Molecular genetics of dementia: the Good, the Bad and the New

Rita João Louro Guerreiro

Universidade de Coimbra

2010

Dissertação apresentada à

Faculdade de Medicina da Universidade de Coimbra

para prestação de provas de Doutoramento em Ciências Biomédicas

Segue o teu destino,
Rega as tuas plantas,
Ama as tuas rosas.
O resto é a sombra
De árvores alheias.

A realidade
Sempre é mais ou menos
Do que nós queremos.
Só nós somos sempre
Iguais a nós-próprios.

Suave é viver só.
Grande e nobre é sempre
Viver simplesmente.
Deixa a dor nas aras
Como ex-voto aos deuses.

Vê de longe a vida.
Nunca a interrogues.
Ela nada pode
Dizer-te. A resposta
Está além dos deuses.

Mas serenamente
Imita o Olimpo
No teu coração.
Os deuses são deuses
Porque não se pensam.

Ricardo Reis

AGRADECIMENTOS/ACKNOWLEDGEMENTS

Agradeço à Fundação para a Ciência e Tecnologia pela concessão da Bolsa de Doutoramento (SFRH/BD/27442/2006) que me permitiu desenvolver o trabalho conducente a esta tese.

À Professora Doutora Catarina Resende Oliveira, minha orientadora, por ter acreditado em mim e no meu trabalho; pelo incentivo, confiança e amizade demonstrados ao longo destes anos. Por sempre ter sido e continuar a ser o exemplo que sigo a nível profissional e pessoal.

A todos os meus professores, especialmente ao Prof. Amaro, à Professora Zaida Chieira e Dra. Cristina Girão, o meu sincero agradecimento por todos os ensinamentos durante a minha vida académica.

Agradeço a todos aqueles que me incentivaram na minha vida profissional, especialmente ao Prof. Preto Ribeiro pela sua atenção e consideração pelo meu trabalho.

À Dra. Isabel Santana, à Maria Helena, à Dra. Teresa, ao Rui, à D. Lurdes, à Inês, à Sandra, à Sílvia, à Lia e a todos os que directa ou indirectamente ajudaram ao desenvolvimento, em Portugal, de partes essenciais deste trabalho, um bem hajam.

Agradeço a toda a minha família o incentivo e a paciência com que sempre encararam o prolongamento dos meus estudos. À Nita, Eduardo e Martinha um muito obrigada por me terem acolhido. Aos meus pais agradeço terem-me permitido, de tantas formas essenciais, escolher e seguir o caminho que me trouxe até aqui. À minha mãe agradeço o encurtar diário da distância que nos tem separado, ao meu pai as actualizações da realidade e aos meus avós por olharem sempre por mim. Ao Zé Miguel agradeço a presença constante nos bons e nos maus momentos, a ajuda e a partilha de uma vida.

I thank to my advisor, John Hardy, for all the teaching, the presence (even from far away) and the good science he exposed me to. Today, I feel blessed to have learned from such a grand scientist, but above all, to have this great person as a friend.

To my adopted advisor, Andrew Singleton, I thank for all his patience, dedication and interest in my work, over these years.

To all my colleagues at LNG, without whom it would have been impossible to complete this work: Alice, Melissa, Sampath, Angela, Karen, Cheng-Song, David, Ian, Joyce, Sean, Matt(s), Monica, Marcel, Sasha, Mike, Mark, Ibardo, Jamie, Raphael, Jordi, Cynthia, Huaibin, Alex, Joan, Azad, Xian and Kim, thank you! To my friends Erinn, Nikki, Javi, Mar, Elisa and Loukia I thank for all the great moments, for all the sharing, enlightening discussions and for making my life so much easier and happy.

I thank to all my co-authors and collaborators, especially to Simon Mead and Jordi Clarimon, for the good work, good discussions and good results accomplished over these years.

POPH QREN QUADRO DE REFERÊNCIA ESTRATÉGICO NACIONAL PORTUGAL 2007.2013 UNIÃO EUROPEIA Fundo Social Europeu

LIST OF ORIGINAL PUBLICATIONS

Whole genome analysis in a consanguineous family with early onset Alzheimer's disease.
Clarimón J, Djaldetti R, Lleó A, **Guerreiro RJ**, Molinuevo JL, Paisán-Ruiz C, Gómez-Isla T, Blesa R, Singleton A, Hardy J.
Neurobiol Aging. 2008 Apr 1.
PMID: 18387709

Novel progranulin mutation: screening for PGRN mutations in a Portuguese series of FTD/CBS cases.
Guerreiro RJ, Santana I, Bras JM, Revesz T, Rebelo O, Ribeiro MH, Santiago B, Oliveira CR, Singleton A, Hardy J.
Mov Disord. 2008 Jul 15;23(9):1269-73.
PMID: 18464284

Genetic screening of Alzheimer's disease genes in Iberian and African samples yields novel mutations in presenilins and APP.
Guerreiro RJ, Baquero M, Blesa R, Boada M, Brás JM, Bullido MJ, Calado A, Crook R, Ferreira C, Frank A, Gómez-Isla T, Hernández I, Lleó A, Machado A, Martínez-Lage P, Masdeu J, Molina-Porcel L, Molinuevo JL, Pastor P, Pérez-Tur J, Relvas R, Oliveira CR, Ribeiro MH, Rogaeva E, Sa A, Samaranch L, Sánchez-Valle R, Santana I, Tàrraga L, Valdivieso F, Singleton A, Hardy J, Clarimón J.
Neurobiol Aging. 2008 Jul 28.
PMID: 18667258

