis between 20p and the amyloid precursor protein region. Am J Hum Genet 71:154-161
19. Myers A, Holmans P, Marshall H, Kwon J, Meyer D, Ramic D, Shears S, Booth J, DeVrieze FW, Crook R, *et al.* (2000) Susceptibility locus for Alzheimer's disease on chromosome 10. Science 290:2304-2305
20. 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
21. Blacker D, Bertram L, Saunders AJ, Moscarillo TJ, Albert MS, Wiener H, Perry RT, Collins JS, Harrell LE, Go RC, *et al.* (2003) Results of a high-resolution genome screen of 437 Alzheimer's disease families. Hum Mol Genet 12:23-32
22. Goddard KA, Olson JM, Payami H, van der Voet M, Kuivaniemi H, Tromp G (2004) Evidence of linkage and association on chromosome 20 for late-onset Alzheimer disease. Neurogenetics 5:121-128
23. Holmans P, Hamshere M, Hollingworth P, Rice F, Tunstall N, Jones S, Moore P, Wavrant DeVrieze F, Myers A, Crook R, *et al.* (2005) Genome screen for loci influencing age at onset and rate of decline in late onset Alzheimer's disease. Am J Med Genet B Neuropsychiatr Genet 135:24-32
24. 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
25. Hiltunen M, Mannermaa A, Thompson D, Easton D, Pirskanen M, Helisalmi S, Koivisto AM, Lehtovirta M, Ryynanen M, Soininen H (2001) Genome-wide linkage disequilibrium mapping of late-onset Alzheimer's disease in Finland. Neurology 57:1663-1668
26. 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
27. Wijsman EM, Daw EW, Yu CE, Payami H, Steinbart EJ, Nochlin D, Conlon EM, Bird TD, Schellenberg GD (2004) Evidence for a novel late-onset Alzheimer disease locus on chromosome 19p13.2. Am J Hum Genet 75:398-409
28. Ashley-Koch AE, Shao Y, Rimmler JB, Gaskell PC, Welsh-Bohmer KA, Jackson CE, Scott WK, Haines JL, Pericak-Vance MA (2005) An autosomal genomic screen for dementia in an extended Amish family. Neurosci Lett 379:199-204
29. Lee JH, Cheng R, Schupf N, Manly J, Lantigua R, Stern Y, Rogaeva E, Wakutani Y, Farrer L, St George-Hyslop P, *et al.* (2007) The association between genetic variants in *SORL1* and Alzheimer disease in an urban, multiethnic, community-based cohort. Arch Neurol 64:501-506
30. Varilo T, Peltonen L (2004) Isolates and their potential use in complex gene mapping efforts. Curr Opin Genet Dev 14:316-323
31. Pardo LM, MacKay I, Oostra B, van Duijn CM, Aulchenko YS (2005) The effect of genetic drift in a young genetically isolated population. Ann Hum Genet 69:288-295
32. Escamilla MA (2001) Population isolates: their special value for locating genes for bipolar disorder. Bipolar Disord 3:299-317
33. Helgason A, Yngvadottir B, Hrafnkelsson B, Gulcher J, Stefansson K (2005) An Icelandic example of the impact of population structure on association studies. Nat Genet 37:90-95

34. Lee JH, Cheng R, Santana V, Williamson J, Lantigua R, Medrano M, Arriaga A, Stern Y, Tycko B, Rogaeva E, *et al.* (2006) Expanded genomewide scan implicates a novel locus at 3q28 among Caribbean hispanics with familial Alzheimer disease. Arch Neurol 63:1591-1598
35. Sleegers K, Roks G, Theuns J, Aulchenko YS, Rademakers R, Cruts M, van Gool WA, Van Broeckhoven C, Heutink P, Oostra BA, *et al.* (2004) Familial clustering and genetic risk for dementia in a genetically isolated Dutch population. Brain 127:1641-1649
36. Geerlings MI, Jonker C, Bouter LM, Ader HJ, Schmand B (1999) Association between memory complaints and incident Alzheimer's disease in elderly people with normal baseline cognition. Am J Psychiatry 156:531-537
37. Tierney MC, Szalai JP, Snow WG, Fisher RH, Nores A, Nadon G, Dunn E, St George-Hyslop PH (1996) Prediction of probable Alzheimer's disease in memory-impaired patients: A prospective longitudinal study. Neurology 46:661-665
38. Jonker C, Geerlings MI, Schmand B (2000) Are memory complaints predictive for dementia? A review of clinical and population-based studies. Int J Geriatr Psychiatry 15:983-991
39. Ando J, Ono Y, Wright MJ (2001) Genetic structure of spatial and verbal working memory. Behav Genet 31:615-624
40. McClearn GE, Johansson B, Berg S, Pedersen NL, Ahern F, Petrill SA, Plomin R (1997) Substantial genetic influence on cognitive abilities in twins 80 or more years old. Science 276:1560-1563
41. Swan GE, Reed T, Jack LM, Miller BL, Markee T, Wolf PA, DeCarli C, Carmelli D (1999) Differential genetic influence for components of memory in aging adult twins. Arch Neurol 56:1127-1132
42. Lee JH, Flaquer A, Stern Y, Tycko B, Mayeux R (2004) Genetic influences on memory performance in familial Alzheimer disease. Neurology 62:414-421
43. Aulchenko YS, Heutink P, Mackay I, Bertoli-Avella AM, Pullen J, Vaessen N, Rademaker TA, Sandkuijl LA, Cardon L, Oostra B, *et al.* (2004) Linkage disequilibrium in young genetically isolated Dutch population. Eur J Hum Genet 12:527-534
44. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM (1984) 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
45. Ciullo M, Bellenguez C, Colonna V, Nutile T, Calabria A, Pacente R, Iovino G, Trimarco B, Bourgain C, Persico MG (2006) New susceptibility locus for hypertension on chromosome 8q by efficient pedigree-breaking in an Italian isolate. Hum Mol Genet 15:1735-1743
46. Sleegers K, Brouwers N, Gijselinck I, Theuns J, Goossens D, Wauters J, Del-Favero J, Cruts M, van Duijn CM, Van Broeckhoven C (2006) APP duplication is sufficient to cause early onset Alzheimer's dementia with cerebral amyloid angiopathy. Brain 129:2977-2983
47. Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16:1215
48. Aulchenko YS, Bertoli-Avella AM, van Duijn CM (2005) A method for pooling alleles from different genotyping experiments. Ann Hum Genet 69:233-238
49. O'Connell JR, Weeks DE (1998) PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet 63:259-266
50. Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347-1363
51. Liu F, Elefante S, van Duijn CM, Aulchenko YS (2006) Ignoring Distant Genealogic Loops Leads to False-positives in Homozygosity Mapping. Ann Hum Genet 70:965-970
52. Fisher RA (1932) Statistical Methods for Research Workers. (Oliver and Boyd, London)

53. Myers A, Wavrant De-Vrieze F, Holmans P, Hamshere M, Crook R, Compton D, Marshall H, Meyer D, Shears S, Booth J, *et al.* (2002) Full genome screen for Alzheimer disease: stage II analysis. Am J Med Genet 114:235-244
54. Kalmijn S, van Boxtel MP, Ocke M, Verschuren WM, Kromhout D, Launer LJ (2004) Dietary intake of fatty acids and fish in relation to cognitive performance at middle age. Neurology 62:275-280
55. Dermaut B, Theuns J, Sleegers K, Hasegawa H, Van den Broeck M, Vennekens K, Corsmit E, St George-Hyslop P, Cruts M, van Duijn CM, *et al.* (2002) The gene encoding nicastrin, a major gamma-secretase component, modifies risk for familial early-onset Alzheimer disease in a Dutch population-based sample. Am J Hum Genet 70:1568-1574
56. Robey FA, Jones KD, Tanaka T, Liu TY (1984) Binding of C-reactive protein to chromatin and nucleosome core particles. A possible physiological role of C-reactive protein. J Biol Chem 259: 7311-7316
57. Despriet DD, Klaver CC, Witteman JC, Bergen AA, Kardys I, de Maat MP, Boekhoorn SS, Vingerling JR, Hofman A, Oostra BA, *et al.* (2006) Complement factor H polymorphism, complement activators, and risk of age-related macular degeneration. Jama 296:301-309
58. Poduslo SE, Yin X, Hargis J, Brumback RA, Mastrianni JA, Schwankhaus J (1999) A familial case of Alzheimer's disease without tau pathology may be linked with chromosome 3 markers. Hum Genet 105:32-37
59. Papassotiropoulos A, Stephan DA, Huentelman MJ, Hoerndli FJ, Craig DW, Pearson JV, Huynh KD, Brunner F, Corneveaux J, Osborne D, *et al.* (2006) Common Kibra alleles are associated with human memory performance. Science 314:475-478
60. 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
61. Yamada H, Dahl ML, Viitanen M, Winblad B, Sjoqvist F, Lannfelt L (1998) No association between familial Alzheimer disease and cytochrome P450 polymorphisms. Alzheimer Dis Assoc Disord 12: 204-207
62. Buyske S, Bates ME, Gharani N, Matise TC, Tischfield JA, Manowitz P (2006) Cognitive traits link to human chromosomal regions. Behav Genet 36:65-76

Chapter 5

The Apolipoprotein E Gene and its Age Specific Effects on Cognitive Function

ABSTRACT

The E4 allele of the apolipoprotein E gene (*APOE*) is a well-established determinant of Alzheimer's disease and cognitive function. We studied the age-specific effects of the *APOE* E4 allele on cognitive function in a series of 2208 related individuals from a family-based study conducted in an isolated population in the Southwest part of The Netherlands. The effect of the E4 allele on cognitive function was evaluated using standard quantitative genetic analysis under a polygenic model, adjusted for cardiovascular risk factors. We found a significant association between the *APOE* E4 allele and reduced scores on the Adult Verbal Learning Test (AVLT) in persons aged 50 years and older (AVLT short-term memory P = 0.01, AVLT learning P = 0.001, AVLT delayed recall P = 0.01 and memory compound score P = 0.001). The effect of *APOE* E4 is most pronounced on learning ability, starting as early 40 years. The *APOE* E4 allele is also strongly associated to cholesterol levels and atherosclerosis. This association did not explain the effect of *APOE* on cognitive function. Our study suggests that *APOE* E4 is an important determinant of vascular and neurological pathology at late age.

INTRODUCTION

The epsilon4 allele of the apolipoprotein E gene (*APOE* E4) is the most important genetic risk factor for Alzheimer's disease (AD)[1]. Although its role in AD has long been known, recently a commercial genetic test for *APOE* E4 was launched to predict the risk of AD (http://www.labtestsonline.org)[2]. However, *APOE* E4 has also an established effect on lipid levels and through this on the risk and progression of atherosclerosis. Athereosclerosis and hypertension have been implicated in the AD and may partly explain the effects of *APOE* E4 on cognitive function [1]. If *APOE* E4 leads through neurodegeneration in part through vascular pathology, this pathway opens the opportunity of clinical counselling of carriers by screening for vascular pathology. A crucial question in this respect is at which age pathology starts. It has been suggested that *APOE* E4 has clinically important effects on cognition in those who do not have signs or symptoms clinical AD. An extensive meta-analysis of all studies conducted in the period 1993-2004 showed evidence for a role of *APOE* E4 in cognitive function in non-demented people over 50 years[3]. *APOE* E4 was significantly related to reduced global cognitive functioning, episodic memory and executive function in a dose-dependent way, whereas no significant effects were seen for primary memory[3]. Although most studies focused on individuals aged 50 years and over, there is some evidence that with increasing age the effect of *APOE* E4 on cognition decreased. However, this trend was far from statistically significant in the meta-analysis. Animal studies provided significant evidence that apolipoprotein E has effects on early brain development[4], suggesting that *APOE* E4 may impact early cognitive reserve. For humans, the evidence supporting early effects of *APOE* on cognitive function is scarce and findings have been contradictory.

In the present study, we evaluated the effects of the *APOE* E4 allele on specific cognitive domains and vascular pathology over a wide age-range in a 3 generation family-based study. This design provides a powerful setting to address age specific effects of *APOE* E4 in a genetically and environmentally homogeneous background.

METHODS

Study population

The Erasmus Rucphen Family (ERF) cohort, which is part of the Genetic Research in Isolated Population (GRIP) program, is a family-based study that includes inhabitants of a genetically isolated community in the south-western area of the Netherlands5. ERF aims to investigate the genetic origins of complex disorders and traits. The study population essentially consists of one extended family of descendants from 20 related couples that lived in the isolate between 1850 and 1900 and had at least 6 children. With relatively limited migration until the last few decades, the isolate now includes approximately 20,000 inhabitants. All data were collected

between 2002 and 2005. The Medical Ethical Committee of the Erasmus Medical Center Rotterdam approved the study and informed consent was obtained from all participants.

Data collection

Participants underwent extensive medical and neuropsychological examinations at the ERF research centre. The examinations included the determination of cardiovascular risk factors, such as serum total cholesterol, high-density lipoprotein (HDL) cholesterol, triglycerides, systolic and diastolic blood pressure, and common carotid intima media thickness (IMT). Serum markers were determined using an automated enzymatic procedure (Boehringer Mannheim System). Blood pressure was measured twice on the right arm in a sitting position after at least five minutes rest, using an automated device (OMRON 711); the average of the two values was used for analysis. IMT was evaluated using ultrasonography according to previously applied protocols[6-8]. The outcome variable was defined as the mean IMT of the near and far wall of both common carotid arteries. The battery of neuropsychological tests included the Dutch version of the Auditory Verbal Learning Test (AVLT)[9], the Trail Making Test (TMT)[10], the Stroop colour-word test[11], the verbal fluency test[10] and the block design subtest of the Weschler Adult Intelligence Scale (WAIS)[12]. These tests were chosen to screen for cognitive deficits related to AD and other dementias[13] and cover different cognitive domains (Table 1). We assessed the general reading ability of the participants with the Dutch Adult Reading Test (DART)[14]. We also computed compound scores for memory performance, executive function and over-all cognitive function (Table 1), by averaging the z-transformed scores of several cognitive tests[15]. The z-scores were calculated based on the direction of the measurement of test performance. For tests where higher scores indicate better performance (AVLT and WAIS tests), $z = (x - \bar{x})/sd$; otherwise (TMT and Stroop), $z = (\bar{x} - x)/sd$. In this way, higher compound scores indicate better performance.

Finally, the education level attained by the subjects within the Dutch educational system was determined according to eight ordinal categories from primary school to university[16].

Genotyping

Genomic DNA was extracted from whole blood samples using the salting out method[17]. Samples were genotyped for the *APOE* C112R (E4 allele) and *APOE* R158C (E2 allele) polymorphisms with a Taqman allelic discrimination Assay-By-Design (Applied Biosystems, Foster City, CA). The assays utilized 5 nanograms of genomic DNA and 2 microliter reaction volumes. The amplification and extension protocol included an initial activation step of 10 min at 95 degrees, which preceded 40 cycles of denaturation at 95 degrees for 15 seconds and annealing and extension at 50 degrees for 60 seconds. Allele-specific fluorescence was analysed on an ABI Prism 7900HT Sequence Detection System with SDS v 2.1 (Applied Biosystems, Foster City, CA).

Table 1. Description of the cognitive tests derived from the neuropsychological battery

Neuropsychological test	Cognitive domain	Task description	Score definition	Reference
Individual tests				
AVLT Short-term memory	Short-term memory	Recall immediately after presentation of 15 words	Number of correctly recalled words	Saan and Deelman 1986
AVLT Learning	Learning	Recall after 2nd to 5th presentation	Total number of correctly recalled words	Saan and Deelman 1986
AVLT Delay	Delayed recall	Recall after 30 minutes	Number of correctly recalled words	Saan and Deelman 1986
AVLT Recognition	Recognition	Recognize words from a list	Number of correctly recognized and rejected words	Saan and Deelman 1986
WAIS Verbal fluency	Semantic fluency and phonological fluency	Mention words fitting a frame (semantic & phonological)	Number of correctly mentioned words	Wechsler 2000
Trail-making test (TMT)	Cognitive flexibility	Connect numbers (A) and together with letters in ascending order (B)	Ratio of time in seconds to complete part B over part A	Reitan 1955
Stroop	Susceptibility to interference	Read colors (card 2) which are wrongly named (card 3)	Ratio of time in seconds to complete the card 3 over card 2	Hammes 1978
WAIS Block design	Visuoconstructive abilites	Place blocks according to reference	Number of replicated blocks	Wechsler 2000
Compound scores				
Memory performance			Average of z-transformation of AVLT short, learning, delay, recognition	
Executive function			Average of z-transformation of WAIS verbal fluency, TMT, Stroop	
Overall cognitive function			Average of z-transformation of all tests	

Statistical analysis

A considerable proportion of participants failed to complete the TMT part B test (N = 171, 7.9%), while some failed to complete the Stroop card III (N = 16, 0.7%) and WAIS block design tests (N = 64, 2.9%) within the time limit. We imputed their scores based on correlations between sex, age, and education level. We grouped the APOE genotypes based on the number of E4 alleles in a dose-dependent manner 3, zero (E2/E2, E2/E3 and E3/E3 genotypes), one (E2/E4 and E3/E4 genotypes) and two copies (E4/E4 genotype). General characteristics of the study population among the genotypic groups were compared using the one-way ANOVA test for continuous variables and the chi-square test for dichotomous variables as implemented in SPSS V.11.0 (SPSS Inc. Chicago IL). The observed frequencies of the APOE genotypes were tested for deviations from Hardy-Weinberg equilibrium using the exact test for multiple alleles[18].

To evaluate the effect of the E4 allele on cognitive functioning and adjust for family relationships, we performed the variable screening analysis under the polygenic model using the

SOLAR software package version 4.1.0[19]. SOLAR was chosen for its power in discriminating the genetic and environmental effects by utilizing all of the information that is provided by large, complex pedigrees. The effect of *APOE* genotype on cognitive tests was estimated by including *APOE* genotype (0, 1, or 2 number of E4 alleles) as a covariate in the model, adjusted for other covariates including age, age-squared, sex, education, inbreeding, DART score, and cardiovascular risk factors (total cholesterol, triglycerides, IMT, and systolic and diastolic blood pressure). Inbreeding coefficients were computed based on all available genealogical information for the GRIP population (N = 107,091) using PEDIG software[20]. In addition, we investigated the interaction between *APOE* and age using a multiplicative model. Before SOLAR analyses, scores from the cognitive tests were normalized using a general rank-transformation[21].

To illustrate the age-specific effect of *APOE* on cognitive function, we smoothed the distribution of cross-sectional test scores across age, using locally weighted regression, or the LOESS smoother, implemented in the software package SigmaPlot version 8.02[22].

RESULTS

Information on both *APOE* genotype and cognitive tests is available for 2208 ERF participants in our study. We excluded 65 individuals who were illiterate, blind, deaf, retarded or who reported having a brain tumor, stroke or severe brain damage. The frequencies of *APOE* alleles were 4.8% for the E2 allele, 74.1% for the E3 allele and 21.1% for the E4 allele. The allele and genotype distributions followed Hardy-Weinberg equilibrium (P = 0.64). There were no significant differences in age, sex, education level and blood pressure between *APOE* genotype groups (Table 2). Heterozygous and homozygous *APOE* E4 carriers had thicker IMT compared to non-carriers (P = 0.05, Table 2). Serum levels of total cholesterol ($P = 1.75 \times 10^{-7}$)

Table 2. Characteristics per APOE genotype

	***APOE* E4**						
Characteristics	**0 (n=1342)**		**1 (n=699)**		**2 (n=102)**		**P-value**
Age (years)	49.0	14.9	49.3	14.4	48.8	13.7	0.85
Gender (% male)	42.9		43.9		46.2		0.76
Body mass index (kg/m2)	27.0	4.6	26.8	4.7	27.5	4.8	0.40
Education	3.24	0.05	3.08	0.07	3.00	0.18	0.07
IMT(mm)	0.81	0.21	0.84	0.21	0.83	0.17	**0.05**
Systolic blood pressure (mmHg/cm)	140.2	20.4	141.0	21.0	141.44	18.25	0.42
Diastoic blood pressure (mmHg/cm)	80.0	10.4	80.3	10.3	82.24	9.67	0.13
Fasting glucose (mmol/l)	4.62	1.03	4.56	0.87	4.62	0.88	0.35
Serum cholesterol (mmol/l)	5.49	1.07	5.70	1.10	5.87	1.22	**1.75E-07**
Serum Triglycerides (mmol/l)	1.30	0.75	1.41	0.82	1.62	0.95	**4.80E-05**
Serum HDL (mmol/l)	1.30	0.36	1.25	0.35	1.21	0.34	**4.81E-05**

Values presented are means and standard deviations or percentages

Table 3. Cardiovascular factors and cognitive function

	Cholesterol		Triglycerides		HDL		IMT		SBP		DBP	
Cognitive domain	**beta**	**se**	**beta**	**se**	**beta**	**se**	**beta**	**se**	**beta**	**se**	**beta**	**se**
Individual tests												
AVLT Short-term memory	0.05	0.0	-0.03	0.0	0.17	0.1	-0.40	0.2	-0.23	0.1	-0.33	0.2
AVLT Learning	0.19	0.1	0.17	0.2	0.47	0.5	-3.52	1.1 **	-1.33	0.6	-0.77	0.7
AVLT Delay	0.08	0.1	0.11	0.1	0.07	0.2	-0.65	0.4	0.06	0.2	-0.17	0.3
AVLT Recognition	0.09	0.0	0.10	0.1 *	-0.13	0.1	-1.01	0.3 **	0.17	0.2	0.59	0.2
WAIS Verbal fluency	0.44	0.3	0.07	0.4	0.36	0.9	-10.03	2.3 ***	-4.47	1.3	-0.53	1.5
TMT	0.04	0.0	-0.01	0.0	-0.01	0.1	0.00	0.2	-0.15	0.1	-0.28	0.2
Stroop	0.00	0.0	0.00	0.0	-0.03	0.0	0.03	0.1	0.19	0.0 ****	0.01	0.1
WAIS Block design	0.12	0.2	-0.05	0.3	0.03	0.7	-1.94	1.8	-0.83	1.0	-1.26	1.2
Compound scores												
Memory performance	0.03	0.0	0.02	0.0	0.03	0.0	-0.32	0.1 **	-0.05	0.1	-0.02	0.1
Executive function	0.00	0.0	0.00	0.0	0.04	0.0	-0.21	0.1	-0.20	0.0 **	0.06	0.1
Overall cognitive function	0.01	0.0	0.01	0.0	0.03	0.0	-0.26	0.1 **	-0.10	0.0	0.00	0.0

Betas and se for SBP and DBP were multiplied by 100; P values adjusted for age, sex, inbreeding, education, and family relationship; * P value < 0.05; ** P value < 0.001; *** P value < 0.0001; **** P value < 0.00001.

and triglycerides ($P = 4.80\times10^{-5}$) significantly increased and serum HDL levels significantly decreased ($P = 4.81\times10^{-5}$) with an increasing number of *APOE* E4 alleles

We studied the relationship between cardiovascular factors and cognitive function (Table 3). Serum levels of triglycerides were significantly associated with AVLT recognition (P = 0.04). There was significant and consistent evidence for association between IMT and multiple cognitive domains (AVLT learning, P = 0.01; AVLT recognition, P = 0.01; WAIS verbal fluency, P = 0.0001; memory compound score, P = 0.01; and over-all cognitive function compound score, P = 0.01). Systolic blood pressure was significantly associated with the Stroop test (P = 0.00001) and executive function compound score (P = 0.01). Adjustment for *APOE* status had little influence on the relationship between vascular risk factors and cognitive function.