Extended tracts of homozygosity identify novel candidate genes associated with late-onset Alzheimer's disease.
Nalls MA*, **Guerreiro RJ*,** Simon-Sanchez J, Bras JT, Traynor BJ, Gibbs JR, Launer L, Hardy J, Singleton AB.
Neurogenetics. 2009 Jul;10(3):183-90.
PMID: 19271249

The heritability and genetics of frontotemporal lobar degeneration
Rohrer JD, **Guerreiro R**, Vandrovcova J, Uphill J, Reiman D, Beck J, Isaacs AM, Authier A, Ferrari R, Fox NC, Mackenzie IR, Collinge J, Warren JD, de Silva R, Holton J, Revesz T, Hardy J, Mead S, Rossor MN.
Neurology. 2009 Nov 3;73(18):1451-6.
PMID: 19884572

Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease.
Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, Pahwa JS, Moskvina V, Dowzell K, Williams A, Jones N, Thomas C, Stretton A, Morgan AR, Lovestone S, Powell J, Proitsi P, Lupton MK, Brayne C, Rubinsztein DC, Gill M, Lawlor B, Lynch A, Morgan K, Brown KS, Passmore PA, Craig D, McGuinness B, Todd S,

Holmes C, Mann D, Smith AD, Love S, Kehoe PG, Hardy J, Mead S, Fox N, Rossor M, Collinge J, Maier W, Jessen F, Schürmann B, van den Bussche H, Heuser I, Kornhuber J, Wiltfang J, Dichgans M, Frölich L, Hampel H, Hüll M, Rujescu D, Goate AM, Kauwe JS, Cruchaga C, Nowotny P, Morris JC, Mayo K, Sleegers K, Bettens K, Engelborghs S, De Deyn PP, Van Broeckhoven C, Livingston G, Bass NJ, Gurling H, McQuillin A, Gwilliam R, Deloukas P, Al-Chalabi A, Shaw CE, Tsolaki M, Singleton AB, **Guerreiro R**, Mühleisen TW, Nöthen MM, Moebus S, Jöckel KH, Klopp N, Wichmann HE, Carrasquillo MM, Pankratz VS, Younkin SG, Holmans PA, O'Donovan M, Owen MJ, Williams J.
Nat Genet. 2009 Oct;41(10):1088-93.
PMID: 19734902

A thorough assessment of benign genetic variability in GRN and MAPT
Guerreiro RJ, Washecka N, Hardy J, Singleton A.
Hum Mutat. 2009 Dec 17.
PMID: 20020531

Genetic variability in CLU and its association with Alzheimer's Disease
Guerreiro RJ, Beck J, Gibbs JR, Santana I, Rossor MN, Schott JM, Nalls MA, Ribeiro H, Santiago B, Fox NC, Oliveira CR, Collinge J, Mead S, Singleton A, Hardy J.
PLoS One. 2010 Mar 3;5(3):e9510

The genetic architecture of Alzheimer's Disease: beyond APP, PSENs and APOE
Guerreiro RJ, Gustafson DR and Hardy J.
(Submitted)

Duplications of APP – but not PRNP – are a significant cause of early onset dementia in a large UK referral series
McNaughton D*, Knight W*, **Guerreiro R**, Ryan N, Poulter M, Nicholl D, Hardy J, Revesz T, Lowe J, Rossor M, Collinge J, Mead S.
(Submitted)

ABBREVIATIONS

AAO	Age at onset
ABI	Applied Biosystems
AD	Alzheimer's disease
APPd	APP duplication
Aß	Amyloid ß
bvFTD	Behavioral variant of frontotemporal dementia
CAA	Cerebral amyloid angiopathy
CADASIL	Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy
CBS	Corticobasal syndrome
cDNA	Complementar DNA
CEPH-HGDP	Centre d'Etude du Polymorphisme Humain – Human Genome Diversity Cell Line Panel
CEU	Caucasian/European
Chr	Chromosome
CI	Confidence interval
CNV	Copy number variant
CSF	Cerebrospinal fluid
cSNP	Coding SNP
Ctrl	Control
DA	Doença de Alzheimer
DFT	Demência frontotemporal
DFT-DNM	Demência frontotemporal associada a doença do neurónio motor
DFT-SUP	Demência frontotemporal com patologia TDP43 negativa e patologia sistema ubiquitina-proteassoma positiva
DFT-TDP tipo3	Demência frontotemporal com patologia TDP43 positiva, tipo 3
DFTvc	Demência frontotemporal variante comportamental
DNA	Deoxyribonucleic acid
DSMIV	Diagnostic and statistical manual of mental disorders – version IV
ECCAC	European collection of cell cultures
EOAD	Early onset Alzheimer's disease
eQTL	Expression quantitative trait locus
Exon-qPCR	Real-time quantitative PCR
Fam hist	Family history
FLAIR	Fluid-attenuated inversion recovery
Fm-qPCR	Fluorescent Microsatellite Quantitative PCR
FTD-MND	Frontotemporal dementia associated with motor neuron disease
FTLD	Frontotemporal lobar degeneration
FTLD-U	FTLD with tau-negative ubiquitin-positive inclusions
FTLD-UPS	FTLD with TDP43-negative and ubiquitin proteosome system-positive inclusions