Table 4 presents the effect of the *APOE* E4 allele on cognitive tests. There was a borderline significant association between *APOE* E4 and AVLT learning (P = 0.07), which became significant (P = 0.05) when adjusting for cardiovascular factors. Test scores generally showed a non-significant trend of poorer performance with an increasing number of E4 alleles. Adjusting for cardiovascular factors had little influence on these results.

When studying cognitive function, there was significant evidence for interaction between *APOE* E4 and age. The interaction term of age and *APOE* E4 was significant for AVLT short-term memory ($P_{interaction} = 0.01$), AVLT learning ($P_{interaction} = 0.05$), and memory compound score ($P_{interaction} = 0.01$), while for AVLT delayed recall ($P_{interaction} = 0.09$) and AVLT recognition ($P_{interaction} = 0.07$), the evidence was borderline significant.

When stratifying the data by age (Table 5), the E4 allele was significantly associated with poorer memory performance in those over 50 years of age (AVLT short-term memory P =

Table 4. Effect of APOE genotype on cognitive tests

Cognitive domain	APOE*E4							
	0 (n=1342)		1 (n=699)		2 (n=102)			
	mean	se	mean	se	mean	se	P1	P2
Individual tests								
AVLT Short-term memory	4.3	0.05	4.3	0.07	4.1	0.17	0.23	0.19
AVLT Learning	33.0	0.25	32.8	0.35	30.9	0.99	0.07	**0.05**
AVLT Delay	7.5	0.08	7.5	0.11	6.8	0.29	0.19	0.19
AVLT Recognition	27.8	0.06	27.8	0.09	27.7	0.20	0.68	0.96
WAIS Verbal fluency	61.5	0.51	61.0	0.71	60.8	1.83	0.80	0.98
TMT	2.7	0.03	2.7	0.04	2.7	0.10	0.64	0.64
Stroop	1.7	0.01	1.7	0.02	1.7	0.07	0.31	0.31
WAIS Block design	27.6	0.41	27.5	0.58	27.0	1.52	0.84	0.91
Compound scores								
Memory performance	0.00	0.02	-0.01	0.03	-0.16	0.08	0.14	0.15
Executive function	-0.01	0.02	-0.04	0.03	-0.02	0.08	0.71	0.92
Overall cognitive function	-0.01	0.02	-0.02	0.03	-0.10	0.07	0.23	0.31

P1: adjusted for age, sex, inbreeding, education, DART, and family relationship; P2: additionally adjusted for total cholesterol, triglycerides, IMT, and systolic and diastolic blood pressure.

Table 5. Effect of APOE genotype on cognitive tests by age category

Cognitive domain	0 E4		1 E4		2 E4		P1	P2
	mean	se	mean	se	mean	se		
<=50 years								
AVLT Short-term memory	5.0	0.1	5.0	0.1	4.8	0.2	0.64	0.59
AVLT Learning	37.2	0.3	37.6	0.4	35.5	1.3	0.72	0.89
AVLT Delay	8.7	0.1	8.9	0.1	8.2	0.4	0.63	0.99
AVLT Recognition	28.6	0.1	28.8	0.1	28.5	0.2	0.22	0.10
WAIS Verbal fluency	69.2	0.6	68.3	0.9	65.4	2.6	0.55	0.57
TMT	2.5	0.0	2.5	0.1	2.5	0.1	0.34	0.60
Stroop	1.6	0.0	1.6	0.0	1.7	0.1	0.99	0.47
WAIS Block design	35.6	0.6	35.1	0.8	33.5	2.4	0.97	0.73
Memory performance	0.4	0.0	0.4	0.0	0.3	0.1	0.43	0.54
Executive function	0.3	0.0	0.3	0.0	0.2	0.1	0.74	0.75
Overall cognitive function	0.4	0.0	0.4	0.0	0.3	0.1	0.66	0.83
>50 years								
AVLT Short-term memory	3.5	0.1	3.5	0.1	3.3	0.2	**0.01**	**0.01**
AVLT Learning	28.3	0.3	27.6	0.4	26.3	1.1	**0.001**	**0.003**
AVLT Delay	6.1	0.1	6.1	0.2	5.4	0.3	**0.01**	**0.04**
AVLT Recognition	27.0	0.1	26.7	0.2	26.9	0.3	0.07	0.18
WAIS Verbal fluency	52.7	0.7	52.9	1.0	56.2	2.4	0.99	0.57
TMT	2.9	0.0	2.9	0.1	2.9	0.2	0.56	0.70
Stroop	1.9	0.0	1.9	0.0	1.8	0.0	0.45	0.24
WAIS Block design	18.7	0.4	19.2	0.5	20.4	1.4	0.55	0.36
Memory performance	-0.4	0.0	-0.5	0.0	-0.6	0.1	**0.001**	**0.01**
Executive function	-0.4	0.0	-0.4	0.0	-0.2	0.1	0.66	0.77
Overall cognitive function	-0.4	0.0	-0.5	0.0	-0.5	0.1	**0.01**	0.07

0.01, AVLT learning P = 0.001, AVLT delayed recall P = 0.01 and memory compound score P = 0.001). Adjusting for cardiovascular factors had little influence on these effects (Table 5). In younger subjects (≤ 50 years of age), none of the tests were significantly associated to cognitive function (Table 5).

To illustrate the age-specific effect of *APOE* on memory performance, we plotted the smoothed distribution of test scores across age (Figure 1). Figure 1A shows that *APOE* has little influence on AVLT short-term memory. Only after the age of 65 years does some effect of the genotype become apparent. *APOE* genotype seems to have the most pronounced effect on AVLT learning. The effect starts around age 40 years (Figure 1B). The effects on AVLT delayed recall and memory compound score are less pronounced, but there is a trend towards poorer cognitive performance with increasing number of *APOE* E4 alleles (Figure 1C and 1D).

Because a considerable number of people could not complete the TMT-B test (N = 141), we investigated the distribution of these missing scores. The E4 allele was significantly associated with the proportion of people who could not complete the TMT-B test (none E4 = 5.4%, one E4 = 8.5%, and two E4 = 9.2%, P = 0.02). This effect was more pronounced in women (P = 0.004).

DISCUSSION

In this study, we found a significant association between the *APOE* E4 allele and reduced memory performance in persons aged 50 years and older. This effect is independent of the effect of *APOE* on cardiovascular factors. In our analyses of cognitive function there is significant evidence for interaction between *APOE* E4 and age. The effect of *APOE* E4 increases significantly with age, particularly in terms of learning ability. As expected *APOE* E4 was strongly related to lipid levels and atherosclerosis, while serum levels of triglycerides, blood pressure and atherosclerosis were significantly associated to cognitive function. Additional adjustment for *APOE* status had little influence on the relationship between vascular risk factors and cognitive function.

The extensive meta-analysis of all studies conducted in the period 1993-2004 showed that *APOE* E4 was significantly related to reduced global cognitive functioning, episodic memory and executive function in a dose-dependent way, whereas no significant effects were seen for primary memory[3]. In contrast, in the present study we see a consistent and significant association to memory. Our findings are in agreement with a recent prospective, population-based study in 5804 subjects aged 70-80 years. That study showed that E4 carriers had significantly poorer performance in immediate and delayed recall at baseline as well as greater decline during the 3.2 years of follow up[23]. The effect of *APOE* was less pronounced on attention and processing cognitive domains. Another study of 611 elderly clergymen showed that the *APOE* E4 allele had a pronounced influence on declines in episodic memory[24]. This study also used

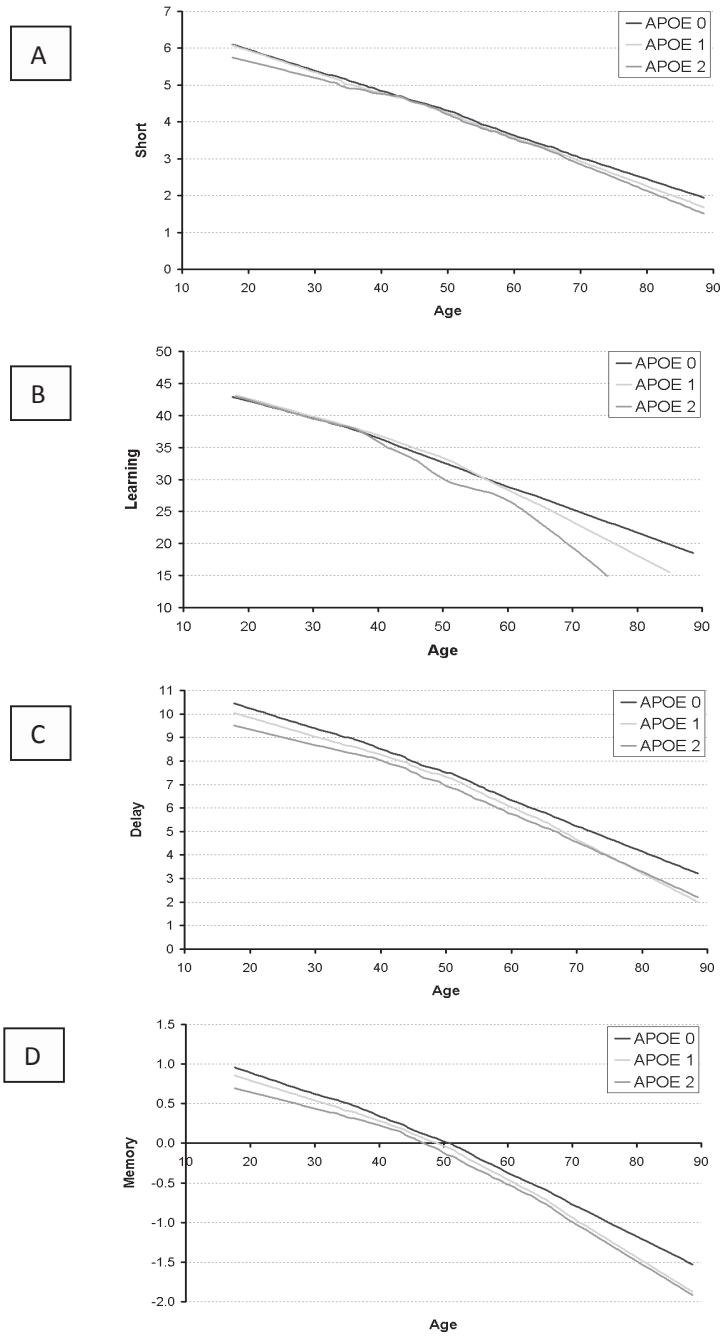


Figure 1. Age specific effect of the *APOE* E4 allele on memory performance. The distribution of test scores (y-axis) was smoothed across age (x-axis) using the LOESS smoother implemented in the software package SigmaPlot version 8.02[22]. A: AVLT short-term memory; B: AVLT learning; C: AVLT delayed recall; D: memory compound score.

a compound score, including word list memory, recall, recognition, immediate and delayed recall, which is comparable to our compound score for memory. A number of smaller studies found a relation between *APOE* E4 and memory performance[25,26]. Furthermore, a family-based study of relatives of AD patients showed an effect of *APOE* E4 on memory in those not yet affected[27]. Finally, episodic memory loss is a key characteristic of AD[28-30] and several epidemiological studies found that measures of delayed recall and learning are predictive of the risk for developing dementia[31-33].

Most studies on *APOE* and cognition in humans have focused on the elderly. Animal studies, however, have demonstrated that apolipoprotein E has a role in early brain development[4]. In our study, the effect of *APOE* was not significant in people younger than 50 years of age. Of interest, *APOE* genotype showed some evidence for an early effect on learning ability. The effect starts in early middle age, at around 40 years.

Cardiovascular factors may potentially be an intermediate feature explaining part of the association between *APOE* and cognitive function. As expected, we observed a strong association between *APOE* and serum levels of total cholesterol, triglycerides, and HDL. Systolic blood pressure and the presence of atherosclerosis as measure by IMT were significantly and consistently associated with multiple cognitive domains. Although the relationship between blood pressure and Alzheimer's disease is only observed in prospective studies[34,35], also other studies have found a strong relationship between cognitive function to blood pressure[36] and atherosclerosis[37,38]. In line with the studies on Alzheimer's disease that suggest the effect of *APOE* on the risk of disease is determined primarily by the effect on lipid metabolism with the brain, the association between *APOE* and memory performance remained significant after adjusting for serum levels of total cholesterol and triglycerides, IMT, and systolic and diastolic blood pressure. Also other estimates for the relation between *APOE* E4 and cognitive function did not change when adjusting for vascular pathology. This indicates that the effect of *APOE* on cognitive functioning is not likely determined by the effect of *APOE* on cardiovascular factors. At the same time, the additional adjustment for *APOE* status had little influence on the relationship between vascular risk factors and cognitive function. The finding implies that measuring *APOE* will not be clinically relevant for preventive strategies targeting the relationship between vascular risk factors and cognitive function.

In summary, *APOE* E4 is associated with poorer memory performance in older people. The effect of *APOE* E4 increases significantly with age and is independent of vascular pathology. The effect of *APOE* E4 on learning ability starts as early as the age of 40 years. In light of the commercial test recently made available for *APOE* genotyping, our findings suggest that those who take the test should be informed not only about the risk of AD but also about the effect of *APOE* genotype on cognitive function and vascular pathology. Whether or not the test is clinically useful remains to be determined in further studies[2]. Our findings clearly show that independent of the test outcome, management of vascular problems will be crucial for maintenance of cognitive function.

REFERENCE

1. Sleegers K, Roks G, Theuns J, Aulchenko YS, Rademakers R, Cruts M, van Gool WA, Van Broeckhoven C, Heutink P, Oostra BA, *et al.* (2004) Familial clustering and genetic risk for dementia in a genetically isolated Dutch population. Brain 127:1641-1649
2. Patterson C, Feightner JW, Garcia A, Hsiung GY, MacKnight C, Sadovnick AD (2008) Diagnosis and treatment of dementia: 1. Risk assessment and primary prevention of Alzheimer disease. Cmaj 178: 548-556
3. Small BJ, Rosnick CB, Fratiglioni L, Backman L (2004) Apolipoprotein E and cognitive performance: a meta-analysis. Psychol Aging 19:592-600
4. Kitamura HW, Hamanaka H, Watanabe M, Wada K, Yamazaki C, Fujita SC, Manabe T, Nukina N (2004) Age-dependent enhancement of hippocampal long-term potentiation in knock-in mice expressing human apolipoprotein E4 instead of mouse apolipoprotein E. Neurosci Lett 369:173-178
5. Aulchenko YS, Heutink P, Mackay I, Bertoli-Avella AM, Pullen J, Vaessen N, Rademaker TA, Sandkuijl LA, Cardon L, Oostra B, *et al.* (2004) Linkage disequilibrium in young genetically isolated Dutch population. Eur J Hum Genet 12:527-534
6. Gozna ER, Marble AE, Shaw A, Holland JG (1974) Age-related changes in the mechanics of the aorta and pulmonary artery of man. J Appl Physiol 36:407-411
7. Nikol S, Isner JM, Pickering JG, Kearney M, Leclerc G, Weir L (1992) Expression of transforming growth factor-beta 1 is increased in human vascular restenosis lesions. J Clin Invest 90:1582-1592
8. Ohji T, Urano H, Shirahata A, Yamagishi M, Higashi K, Gotoh S, Karasaki Y (1995) Transforming growth factor beta 1 and beta 2 induce down-modulation of thrombomodulin in human umbilical vein endothelial cells. Thromb Haemost 73:812-818
9. Saan R, Deelman B (1986) De 15-Woordentests A en B. (Een voorlopige bandleiding) [Internal publication]. Groningen: Section neuropsychology, Academic Hospital Groningen.
10. Reitan RM (1955) The relation of the Trail Making Test to organic brain damage. Journal of Consulting Psychology 19:393-394
11. Hammes J (1978) Stroop Kleur-woord Test: Dutch Manual. Swets and Zeitlinger BV: Lisse, The Netherlands
12. Wechsler D (2000) Wechsler adult intelligence scale 3rd (WAIS-III): test Manual (Dutch version). New York: Psychological Corporation
13. Estevez-Gonzalez A, Kulisevsky J, Boltes A, Otermin P, Garcia-Sanchez C (2003) Rey verbal learning test is a useful tool for differential diagnosis in the preclinical phase of Alzheimer's disease: comparison with mild cognitive impairment and normal aging. Int J Geriatr Psychiatry 18:1021-1028
14. Schmand B, Lindeboom J, Van Harskamp F (1992) De Nederlandse Leestest voor Volwassenen. [The Dutch adult reading test] Lisse: Swets and Zeitlinger.
15. Prins ND, Den Heijer T, Hofman A, Koudstaal PJ, Jolles J, Clarke R, Breteler MM (2002) Homocysteine and cognitive function in the elderly: the Rotterdam Scan Study. Neurology 59:1375-1380
16. Van der Elst W, van Boxtel MP, van Breukelen GJ, Jolles J (2005) Rey's verbal learning test: normative data for 1855 healthy participants aged 24-81 years and the influence of age, sex, education, and mode of presentation. J Int Neuropsychol Soc 11:290-302
17. Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16:1215
18. Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361-372