FTLD/CBS	Frontotemporal lobar degeneration / Corticobasal syndrome
gDNA	Genomic DNA
GEO	Gene expression omnibus
GNT	Graded naming test
GRIP	Genetic research in isolated populations
GWAS	Genome wide association study
HCHWA-D	Hereditary recurrent cerebral hemorrhages with amyloidosis of the Dutch type
HL	Hidrophilic loop
IBMPFD	Inclusion body myopathy and Paget's disease of the bone
ICH	Intracranial hemorrhage
IGV	Illumina genome viewer
kb	Kilobase
LD	Linkage disequilibrium
LH	Luteinizing hormone
LOAD	Late onset Alzheimer's disease
LOD	Log 10 of the odds
LPA	Logopenic/phonological variant of primary progressive aphasia
MAC	Membrane attack complex
MAF	Minor allele frequency
Mb	Megabase
MCI	Mild cognitive impairment
MGB	Minor groove binding
MMSE	Mini-mental state examination
MRI	Magnetic resonance imaging
mRNA	Messenger RNA
NCBI	National Center for Biotechnology Information
NINCDS-ADRDA	National Institute of Neurological Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association
OR	Odds ratio
PCR	Polymerase chain reaction
PNFA	Progressive nonfluent aphasia
PPA	Primary progressive aphasia
PSP	Progressive supranuclear palsy
RMTw/RMTf	Recognition memory test for words and faces
RNA	Ribonucleic acid
ROH	Region of homozygosity
SD	Standard deviation
SemD	Semantic dementia
SNP	Single nucleotide polymorphism

TGen	Translational Genomics Research Institute
TM	Tramsmembrane
UK	United Kingdom
USA	United States of America
VIQ/PIQ	Verbal and performance subscales of WAIS-R

RESUMO

Hoje em dia é claro que patologias como a Doença de Alzheimer (DA) e Demência frontotemporal (DFT) são doenças neurodegenerativas complexas com uma marcada componente genética. Mutações em três genes foram identificadas como a causa de DA familiar de início precoce (*APP*, *PSEN1* e *PSEN2*). No entanto, a forma de DA mais comum é esporádica e de início tardio. A componente genética desta forma tardia da doença tem sido alvo de um elevado numero de estudos, uma vez que apenas se conhece um factor de risco genético robustamente associado com a doença (*APOE*).

Mutações em cinco genes e um locus genético estão na base do desenvolvimento de DFT. Destes, dois genes são causas relativamente comuns da doença (*MAPT* e *GRN*) e três são raros (*VCP*, *CHMP2B* e *TARDP*).

Na primeira parte deste trabalho (capítulo 4), diferentes amostras de DA e DFT foram caracterizadas molecularmente. Este estudo levou à identificação de diversas mutações novas, bem como de mutações anteriormente descritas. Em ambos os casos, os resultados desta análise têm uma elevada relevância em termos clínicos por diversas razões: 1) estas foram as primeiras mutações descritas nos genes associados com DA e DFT na população portuguesa; 2) é agora possível estabelecer uma estimativa da percentagem de casos na população portuguesa com alterações nestes genes; 3) nos casos particulares das famílias onde foram encontradas mutações, os doentes e familiares podem agora ser encaminhados para consultas de aconselhamento genético.

Na DA, a pesquisa de mutações nos genes das presenilinas e *APP* num total de 231 doentes da Península Ibérica com um diagnóstico clínico de DA com início precoce, identificou três novas mutações no gene *PSEN1*, uma nova mutação no gene *PSEN2* e uma nova mutação no gene *APP*. Quatro mutações descritas anteriormente foram também identificadas no gene *PSEN1*.

Para além do elevado numero de mutações pontuais descritas nos genes acima referidos, outro tipo de variações genéticas foram recentemente associadas com a DA: alterações estruturais no cromossoma 21, mais especificamente, duplicações do gene *APP*. O estudo de 1531 doentes com

demência provenientes do Reino Unido, revelou cinco duplicações deste gene. Neste estudo foram usadas diferentes metodologias por forma a avaliar o seu desempenho na identificação destas alterações genéticas. As metodologias usadas (PCR em tempo real e PCR quantitativo com microssatelites de fluorescência) tiveram uma performance semelhante individualmente, mas mostraram ser complementares em termos de sensibilidade e especificidade (tendo como referência os *arrays* de genotipagem de alta densidade comercializados pela empresa Illumina). Uma das cinco duplicações foi encontrada num doente sem história familiar de DA. Esta alteração genética revelou-se peculiar, sendo caracterizada por ser uma duplicação extensa e descontínua do cromossoma 21, consistente com uma translocação não balanceada excluindo a região crítica do síndrome de Down. A nível clínico, os doentes com duplicações típicas apresentaram convulsões generalizadas como sintoma proeminente coincidente com, ou precedendo (em décadas numa família) o início do declínio cognitivo. Este facto sublinhou a importância de expandir o fenótipo geralmente associado a estas alterações genéticas de forma a incluir convulsões e a ausência de história familiar de demência. Em conclusão, duplicações do gene *APP* foram identificadas como uma causa significativa de demência de início precoce no Reino Unido.