19. Almasy L, Blangero J (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet 62:1198-1211
20. Boichard D (2002) PEDIG: a FORTRAN package for pedigree analysis studied for large populations. Proceeding of the 7th World Congress of Genet Appl Livest Prod, Montpellier, France 28-13
21. Huberty C (1994) Applied Discriminant Analysis. Wiley, New York
22. Dagum E, Luati A (2001) A study of the asymmetric and symmetric weights of Kernel smoothers and their spectral properties. Estadistica J Interamerican Statist Inst 53:215-258
23. Packard CJ, Westendorp RG, Stott DJ, Caslake MJ, Murray HM, Shepherd J, Blauw GJ, Murphy MB, Bollen EL, Buckley BM, *et al.* (2007) Association between apolipoprotein E4 and cognitive decline in elderly adults. J Am Geriatr Soc 55:1777-1785
24. Wilson RS, Schneider JA, Barnes LL, Beckett LA, Aggarwal NT, Cochran EJ, Berry-Kravis E, Bach J, Fox JH, Evans DA, *et al.* (2002) The apolipoprotein E epsilon 4 allele and decline in different cognitive systems during a 6-year period. Arch Neurol 59:1154-1160
25. Bondi MW, Salmon DP, Galasko D, Thomas RG, Thal LJ (1999) Neuropsychological function and apolipoprotein E genotype in the preclinical detection of Alzheimer's disease. Psychol Aging 14: 295-303
26. Wehling E, Lundervold AJ, Standnes B, Gjerstad L, Reinvang I (2007) *APOE* status and its association to learning and memory performance in middle aged and older Norwegians seeking assessment for memory deficits. Behav Brain Funct 3:57
27. Lee JH, Flaquer A, Stern Y, Tycko B, Mayeux R (2004) Genetic influences on memory performance in familial Alzheimer disease. Neurology 62:414-421
28. Smith JD (2002) Apolipoproteins and aging: emerging mechanisms. Ageing Res Rev 1:345-365
29. Huang Y (2006) Apolipoprotein E and Alzheimer disease. Neurology 66:S79-85
30. Weisgraber KH, Mahley RW (1996) Human apolipoprotein E: the Alzheimer's disease connection. Faseb J 10:1485-1494
31. Tierney MC, Yao C, Kiss A, McDowell I (2005) Neuropsychological tests accurately predict incident Alzheimer disease after 5 and 10 years. Neurology 64:1853-1859
32. Albert MS, Moss MB, Tanzi R, Jones K (2001) Preclinical prediction of AD using neuropsychological tests. J Int Neuropsychol Soc 7:631-639
33. Grober E, Kawas C (1997) Learning and retention in preclinical and early Alzheimer's disease. Psychol Aging 12:183-188
34. Muller M, Tang MX, Schupf N, Manly JJ, Mayeux R, Luchsinger JA (2007) Metabolic syndrome and dementia risk in a multiethnic elderly cohort. Dement Geriatr Cogn Disord 24:185-192
35. Qiu C, von Strauss E, Fastbom J, Winblad B, Fratiglioni L (2003) Low blood pressure and risk of dementia in the Kungsholmen project: a 6-year follow-up study. Arch Neurol 60:223-228
36. Knecht S, Wersching H, Lohmann H, Bruchmann M, Duning T, Dziewas R, Berger K, Ringelstein EB (2008) High-normal blood pressure is associated with poor cognitive performance. Hypertension 51:663-668
37. Chang EH, Rigotti A, Huerta PT (2007) Age-related influence of the HDL receptor SR-BI on synaptic plasticity and cognition. Neurobiol Aging
38. Komulainen P, Kivipelto M, Lakka TA, Hassinen M, Helkala EL, Patja K, Nissinen A, Rauramaa R (2007) Carotid intima-media thickness and cognitive function in elderly women: a population-based study. Neuroepidemiology 28:207-213

Chapter 6

A Study of the *SORL1* Gene in Alzheimer's Disease and Cognitive Function

ABSTRACT

Several studies have investigated the role of the neuronal sortilin-related receptor (*SORL1*) gene in Alzheimer's disease (AD), but findings have been inconsistent. We conducted a study of 7 single nucleotide polymorphisms (SNPs), rs668387, rs689021, rs641120, rs1699102, rs3824968, rs2282649, and rs1010159 that were associated to AD in earlier studies. We tested for association with AD and cognitive function in 6741 participants of the Rotterdam Study and in 2883 individuals from the Erasmus Rucphen Family (ERF) study. We performed meta-analyses on AD using our data together with those of previous studies in Caucasians. Further, we studied up to 76 SNPs in a 400 kb region covering the gene to evaluate the evidence of other genetic variants that may be associated with AD or cognitive function. There was no significant evidence for association between *SORL1* SNPs and incident AD patients in the Rotterdam Study. When meta-analyzing our data with those of others, six out of seven SNPs remained borderline significant. However, removing the first study reporting association from the meta-analysis resulted in non-significant odds ratios for all SNPs, suggesting that the initial study may have overestimated the effects. SNPs rs668387, rs689021, and rs641120 were borderline significantly associated with cognitive function in two independent Dutch cohorts, but in an opposite direction. Testing for association using dense SNPs in the *SORL1* gene did not reveal significant association with AD, or with cognitive function when adjusting for multiple testing. In conclusion, our data do not support that genetic variants in *SORL1* are related to risk of AD.

INTRODUCTION

Two clusters of single nucleotide polymorphisms (SNPs) in the gene encoding the sortilin-related receptor, low-density lipoprotein receptor class A repeat (*SORL1*) have been associated with Alzheimer's disease (AD)[1]. These two clusters in the 3' and 5' end of the gene include seven SNPs, rs668387, rs689021, rs641120, rs1699102, rs3824968, rs2282649 and rs1010159 in the *SORL1* gene. To date, there have been 8 studies studying various populations aiming to replicate this finding but so far the evidence is not conclusive[2-7]. An online meta-analysis of all studies suggests significant but small effects with summary odds ratios (ORs) ranging from 1.03 to 1.14 for the risk alleles and 0.91 to 0.94 for the protective alleles (http://www.alzgene.org). In the present study, we tested these 7 SNPs for association with AD and cognitive function in 6741 participants of the Rotterdam Study, and tested the association with cognitive function in 2883 individuals from the Erasmus Rucphen Family study. Further, we performed a meta-analysis on AD using our data together with those of previous studies in Caucasians. Finally, we studied a set of dense SNPs covering the *SORL1* gene in both populations to evaluate the presence of other genetic variants that may be associated with AD or cognitive function.

METHODS

Rotterdam Study

The Rotterdam Study is a population-based prospective study of 7,983 subjects aged 55 years and older residing in Ommoord, a suburb of Rotterdam, that aims to assess the occurrence and determinants of chronic diseases[8,9]. The population is an outbred population, predominantly of Dutch origin. At baseline (1990-1993) and during three follow-up rounds (1993-1995, 1999-2000, 2001-2002) participants were invited to visit the research center for a clinical examination and had their blood drawn. The Medical Ethics Committee of the Erasmus Medical Center approved the study protocol, and all participants provided written informed consent.

Erasmus Rucphen Family Study

The Erasmus Rucphen Family (ERF) study is part of the Genetic Research in Isolated Population (GRIP) program. ERF is a family-based study that includes inhabitants of a genetically isolated community in the south-western area of the Netherlands [10]. The study population essentially consists of one extended family of descendants from 20 related couples that lived in the isolate between 1850 and 1900 and had at least 6 children baptized in the community church. The isolate now includes approximately 20,000 inhabitants. Genealogical relationships of the inhabitants are well documented from the mid-18th century to the present.

All data were collected between 2002 and 2005. The population shows increased linkage disequilibrium[10] and decreased genetic complexity[11], compared to outbred populations. The Medical Ethics Committee of the Erasmus Medical Center Rotterdam approved the study and informed consent was obtained from all participants.

Data Collection

In the Rotterdam Study, persons suffering from AD at baseline, or persons who did not have their blood drawn, were excluded. This left a total of 6741 persons for the current analyses (Figure 1). Follow-up for Alzheimer's disease was conducted according to a previously described protocol[12]. This included two-step assessments for Alzheimer's disease during the follow-up rounds and continuous monitoring through the GP and hospital records throughout the follow-up period. Follow-up for AD was complete until January 1st 2005, during which 490 persons developed AD. During the third follow-up round, a subset of 2583 non-demented participants (Figure 1) from the Rotterdam Study underwent neuropsychological testing, including the Stroop Color and Word Test[13], the Letter-Digit Substitution Task[14], a verbal fluency task[15], and a 15-item verbal learning test[16].

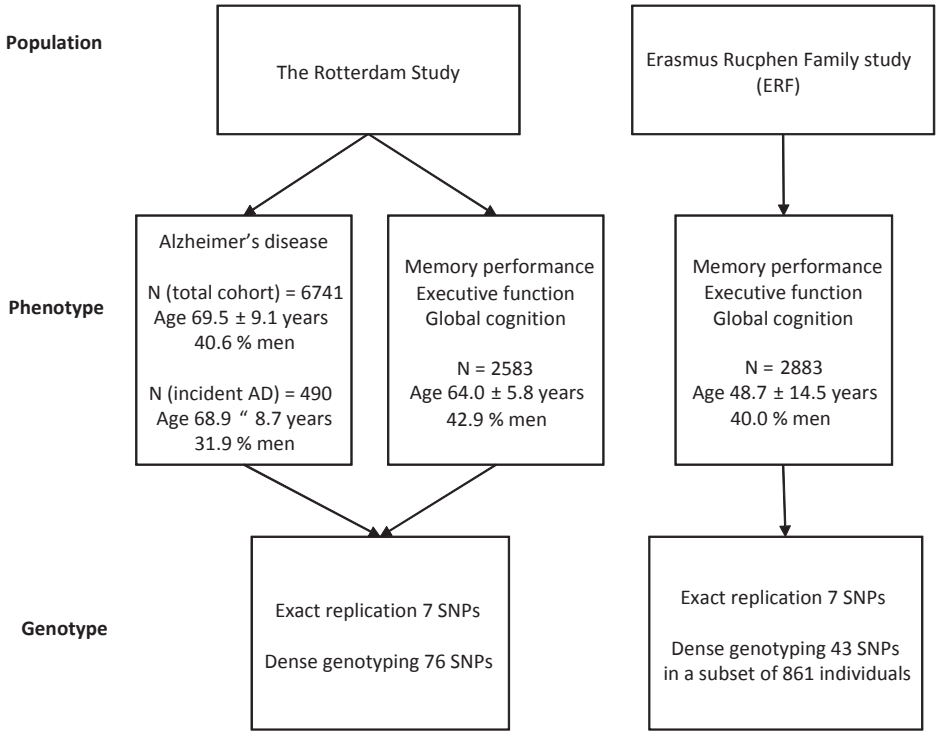


Figure 1. The design of the study

In the ERF study, a total of 2883 participants (Figure 1) underwent neuropsychological testing, including the Trail Making Test (TMT)[17], the Stroop Color and Word Test[13,18], a verbal fluency test, a 15-item verbal learning test, and the block design subtest of the Weschler Adult Intelligence Scale (WAIS)[19].

We computed compound scores for memory performance, which consisted of immediate recall and delayed recall for the Rotterdam Study, and of immediate recall, learning, delayed recall and recognition for the ERF study. The composite score for, executive function consisted of the Stroop Color and Word Test, the Letter-Digit Substitution Task, and verbal fluency for the Rotterdam Study, and of the Stroop Color and Word Test, TMT and verbal fluency for the ERF study. Finally, global cognitive function was computed by taking the average of scores for all cognitive tests. The computation of these compound scores have been described previously[6,20].

Genotyping

Seven SNPs in the *SORL1* gene were genotyped (rs668387, rs689021, rs641120, rs1699102, rs3824968, rs2282649, and rs1010159) using TaqMan allelic discrimination genotyping assays (Applied Biosystems). Primer and probe sequences for these SNPs are available from the manufacturer. For dense-typing, 76 SNPs from the 120.7- 121.1 Mb region of chromosome 11 covering the *SORL1* gene were available from the 550K array of the Illumina Infinium whole-genome genotyping assay for the Rotterdam Study. For the ERF study, a subset of the 76 SNPs (43 SNPs), was available in the same genomic region from the 318K array of the Illumina Infinium whole-genome genotyping assay (HumanHap300-2). This array was genotyped in a subset of 861 ERF individuals. Microarray-based genotyping according to the manufacturer's instructions was performed at the Leiden Genome Technology Center of the Leiden University Medical Center.

Statistical Analysis

To test the association with incident AD in the Rotterdam Study, we estimated the hazard ratios (HRs), adjusted for age and sex, using Cox Proportional-Hazards models. In the Cox model, those homozygous for the rare alleles and heterozygous were compared with those homozygous for the common alleles.

For the meta-analyses we searched PubMed (www.ncbi.nlm.nih.gov), Huge Navigator[21] and AlzGene database (www.alzgene.org) for genetic case-control association studies on the *SORL1* gene and AD published before the 1st of September 2008 using the keywords "Alzheimer" and "*SORL1*". Twelve genetic association studies were found[1-7,22]. We excluded the study of AD in adults with Down syndrome[2] and the studies of Webster[5] and Meng[4], which both used the same TGEN data set as described in Reiman's study[2]. Because the Rotterdam Study and the ERF studies are of Caucasian origin, we excluded a study based on the Chinese population[22] and considered only the Caucasian subgroups from the selected papers[1-3,6,7].

Per-allele ORs were derived for all included data sets, using the common allele as reference. Summary ORs were calculated using random effects meta-analyses with 95% confidence intervals as outlined by DerSimonian and Laird[23]. The degree of heterogeneity between the study results was assessed using the I^2 statistic[24]. For the SNPs out of Hardy-Weinberg Equilibrium (HWE) in individual studies, we calculated the ORs including and excluding the populations not in HWE. The meta-analyses were conducted in R (www.r-project.org), using the R library rmeta version 2.14.

To test the association between each SNP and cognitive function, we used a general linear model adjusted for age and sex, where the SNP genotype was coded as 0, 1, or 2 , representing the number of the minor alleles. For the ERF study, phase information was derived based on known genealogic relationships. Haplotypes consisting of all 7 selected SNPs were inferred using the software package SimWalk2[25]. To reduce computational time, extended ERF pedigrees were split into 20-bit pedigrees using PedCut version 1.18[6] prior to haplotype analysis. Mendelian errors and ambiguous haplotypes were removed. Inferred haplotypes were considered as a fixed factor in a general linear model. We used the genomic control method[26,27] to adjust for family relationships between ERF participants and to adjust for population substructure in the Rotterdam Study. The inflation factor, lambda, was estimated separately for each trait based on the Illumina Infinium Assay (318K in the ERF and 550K in the Rotterdam Study).

To adjust for the multiple testing of correlated markers in the selected region, we derived the empirical distribution of the test statistic after 10,000 region-wide permutations. Association tests, permutation, and genomic control analysis were performed using functions implemented in the R library GenABEL version 1.3-5[22].

RESULTS

The study population characteristics are illustrated in Figure 1. Among the 6741 participants (mean age 69.5±9.1, 40.6% men) of the Rotterdam Study, 490 (mean age 68.9±8.7, 31.9% men) developed AD. ERF participants were on average younger (mean age 48.7±14.5, 40.0% men).

There was no significant evidence for association between *SORL1* SNPs and incident AD in the Rotterdam Study. Genotypic HRs ranged from 0.91 to 1.16 (Table 1). When pooling our data with that from previous published studies[1-3,6,7], six out of seven SNPs remained borderline significantly associated with AD (Figure 2). The six SNPs were rs668387 ($OR_{T\ vs.\ C}$ = 0.93, 95% CI: 0.88-0.99), rs689021 ($OR_{A\ vs.\ G}$ = 0.93, 95% CI: 0.87-1.00), rs1699102 ($OR_{C\ vs.\ T}$ = 1.07, 95% CI: 1.00-1.14), rs3824968 ($OR_{T\ vs.\ A}$ = 1.11, 95% CI: 1.01-1.23), rs2282649 ($OR_{T\ vs.\ C}$ = 1.09, 95% CI: 1.02-1.17), and rs1010159 ($OR_{C\ vs.\ T}$ = 1.06, 95% CI: 1.00-1.12). One SNP, rs3824968, showed significant heterogeneity (I^2 = 29.4, df = 11, P = 0.01). After excluding the North European

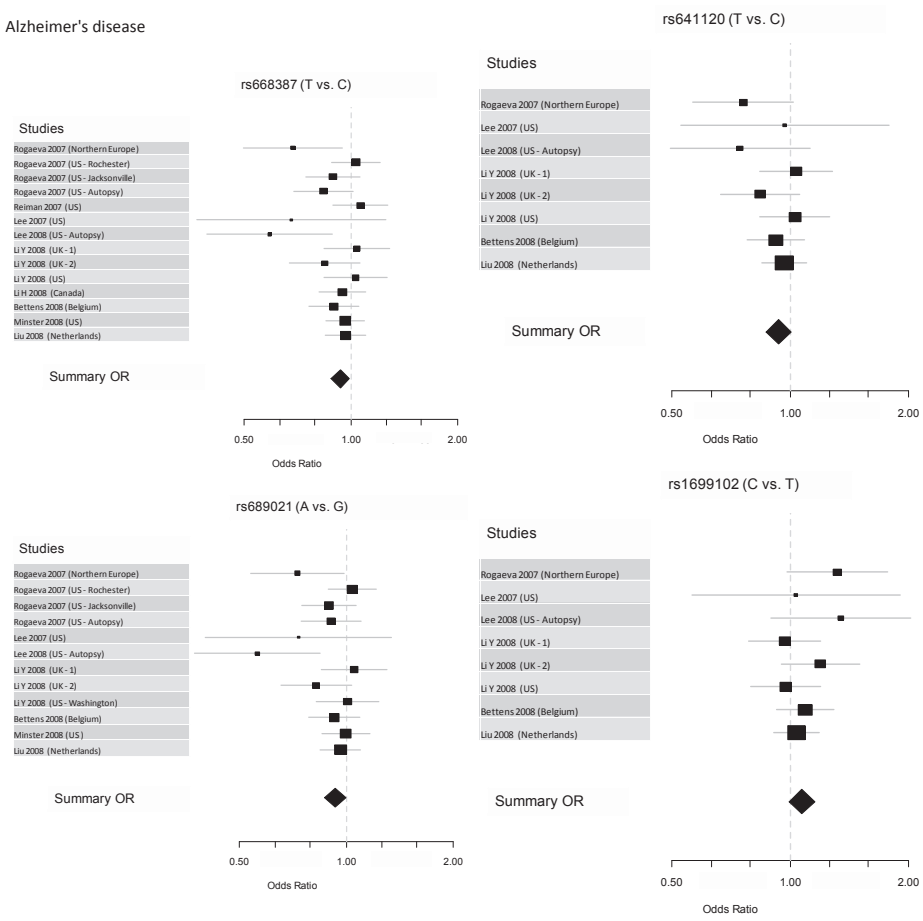


Figure 2. Forest plot of meta analysis of 7 SNPs in the SORL1 gene in association with Alzheimer's disease

sample from the original study[1], no significant heterogeneity was detected ($I^2 = 15.7$, df = 10, P = 0.11), resulting in non-significant summary ORs for all seven SNPs. Excluding the studies in which the genotype was out of HWE, did not alter these results.

Table 1. Association of SORL1 SNPs with Alzheimer's disease in the Rotterdam Study

				95% CI		Homozygous		95% CI	
SNP	Reference	Heterozygous	HR	Lower	Upper	minor allele	HR	Lower	Upper
rs668387	CC	CT	0.95	0.75	1.20	TT	1.03	0.81	1.32
rs689021	GG	AG	0.93	0.76	1.14	AA	0.95	0.75	1.22
rs641120	CC	CT	0.91	0.74	1.11	TT	0.97	0.75	1.24
rs1699102	TT	CT	0.97	0.80	1.17	CC	1.00	0.72	1.36
rs3824968	AA	AT	0.96	0.80	1.16	TT	1.16	0.85	1.57
rs2282649	CC	CT	0.95	0.78	1.15	TT	1.01	0.74	1.38
rs1010159	TT	CT	0.98	0.81	1.19	CC	1.12	0.85	1.49

Reference genotype consists of homozygous major alleles

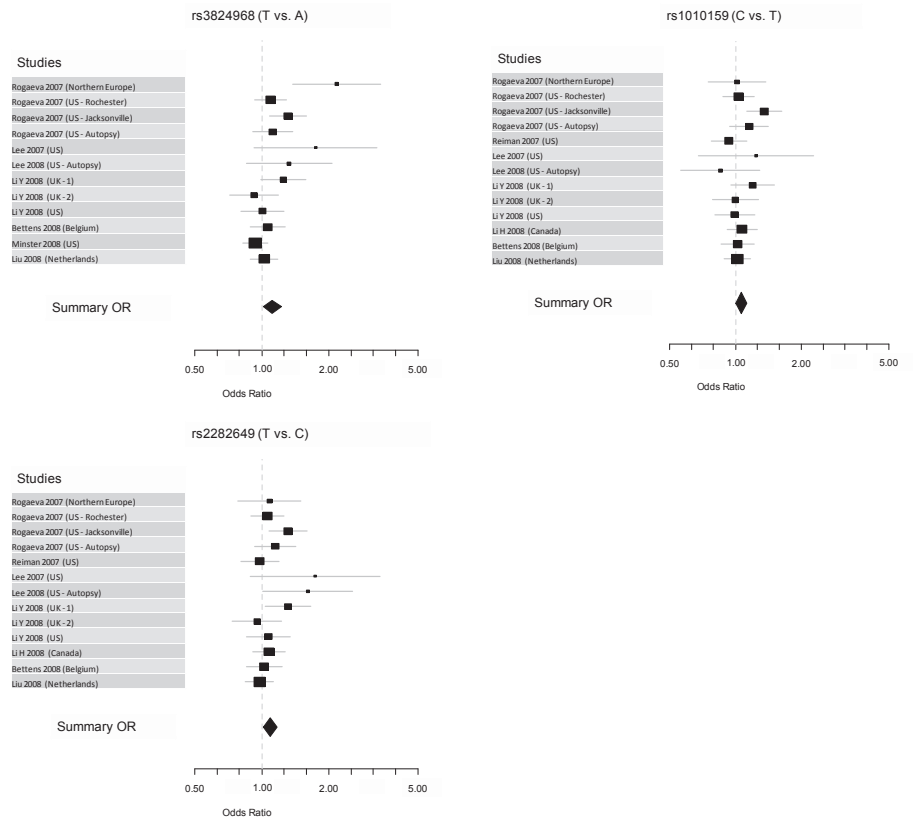


Figure 2. (continued)

Next, we tested the association between the 7 SNPs and cognitive function in both Dutch populations. SNPs rs668387, rs689021, and rs641120 were borderline significantly associated with memory compound scores in both the Rotterdam Study and the ERF study (Table 2). However, the allelic effect was in an opposite direction: the minor alleles of rs668387,

Table 2. Association of SORL1 SNPs with cognitive function in two Dutch populations

	Rotterdam Study (N = 2583)								Erasmus Rucphen Family Study (N = 2883)							
			Memory		Executive		Global				Memory		Executive		Global	
SNP	MA	MAF	beta	P	beta	P	beta	P	MA	MAF	beta	P	beta	P	beta	P
rs668387	T	0.46	-0.05	**0.04**	-0.01	0.63	-0.03	0.16	T	0.41	0.04	**0.02**	0.03	0.08	0.03	**0.03**
rs689021	A	0.47	-0.05	**0.03**	-0.01	0.53	-0.03	0.12	A	0.41	0.04	**0.02**	0.03	0.08	0.03	**0.03**
rs641120	T	0.47	-0.05	**0.04**	-0.01	0.67	-0.03	0.16	T	0.41	0.04	**0.02**	0.02	0.16	0.03	**0.04**
rs1699102	C	0.32	0.00	0.89	0.02	0.31	0.01	0.54	C	0.35	-0.02	0.34	0.03	0.06	0.00	0.83
rs3824968	T	0.32	0.03	0.31	0.02	0.25	0.03	0.17	T	0.34	-0.02	0.44	0.04	**0.02**	0.01	0.57
rs2282649	T	0.32	0.01	0.69	0.03	0.25	0.02	0.37	T	0.32	-0.01	0.67	0.05	**0.003**	0.02	0.28
rs1010159	C	0.35	0.03	0.22	0.02	0.27	0.03	0.13	C	0.37	-0.01	0.47	0.04	**0.03**	0.01	0.61

MA, minor allele; MAF, minor allele frequency; P values smaller than 0.05 are indicated in bold.

Table 3. Haplotype association with cognitive function in ERF population

			Memory		Executive		Global	
Haplotype	**N**	**%**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
CGCTTCT	1420	29.35	0.00	0.87	-0.04	0.75	-0.01	0.72
TATTTCT	1068	22.08	0.04	0.82	0.00	0.70	0.02	0.68
CGCCATC	789	16.31	-0.04	0.90	0.01	0.72	-0.02	0.73
TATCATC	435	8.99	0.04	0.83	0.05	0.68	0.05	0.68
CGCTTCC	140	2.89	-0.08	0.79	-0.05	0.63	-0.06	0.62
TATCTCT	87	1.80	-0.01	0.80	-0.08	0.66	-0.06	0.64
Others	899	18.58	0.00	0.77	-0.05	0.67	-0.03	0.62
P value			0.15		0.13		0.18	

Ambiguous haplotypes were eliminated; All haplotypes with less than 1% frequency were pooled.

rs689021, and rs641120 (T, A, T) were associated with lower scores on memory tests in the Rotterdam Study, but with higher scores in the ERF Study (Table 2). Three other SNPs rs3824968, rs2282649, and rs1010159 were significantly associated with executive function compound scores in the ERF Study. This association is, however, inconsistent with the findings of the meta-analysis on AD, because the alleles associated with better executive function (T for rs3824968, T, for rs2282649, and C for rs1010159) were found to be the risk alleles of AD in the meta-analysis. In addition, we did not observe a significant association between haplotypes and any cognitive trait (Table 3).

Finally, we analyzed the dense SNP series covering the *SORL1* region (figure 3). Although there is some marginal evidence for association, none of the tested SNPs was significantly associated with neither AD, nor with cognitive function after adjusting for multiple testing.

DISCUSSION

In this study we failed to replicate previous findings on the association between *SORL1* gene and AD. Our meta-analysis showed that removing the data of the first published study resulted in non-significant summary ORs for all SNPs. The association for rs668387, rs689021, and rs641120 with memory was in an opposite direction in two Dutch populations. The association for rs3824968, rs2282649, and rs1010159 with executive function observed in the ERF study was contradictive with the results from the meta-analysis of AD. When adjusting for multiple testing no other tested SNPs within or flanking the *SORL1* gene were significantly associated with AD, or with cognitive function.

Our results are in line with recent large studies with similar sample sizes[3,6,7] and a recent study of *SORL1* using genome wide association data yielded weak evidence for association of SORL to AD for four SNPs (rs2101756, rs11218313, rs626885, and rs7131432) that is far from genome wide significant[5]. The current study by itself may have limited power to detect an OR below 1.24 (alpha = 0.05; power = 0.80). However, also the findings of our meta-analysis are

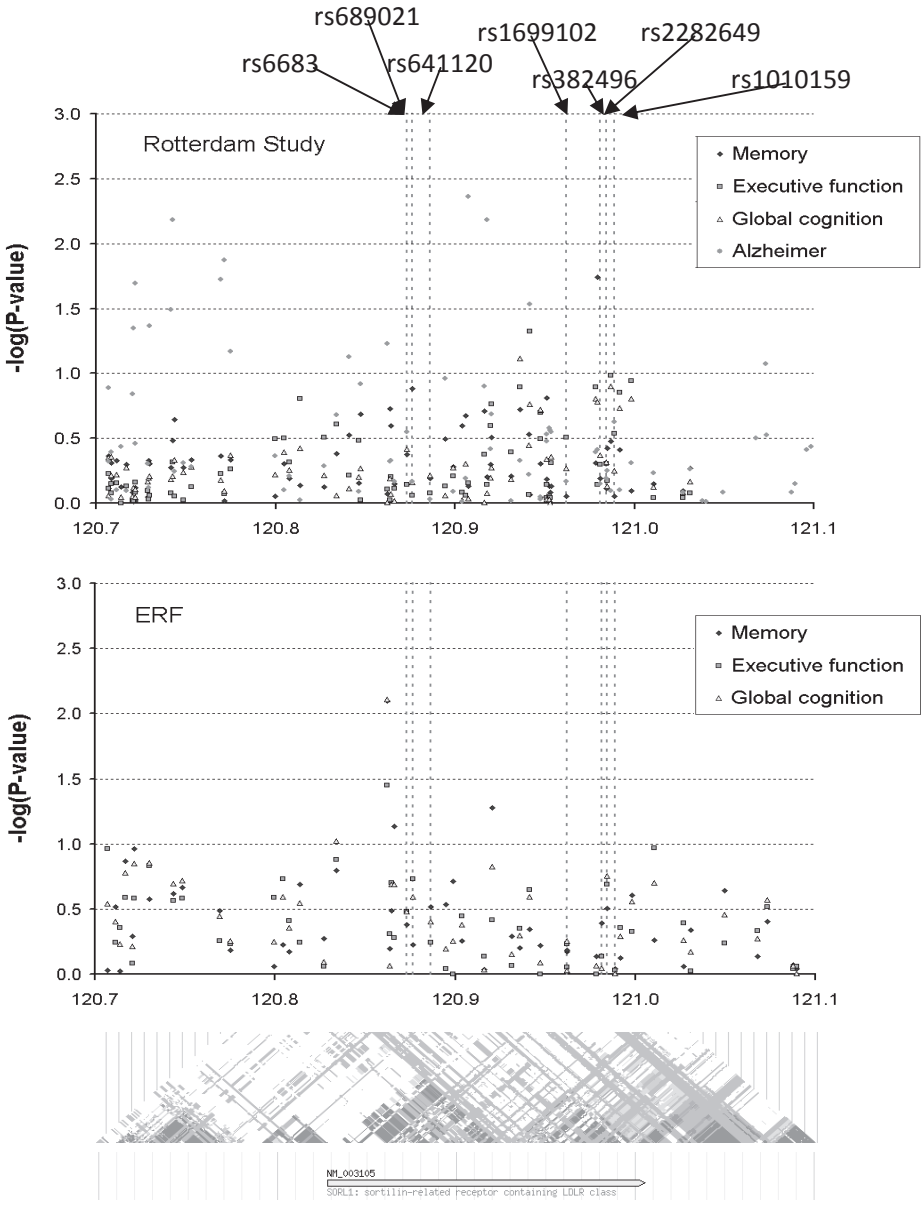


Figure 3. Dense SNPs flanking the SORL1 gene in association with AD and cognitive function. The minus base-10 logarithm transformed P-values (Y-axis) are plotted against the physical positions of each SNP (X-axis). Dashed vertical lines indicate the 7 studied SNPs, from left to right are rs668387, rs689021, rs641120, rs1699102, rs3824968, rs2282649, rs1010159. Haplotype blocks according to HapMap Europeans and known genes in the region are aligned below.

far from convincing. When adding our data to the previous ones in Caucasian populations, the ORs became less significant and move towards the null hypothesis of no association. The

evidence of association was mainly driven by the first three studies from the same research group[1-3], where the odds ratios exceeded the genetic effect estimated by meta-analysis of the remaining studies. Removing the original study from the meta-analysis, results in non-significant ORs for all SNPs. Notably, even after excluding the original study, the remaining 6,000 cases and 11,000 controls will still have over 95% power to detect a small OR of 1.1 at a 5% false positive rate. Thus, the association of seven *SORL1* with AD when meta-analyzing all data available on Caucasians is most likely partly explained by the 'winner's curse' phenomenon, i.e., remarkably strong effects the initial study of Rogaeva *et al.*[1,28,29].

In addition to AD, we also considered cognitive function as an alternative outcome. Using a cognitive compound score that is based on a z-score transformation of a number of well-validated tests, we previously detected the association with *APOE* in non-demented elderly[6], providing empirical support for this method in finding genetic associations with AD. In the current study, the findings on cognitive function were, however, inconsistent in two Dutch populations and contradictive to the findings of our meta-analysis. Thus, also these analysis do not support a role of *SORL1* in AD.

The reported association of *SORL1* with AD may also be explained by causal SNPs that are in linkage disequilibrium with the 7 SNPs targeted initially reported. To evaluate this possibility we further studied a dense set of SNPs within a 400 kb region covering the *SORL1* gene. These SNPs capture all haplotype blocks within or overlapping the gene. Although there is some marginal evidence for association, when adjusted for multiple testing no SNP was significantly associated. Other studies also have genotyped a dense series of SNPs to evaluate additional genetic variants in the *SORL1* gene[1,2,4-6]. None of these studies could show significant evidence of association with AD when adjusted for multiple testing.

In conclusion, we did not find a significant effect of *SORL1* on the risk of AD when adjusting for multiple testing. Clearly, the marginal evidence for association may be interpreted as suggestive evidence for association. However, the recent successes in genome wide association studies have shown that adjusting for multiple testing and consistent replication of findings are the keys to success. So far, the findings on SORL are far from genome wide significant. Deep sequencing of the *SORL1* gene may reveal rare variants that explain the two clusters of SNPs associated to AD observed in some populations.

Web resources

R library GenABEL for SNP association analysis and PedCut for breaking large pedigrees, http://mga.bionet.nsc.ru/soft/index.html

Simwalk2 for haplotype inference, http://watson.hgen.pitt.edu/register.

The AlzGene Database for the meta-analysis of the *SORL1* gene, http://www.alzgene.org.

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/.

Alzgene database, http://www.alzgene.org

Huge navigator, http://www.hugenavigator.net

REFERENCES

1. Lee JH, Cheng R, Schupf N, Manly J, Lantigua R, Stern Y, Rogaeva E, Wakutani Y, Farrer L, St George-Hyslop P, *et al.* (2007) The association between genetic variants in *SORL1* and Alzheimer disease in an urban, multiethnic, community-based cohort. Arch Neurol 64:501-506
2. Coon KD, Myers AJ, Craig DW, Webster JA, Pearson JV, Lince DH, Zismann VL, Beach TG, Leung D, Bryden L, *et al.* (2007) A high-density whole-genome association study reveals that *APOE* is the major susceptibility gene for sporadic late-onset Alzheimer's disease. J Clin Psychiatry 68:613-618
3. Bettens K, Brouwers N, Engelborghs S, De Deyn PP, Van Broeckhoven C, Sleegers K (2008) *SORL1* is genetically associated with increased risk for late-onset Alzheimer disease in the Belgian population. Hum Mutat 29:769-770
4. Meng Y, Lee JH, Cheng R, St George-Hyslop P, Mayeux R, Farrer LA (2007) Association between *SORL1* and Alzheimer's disease in a genome-wide study. Neuroreport 18:1761-1764
5. Webster JA, Myers AJ, Pearson JV, Craig DW, Hu-Lince D, Coon KD, Zismann VL, Beach T, Leung D, Bryden L, *et al.* (2008) *SORL1* as an Alzheimer's disease predisposition gene? Neurodegener Dis 5: 60-64
6. Axenovich TI, Zorkoltseva IV, Liu F, Kirichenko AV, Aulchenko YS (2008) Breaking loops in large complex pedigrees. Hum Hered 65:57-65
7. Minster RL, DeKosky ST, Kamboh MI (2008) No association of *SORL1* SNPs with Alzheimer's disease. Neurosci Lett 440:190-192
8. Hofman A, Grobbee DE, de Jong PT, van den Ouweland FA (1991) Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. Eur J Epidemiol 7:403-422
9. Gonzalez-Zuloeta Ladd AM, Liu F, Houben MP, Arias Vasquez A, Siemes C, Janssens AC, Coebergh JW, Hofman A, Janssen JA, Stricker BH, *et al.* (2007) IGF-1 CA repeat variant and breast cancer risk in postmenopausal women. Eur J Cancer 43:1718-1722
10. Aulchenko YS, Heutink P, Mackay I, Bertoli-Avella AM, Pullen J, Vaessen N, Rademaker TA, Sandkuijl LA, Cardon L, Oostra B, *et al.* (2004) Linkage disequilibrium in young genetically isolated Dutch population. Eur J Hum Genet 12:527-534
11. Pardo LM, MacKay I, Oostra B, van Duijn CM, Aulchenko YS (2005) The effect of genetic drift in a young genetically isolated population. Ann Hum Genet 69:288-295
12. Ott A, Breteler MM, van Harskamp F, Stijnen T, Hofman A (1998) Incidence and risk of dementia. The Rotterdam Study. Am J Epidemiol 147:574-580
13. Houx PJ, Jolles J, Vreeling FW (1993) Stroop interference: aging effects assessed with the Stroop Color-Word Test. Exp Aging Res 19:209-224
14. Lezak MD (1984) Neuropsychological assessment in behavioral toxicology--developing techniques and interpretative issues. Scand J Work Environ Health 10 Suppl 1:25-29
15. Welsh KA, Butters N, Mohs RC, Beekly D, Edland S, Fillenbaum G, Heyman A (1994) The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part V. A normative study of the neuropsychological battery. Neurology 44:609-614
16. Bleecker ML, Bolla-Wilson K, Agnew J, Meyers DA (1988) Age-related sex differences in verbal memory. J Clin Psychol 44:403-411
17. Reitan RM (1955) The relation of the Trail Making Test to organic brain damage. Journal of Consulting Psychology 19:393-394
18. Hammes J (1978) Stroop Kleur-woord Test: Dutch Manual. Swets and Zeitlinger BV: Lisse, The Netherlands

19. Wechsler D (2000) Wechsler adult intelligence scale 3rd (WAIS-III): test Manual (Dutch version). New York: Psychological Corporation
20. Ikram MA, Vrooman HA, Vernooij MW, Heijer TD, Hofman A, Niessen WJ, van der Lugt A, Koudstaal PJ, Breteler MM (2008) Brain tissue volumes in relation to cognitive function and risk of dementia. Neurobiol Aging
21. Li H, Wetten S, Li L, St Jean PL, Upmanyu R, Surh L, Hosford D, Barnes MR, Briley JD, Borrie M, *et al.* (2008) Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. Arch Neurol 65:45-53
22. Amin N, van Duijn CM, Aulchenko YS (2007) A genomic background based method for association analysis in related individuals. PLoS ONE 2:e1274
23. DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. Control Clin Trials 7:177-188
24. Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. Bmj 327:557-560
25. Sobel E, Lange K (1996) Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. Am J Hum Genet 58:1323-1337
26. Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55:997-1004
27. Steinthorsdottir V, Thorleifsson G, Reynisdottir I, Benediktsson R, Jonsdottir T, Walters GB, Styrkarsdottir U, Gretarsdottir S, Emilsson V, Ghosh S, *et al.* (2007) A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. Nat Genet 39:770-775
28. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG (2001) Replication validity of genetic association studies. Nat Genet 29:306-309
29. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet 33:177-182

Chapter 7

The *GAB2* Gene and the Risk of Alzheimer's Disease: Replication and Meta-Analysis

ABSTRACT

In a recent genome-wide association study, the *GAB2*-gene has been suggested to modify the risk of late-onset Alzheimer's disease (AD) among *APOE* E4 carriers. However, replication data are scarce and inconsistent. In a population-based cohort study (N = 5507; age>55) with 443 incident AD cases we sought to replicate the association between rs4945261 and AD. Because we used high-density genotyping of *GAB2*, we also investigated several other polymorphisms within and around this gene. Furthermore, we performed a meta-analysis with all previously published studies. We found that rs4945261 was associated with AD among *APOE* E4 carriers (P = 0.02), but not among non-carriers (P = 0.26). Fifteen of the 20 remaining polymorphisms within *GAB2* and several polymorphisms in the 250 kbp region surrounding *GAB2* were also associated with AD among *APOE* E4 carriers and only one among non-carriers. For rs2373115 meta-analysis with published studies yielded an odds-ratio of 1.58 (1.17-2.14) with $P = 3.0\times10^{-3}$ among *APOE*ε4-carriers and 1.09 (0.97-1.23) with P = 0.16 among non-carriers. For rs4945261 the pooled odds-ratio was 1.75 (1.21-2.55) with $P = 3.0\times10^{-3}$ among *APOE*ε4-carriers and 1.20 (1.01-1.41) with P = 0.03 among non-carriers. We found the *GAB2* gene to be associated with AD. When taken together with published data, our data suggest *GAB2* to modify the risk of AD in *APOE* E4-carriers.

INTRODUCTION

The quest for finding genes that are related to Alzheimer's disease (AD) has turned towards using high throughput genotyping analysis, in which thousands of polymorphisms can be studied concomitantly. Several genome-wide association studies have been conducted for AD so far and most found a consistent and strong hit in or around the Apolipoprotein E (*APOE*) gene, confirming earlier linkage and candidate gene studies[1-3]. Reiman *et al.* carried out a genome-wide association analysis after stratification by the *APOE*ε4-allele[4]. They showed that the gene encoding GRB-associated binding protein 2 (*GAB2*) was associated with AD in *APOE* E4-carriers, but not in non-carriers. Replication data are scarce and inconsistent: Chapuis *et al.*[5] failed to find an association between *GAB2* with AD in three independent study samples, either in *APOE*ε4-carriers or non-carriers. Li *et al.*[3] also did not confirm this finding in their genome-wide association study, but did not stratify by *APOE* E4 status. However, in a Belgian sample Sleegers *et al.*[6] did find an association, and finally Miyashita *et al.*[7] could not replicate this association in a Japanese population.

In the population-based Rotterdam Study, we sought to replicate the association between *GAB2* and AD. We investigated rs4945261, which was one of the SNPs in the original report. Moreover, because we used high-density genotyping, we also investigated other SNPs within *GAB2* and within a 250 kbp region surrounding the gene. Finally, we conducted a meta-analysis using random-effects pooling for rs4945261 and rs2373115 based on data from previously published studies[3-6]. We searched PubMed using the key-words *GAB2*, Alzheimer's disease, and dementia. We also sought through reference lists of previous papers and queried the AlzGene database (www.alzgene.org). We did not find any other studies and restricted our current meta-analysis to Caucasian populations.

METHODS

Study population

The Rotterdam Study is a prospective population-based cohort study of 7,983 Caucasian participants (aged 55 years and over) living in Ommoord, a district of Rotterdam, The Netherlands[8]. The study investigates determinants of chronic diseases in the elderly, including AD. Persons gave written informed consent to participate and the study was approved by the institutional medical-ethics committee. At baseline (1990-1993) participants were interviewed and underwent physical examination and blood sampling. For the present study, only persons who were non-demented at baseline were eligible (n=7,046). No overlap exists between our study population and the Dutch sample reported on in the original report[4].

Genotyping

Only participants with proper quality DNA-samples (n=6,449) were considered for genotyping using the version 3 Illumina-Infinium-II HumanHap550SNP chip-array as part of a large project on genetics of complex diseases. Genotyping procedures were followed according to manufacturer's protocol[9]. After quality control 5,974 persons remained with proper genotyped data. No population stratification was present in this sample[9]. For the current report we extracted data on rs4945261. We also extracted other SNPs that were located in *GAB2* (total of 20 SNPs) or within a 250-kbp region surrounding the gene (94 SNPs), and were in Hardy-Weinberg equilibrium ($P > 0.001$). In order to pool our data with all previous studies we imputed allelic data for rs2373115 based on the local linkage disequilibrium structure using the MACH-imputation software (http://www.sph.umich.edu/csg/abecasis/MACH). The quality of imputation was 99.8%.

APOE genotyping was performed on coded samples without knowledge of the other measurements as described elsewhere[10], and was unavailable in 467 persons mostly due to technical reasons leaving a total of 5,507 persons available in the current analysis.