Em DFT, a análise molecular dos genes *MAPT* e *GRN* num grupo de doentes do Reino Unido permitiu estudar a herditabilidade nos diferentes subtipos clínicos da doença. Os resultados deste estudo revelaram um elevado grau de herditabilidade em DFT. No entanto, diferentes formas da doença foram associadas a níveis de herditabilidade diferentes, apresentando a DFTvc o grau mais elevado e as variantes DFT-DNM e síndromes da fala (especialmente demência semântica) o menor grau de herditabilidade. Adicionalmente, foram encontradas mutações nos genes *MAPT* em 8.9% e *GRN* em 8.4% do grupo em estudo. Dos restantes doentes sem mutações nestes genes mas com uma história familiar acentuada, sete têm confirmação patológica podendo dividir-se em dois grupos: DFT-TDP tipo 3 sem mutações no gene *GRN* (seis doentes) e DFT-SUP (um doente). Estes resultados indicam nitidamente que, apesar de mutações nos genes *MAPT* e *GRN* serem a causa da doença numa

proporção substancial dos casos familiares de DFT, existem outros genes ainda por identificar, especialmente em doentes com DFT-TDP tipo 3 sem mutações na *GRN*.

A sequenciação do gene *GRN* numa série consecutiva de doentes portugueses com demência frontotemporal/degenerescência corticobasal levou à identificação de duas mutações: uma inserção patogénica nova (p.Gln300GlnfsX61) e uma variante pontual anteriormente descrita, cuja patogeneicidade não é clara (p.T182M). Dos 36 casos índex incluídos nesta série, apenas um apresentou uma mutação patogénica (3% da série estudada), indicando que mutações neste gene não são uma causa frequente de DFT na população portuguesa.

Na segunda parte deste trabalho (capítulo 5), o mesmo tipo de análise descrita no capítulo 4 foi efectuada em controlos, com o intuito de avaliar a patogeneicidade de algumas das variantes descritas neste capítulo. Por forma a determinar a extensão de variabilidade normal nos genes *APP*, *PSEN1* e *PSEN2* foram analisadas amostras de 121 controlos idosos da Península Ibérica e uma série de 130 indivíduos de sete populações Africanas, pertencentes ao painel do Centre d'Etude du Polymorphisme Humain-Human Genome Diversity. Nesta série Africana foram encontradas cinco novas variantes não sinónimas nos três genes, bem como variantes anteriormente descritas, incluindo a variante p.R62H no gene *PSEN2* previamente descrita como associada com a DA. Muitos dos estudos que identificaram mutações nestes genes em doentes e famílias com DA não estudaram controlos e não confirmaram a nível funcional a patogeneicidade das variantes encontradas. Adicionalmente, diferentes estudos classificaram como mutações patogénicas, variantes que posteriormente se concluiu serem benignas, o que claramente enfatiza a importância de discernir entre estes dois tipos de variantes genéticas. Por forma a tornar esta avaliação mais fácil, propusemos e usámos um algoritmo que permite sistematizar e classificar a patogeneicidade das variantes encontradas nos genes das presenilinas.

A mesma questão foi para nós óbvia em relação às variantes encontradas em doentes de DFT. Desta forma, os genes *GRN* e *MAPT* foram sequenciados em 282 amostras do painel do Centre d'Etude du Polymorphisme Humain-Human Genome Diversity, por forma a identificar variantes benignas. Neste

estudo foram identificadas dezasseis alterações genéticas não sinónimas diferentes, das quais, onze são novas. Os resultados obtidos neste estudo questionam seriamente resultados obtidos anteriormente por outros grupos relativos a patogeneicidade de variantes pontuais no gene *GRN*. Adicionalmente, este estudo identificou novas variantes genéticas nos genes *GRN* e *MAPT* que poderiam facilmente ser classificadas como patogénicas, caso tivessem sido encontradas em doentes.

Na terceira parte deste trabalho (capítulo 6), são descritas diferentes abordagens ao uso de tecnologias recentemente desenvolvidas (plataformas de genotipagem de elevada densidade), com o objectivo de identificar genes e factores de risco envolvidos na DA. Desta forma, foi feita uma avaliação do papel da homozigosidade genética na doença de Alzheimer e participámos no maior estudo de associação analisando todo o genoma em DA.

Regiões extensas de homozigosidade podem hoje em dia ser facilmente localizadas permitindo a identificação de variantes de risco recessivas e raras que contribuam para a doença. A comparação de diferentes métricas associadas a estas regiões de homozigosidade numa população de 837 doentes de Alzheimer de início tardio e 550 controlos saudáveis permitiu-nos confirmar que este tipo de regiões são comuns no genoma de populações "outbred" e concluir que regiões de homozigosidade extensas especificas podem ser consideradas um factor de risco para o desenvolvimento de DA. Assim sendo, identificámos uma região de homozigosidade no cromossoma 8 significativamente associada com DA de início tardio. Esta região contem sete genes, dos quais, *STAR*, *EIF4EBP1* e *ADRB3* foram considerados os candidatos biologicamente mais plausíveis. Estes resultados sugerem pela primeira vez, uma componente recessiva na etiologia da Doença de Alzheimer de início tardio.