Ascertainment of incident AD

The diagnosis of incident AD was made following a three-step protocol[11]. At baseline (1990-1993) and during three follow-up visits (1993-1994, 1997-1999, 2002-2004) two brief tests of cognition (MMSE and Geriatric Mental State schedule (GMS)) were used to screen all subjects. Screen-positives (MMSE score<26 or GMS>0) underwent the Cambridge examination for mental disorders of the elderly (Camdex). Persons suspected of having dementia were examined by a neuropsychologist if additional neuropsychological testing was required for diagnosis. When available, imaging data were used. In addition, the total cohort was continuously monitored for incident AD through computerized linkage between the study database and digitalized medical records from general practitioners and the Regional Institute for Outpatient Mental Health Care. The diagnosis of AD was made in accordance with internationally accepted criteria by a panel of a neurologist, neuropsychologist and research physician. Follow-up was complete until January 1st, 2005.

Statistical analysis

We used the allelic χ^2-test with one degree-of-freedom to investigate the association between rs4945261 and AD, before and after stratification by *APOE* E4 status. We used a threshold of $p=0.05$ for statistical significance, because our aim was to replicate previous genome-wide findings. A similar approach was used when investigating the remaining 20 SNPs within *GAB2*. However, in this instance we also assessed multiple testing by calculating false-discovery rates[12] and by permutation testing. Subsequently, we analyzed the 94 SNPs in the region surrounding *GAB2*. We further explored these associations by using Cox'-proportional hazards models and adjusting for age, sex, and time-to-event. Finally, we tested for interaction by adding an interaction term SNP*E4 status to the models.

RESULTS

Table 1 shows the characteristics of the study population. *APOE* E4 carriers were younger than non-carriers and as expected had a shorter follow-up time with a larger percentage of incident AD cases. Table 2 shows the association of SNPs in *GAB2* with AD. Rs4945261, the only SNP similar to the discovery report, showed a significant association with AD (p=0.02). Stratification by the *APOE* E4 allele showed that the association was particularly marked in carriers of the *APOE* E4 allele. In non-carriers no significant association with AD was seen.

Table 1. Characteristics of the study population

	Total	E4 non-carriers	E4 carriers	p-value
N	5,507	3,958	1,549	
Age	68.9 (8.7)	69.1 (8.8)	68.4 (8.4)	0.01*
Women	3,215 (58%)	2,322 (59%)	893 (58%)	0.67**
Mean follow-up	9.24 (3.21)	9.31 (3.17)	9.07 (3.32)	<0.01†
Incident AD cases	443 (8%)	249 (6%)	194 (13%)	<0.01†

Values are numbers (percentages) or means (standard deviation). The p-values are for the difference between *APOE* E4 carriers and non-carriers; * sex-adjusted; ** age-adjusted; † age and sex-adjusted.

Table 2. Association between polymorphisms in GAB2 and Alzheimer's disease

SNP	Position	MA	MAF	RA	Overall		E4 non-carriers		E4 carriers	
					p-value	OR (95% CI)	p-value	OR (95% CI)	p-value	OR (95% CI)
rs2450135	77605643	A	0.03	G	0.06	0.73 (0.52-1.02)	0.16	0.73 (0.47-1.13)	0.17	0.69 (0.41-1.17)
rs1318241	77608440	A	0.14	G	0.02	1.28 (1.04-1.59)	0.27	1.16 (0.89-1.53)	0.02	1.51 (1.08-2.11)
rs2450129	77618033	G	0.15	A	0.02	1.29 (1.04-1.59)	0.27	1.17 (0.89-1.53)	0.01	1.52 (1.09-2.13)
rs731600	77640781	G	0.15	A	0.02	1.28 (1.04-1.58)	0.27	1.17 (0.89-1.53)	0.02	1.50 (1.07-2.08)
rs1893447	77650830	G	0.15	A	0.02	1.29 (1.05-1.59)	0.24	1.18 (0.90-1.54)	0.01	1.52 (1.09-2.12)
rs2511175	77652729	G	0.15	A	0.02	1.28 (1.04-1.58)	0.26	1.17 (0.89-1.54)	0.02	1.49 (1.07-2.07)
rs1981405	77653856	A	0.11	G	0.04	1.29 (1.01-1.64)	0.22	1.22 (0.89-1.67)	0.06	1.44 (0.99-2.09)
rs7927923	77657062	G	0.19	A	0.41	1.08 (0.90-1.29)	0.97	1.00 (0.80-1.26)	0.18	1.21 (0.92-1.61)
rs4945261	**77667908**	**A**	**0.15**	**G**	**0.02**	**1.28 (1.04-1.58)**	**0.26**	**1.17 (0.89-1.54)**	**0.02**	**1.49 (1.07-2.07)**
rs7107174	77675584	A	0.14	G	0.02	1.28 (1.03-1.58)	0.26	1.17 (0.89-1.54)	0.02	1.48 (1.06-2.07)
rs4944196	77686379	A	0.15	G	0.02	1.29 (1.04-1.59)	0.19	1.20 (0.91-1.58)	0.03	1.45 (1.05-2.02)
rs6592772	77693211	C	0.15	A	0.02	1.29 (1.05-1.60)	0.20	1.19 (0.91-1.57)	0.02	1.49 (1.07-2.08)
rs10899469	77695961	G	0.15	A	0.02	1.29 (1.04-1.59)	0.23	1.18 (0.90-1.55)	0.02	1.50 (1.08-2.09)
rs11237451	77703107	G	0.19	A	0.29	1.10 (0.92-1.32)	0.65	1.06 (0.84-1.34)	0.22	1.19 (0.90-1.57)
rs2292572	77730512	A	0.15	C	0.01	1.32 (1.07-1.63)	0.13	1.23 (0.94-1.62)	0.01	1.51 (1.09-2.09)
rs10501426	77734770	A	0.15	G	0.01	1.31 (1.07-1.62)	0.13	1.23 (0.94-1.62)	0.02	1.48 (1.07-2.05)
rs11601726	77745687	G	0.12	A	0.95	1.01 (0.82-1.24)	0.80	0.96 (0.73-1.27)	0.72	1.06 (0.76-1.48)
rs11603112	77751139	A	0.15	G	0.01	1.34 (1.08-1.65)	0.14	1.23 (0.94-1.62)	0.01	1.55 (1.11-2.16)
rs2373115	***77768798***	***A***	***0.15***	***C***	***0.01***	***1.32 (1.07-1.63)***	***0.10***	***1.26 (0.96-1.65)***	***0.02***	***1.46 (1.06-2.01)***
rs7112234	77780118	A	0.15	G	0.01	1.33 (1.08-1.64)	0.10	1.26 (0.96-1.66)	0.02	1.47 (1.07-2.04)
rs7941639	77794607	A	0.14	G	0.03	1.26 (1.02-1.57)	0.42	1.12 (0.85-1.48)	0.02	1.53 (1.08-2.17)
rs10899496	77801479	G	0.15	A	0.01	1.34 (1.09-1.65)	0.08	1.28 (0.97-1.68)	0.02	1.48 (1.07-2.05)

Odds ratios are unadjusted and calculated using the allelic $\chi 2$ test; MA, Minor allele; MAF, Minor allele frequency; RA, Risk allele; OR, odds ratio; CI, Confidence interval; SNPs in bold were also genotyped by Reiman et al.; The SNP in italic was imputed.

Table 3. Statistics for the association between polymorphisms in GAB2 and incident Alzheimer's disease, stratified by *APOE* E4 status

A. *APOE* E4 carriers				
SNP	Unadjusted	FDR*	Adjusted FDR**	Empirical P ***
rs11603112	0.0109	0.0290	0.0050	0.0672
rs1893447	0.0134	0.0290	0.0050	0.0787
rs2450129	0.0141	0.0290	0.0050	0.0821
rs2292572	0.0147	0.0290	0.0050	0.0866
rs1318241	0.0162	0.0290	0.0050	0.0918
rs10899469	0.0164	0.0290	0.0050	0.0922
rs7941639	0.0176	0.0290	0.0050	0.0941
rs731600	0.0177	0.0290	0.0050	0.0963
rs10501426	0.0184	0.0290	0.0050	0.0978
rs10899496	0.0184	0.0290	0.0050	0.0978
rs6592772	0.0185	0.0290	0.0050	0.0986
rs2511175	0.0194	0.0290	0.0050	0.1036
rs4945261	0.0194	0.0290	0.0050	0.1036
rs7112234	0.0199	0.0290	0.0050	0.1042
rs7107174	0.0207	0.0290	0.0050	0.1160
rs4944196	0.0257	0.0337	0.0058	0.1374
rs1981405	0.0574	0.0709	0.0121	0.2652
rs2450135	0.1710	0.1976	0.0341	0.6026
rs7927923	0.1788	0.1976	0.0356	0.6276
rs11237451	0.2231	0.2343	0.0440	0.6833
rs11601726	0.7215	0.7215	0.1295	0.9992

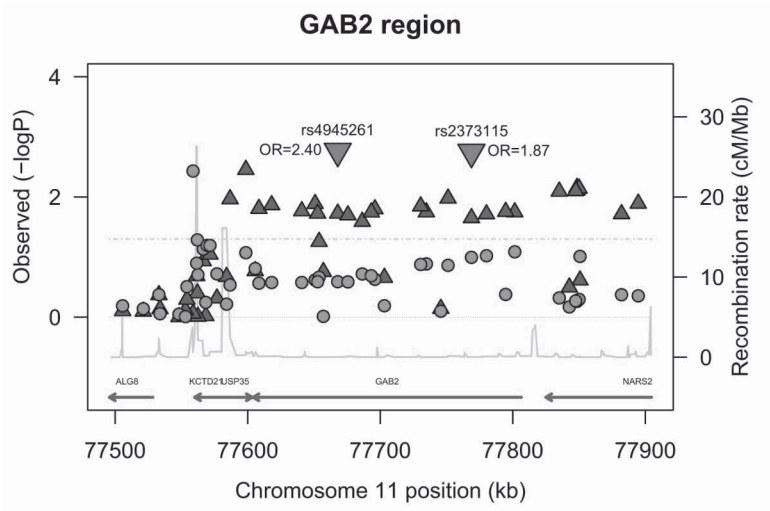


Figure 1. The association between polymorphisms surrounding the GAB2 gene and Alzheimer's disease. P-values are for the allelic χ2 test. Blue triangles indicate p-values for *APOE* E4 carriers. Red circles indicate P values for *APOE* E4 non-carriers. Purple triangles indicate P values for the meta-analysis in *APOE* E4 carriers. Pink dot-dashed line indicates the threshold for P value = 0.05. Light blue line indicates the estimated recombination rates reflecting the local linkage disequilibirum (LD) structure. Known genes are aligned along their genomic position.

Table 3 (continued)

B. *APOE* E4 non-carriers				
SNP	Unadjusted	FDR*	Adjusted FDR **	Empirical P***
rs10899496	0.0824	0.3365	0.0995	0.3366
rs7112234	0.0954	0.3365	0.0995	0.3750
rs10501426	0.1303	0.3365	0.0995	0.4728
rs2292572	0.1337	0.3365	0.0995	0.4817
rs11603112	0.1382	0.3365	0.0995	0.4883
rs2450135	0.1568	0.3365	0.0995	0.5417
rs4944196	0.1916	0.3365	0.0995	0.6258
rs6592772	0.2060	0.3365	0.0995	0.6435
rs1981405	0.2151	0.3365	0.0995	0.6605
rs10899469	0.2370	0.3365	0.0995	0.6921
rs1893447	0.2426	0.3365	0.0995	0.6965
rs7107174	0.2569	0.3365	0.0995	0.7329
rs4945261	0.2574	0.3365	0.0995	0.7330
rs2511175	0.2597	0.3365	0.0995	0.7337
rs731600	0.2664	0.3365	0.0995	0.7416
rs2450129	0.2665	0.3365	0.0995	0.7423
rs1318241	0.2724	0.3365	0.0995	0.7642
rs7941639	0.4175	0.4870	0.1448	0.9150
rs11237451	0.6476	0.7157	0.2080	0.9937
rs11601726	0.8011	0.8412	0.2452	0.9999
rs7927923	0.9743	0.9743	0.2832	1.0000

Statistics are based on the 21 genotyped SNPs in GAB2. SNPs are ordered by the unadjusted p-value; * Using the method described by Benjamini and Hochberg[12]; ** calculated using the R-package 'fdrtool'[17], which additionally adjusts for the estimated proportion of true null associations; *** obtained after 10,000 permutations

Of the 20 remaining SNPs in *GAB2* 15 also showed a significant ($P < 0.05$) association as well as the imputed SNP rs2373115 (Table2). Table 3 shows that although the P values would not survive multiple-testing correction for 21 SNPs, the probability that these findings are false-discoveries is very small. In the Figure 1 P values for all SNPs within 250 kbp of the *GAB2* gene are plotted against their respective genomic position and stratified by *APOE* E4-allele. Among *APOE* E4-carriers various SNPs that were in high linkage disequilibrium (LD) with SNPs within *GAB2* had a p value < 0.05. Non-significant SNPs were located further away from *GAB2* across recombination sites and in other LD-blocks (Figure 1). In contrast, in non-carriers only one SNP in the whole region located outside *GAB2* had a P value < 0.05.

Table 4 shows hazard-ratios, adjusted for age, sex and time-to-event. The associations among non-carriers hardly changed, but among *APOE* E4 carriers the associations attenuated slightly. Nevertheless, eleven of the 22 SNPs were still significant among *APOE* E4 carriers, whereas several others SNPs were borderline significant. Finally, the interaction term for SNP*E4 was not significant for any SNP.

Meta-analysis with published studies showed a pooled random-effects odds ratio for rs2373115 among *APOE* E4 carriers of 1.58 (95%CI 1.17-2.14) with $P = 3.0 \times 10^{-3}$. For rs4945261 the meta-analysis showed among *APOE*ε4-carriers a random-effects odds ratio of 1.75 (1.21-

Table 4. Association between polymorphisms in GAB2 and Alzheimer's disease using Cox'-proportional hazards models

SNP	Position	Hazard ratios (95% CI)	
		E4 non-carriers	E4 carriers
rs2450135	77605643	0.74 (0.50-1.11)	0.76 (0.48-1.20)
rs1318241	77608440	1.16 (0.90-1.51)	1.38 (1.00-1.90)
rs2450129	77618033	1.17 (0.90-1.52)	1.38 (1.00-1.90)
rs731600	77640781	1.17 (0.90-1.51)	1.36 (0.99-1.87)
rs1893447	77650830	1.17 (0.90-1.51)	1.37 (1.00-1.88)
rs2511175	77652729	1.17 (0.90-1.52)	1.36 (0.99-1.87)
rs1981405	77653856	1.22 (0.90-1.64)	1.32 (0.92-1.88)
rs7927923	77657062	0.99 (0.79-1.23)	1.11 (0.86-1.44)
rs4945261	**77667908**	**1.17 (0.90-1.52)**	**1.36 (0.99-1.87)**
rs7107174	77675584	1.17 (0.90-1.53)	1.35 (0.99-1.85)
rs4944196	77686379	1.20 (0.92-1.55)	1.32 (0.97-1.80)
rs6592772	77693211	1.20 (0.92-1.55)	1.37 (0.99-1.88)
rs10899469	77695961	1.18 (0.91-1.54)	1.38 (1.00-1.89)
rs11237451	77703107	1.03 (0.82-1.29)	1.10 (0.85-1.42)
rs2292572	77730512	1.21 (0.93-1.57)	1.38 (1.02-1.88)
rs10501426	77734770	1.21 (0.93-1.57)	1.36 (1.01-1.85)
rs11601726	77745687	0.97 (0.74-1.26)	1.14 (0.84-1.55)
rs11603112	77751139	1.19 (0.92-1.55)	1.41 (1.03-1.92)
rs2373115	***77768798***	***1.23 (0.95-1.61)***	***1.36 (1.00-1.85)***
rs7112234	77780118	1.23 (0.95-1.61)	1.36 (1.01-1.85)
rs7941639	77794607	1.12 (0.86-1.45)	1.44 (1.04-2.00)
rs10899496	77801479	1.24 (0.95-1.61)	1.37 (1.01-1.86)

Adjusted for age, sex, and time to event; SNPs in bold were also genotyped by Reiman et al.[4]; The SNP in italic was imputed.

2.55) with $P = 3.0\times10^{-3}$. Among non-carriers the odds ratios were 1.09 (0.97-1.23) with P = 0.16 for rs2373115 and 1.20 (1.01-1.41) with P = 0.03 for rs4945261 (Figure 2).

DISCUSSION

In this population-based cohort study we found that rs4945261 was associated with AD in persons carrying the *APOE* E4 allele. In non-carriers no significant association was found. Furthermore, we also found that several other SNPs within and around *GAB2* were associated with AD in *APOE* E4-allele carriers but not in non-carriers, though these would not survive multiple-testing correction. To our knowledge this is the first study to longitudinally investigate the association between *GAB2* and AD. The population-based design limits the possibility of selection biases often seen in case-control studies. A possible limitation is that in some cases the diagnosis of AD might have been misclassified. However, such misclassification is likely to be random and would therefore lead to an underestimation of the true effect. Another consideration is that apart from rs4945261 the other SNPs were different from

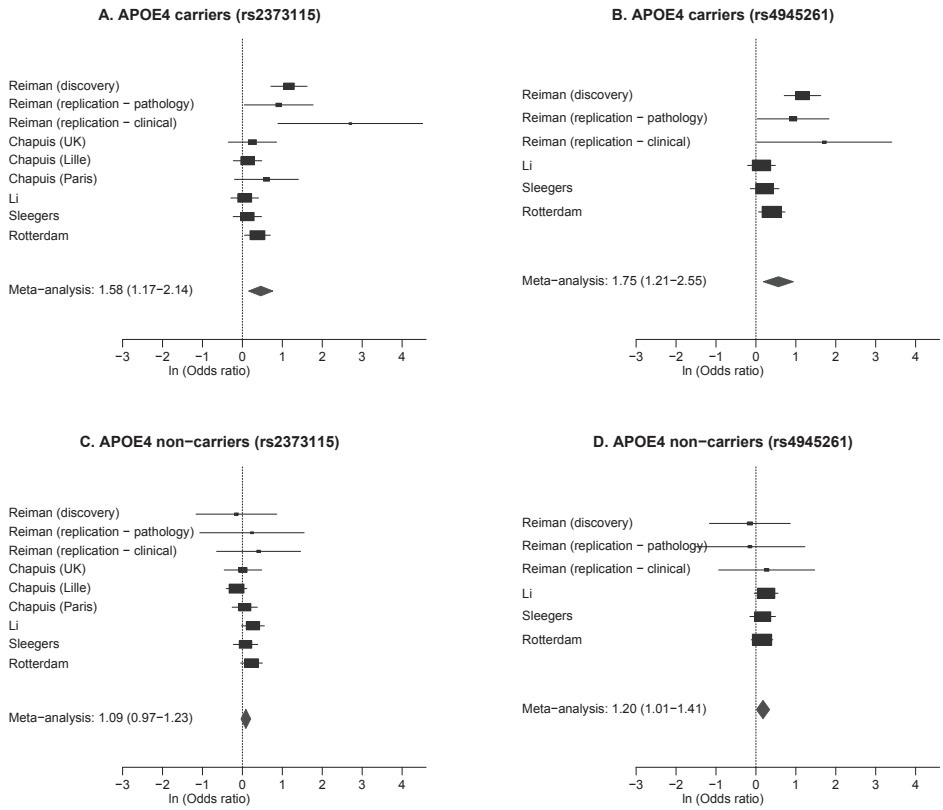


Figure 2. Meta-analysis of published studies on the association between polymorphisms in *GAB2* and Alzheimer's disease, stratified by the *APOE* E4 allele

previous reports. However, additional genotyping is unlikely to change our results given the density of SNPs we studied and the low recombination rate in *GAB2* (Figure 1). Moreover, genotyping different SNPs can be regarded as contributing to fine-mapping the *GAB2*-gene and its association with AD. Our associations would not have survived stringent multiple testing correction for 21 SNPs. However, given the strong LD between SNPs and the low prior probability of these findings being false-positive, standard multiple testing could be considered overly conservative. More importantly, the meta-analysis also points towards a positive association. Thus far, three studies have failed to replicate the initial findings[3,5,7] and only one confirmed the association[6]. In line with the initial study, we found that *GAB2* alleles were associated with AD only among *APOE* E4 carriers, and not in non-carriers. Pooling our data with previously published data showed highly significant associations with odds ratios of 1.58 and 1.75.

GAB2 is a protein involved in various pathways, some of which involve AD-related tau processing[13-15]. Indeed, Reiman *et al.* also found that *GAB2* expression was associated with protection from neurofibrillary tangle formation[13-15]. Moreover, *GAB2* is expressed together

with other potential AD-related genes[16]. However, the exact mechanism of interaction with the *APOE* gene is still unknown. Future research should focus on disentangling the exact interactive mechanism as well as high-density sequencing of *GAB2* to find the possible causative variant.

In conclusion, we found *GAB2* to be associated with AD. Together with previous data, this suggests *GAB2* as a novel gene modifying the risk of AD in *APOE* E4 carriers.

REFERENCES

1. Grupe A, Abraham R, Li Y, Rowland C, Hollingworth P, Morgan A, Jehu L, Segurado R, Stone D, Schadt E, *et al.* (2007) Evidence for novel susceptibility genes for late-onset Alzheimer's disease from a genome-wide association study of putative functional variants. Hum Mol Genet 16:865-873
2. Coon KD, Myers AJ, Craig DW, Webster JA, Pearson JV, Lince DH, Zismann VL, Beach TG, Leung D, Bryden L, *et al.* (2007) A high-density whole-genome association study reveals that *APOE* is the major susceptibility gene for sporadic late-onset Alzheimer's disease. J Clin Psychiatry 68:613-618
3. Li H, Wetten S, Li L, St Jean PL, Upmanyu R, Surh L, Hosford D, Barnes MR, Briley JD, Borrie M, *et al.* (2008) Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. Arch Neurol 65:45-53
4. Reiman EM, Webster JA, Myers AJ, Hardy J, Dunckley T, Zismann VL, Joshipura KD, Pearson JV, Hu-Lince D, Huentelman MJ, *et al.* (2007) *GAB2* alleles modify Alzheimer's risk in *APOE* epsilon4 carriers. Neuron 54:713-720
5. Chapuis J, Hannequin D, Pasquier F, Bentham P, Brice A, Leber I, Frebourg T, Deleuze JF, Cousin E, Thaker U, *et al.* (2008) Association study of the *GAB2* gene with the risk of developing Alzheimer's disease. Neurobiol Dis 30:103-106
6. Sleegers K, Bettens K, Brouwers N, Engelborghs S, van Miegroet H, De Deyn PP, Van Broeckhoven C (2008) Common variation in GRB-associated Binding Protein 2 (*GAB2*) and increased risk for Alzheimer dementia. Hum Mutat
7. Miyashita A, Arai H, Asada T, Imagawa M, Shoji M, Higuchi S, Urakami K, Toyabe S, Akazawa K, Kanazawa I, *et al.* (2008) *GAB2* is not associated with late-onset Alzheimer's disease in Japanese. Eur J Hum Genet
8. Hofman A, Breteler MM, van Duijn CM, Krestin GP, Pols HA, Stricker BH, Tiemeier H, Uitterlinden AG, Vingerling JR, Witteman JC (2007) The Rotterdam Study: objectives and design update. Eur J Epidemiol 22:819-829
9. Richards JB, Rivadeneira F, Inouye M, Pastinen TM, Soranzo N, Wilson SG, Andrew T, Falchi M, Gwilliam R, Ahmadi KR, *et al.* (2008) Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. Lancet 371:1505-1512
10. Slooter AJ, Cruts M, Kalmijn S, Hofman A, Breteler MM, Van Broeckhoven C, van Duijn CM (1998) Risk estimates of dementia by apolipoprotein E genotypes from a population-based incidence study: the Rotterdam Study. Arch Neurol 55:964-968
11. Ott A, Breteler MM, van Harskamp F, Stijnen T, Hofman A (1998) Incidence and risk of dementia. The Rotterdam Study. Am J Epidemiol 147:574-580
12. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc B 57:289-300
13. Pratt JC, Igras VE, Maeda H, Baksh S, Gelfand EW, Burakoff SJ, Neel BG, Gu H (2000) Cutting edge: *GAB2* mediates an inhibitory phosphatidylinositol 3'-kinase pathway in T cell antigen receptor signaling. J Immunol 165:4158-4163
14. Zompi S, Gu H, Colucci F (2004) The absence of Grb2-associated binder 2 (*GAB2*) does not disrupt NK cell development and functions. J Leukoc Biol 76:896-903
15. Koncz G, Bodor C, Kovesdi D, Gati R, Sarmay G (2002) BCR mediated signal transduction in immature and mature B cells. Immunol Lett 82:41-49
16. Li KC, Liu CT, Sun W, Yuan S, Yu T (2004) A system for enhancing genome-wide coexpression dynamics study. Proc Natl Acad Sci U S A 101:15561-15566
17. Strimmer K (2008) A unified approach to false discovery rate estimation. BMC Bioinformatics 9:303

Chapter 8

General Discussion

Alzheimer's disease is (AD) a heterogeneous and complex disorder and finding genes involved in its pathophysiology has been proven to be a challenge. Four genes are known to be implicated in the onset of AD, the amyloid precursor protein gene (*APP*)[1], presenilin-1 (*PSEN-1*)[2-6] and presenilin-2 (*PSEN-2*)[7,8], and the apolipoprotein E gene (*APOE*)[9-12]. After decades of genetic research, numerous genes have been studied as potential AD susceptibility genes. A large meta-analysis from the AlzGene database representing 1055 polymorphisms and 355 genes reported in the literature as at August 2006 revealed the following 13 additional potential AD-susceptibility genes: angiotensin I converting enzyme (*ACE*); cholinergic receptor nicotinic beta 2 (*CHRNB2*); cystatin C (*CST3*); estrogen receptor 1 (*ESR1*); glyceraldehyde-3-phosphate dehydrogenase spermatogenic (*GAPDHS*); insulin-degrading enzyme (*IDE*); 5,10-methylenetetrahydrofolate reductase (MTHFR); nicastrin (*NCSTN*); prion protein (*PRNP*); *PSEN-1*; transferrin (*TF*); mitochondrial transcription factor A (*TFAM*); and tumor necrosis factor (*TNF*)[13]. All of these genes are associated with relevant biological mechanisms and pathways but whether they are significant markers for AD still needs to be further elucidated. It is important to note that in the meta analysis, none of the P-values approaches the level of genome-wide significance as required from genome-wide association studies. Recent reports from individual studies reveal significant associations with the sortilin-related receptor (*SORL1*)[14] and glycine-rich protein 2–associated binding protein 2 (*GAB2*)[15] on chromosome 11, death-associated protein kinase 1 (*DAPK1*)[16] and adenosine triphosphate–binding cassette transporter 1 subfamily A (*ABCA1*) on chromosome 9[17], and low-density lipoprotein receptor–related protein 6 (*LRP6*) on chromosome 12[18]. All of these putative variants still lack of replication in large representative populations but have relevance to neuropathlogical mechanisms and pathways that may be associated with AD pathogenesis[19]. A summary of these listed genes is given in Table 1. It becomes clearer that a number of genes are no longer significant in the updated meta-analysis, such as *ESR1*, *IDE*, *NCSTN*, *TNF*.

This thesis describes an investigation of genetic susceptibility to AD and cognitive function. We discuss solutions for several theoretical and practical challenges which we encountered during the investigation. Further, we conducted our research in a genetically isolated population. We started with a whole genome screen and then zoomed into regions of interest. We further performed three candidate gene studies. Below the main findings presented in the chapters of this thesis are discussed briefly, and put into perspective of each other and of previous findings.

METHODOLIGICAL ISSUES

Linkage analysis in a genetic isolate population

Most studies described in this thesis were conducted in a recently genetically isolated population studied in the Genetic Research in Isolated Populations (GRIP) program. Advantages of this population for finding genes for complex disorders include (1) increased linkage disequi-

Table 1. Meta analysis of candidate genes associated with Alzheimer's disease in previous case-control studies

		Nbr.		Frequency		Allelic	
Gene	SNP	studies	Allele	AD	CTR	OR	95% CI
APOE		38	E4	0.38	0.14	3.68	[3.31, 4.11]
ACE	rs1799752	41	insertion	0.48	0.47	1.08	[1.00, 1.17]
CHRNB2	rs4845378	4	T	0.07	0.10	0.67	[0.50, 0.90]
CST3	rs5030707	4	C	0.19	0.16	1.23	[1.03, 1.47]
ESR1	rs2234693	16	P	0.43	0.45	1.10	[0.97, 1.24]
GAPDHS	rs4806173	4	G	0.36	0.39	0.87	[0.75, 1.00]
IDE	rs2251101	13	C	0.26	0.27	0.98	[0.89, 1.07]
MTHFR	rs1801133	24	T	0.41	0.42	1.11	[1.02, 1.21]
NCSTN	rs12239747	5	G	0.04	0.04	0.99	[0.74, 1.34]
PRNP	rs1799990	13	G	0.22	0.26	0.91	[0.83, 0.99]
PSEN1	rs165932	41	G	0.41	0.42	0.92	[0.86, 1.00]
TF	rs1049296	14	C2	0.20	0.18	1.18	[1.04, 1.33]
TFAM	rs2306604	5	G	0.41	0.46	0.82	[0.72, 0.94]
TNF	rs1800629	8	A	0.16	0.15	1.07	[0.85, 1.34]
SORL1	rs668387	15	T	0.43	0.44	0.92	[0.84, 1.00]
GAB2	rs2373115	7	T	0.15	0.18	0.79	[0.67, 0.94]
DAPK1	rs4878104	7	T	0.35	0.38	0.88	[0.82, 0.95]
ABCA1	rs2230806	13	A	0.28	0.28	1.00	[0.90, 1.12]

librium (LD) [20] and (2) reduced genetic complexity [21]. Another merit of this population is the availability of extensive genealogic information, which has been systematically collected and computerized. The latest release of the GRIP genealogic database holds information on more than 110,000 individuals across 23 generations (oldest birth year 1300 and youngest birth year 2006). All the individuals can be connected to a single, large, and complex pedigree that is characterized by multiple distant consanguineous loops. The complexity of this pedigree is illustrated in Figure 1.

For late onset or sporadic AD, it is not easy to find clusters of closely related patients even in genetically isolated populations, which is likely explained by disease alleles with incomplete penetrance. However, with profound genealogic information available, it is possible to identify clusters of distantly related patients. Because overall haplotype sharing between distantly related individuals is rare, any observed sharing becomes valuable. An example of the power of large pedigrees consisting of distantly related patients was provided by a linkage analysis of pituitary adenoma predisposition in northern Finland. Significant linkage signals were obtained with a one nine-generation pedigree containing only six patients[22]. Another successful example is the identification of a novel locus for autosomal recessive early-onset parkinsonism using 4 distantly related patients[23]. Therefore, known relationships between individuals over many generations might provide a solid base for linkage analysis, even of complex disorders, such as late onset AD.

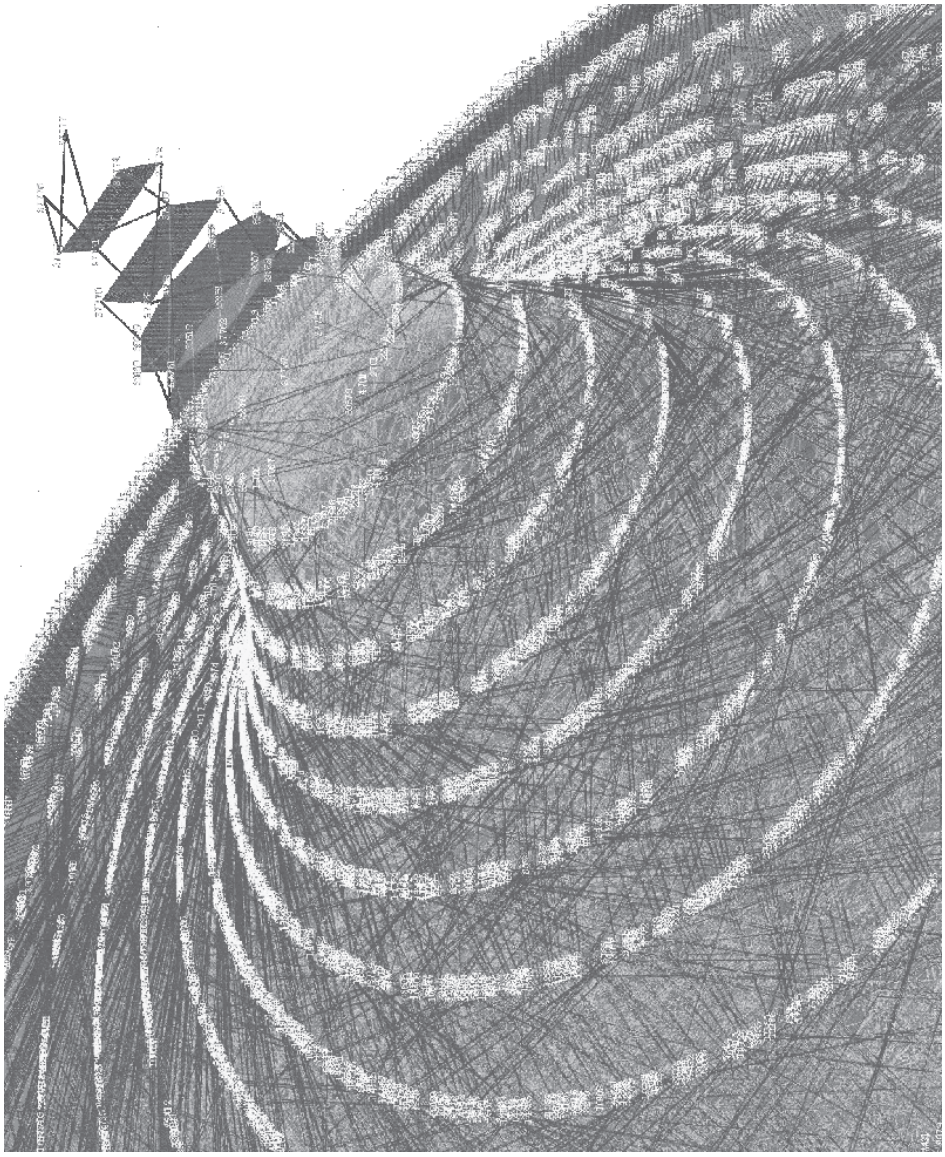


Figure 1. A pedigree consisting of subjects from the GRIP population

Analysis of large and complex pedigrees in linkage analysis is computationally complex. Currently, there are two algorithms used for computing exact multipoint likelihoods described by Elston-Stewart[24] and Lander-Green-Kruglyak[25,26]. The key computational challenge is the determination of the inheritance vector for all meiosis, whether a grandmaternal or grandpaternal allele is transmitted from parent to offspring. For analysis of large pedigrees, the Lander-Green-Kruglyak algorithm, which enumerates the probabilities of all possible states of the inheritance vector, becomes intractable because the number of possible states

increases exponentially with pedigree size. The computational complexity of the Lander-Green-Kruglyak algorithm increases linearly with the number of markers, thus making it ideal for analyzing data from a genome-wide scan in small to medium sized pedigrees. The Elston-Stewart algorithm uses the technique of 'peeling' to traverse the pedigree one nuclear family at a time and 'clipping' to trim off each nuclear family by computing the conditional probability of one of its members, based on all the information of the other members of the family. Peeling and clipping reduce the basic unit of computation in a pedigree to the nuclear family, and thus, accelerate the overall likelihood computation for a single marker. However, the 'peeling' fails in the presence of multiple consanguineous loops, as in the case of large and complex pedigrees ascertained from genetically isolated populations (**Figure 1**). Further, the complexity of the Elston-Stewart algorithm is exponential for the number of markers. Thus, this method is most suited for studying a region of interest on a chromosome containing a limited number of markers in pedigrees without loops.

Efforts have been made to extend the computational boundaries of both algorithms, such as the algorithmic improvements taking advantage of symmetries in the Lander-Green algorithm[27,28], and the technique of set-recoding and fuzzy inheritance in the Elston-Stewart algorithm[29,30]. Still, there are no exact solutions for analysis of large and complex pedigrees with large numbers of markers. For linkage analysis of larger pedigrees, approximate Markov chain Monte Carlo (MCMC) sampling is generally used[31-33]. The MCMC algorithm converges nearly to the exact solution with sufficiently sampling. However, the time required to obtain an accurate solution can be very long for MCMC samplers, and some user experience is required to determine when an MCMC run has converged. Furthermore, the recently available dense sets of SNP markers also increase the sampling time to observe sufficient number of recombination events. In addition, with dense marker sets, it is critical to consider linkage disequilibrium (LD) between nearby markers. Only one Lander-Green-Kruglyak based program, MERLIN, can handle LD between markers. The MCMC algorithm has been extended to allow for LD but this is still experimental[34]. The two most commonly used MCMC programs, MORGAN and SIMWALK2, cannot yet handle LD.

A common approach to reduce the computational complexity is to split large pedigrees into smaller, and thus computable units[35,36]. The existing pedigree-cutting methods do not specifically consider the pedigree bit-size as a parameter and thus often result in sub-pedigrees that are still too complex for the Lander-Green-Kruglyak algorithm based linkage analysis. In **chapter 3**, we present a recursive pedigree-splitting method that, within a user supplied bit-size limit, identifies sub-pedigrees having the maximal number of patients who share a common ancestor. Fast grouping is achieved by prioritizing relatives using kinship. Compared to the current existing pedigree cutting methods, this algorithm guarantees that the derived sub-pedigrees can be directly and efficiently analyzed by software implementing the Lander-Green-Kruglyak algorithm.

Another approach to reduce the computational complexity of linkage analysis is to break all consanguineous loops in large pedigrees, and thus enable Elston-Stewart algorithm based linkage analysis[37]. Due to ignoring loops, this method results in a loss of genetic information and thus inflates type-I error rate as well as decreases the power to detect linkage. To minimize this loss, an optimal set of loop breakers has to be selected. For pedigrees in which any person married no more than one other person, the problem of selecting loop breakers with a minimum lost of information was solved[38]. For pedigrees with multiple marriages for which this algorithm is not applicable, several heuristical algorithms for the selection of loop breakers has been proposed[38-40]. These algorithms, however, do not guarantee the optimal selection of loop breakers. We developed an algorithm for the automatic selection of loop breakers and guarantee the minimal loss of the total relationship between measured individuals[41]. Our loop-breaking algorithm provides another option for linkage analysis of large complex pedigrees with multiple loops and incomplete genotypic and phenotypic information. This approach is particularly useful for studying a region of interest while taking into account the inheritance information provided by the whole pedigree.

A drawback of any pedigree-splitting or loop-breaking method is that it unavoidably underestimates the likelihood of haplotype sharing under the null hypothesis of no linkage, and, thus, increases the probability of false positive linkage signals[42]. This is actually a more general problem for all parametric linkage analyses conducted in genetically isolated populations. In **chapter 2**, we illustrated this problem using homozygosity mapping as an example. With the extensive genealogic information available in the GRIP population, we were able to show that the degree of consanguinity may be seriously underestimated when only the shortest consanguineous loops were considered in the analysis. Although the contribution of each distant loop to inbreeding may be very small, hundreds, and even thousands, of such distant loops may exist, and they together may contribute substantially to the inbreeding. We quantified the effect of underestimation of inbreeding on the false positive rate, and showed that the frequency of false positive conclusions may be seriously inflated. We then proposed a simple solution by constructing hypothetical loops based on patients' true inbreeding values. However, this solution may not work for other types of parametric linkage analysis using sub-pedigrees derived from the pedigree splitting algorithm. When more than two patients are related with each other, the hypothetical relationships between patients may not fit in a pedigree. Therefore we proposed estimating type-I error rate empirically by means of simulation as a general guild line for linkage analysis conducted in genetically isolated populations. As described in **chapter 2** and **chapter 4**, marker inheritance should be simulated using the large pedigree, and following linkage analysis should be conducted using the sub-pedigrees or zero loop pedigrees.

Association analysis in genetically isolated populations

We discussed the risk of inflation of type-I error rates in linkage analysis due to ignoring or breaking pedigrees. This is also true for association analysis without taking into account individual relationships. Many statistical methods for family based association analysis have been developed. The most popular of these family-based tests of association is the transmission/disequilibrium test (TDT), which is a test of linkage in the presence of allelic association[43]. The TDT method has been further developed to allow more general form of pedigrees[44] and improved statistical power by taking into account sibling controls[45], and allow analysis of genetic imprinting or dominance[46]. Methods for family-based association tests of quantitative traits were also developed[47,48] and improved[27,49,50]. There are also other methods for family-based association analysis[51-58]. Due to the availability of dense genome-wide single nucleotide polymorphisms (SNPs), increasing interests goes to methods that do not require the knowledge of pedigree structure, such as the methods of genomic control[59], STRUCTURE[60] and EIGENSTRAT[61], which are based on population data rather than pedigree data. These methods were initially developed to correct for population stratification and later proven to be very useful for genome-wide association (GWA) analysis in related individuals. Recently, our group developed a fast powerful method for GWA analysis of quantitative traits in samples of related individuals, which does not require precise knowledge of pedigree structure[62]. The methods testing for association and correcting for population stratification are in general more powerful compared to TDT based methods and less computationally complex compared with linkage analysis.

EMPIRICAL STUDIES

Genome wide linkage analysis of AD

In **Chapter 4** we aimed to conduct a genome-wide linkage analysis for late onset AD in the GRIP population. We applied the pedigree splitting algorithm described in **chapter 3** to split a large pedigree including 4,645 people from whom 112 were late onset AD patients. We could assign 103 patients to 35 pedigrees using bit-size limit of 35 in about 11 minutes. Empirical LOD score threshold for 5% genome-wide significance was estimated to be 3.64 by means of simulation. Using 402 microsatellite markers over the genome, we detected evidence for linkage for previously established loci on chromosomes 1q21-25 [63-65] and 10q22-24 [66-70]. We also identified a novel locus at chromosome 3q23 that was significantly linked to AD. We followed up these regions by association test of dense SNPs with cognitive function and found region-wide significant association.