Relativamente ao estudo da homozigosidade na doença de Alzheimer de início precoce analisámos uma família Israelita com dois irmãos com DA de início precoce cujos pais eram neurologicamente saudáveis e primos em primeiro grau (ambos os pais morreram depois dos 90 anos de idade). A análise molecular dos genes *PSEN1*, *PSEN2*, *APP*, *MAPT*, *GRN* e *PRNP* não revelou qualquer mutação

nos irmãos. Uma vez que a doença nesta família apresentava um padrão consistente com uma herditabilidade autossómica recessiva, recorrendo a plataformas de genotipagem de elevada densidade, identificámos todas as regiões homozigóticas idênticas por descendência em ambos os irmãos. Este estudo resultou no primeiro catálogo de autozigosidade em DA de início precoce e os resultados obtidos sugerem que as regiões identificadas são excelentes *loci* para a ocorrência de uma lesão genética recessiva causadora da doença nesta família.

Uma abordagem diferente na nossa tentativa de identificar genes novos envolvidos na DA levou-nos a participar num estudo de associação analisando polimorfismos em todo o genoma. Este estudo englobou duas fases e incluiu mais de 16000 amostras, sendo o estudo deste género com maior poder estatístico até hoje. Na primeira fase deste estudo foram estudados 3941 casos e 7848 controlos, tendo sido replicada a associação com o *locus APOE* (sendo o polimorfismo mais significante: rs2075650, $P=1.8x10^{-157}$). Adicionalmente, foram identificadas associações significativas com dois outros *loci*: no gene *CLU* (rs11136000, $P=1.4x10^{-9}$) e na região 5' não traduzida do gene *PICALM* (rs3851179, $P=1.9x10^{-8}$). Estas associações foram replicadas na segunda fase deste estudo (2023 casos e 2340 controlos) gerando provas convincentes para o envolvimento destes genes na DA (resultados da análise conjunta das duas fases: rs11136000, $P=8.5x10^{-10}$, odds ratio = 0.86; rs3851179, $P=1.3x10^{-9}$, odds ratio = 0.86). A associação mais forte, estabelecida fora do *locus* da *APOE*, foi descrita relativamente ao polimorfismo intrónico rs11136000 no gene *CLU*. Por forma a melhor caracterizar esta associação, sequenciámos a região codificante deste gene num total de 495 casos de Alzheimer e 330 controlos. Nos casos em que as variantes foram encontradas em mais do que um indivíduo, as frequências genotípicas foram comparadas entre os dois grupos em estudo. Foram identificadas vinte variantes genéticas em casos e controlos, incluindo uma mutação *nonsense* num indivíduo controlo, o que indica que a patogenecidade de variantes encontradas neste gene deve ser cuidadosamente avaliada. Por forma a determinar se variantes comuns no gene *CLU* têm efeito na expressão de transcritos de mRNA foi realizada uma análise da expressão de *loci*

quantitativos (eQTL), não tendo sido observadas associações significativas em eQTLs para os SNPs previamente associados com a Doença de Alzheimer.

SUMMARY

Alzheimer's disease and frontotemporal lobar degeneration are complex neurodegenerative disorders with a clear genetic component.

Three genes (*APP*, *PSEN1* and *PSEN2*) have been identified as the cause of early onset familial AD. The commonest form of the disease is, however, a sporadic one presenting itself in later stages of life. The genetic component of this late onset form of AD has been the target of a large number of studies, since only one genetic risk factor (APOE) has been consistently associated with the disease.

Five genes and one confirmed genetic locus are currently known to cause FTLD of which two genes are relatively common (*MAPT* and *GRN*) and three are rare causes (*VCP*, *CHMP2B* and *TARDP*).

In the first part of this work (chapter 4), the molecular characterization of different AD and FTLD samples lead us to the identification of several new and already known mutations.

In AD, screening for mutations in presenilins and *APP* in a total of 231 patients from the Iberian peninsular with a clinical diagnosis of early-onset AD, identified three novel mutations in *PSEN1*, one novel mutation in *PSEN2*, and a novel mutation in the *APP* gene. Four previously described mutations in *PSEN1* were also found.

The screening of 1531 dementia probands for *APP* duplications uncovered five *APP* duplications. To perform this analysis, different methodologies were used: real-time quantitative PCR and fluorescent microsatellite quantitative PCR behaved in a similar way, individually; but were complementary in terms of sensitivity and specificity (based on using the Illumina array as the gold standard assay). One of the five APP duplications was found in a patient with no known family history. This was an extensive and non-contiguous duplication on chromosome 21, consistent with an unbalanced translocation that did not include the Down's syndrome critical region. Clinically, seizures were found to be prominent in the other typical *APP* duplications found, occurring at presentation or predating cognitive decline by decades in one family. This stressed the fact that the recognized phenotype associated with *APP* duplications should be expanded to include early seizures and apparently sporadic disease which, in part, may be caused by a novel mutational

mechanism. Overall, *APP* duplications were found to be a significant cause of early onset dementia in the UK.

In FTLD the molecular screening of *MAPT* and *GRN* in a large cohort from the UK allowed the study of heritability in the different clinical subtypes of this disorder, revealing that FTLD is a highly heritable disorder but that heritability varies between the different syndromes with the behavioral variant being the most heritable and FTD-MND and the language syndromes (particularly semantic dementia) the least heritable. Mutations were found in *MAPT* (8.9% of the cohort) and *GRN* (8.4%). Of the remaining patients without mutations but with a strong family history, seven had pathological confirmation, falling into two groups: type 3 FTLD-TDP without *GRN* mutations (six) and FTLD-UPS (one). These results show that whilst *MAPT* and *GRN* mutations account for a substantial proportion of familial cases, there are other genes yet to be discovered (particularly in patients with type 3 FTLD-TDP without *GRN* mutations).

Furthermore, the sequencing of *GRN* in a consecutive series of 46 FTLD/CBS Portuguese patients lead us to the identification of two mutations: a novel pathogenic insertion (p.Gln300GlnfsX61) and a previously described point variant (p.T182M) of unclear pathogenicity. Pathogenic mutations in the *GRN* gene were found in one of the 36 probands studied (3% of the probands in our series) who had a corticobasal syndrome presentation, indicating that in the Portuguese population, mutations in this gene are not a major cause of FTLD.

In the second part of this work (chapter 5) and in order to evaluate the pathogenicity of several of the variants found to be associated with AD and FTLD in chapter 4, the same type of analysis was carried in control individuals. To determine the extent of normal variability in *APP*, *PSEN1* and *PSEN2*, DNA samples from 121 elderly healthy controls from the Iberian peninsular, and a set of 130 individuals from seven African populations belonging to the Centre d'Etude du Polymorphisme Humain-Human Genome Diversity Panel were analyzed. In the latter series, we found five new non-synonymous changes in all three genes and a presenilin 2 variant (p.R62H) that had been previously related to AD. However, for some of the mutations found in AD cases, the pathologic consequence

continued uncertain. To address this issue we proposed and used a systematic algorithm to classify the putative pathogenicity of AD mutations. The previous miss-assignment of pathogenicity to benign variants clearly stressed the importance of discerning between disease causing mutations and benign variants with no pathogenic effect on the function of the respective protein. The same issue was obvious for the variants found in FTLD cases. To address this, we sequenced *GRN* and *MAPT* in 282 samples from the Centre d'Etude du Polymorphisme Humain- Human Genome Diversity Cell Line Panel, in order to identify benign variants that could otherwise be mistaken for pathogenic mutations. We identified sixteen different non-synonymous changes, eleven of which were novel variants. This particular study added to the classification of variants within *MAPT* and *GRN*, and, identified novel variants that could easily be mistaken as pathogenic mutations in future studies.

In the third part of this work (chapter 6) we tried to identify new genes associated with Alzheimer's disease. To do this, we took advantage of recently developed high-density genotyping platforms to assess the role of homozygosity in early and late onset AD and participated in the largest AD GWAS to date. Regions of extended homozygosity can now be accurately located allowing for the identification of rare recessive risk variants contributing to disease. The comparison of measures of extended homozygosity (greater than 1 Mb in length) in a population of 837 late-onset Alzheimer's disease cases and 550 controls allowed the confirmation that extended runs of homozygosity are common in outbred populations and the conclusion that specific large tracts of homozygosity are a risk factor for Alzheimer's disease. We identified one homozygous region on chromosome 8 significantly associated with LOAD. This region contains seven genes from which the most biologically plausible candidates are *STAR*, *EIF4EBP1*, and *ADRB3*. These results suggested a recessive component to the etiology of LOAD.

In EOAD we studied the clinical characteristics of an Israeli family comprising two affected siblings with EOAD born to neurologically healthy parents who were first cousins (both parents died after 90 years old). Sequence analysis of *PSEN1*, *PSEN2*, *APP*, *MAPT*, *GRN*, and *PRNP* failed to reveal any

mutations in the affected siblings. Since the disease in this family was consistent with an autosomal recessive mode of inheritance we identified all homozygous regions identical by descent in both siblings, by high-density SNP genotyping. This study provided the first catalog of autozygosity in EOAD and the results obtained suggested that the regions identified are excellent candidate loci for a recessive genetic lesion causing this disease.

A different approach in our attempt to identify new genes in AD directed us to participate in a large two-stage genome-wide association study involving over 16,000 individuals (the most powerful AD GWAS to date). In stage 1 (3,941 cases and 7,848 controls), the established association with the *APOE* locus (most significant SNP, rs2075650, $P = 1.8\times10^{-157}$) was replicated and the genome-wide significant association with SNPs at two loci not previously associated with the disease: at the *CLU* gene (rs11136000, $P = 1.4\times10^{-9}$) and 5' to the *PICALM* gene (rs3851179, $P = 1.9\times10^{-8}$), was observed. These associations were replicated in stage 2 (2,023 cases and 2,340 controls), producing compelling evidence for association with Alzheimer's disease in the combined dataset (rs11136000, $P = 8.5\times10^{-10}$, odds ratio = 0.86; rs3851179, $P = 1.3\times10^{-9}$, odds ratio = 0.86).