In the region of chromosome 1 there are two obvious candidate genes for AD, the C-reactive protein (*CRP*) gene[71] and the nicastrin (NCSTN) gene[72]. We sequenced these genes for exons and exon-intron boundaries but did not find any mutation. For the region on chromosome

3, we found that one SNP (rs952797) was significantly associated with cognitive function in 197 unrelated subjects from GRIP population. This SNP is 126 kb downstream of the spell out the name (*NMNAT3)* gene and 131 kb upstream of the spell out the name (*CLSTN2)* gene. It has been reported recently that SNP rs6439886, which is a common T → C substitution within the first intron of *CLSTN2*, was significantly associated with memory performance[73]. We sequenced all exons of this gene in 18 patients but the result so far is not conclusive. We are currently sequencing more patients for this gene.

We also found suggestive evidence for linkage to chromosome 11q25. This linkage peak is about 23 cM downstream to the spell out (*SORL1)* gene, which was recently reported to be significantly associated with AD[14]. We specifically tested the association between polymorphisms flanking *SORL1* gene and cognitive function in 197 unrelated subjects but failed to detect a significant association, suggesting that other gene(s) may explain our linkage peak (**chapter 4**). Furthermore, we conducted a replication study including more subjects from Erasmus Rouphen Family (ERF) study and Rotterdam and failed to replicate the initial findings of *SORL1* gene being associated with late onset AD (**chapter 6**).

Candidate gene studies

APOE gene is so far the most important gene for AD. In **chapter 5**, we studied the age-specific effects of the *APOE* E4 allele on cognitive function and vascular pathology in a series of 2208 related individuals from a family-based study conducted in ERF. We found a significant association between the *APOE* E4 allele and reduced memory performance in persons aged 50 years and older. This effect was independent of the effect of *APOE* gene on cardiovascular factors. In our analyses of cognitive function there was significant evidence for interaction between *APOE* E4 allele and age. The effect of *APOE* E4 allele increases significantly with age, particularly in terms of learning ability. As expected *APOE* E4 allele was strongly related to lipid levels and atherosclerosis, while serum levels of triglycerides, blood pressure and atherosclerosis were significantly associated to cognitive function. Additional adjustment for *APOE* gene status had little influence on the relationship between vascular risk factors and cognitive function.

In **chapter 6** we studied extensively the association of genetic variants in *SORL1* gene with AD and cognitive function. Although our study was carefully designed and well powered, we could not replicate the association. Nor did we find a consistent and significant association between *SORL1* gene and cognitive function. When adding our data to the previous ones in Caucasian populations, the ORs became less significant and move towards the null hypothesis of no association. The evidence of association was mainly driven by the first three studies from the same research group 1-3, where the odds ratios exceeded the genetic effect estimated by meta-analysis of the remaining studies. Removing the original study from the meta-analysis, results in non-significant ORs for all SNPs. Notably, even after excluding the original study, the remaining 6,000 cases and 11,000 controls will still have over 95% power

to detect a small OR of 1.1 at a 5% false positive rate. Thus, the association of seven SORL1 with AD when meta-analyzing all data available on Caucasians is most likely partly explained by the 'winner's curse' phenomenon[74].

In the first genome wide association in AD, Reiman et al. reported that a haplotype encompassing 6 polymorphisms of the *GAB2* gene was associated with the risk of developing AD in 527 case and 117control *APOE* E4 carriers[15]. In **chapter 7**, we tried to replicate this finding using high-density genotyping surrounding the *GAB2* gene in subjects from the Rotterdam Study. Eighteen of 22 polymorphisms in *GAB2* were significantly associated with AD among *APOE* E4 carriers, yet none among non-carriers. Of the 50 polymorphisms in the 100kbp-region surrounding *GAB2* gene, 26 polymorphisms were associated with AD among *APOE* E4 carriers and only one among non-carriers. A point of concern is that all SNPs had similar P values and the genome wide significance was not reached, even when pooling with previous studies. However, since this study is targeted replication, threshold of significance may be less stringent. This study is the first replication of polymorphisms in *GAB2* gene in relation with AD. *GAB2* gene is a protein involved in various signaling pathways including AD -related tau processing. Indeed, Reiman et al. also found that *GAB2* gene expression was associated with protection from neurofibrillary tangle formation. Future research may focus on disentangling the exact interactive mechanism between *GAB2* gene and *APOE* gene as well as finding the possible causative mutation.

SUGGESTIONS FOR FUTURE RESEARCH

Although most previous studies of AD were using disease status as the phenotype, testing for association or linkage with other traits such as cognitive function, biomarkers including neuroimaging, and neuropathological features may have considerable merit. Because of the need for standardization across different studies, considerable effort goes into unifying the method for clinical diagnosis of AD. However, establishing an appropriate control continues to be a challenge due to the late-onset nature of the disease and lack of pathology confirmation. Furthermore, most genetic defects may be more directly involved in neurobiological and biochemical processes, which ultimately lead to clinical AD. Therefore, strategies to use endophenotypes as the research targets may minimize the effects of misclassification of disease status.

Until recently, most studies have been based on linkage analyses and candidate gene analysis. The completion of the Human Genome Project[75], along with the advances in high-throughput, high-density genotyping technology have led to a quick increase in the number of studies examining a large number of SNPs simultaneously in hypothesis-independent designs. Genome wide association (GWA) studies have emerged as an effective tool for identifying genetic contributions to complex diseases. In contrast to linkage analysis this

approach is less susceptible to diagnostic misclassification. There have been a number of success stories for diseases such as macular degeneration[76-78] and diabetes mellitus[79]. GWA studies of AD have been disappointing. A GWA study of neuropathologically confirmed AD cases and control subjects showed a SNP rs4420638 in LD to the *APOE* gene reached genome wide significance but no other variants or genes were identified[80]. Given the fact that a large proportion (~55%) of the disease must be explained by genetic effects other than that of *APOE*[81], this implicates that multiple unidentified genetic variants must have small effects and large sample sizes are needed to identify them.

So far the putative variants of AD show very modest effect sizes, most of which with odds ratios less than 1.5 and lack of solid replication in large cohorts (Table 1). It therefore typically requires sufficiently a large sample size over 5,000 cases and the same amount of controls that allows novel gene discovery[82], and even larger for replication. Huge efforts have been made to achieve this, such as those funded through the Genetic Association Information Network (http://www.fnih.org/GAIN2/home_new.shtml) and the Wellcome Trust Case Control Consortium (http://www.wtccc.org.uk/), that have collected 2000 to more than 10 000 patients and controls. Such efforts are on the way for AD.

Another way to acquire powerful studies is to conduct genome wide meta-analysis. Worldwide collaborations on data sharing policies are going on. Efficient programs for genome wide meta-analysis were developed. A recently genome-wide meta-analysis of type 2 diabetes successfully identified novel loci associated with the disease[83], which highlight the future of genetic research of complex diseases.

An alternative development that maybe very important for AD research is that of deep sequencing. There are a large number of genes and regions for which the causal variant has remained unknown. These include the *SORL1*, *GAB2*, and *LRP6* genes and genomic regions have been significantly linked to AD, including chromosome 1p36, 1q23-25, 2q11, 4p16, 5p14, 6q16, 9q22, 10q21-24, 12p13, 14q22, 19q13, 21q22, and Xp22 (www.alzgene.org). By deep sequencing technology in future we may be able to sequence these regions including the regions we have identified at chromosome 3q22-24 fully including all exons and introns. In this way we may identify new variants both common and rare ones. that explain the AD pathology.

REFERENCES

1. Goate A, Chartier-Harlin MC, Mullan M, Brown J, Crawford F, Fidani L, Giuffra L, Haynes A, Irving N, James L, et al. (1991) Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. Nature 349:704-706
2. Mullan M, Houlden H, Windelspecht M, Fidani L, Lombardi C, Diaz P, Rossor M, Crook R, Hardy J, Duff K, et al. (1992) A locus for familial early-onset Alzheimer's disease on the long arm of chromosome 14, proximal to the alpha 1-antichymotrypsin gene. Nat Genet 2:340-342
3. Schellenberg GD, Bird TD, Wijsman EM, Orr HT, Anderson L, Nemens E, White JA, Bonnycastle L, Weber JL, Alonso ME, et al. (1992) Genetic linkage evidence for a familial Alzheimer's disease locus on chromosome 14. Science 258:668-671
4. St George-Hyslop P, Haines J, Rogaev E, Mortilla M, Vaula G, Pericak-Vance M, Foncin JF, Montesi M, Bruni A, Sorbi S, et al. (1992) Genetic evidence for a novel familial Alzheimer's disease locus on chromosome 14. Nat Genet 2:330-334
5. Van Broeckhoven C, Backhovens H, Cruts M, De Winter G, Bruyland M, Cras P, Martin JJ (1992) Mapping of a gene predisposing to early-onset Alzheimer's disease to chromosome 14q24.3. Nat Genet 2:335-339
6. Sherrington R, Rogaev EI, Liang Y, Rogaeva EA, Levesque G, Ikeda M, Chi H, Lin C, Li G, Holman K, et al. (1995) Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. Nature 375:754-760
7. Levy-Lahad E, Wasco W, Poorkaj P, Romano DM, Oshima J, Pettingell WH, Yu CE, Jondro PD, Schmidt SD, Wang K, et al. (1995) Candidate gene for the chromosome 1 familial Alzheimer's disease locus. Science 269:973-977
8. Rogaev EI, Sherrington R, Rogaeva EA, Levesque G, Ikeda M, Liang Y, Chi H, Lin C, Holman K, Tsuda T, et al. (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
9. Strittmatter WJ, Saunders AM, Schmechel D, Pericak-Vance M, Enghild J, Salvesen GS, Roses AD (1993) Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. Proc Natl Acad Sci U S A 90:1977-1981
10. van Duijn CM, de Knijff P, Cruts M, Wehnert A, Havekes LM, Hofman A, Van Broeckhoven C (1994) Apolipoprotein E4 allele in a population-based study of early-onset Alzheimer's disease. Nat Genet 7:74-78
11. Tol J, Roks G, Slooter AJ, van Duijn CM (1999) Genetic and environmental factors in Alzheimer's disease. Rev Neurol (Paris) 155 Suppl 4:S10-16
12. Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, Myers RH, Pericak-Vance MA, Risch N, van Duijn CM (1997) Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. Jama 278:1349-1356
13. 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
14. Rogaeva E, Meng Y, Lee JH, Gu Y, Kawarai T, Zou F, Katayama T, Baldwin CT, Cheng R, Hasegawa H, et al. (2007) The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. Nat Genet
15. Reiman EM, Webster JA, Myers AJ, Hardy J, Dunckley T, Zismann VL, Joshipura KD, Pearson JV, Hu-Lince D, Huentelman MJ, et al. (2007) GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers. Neuron 54:713-720

16. Li Y, Grupe A, Rowland C, Nowotny P, Kauwe JS, Smemo S, Hinrichs A, Tacey K, Toombs TA, Kwok S, et al. (2006) DAPK1 variants are associated with Alzheimer's disease and allele-specific expression. Hum Mol Genet 15:2560-2568
17. Sundar PD, Feingold E, Minster RL, DeKosky ST, Kamboh MI (2007) Gender-specific association of ATP-binding cassette transporter 1 (ABCA1) polymorphisms with the risk of late-onset Alzheimer's disease. Neurobiol Aging 28:856-862
18. De Ferrari GV, Papassotiropoulos A, Biechele T, Wavrant De-Vrieze F, Avila ME, Major MB, Myers A, Saez K, Henriquez JP, Zhao A, et al. (2007) Common genetic variation within the low-density lipoprotein receptor-related protein 6 and late-onset Alzheimer's disease. Proc Natl Acad Sci U S A 104:9434-9439
19. Waring SC, Rosenberg RN (2008) Genome-wide association studies in Alzheimer disease. Arch Neurol 65:329-334
20. Aulchenko YS, Heutink P, Mackay I, Bertoli-Avella AM, Pullen J, Vaessen N, Rademaker TA, Sandkuijl LA, Cardon L, Oostra B, et al. (2004) Linkage disequilibrium in young genetically isolated Dutch population. Eur J Hum Genet 12:527-534
21. Pardo LM, MacKay I, Oostra B, van Duijn CM, Aulchenko YS (2005) The effect of genetic drift in a young genetically isolated population. Ann Hum Genet 69:288-295
22. Vierimaa O, Georgitsi M, Lehtonen R, Vahteristo P, Kokko A, Raitila A, Tuppurainen K, Ebeling TM, Salmela PI, Paschke R, et al. (2006) Pituitary adenoma predisposition caused by germline mutations in the AIP gene. Science 312:1228-1230
23. van Duijn CM, Dekker MC, Bonifati V, Galjaard RJ, Houwing-Duistermaat JJ, Snijders PJ, Testers L, Breedveld GJ, Horstink M, Sandkuijl LA, et al. (2001) Park7, a novel locus for autosomal recessive early-onset parkinsonism, on chromosome 1p36. Am J Hum Genet 69:629-634
24. Elston RC, Stewart J (1971) A general model for the genetic analysis of pedigree data. Hum Hered 21:523-542
25. Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347-1363
26. Lander ES, Green P (1987) Construction of multilocus genetic linkage maps in humans. Proc Natl Acad Sci U S A 84:2363-2367
27. 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
28. Gudbjartsson DF, Thorvaldsson T, Kong A, Gunnarsson G, Ingolfsdottir A (2005) Allegro version 2. Nat Genet 37:1015-1016
29. O'Connell JR, Weeks DE (1995) The VITESSE algorithm for rapid exact multilocus linkage analysis via genotype set-recoding and fuzzy inheritance. Nat Genet 11:402-408
30. O'Connell JR (2001) Rapid multipoint linkage analysis via inheritance vectors in the Elston-Stewart algorithm. Hum Hered 51:226-240
31. Sobel E, Lange K (1996) Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. Am J Hum Genet 58:1323-1337
32. Heath SC (1997) Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. Am J Hum Genet 61:748-760
33. Sung YJ, Thompson EA, Wijsman EM (2007) MCMC-based linkage analysis for complex traits on general pedigrees: multipoint analysis with a two-locus model and a polygenic component. Genet Epidemiol 31:103-114
34. Thomas A (2007) Towards linkage analysis with markers in linkage disequilibrium by graphical modelling. Hum Hered 64:16-26

35. Pankratz VS, Iturria SJ (2001) A pedigree partitioning approach to quantitative trait loci mapping of IgE serum level in the GAW12 Hutterite data. Genet Epidemiol 21 Suppl 1:S258-263
36. Falchi M, Forabosco P, Mocci E, Borlino CC, Picciau A, Virdis E, Persico I, Parracciani D, Angius A, Pirastu M (2004) A genomewide search using an original pairwise sampling approach for large genealogies identifies a new locus for total and low-density lipoprotein cholesterol in two genetically differentiated isolates of Sardinia. Am J Hum Genet 75:1015-1031
37. Stricker C, Fernando R, Elston R (1995) An algorithm to approximate the likelihood for pedigree data with loops by cutting. Theor Appl Genet 91:1054-1063
38. Becker A, Geiger D, Schaffer AA (1998) Automatic selection of loop breakers for genetic linkage analysis. Hum Hered 48:49-60
39. Vitezica ZG, Mongeau M, Manfredi E, Elsen JM (2004) Selecting loop breakers in general pedigrees. Hum Hered 57:1-9
40. Becker A, Gold M (1988) Prediction of an ATP reactive center in the small subunit, gpNu1, of the phage lambda terminase enzyme. J Mol Biol 199:219-222
41. Axenovich TI, Zorkoltseva IV, Liu F, Kirichenko AV, Aulchenko YS (2008) Breaking loops in large complex pedigrees. Hum Hered 65:57-65
42. Liu F, Elefante S, van Duijn CM, Aulchenko YS (2006) Ignoring Distant Genealogic Loops Leads to False-positives in Homozygosity Mapping. Ann Hum Genet 70:965-970
43. 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
44. Martin ER, Kaplan NL, Weir BS (1997) Tests for linkage and association in nuclear families. Am J Hum Genet 61:439-448
45. Curtis D, Sham PC (1995) A note on the application of the transmission disequilibrium test when a parent is missing. Am J Hum Genet 56:811-812
46. Weinberg CR, Wilcox AJ, Lie RT (1998) A log-linear approach to case-parent-triad data: assessing effects of disease genes that act either directly or through maternal effects and that may be subject to parental imprinting. Am J Hum Genet 62:969-978
47. Allison DB (1997) Transmission-disequilibrium tests for quantitative traits. Am J Hum Genet 60: 676-690
48. Rabinowitz D (1997) A transmission disequilibrium test for quantitative trait loci. Hum Hered 47: 342-350
49. Fulker DW, Cherny SS, Sham PC, Hewitt JK (1999) Combined linkage and association sib-pair analysis for quantitative traits. Am J Hum Genet 64:259-267
50. Abecasis GR, Cardon LR, Cookson WO (2000) A general test of association for quantitative traits in nuclear families. Am J Hum Genet 66:279-292
51. Laird NM, Lange C (2006) Family-based designs in the age of large-scale gene-association studies. Nat Rev Genet 7:385-394
52. Slager SL, Schaid DJ (2001) Evaluation of candidate genes in case-control studies: a statistical method to account for related subjects. Am J Hum Genet 68:1457-1462
53. Follmann D, Proschan M, Leifer E (2003) Multiple outputation: inference for complex clustered data by averaging analyses from independent data. Biometrics 59:420-429
54. Tian X, Joo J, Zheng G, Lin JP (2005) Robust trend tests for genetic association in case-control studies using family data. BMC Genet 6 Suppl 1:S107
55. Liang KY, Pulver AE (1996) Analysis of case-control/family sampling design. Genet Epidemiol 13: 253-270

56. Xing G, Xing C, Lu Q, Elston RC (2007) A logistic mixture model for a family-based association study. BMC Proc 1 Suppl 1:S44
57. Horvath S, Xu X, Lake SL, Silverman EK, Weiss ST, Laird NM (2004) Family-based tests for associating haplotypes with general phenotype data: application to asthma genetics. Genet Epidemiol 26: 61-69
58. Lange C, DeMeo DL, Laird NM (2002) Power and design considerations for a general class of family-based association tests: quantitative traits. Am J Hum Genet 71:1330-1341
59. Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55:997-1004
60. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945-959
61. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904-909
62. Amin N, van Duijn CM, Aulchenko YS (2007) A genomic background based method for association analysis in related individuals. PLoS ONE 2:e1274
63. 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
64. Hiltunen M, Mannermaa A, Thompson D, Easton D, Pirskanen M, Helisalmi S, Koivisto AM, Lehtovirta M, Ryynanen M, Soininen H (2001) Genome-wide linkage disequilibrium mapping of late-onset Alzheimer's disease in Finland. Neurology 57:1663-1668
65. Myers A, Wavrant De-Vrieze F, Holmans P, Hamshere M, Crook R, Compton D, Marshall H, Meyer D, Shears S, Booth J, et al. (2002) Full genome screen for Alzheimer disease: stage II analysis. Am J Med Genet 114:235-244
66. Kehoe P, Wavrant-De Vrieze F, Crook R, Wu WS, Holmans P, Fenton I, Spurlock G, Norton N, Williams H, Williams N, et al. (1999) A full genome scan for late onset Alzheimer's disease. Hum Mol Genet 8: 237-245
67. Myers A, Holmans P, Marshall H, Kwon J, Meyer D, Ramic D, Shears S, Booth J, DeVrieze FW, Crook R, et al. (2000) Susceptibility locus for Alzheimer's disease on chromosome 10. Science 290:2304-2305
68. Li YJ, Scott WK, Hedges DJ, Zhang F, Gaskell PC, Nance MA, Watts RL, Hubble JP, Koller WC, Pahwa R, et al. (2002) Age at onset in two common neurodegenerative diseases is genetically controlled. Am J Hum Genet 70:985-993
69. Blacker D, Bertram L, Saunders AJ, Moscarillo TJ, Albert MS, Wiener H, Perry RT, Collins JS, Harrell LE, Go RC, et al. (2003) Results of a high-resolution genome screen of 437 Alzheimer's disease families. Hum Mol Genet 12:23-32
70. Holmans P, Hamshere M, Hollingworth P, Rice F, Tunstall N, Jones S, Moore P, Wavrant DeVrieze F, Myers A, Crook R, et al. (2005) Genome screen for loci influencing age at onset and rate of decline in late onset Alzheimer's disease. Am J Med Genet B Neuropsychiatr Genet 135:24-32
71. Robey FA, Jones KD, Tanaka T, Liu TY (1984) Binding of C-reactive protein to chromatin and nucleosome core particles. A possible physiological role of C-reactive protein. J Biol Chem 259: 7311-7316
72. Yu G, Nishimura M, Arawaka S, Levitan D, Zhang L, Tandon A, Song YQ, Rogaeva E, Chen F, Kawarai T, et al. (2000) Nicastrin modulates presenilin-mediated notch/glp-1 signal transduction and betaAPP processing. Nature 407:48-54
73. Papassotiropoulos A, Stephan DA, Huentelman MJ, Hoerndli FJ, Craig DW, Pearson JV, Huynh KD, Brunner F, Corneveaux J, Osborne D, et al. (2006) Common Kibra alleles are associated with human memory performance. Science 314:475-478

74. Bazerman M, Samuelson W (1983) I won the auction but don't want the price. J Conflict Resolut 27: 618-634
75. International_HapMap_Consortium (2005) A haplotype map of the human genome. Nature 437: 1299-1320
76. Edwards AO, Ritter R, 3rd, Abel KJ, Manning A, Panhuysen C, Farrer LA (2005) Complement factor H polymorphism and age-related macular degeneration. Science 308:421-424
77. Haines JL, Hauser MA, Schmidt S, Scott WK, Olson LM, Gallins P, Spencer KL, Kwan SY, Noureddine M, Gilbert JR, et al. (2005) Complement factor H variant increases the risk of age-related macular degeneration. Science 308:419-421
78. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, et al. (2005) Complement factor H polymorphism in age-related macular degeneration. Science 308:385-389
79. Wellcome_Trust_Case_Control_Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447:661-678
80. Coon KD, Myers AJ, Craig DW, Webster JA, Pearson JV, Lince DH, Zismann VL, Beach TG, Leung D, Bryden L, et al. (2007) A high-density whole-genome association study reveals that APOE is the major susceptibility gene for sporadic late-onset Alzheimer's disease. J Clin Psychiatry 68:613-618
81. Sleegers K, Roks G, Theuns J, Aulchenko YS, Rademakers R, Cruts M, van Gool WA, Van Broeckhoven C, Heutink P, Oostra BA, et al. (2004) Familial clustering and genetic risk for dementia in a genetically isolated Dutch population. Brain 127:1641-1649
82. Wang WY, Barratt BJ, Clayton DG, Todd JA (2005) Genome-wide association studies: theoretical and practical concerns. Nat Rev Genet 6:109-118
83. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, de Bakker PI, Abecasis GR, Almgren P, Andersen G, et al. (2008) Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet 40:638-645

SUMMARY

Alzheimer's disease (AD) is the primary cause of dementia in Western societies, affecting approximately 30% of the population aged 85 years or over. Despite the huge effort in the scientific community to identify the genes involved in AD, the genetic origin of late onset AD and cognitive function is largely unknown. In this thesis I aimed to develop new approach to identify genetic determinants of AD and cognitive function.