The strongest association was reported for the intronic SNP rs11136000 in *CLU*. To further characterize this association we sequenced the coding region of this gene in 495 AD cases and 330 healthy controls. A total of twenty-four variants were found both in cases and controls. For the changes found in more than one individual, the genotypic frequencies were compared between cases and controls. Coding variants were found in both groups (including a nonsense mutation in a healthy subject), indicating that the pathogenicity of variants found in this gene must be carefully evaluated. In order to determine if common variants at the *CLU* locus effect expression of nearby (cis) mRNA transcripts, an expression quantitative loci (eQTL) analysis was performed. No significant eQTL associations were observed for the SNPs previously associated with AD.

CONTENTS

CHAPTER 1

INTRODUCTION

Based on:

1) **Guerreiro RJ**, Gustafson DR and Hardy J (2009) The genetic architecture of Alzheimer's Disease: beyond APP, PSENs and APOE (submitted).

CHAPTER 1 – INTRODUCTION

1.1 – Genetics of Alzheimer's Disease

Alzheimer's disease (AD) is the most common cause of dementia. It is a progressive neurodegenerative disorder with an insidious onset, which typically appears in older individuals, but may affect people as early as the third decade of life [1, 2]. The genetics of Alzheimer's disease is usually described as complex and heterogeneous. Most of the cases are "sporadic" with no apparent familial recurrence of the disease. However, a small percentage of AD cases (1-2% of all cases) have an early onset (EOAD), with symptoms appearing before 65 years of age. In these patients, the disease commonly aggregates within families and typically presents an autosomal dominant pattern of inheritance. Mutations in three genes are known to account for this early onset, familial type of the disease: in the amyloid precursor protein gene (*APP*), in the presenilin 1 gene (*PSEN1*) and in the presenilin 2 gene (*PSEN2*) [2]. In fact, early-onset autosomal dominant disease with an onset under sixty years seems to be completely explained by pathogenic mutations in these three genes. The most common, late onset and sporadic form of the disease remains mostly a genetic conundrum. The only well established genetic risk factor for late onset AD (LOAD) is the E4 allele of apolipoprotein E. It has been estimated that these four established AD genes probably account for less than 30% of the genetic variance of AD, suggesting that numerous additional AD genes may exist [3]. The pursuit of these additional genes has been unproductive until recently. In the last few years, the development of high throughput technologies able to genotype up to one million single nucleotide polymorphisms (SNPs) has revealed some of the genetic players in different complex disorders. With this in mind, in this introductory section, we will consolidate the current knowledge on the role of APP, PSENs and APOE in AD, we will systematically review the most promising loci and variants reported to be associated with AD and we will discuss the recent advances in AD genome wide analysis.

1.1.1 - EOAD Causative Genes

APP

The amyloid precursor protein gene (*APP*, OMIM 104760) consists of eighteen exons and spans more than 290 kb of chromosome 21q21. The APP protein is an ubiquitously expressed, integral type I membrane glycoprotein that, when fully expressed, contains a short cytoplasmic tail, a relative large extracellular portion and a transmembrane domain [4, 5]. It exists as three major alternatively spliced isoforms: APP751 (encoded by exons 1-7, 9-18) and APP770 (encoded by exons 1-18), both containing the Kunitz protease inhibitor domain (encoded by exon 7) within the extracellular portion are the major non-neuronal forms; and APP695 (encoded by exons 1-6 and 9-18), which does not contain the Kunitz domain and is the main isoform found in brain [5]. The proteolytic processing of APP gives rise to amyloidogenic and non-amyloidogenic fragments, depending if the Aβ peptide is a result of this processing, or not. The Aβ peptide is encoded by parts of exons 16 and 17 and is partially contained in both the extracellular and transmembrane domains of the APP protein [2, 4, 5].

The proteolytic processing of APP may occur through two mutually exclusive pathways: the amyloidogenic pathway (fundamentally considered as the pathogenic pathway) and the non-amyloidogenic or constitutive pathway [2, 4, 5]. The cleavage of APP occurs by either alpha or beta-secretases and produces soluble N-terminal fragments plus membrane-bound C-terminal fragments (C83 and C99, respectively). The subsequent cleavage by gamma-secretase results in the release and secretion of a small non-pathogenic peptide (p3) when the first cleavage occurred by alpha-secretase, or Aβ (previous beta-secretase cleavage). Different Aβ peptides may be generated, depending on the exact site where the gamma-secretase cleavage occurred, varying from Aβ 38 to Aβ 43 (depending on the number of amino acids).

Although the exact function of APP is still to be resolved, the intracellular domains produced in both pathways are thought to have different functions such as neurogenesis, transcriptional regulation

and signal transduction [2, 4]. More recently the N-terminal cleavage products of APP were shown to bind the death receptor 6 to activate neurodegeneration in neurons deprived of trophic factors [6]. This study adds to the amyloid cascade hypothesis proposed in 1991 [7], which has been the central theory leading AD research since it was proposed.