In **chapter 1**, a brief introduction to AD is given; the aims and general outline of the research described in this thesis are presented.

In **chapter 2**, the effect of ignoring distant genealogic loops on false-positives in homozygosity mapping was examined. Distant consanguineous loops are often unknown or ignored during homozygosity mapping analysis. This may potentially lead to an increased rate of false-positive linkage findings. We show that failure to take into account the distant loops may lead to a serious underestimation of the degree of consanguinity, especially for people from genetically isolated populations. We also show that converting multiple loops to a hypothetical loop capturing all inbreeding may be a convenient solution to avoid false positive results.

In **chapter 3**, we present an efficient pedigree splitting algorithm. Utilizing large pedigrees in linkage analysis is a computationally challenging task. A common solution is to split large pedigrees into smaller computable sub-units. We present a pedigree-splitting method that, within a user supplied bit-size limit, identifies sub-pedigrees having the maximal number of subjects of interest who share a common ancestor. We show that using a bit-size limit our method can assign more patients to sub-pedigrees than the clique partitioning method, particularly when splitting deep pedigrees where the subjects of interest are scattered in recent generations and are relatively distantly related via multiple genealogic connections. Our pedigree-splitting algorithm and associated software can facilitate genome-wide linkage scans searching for rare mutations in large pedigrees.

In **chapter 4**, we conducted a genome screen for late onset AD in a young genetically isolated Dutch population. We confirmed two previously well described linkage regions for late onset AD on chromosomes 1q21-25 and 10q22-24. We suggest the RGSL2, RALGPS2, and C1orf49 genes at 1q25 and HTR7, MPHOSPH1, and CYP2C cluster on chromosome 10q22-24 may explain the linage to these regions. We identified a new locus that showed significant linkage to chromosome 3q23 markers. For this region we suggest NMNAT3 and CLSTN2 genes may be relevant. Our findings also confirm linkage to chromosome 11q25. We could not confirm *SORL1*, instead, our analysis points to OPCML and HNT genes.

In **chapter 5**, we studied the age specific effects of *APOE* on vascular pathology and cognitive function. We found a significant association between the *APOE* E4 allele and reduced memory performance in persons aged 50 years and older. The effect of *APOE* E4 is most pronounced on learning ability, starting as early 40 years. The *APOE* E4 allele is also strongly

associated to cholesterol levels and atherosclerosis. This association did not explain the effect of *APOE* on cognitive function. Our study suggests that *APOE* E4 is an important determinant of vascular and neurological pathology at late age.

In **chapter 6**, we performed a extensive genomic study for *SORL1* in relation to AD and cognitive function. The *SORL1* gene is one of the most recent genes associated with AD. We failed to replicate previous findings on *SORL1* gene being associated AD. Nor did we find a consistent or significant association between *SORL1* and cognitive function. When meta-analyzing our and previously published data we found that the odds ratios reported by the initial study exceeded the genetic effect estimated by meta-analysis of the remaining studies. Removing the original study from the meta-analysis, results in non-significant ORs for all SNPs. The exact causal variant explaining the previously observed association remained to be identified by deep sequencing.

In **chapter 7**, we performed a replication study for the *GAB2* gene, which was identified to be associated with AD in *APOE* E4 carriers through a genome wide association study (Reiman *et al.* 2007). We could successfully replicate this finding. Our results were consistent with that of Reiman *et al.* and convincing in that multiple SNPs in the *GAB2* gene showed significant association with AD in *APOE* E4 carriers, calling further investigations to disentangle the exact interactive mechanism between *GAB2* and *APOE*.

Finally, in **chapter 8**, we discuss the findings of our studies in the context of those of others.

SAMENVATTING

De ziekte van Alzheimer (AD) is de voornaamste oorzaak van dementie in Westerse samenlevingen, die ongeveer 30% van de bevolking boven de 85 jaar treft. Ondanks de enorme inspanningen in de wetenschappelijke gemeenschap om de bij AD betrokken genen te identificeren, is de genetische basis van AD en cognitief verval op latere leeftijd (late onset AD) grotendeels nog onbekend. Dit proefschrift heeft als doelstelling methoden te ontwilkkelen om genetische factoren van AD te identificeren.

In **hoofdstuk 1** wordt een korte inleiding tot AD gegeven; de doelen en algemene hoofdlijnen van het in dit proefschrift beschreven onderzoek worden gepresenteerd.

In **hoofdstuk 2** wordt het effect bestudeerd van het negeren van ververwijderde genealogische loops op vals-positieven bij het in kaart brengen van de homozygositeit ("homozygosity mapping"). Verre familieverwantschappen zijn vaak onbekend of worden genegeerd tijdens homozygosity mapping analyses. Dit zou in potentie kunnen leiden tot een verhoogde mate van vals-positieve linkage bevindingen. We tonen aan dat het nalaten om rekening te houden met de ververwijderde loops, zou kunnen leiden tot een ernstige onderschatting van de mate van verwantschap, met name voor individuen uit genetisch geïsoleerde populaties. We laten ook zien dat het omzetten van meervoudige loops in een hypothetische loop die alle inteelt omvat, een passende oplossing zou kunnen zijn om vals-positieve resultaten te vermijden.

In **hoofdstuk 3** presenteren we een efficient algoritme om stamboom te splitsen. Het gebruik van grote stambomen in linkage analyses is computationeel veeleisend. Een gangbare oplossing is het opsplitsen van grote stambomen in rekenkundig beter hanteerbare eenheden. We presenteren een stamboom-splitsingsmethode die, binnen een door de gebruiker gedefinieerde bitgrootte, deelstambomen identificeert met het maximale aantal relevante individuen dat een gemeenschappelijke voorouder deelt. We laten zien dat onze methode, door gebruik van een begrensde bitgrootte, meer patienten kan toewijzen aan deelstambomen dan de clique partitioning methode, in het bijzonder bij het splitsen van ver terugvoerende stambomen waar de relevante individuen zijn verspreid over recente generaties en die relatief ver verwant zijn via meervoudige genealogische connecties. Ons algoritme en bijbehorende software kan genoomwijde linkage scans mogelijk maken die zoeken naar zeldzame mutaties in grote stambomen verkregen aan de hand van genetisch geïsoleerde populaties.

In **hoofdstuk 4** hebben we een genoom-screen uitgevoerd voor late onset AD in een jong genetisch geïsoleerde Nederlandse populatie. We bevestigen twee eerder veel beschreven linkage regio's voor late onset AD op chromosomen 1q21-25 en 10q22-24. We opperen dat de RGSL2, RALGPS2 en C1orf49 genen op 1q25 en HTR7, MPHOSPH1 en het CYP2C cluster op chromosoom 10q22-24, de linkage met deze regio's zou kunnen verklaren. We hebben een nieuwe locus geïdentificeerd die significante linkage vertoond met markers op chromosoom

3q23. Voor deze regio opperen we dat de NMNAT3 en CLSTN2 genen relevant zouden kunnen zijn. Onze bevindingen bevestigen ook linkage met chromosoom 11q25. We konden *SORL1* niet bevestigen; in plaats daarvan duidt onze analyse op de OPCML en HNT genen.

In **hoofdstuk 5** hebben we de leeftijd-specifieke effecten van *APOE* op vasculaire pathologie en cognitieve functie bestudeerd. We vonden een significant verband tussen het *APOE* E4 allel en verminderde werking van het geheugen in personen van 50 jaar en ouder. Het effect van *APOE* E4 is het sterkst op het leervermogen, reeds beginnend bij 40 jaar. Het *APOE* E4 allel is ook sterk gerelateerd aan cholesterolspiegels en atherosclerose. Deze associatie verklaarde niet het effect van *APOE* op cognitieve functie. Onze studie suggereert dat *APOE* E4 een belangrijke determinant is voor vasculaire en neurologische pathologie op latere leeftijd.

In **hoofdstuk 6** hebben we een studie uitgevoerd voor *SORL1* in relatie tot AD en cognitieve functie. Het *SORL1* gen is een van de meest recente genen dat in verband is gebracht met AD. We zijn er niet in geslaagd om eerdere bevindingen met betrekking tot het verband tussen *SORL1* en AD te repliceren. Noch hebben we een consistent of significant verband gevonden tussen *SORL1* en cognitieve functie. Toen we onze bevindingen samenvoegden met eerder gepubliceerde data vonden we een significant maar klein effect. Weglating van de oorspronkelijke studie uit de meta-analyse resulteert in niet-significante odds ratio's (ORs) voor alle SNPs. De exacte causale variant die de eerder geobserveerde associatie verklaart, moet nog geïdentificeerd worden door deep-sequencen.

In **hoofdstuk 7** hebben we een replicatiestudie uitgevoerd voor het *GAB2* gen, waarvoor een verband was aangetoond met AD in *APOE* E4 dragers door middel van een genoomwijde associatie studie (Reiman *et al.* 2007). We zijn erin geslaagd deze bevinding met succes te repliceren. Onze resultaten waren in overeenstemming met die van Reiman *et al.* En ze waren overtuigend gezien het feit dat verscheidene SNPs in het *GAB2* gen significante associatie vertoonden met AD in *APOE* dragers. Deze resultaten vragen om verder onderzoek wat kan leiden tot het ontrafelen van het precieze interactieve mechanisme tussen *GAB2* en *APOE*.

Tenslotte bespreken we in **hoofdstuk 8** de bevindingen van onze studies in de context van de bevindingen van anderen.

ACKNOWLEDGMENTS

I would like to thank my promoters professor Conelia van Duijn and professor Ben Oostra and my co-promoter Dr. Yurii Aulchenko. Dear Cock, thank you for offering me the scholarship during my DSc project which helped me to concentrate on my studying. Thank you for giving me the opportunity to work as a phD student in your department. Besides these, I greatly admire that during every lecture, lunch meeting, neural meeting, email, or conversation, you taught me something that contributed to my development as a researcher. Dear Ben, your feedbacks on my papers were always quick. I never forget that you were sitting together with me to submit a scientific paper. Your help was essential in finalizing this thesis. I greatly appreciate all your help. Dear Yurii, you are the one who made all these possible. You taught me how to analysis and interpret data, how to program, and how to write scientific papers, step by step, hand by hand. More importantly, you taught me how to work with precision and discipline. As I immerse myself further in the scientific field, I realize how much I have learned from you. All the "thank-you"s you have heard from me are not enough to truly express my gratitude and appreciation. As I side, I was your student and will always be.

I would like to thank the members of my doctoral committee. Dear professor Manfred Kayser, thank you for the exciting and fruitful collaboration on the eye color project. Thank you for giving me the opportunity to work as a post doc in your department. It is my great honor to have you in my committee. Dear Dr. John van Swieten and professor André G. Uitterlinden, I greatly appreciate the time and energy you invested in my path towards my promotion. Thanks for your participation.

I would like to acknowledge and thank all the co-authors. Dear professor Albert Hofman, thank you for helping me lay the foundation of my epidemiological knowledge. It was a great privilege to have had you as a teacher as well as a co-author. Dear professor Tatiana Axenovich, professor Monique Breteler, Dr. Cecile Janssen, and professor Christine van Broeckhoven: thank you for all your invaluable feedback. Dr Mark Houben: you were my first tutor and your help on the IGF1 paper was crucial. Dr. Stefano Elefante, Dr. Anatoly Kirichenko and Dr. Irina Zorkoltseva: your contribution to my first papers was invaluable. Dr. Angela Gonzalez-Zuloeta Ladd, Dr. Isle Hoppenbrouwers, and Arfan Ikram: it is a great pleasure see our collaborations turned out to be fruitful.

Special thanks to my two companions. Dear Aaron Isaacs and Mojgan Yazdanpanah, since Africa Inn, we have went all the way together through the MSc, DSc, and PhD projects. Remember the old days when we had to fight for limited computer resources … Our daily communications had tremendous meaning to me. Thanks for sharing so much happiness and sadness with me along the way here and now I want to sincerely thank you for your tremendous help in finalizing my phD project.

I am grateful to the participants of the ERGO and GRIP studies and the people in the lab of Department of Epidemiology and Biostatistics: Andy, Bernadette, Debby, Els, Florencia,

Jeanette, Hasna, Leon, Tessa, Hilda , and Petra: you established the data based on which the works described in this thesis becomes possible. I would like to thank the administrative, financial and IT group: Alwin, Annette, Eric, Frank, Kabita, Marcel, Marjolijn, Marti, Michiel, Nano, Natacha, Petra, and Rene. Jeanette, thanks for making every appointment and paperwork goes so smoothly. Solange, thanks for helping me with piles of documents.

I would like to thank all my colleagues at the Department of Epidemiology & Biostatistics for making these years so intellectually gratifying. It was a pleasure to work with you: Anna, Annelous, Behrooz, Dominiek, Esther, Esther de Vries, Ingrid, Iratxe, Jeanine, Leonieke, Linda, Marie-Josee, Marieke, Mark Sie, Nahid, Omer, Regie, Roxana, Slavica, and Stefania. In particular I would like to thank the people with whom I worked closer: Aida, Fakhredin, Fernando, Kristel, Luba, Maaike, Maksim, Pascual, Sandra, and Suzanne. Thank you for your friendship. It was a great pleasure to walk with you on this road. Alejandro, we finally succeed in finding linkage signals after a long struggling. Bingjian (冯丙健), you were the only person in the department to whom I could talk in Chinese. Najaf, thanks for all invitations and your cooking skill was truly impressive. Thank you all for leaving those great memories in my life.

In writing this thesis, Mannis van Oven, Mijke Visser, and Mark Vermeulen have kindly translated the summary to Dutch.

I would like to thank my Chinese colleagues and friends. 周春水，殷冠聪，刘小昊，胡春祥，Andrew Yong，吴磊，郑重，翟伊，刘成华，陆甍，吴婷，谢谢那些与你们共度的快乐时光。我深深感谢山东大学医学院曾经教导过我的王廷础校长，王琰璧校长，王永平女士。谢谢你们长期以来对我的教导和关照，我终于没有辜负你们对我的期望。 我要谢谢北京友谊医院的贺正一主任，卢晓梅大夫，和王春燕老师：谢谢你们在我在医院实习期间的教诲和关照。

I would like to express my deepest gratitude to my family. 亲爱的爸爸刘廷纲和妈妈戴秀中，你们对我的爱和为我做的所有的牺牲是我今生今世无法报答的。亲爱的哥哥刘工谢谢你从小对我的点点呵护和伴我一起度过最美好的童年时光。感谢岳父温树文和岳母方西峰对我的督促和信任。最后，我要把此书献给我的爱妻温蓓。亲爱的宝贝，谢谢你一直以来对我无微不至的关怀，你永远是我生命中最重要的人。

LIST OF PUBLICATIONS

1. Arias-Vasquez A, de Lau L, Pardo L, Liu F, Feng BJ, Bertoli-Avella A, Isaacs A, Aulchenko Y, Hofman A, Oostra B, *et al.* (2007) Relationship of the Ubiquilin 1 gene with Alzheimer's and Parkinson's disease and cognitive function. Neuroscience letters 424:1-5
2. Berends AL, Steegers EA, Isaacs A, Aulchenko YS, Liu F, de Groot CJ, Oostra BA, van Duijn CM (2008) Familial aggregation of preeclampsia and intrauterine growth restriction in a genetically isolated population in The Netherlands. Eur J Hum Genet 16:1437-1442
3. Gonzalez-Zuloeta Ladd AM, Liu F, Houben MP, Arias Vasquez A, Siemes C, Janssens AC, Coebergh JW, Hofman A, Janssen JA, Stricker BH, give all. (2007) IGF-1 CA repeat variant and breast cancer risk in postmenopausal women. Eur J Cancer 43:1718-1722
4. Hoppenbrouwers IA, Liu F, Aulchenko YS, Ebers GC, Oostra BA, van Duijn CM, Hintzen RQ (2008) Maternal transmission of multiple sclerosis in a dutch population. Arch Neurol 65:345-348
5. Ikram MA, Liu F, Oostra B, Hofman A, van Duijn C, Breteler MB (2008) The *GAB2* gene and the risk of Alzheimer's disease: Replication and meta-analysis. Biological Psychiatry. In press
6. Kayser M, Liu F, Janssens AC, Rivadeneira F, Lao O, van Duijn K, Vermeulen M, Arp P, Jhamai MM, van Ijcken WF, et give all. (2008) Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. Am J Hum Genet 82:411-423
7. Liu F, Arias-Vasquez A, Sleegers K, Aulchenko YS, Kayser M, Sanchez-Juan P, Feng BJ, Bertoli-Avella AM, van Swieten J, Axenovich TI, *et al.* (2007) A genomewide screen for late-onset Alzheimer disease in a genetically isolated Dutch population. Am J Hum Genet 81:17-31
8. Liu F, Elefante S, van Duijn CM, Aulchenko YS (2006) Ignoring distant genealogic loops leads to false-positives in homozygosity mapping. Ann Hum Genet 70:965-970
9. Liu F, Ikram MA, Janssen ACJW, Schuur M, de Koning I, Isaacs A, Struchalin M, Uitterlinden AG, den Dunnen JT, Sleegers K, give all. (2008) A study of the *SORL1* gene in Alzheimer's disease and cognitive function. Journal of Alzheimer's Disease (Submitted)
10. Liu F, Kirichenko A, Axenovich TI, van Duijn CM, Aulchenko YS (2008) An approach for cutting large and complex pedigrees for linkage analysis. Eur J Hum Genet 16:854-860
11. Liu F, Pardo LM, Schuur M, Sanchez-Juan P, Isaacs A, Sleegers K, de Koning I, Zorkoltseva IV, Axenovich TI, Witteman JC, give all authors (2008) The apolipoprotein E gene and its age-specific effects on cognitive function. Neurobiology of aging . Neurobiol Aging. In press.
12. Liu F, van Duijn K, Vingerling JR, Hofman A, Uitterlinden AG, Janssen ACJW, Kayser M (2008) Predicting complex phenotypes from genotypes: the case of eye color. Current Biology . In press
13. van den Boogaard MJ, de Costa D, Krapels IP, Liu F, van Duijn C, Sinke RJ, Lindhout D, Steegers-Theunissen RP (2008) The MSX1 allele 4 homozygous child exposed to smoking at periconception is most sensitive in developing nonsyndromic orofacial clefts. Human genetics 124:525-534

ABOUT THE AUTHOR

Fan Liu was born on January 19th, 1976 in Beijing, China. After graduating from the High School attached to Beijing Industry University, he started his medical studies at the Department of Clinic Medicine, Shandong University School of Medicine. He obtained his Bachelor degree in Clinic medicine in 1999. He continued medical training as a medical doctor in the Department of pathology Beijing Friendship Hospital from 1999 to 2001. Later he transferred to work for China Cancer Research Foundation as a clinical trial monitor. In August 2003, he moved to the Netherlands to pursue a Master of Science program in Genetic Epidemiology at Department of Epidemiology & Biostatistics at Erasmus University Medical Center in Rotterdam. He obtained his master's degree in June 2004. He subsequently started a Doctor of Science program and completed it in June 2005. Later that year he started the work described in this thesis towards a Doctor of Philosophy degree. He is currently working as a post-doc in Department of Forensic Molecular Biology at Erasmus University Medical Center.