The identification of Aβ as a metabolic product of APP and the reports of AD families harboring *APP* causative mutations lead to the general concept that Aβ is a key player in the development of AD, and that EOAD mutations influence the properties or ratios of the different Aβ isoforms in the brain [8]. Mutations in *APP* are, however, a rare cause of AD with an estimated frequency of 16% of familial EOAD patients [9]. The first *APP* mutation (p.G693Q) was found in 1991 in a family comprising patients suffering from hereditary recurrent cerebral hemorrhages with amyloidosis of the Dutch type (HCHWA-D) [10]. This mutation induced recurrent hemorrhagic strokes due to excessive Aβ deposition in the cerebral and leptomeningeal blood vessel walls. The elusive role of *APP* in AD genetics was definitely resolved when in 1991 Goate and colleagues found the p.V717I mutation segregating in an EOAD kindred [11]. Several other *APP* missense mutations have been identified in the following years. Now, 24 different pathogenic missense mutations and 6 nonpathogenic or with unclear pathogenicity variants have been described (http://www.molgen.ua.ac.be/ADMutations, assessed on April 27th, 2009). All clearly pathogenic mutations seem to be located in or close to the proteolytic cleavage sites and act by preventing the normal metabolic processing of APP with a consequent deregulation of AB levels.

More recently, a mutation in *APP* (A673V) has been found to cause AD only in the homozygous state, in a family with an apparent recessive mode of inheritance. The mutation was shown to affect APP processing resulting in enhanced β-amyloid production and development of amyloid fibrils in vitro. Additionally, the co-incubation of mutated and wild type peptides conferred instability on Aβ aggregates and inhibited amyloidogenesis and neurotoxicity. This highly amyloidogenic effect of the mutation in the homozygous state and its anti-amyloidogenic effect in the heterozygous state were proposed to account for the autosomal recessive pattern of inheritance present in the family [12].

In addition to missense variants, copy number mutations have recently been identified in autosomal dominant early-onset families. Five French families [13, 14] were first reported to harbor small chromosomal duplications with different break points, but all including the *APP* locus. The prevalence and phenotypic spectrum of *APP* duplications are yet to be fully defined, including mixed phenotypes of AD and/or cerebral amyloid angiopathy (CAA). The estimated frequency in the selected Rovelet-Lecrux cohort was 8%, about half the contribution of missense *APP* mutations to early onset, autosomal dominant AD [9]. A subsequent screen of Finnish cases of autosomal dominant EOAD with prominent CAA also revealed *APP* duplications in some individuals [15-17].

PSEN1 AND PSEN2

The presenilin 1 gene (*PSEN1*, OMIM 104311) comprises thirteen exons and spans over 83 kb in chromosome 14q24.3. The first four exons contain untranslated sequence, and exons 1 and 2 represent alternate transcription initiation sites [18]. The presenilin 2 gene (*PSEN2*, OMIM 600759), spanning over 25 kb on chromosome 1q31-q42, has a very similar genetic structure with twelve exons, from which ten are coding exons and the first two encode the 5-prime untranslated region [19].

Both PSEN1 and PSEN2 are expressed in a multiplicity of tissues including the brain, with higher levels in the cerebellum and the hippocampus and a primarily neuronal expression [20].These two proteins are highly homologous, sharing an overall amino acid sequence identity of 67%. Hydrophobicity plots predicted these to be integral membrane proteins [20] most likely adopting a nine transmembrane structure with a hydrophilic intracellular loop region [21, 22]. PSENs are predominantly located in the endoplasmic reticulum and Golgi compartments, clearly suggesting their involvement in protein processing [23].

PSEN1 and PSEN2 are important components of the multimeric gama-secretase complex that also includes the proteins nicastrin, APH1A and PEN2 [24]. In addition to the role in β-amyloid

production, PSENs have also been found to be involved in an array of biological processes such as modulation of the Notch signaling pathway [25]; cell adhesion [26, 27]; G-protein mediated signal transduction [28] and in the unfolded protein response [29, 30].

The first disease causing mutations in *PSEN1* and *PSEN2* were identified in 1995 [20, 31]. At the present time, 175 pathogenic mutations and 7 non-pathogenic variants or changes with unclear pathogenicity have been identified in *PSEN1*. *PSEN2* harbors a much smaller number of mutations: 14 pathogenic mutations and 9 non-pathogenic or with unclear pathogenicity variants (http://www.molgen.ua.ac.be/ADMutations, accessed on May 2nd, 2009). Overall, AD patients with mutations in this gene present a later age of onset of the disease and a slower disease progression, when compared with patients carrying *APP* or *PSEN1* mutations .The PSENs mutation range encompasses mainly missense mutations scattered all over the proteins with some clustering around transmembrane domains [32, 33]. The integration of PSEN mutations within the amyloid cascade is based on several lines of evidence: 1) *PSEN1* and *PSEN2* mutation carriers present relatively elevated plasma Aβ42 levels compared to Aβ40 [34]; 2) *PSEN1* mutation carriers have a significant increase in Aβ42, but not Aβ40, in brain tissue, when compared to sporadic AD patients [35]; 3) transgenic mouse models overexpressing mutant PSENs present a relatively increased production of longer Aβ species [36]. Still, this disruption of the Aβ ratio by *PSEN1* and *PSEN2* mutations may result from either an increase in Aβ42 levels or from a decrease in Aβ40 levels [37].

1.1.2 – The Genetic Component of LOAD

APOE

The Apolipoprotein E gene (*APOE*, OMIM 107741) contains four exons, three of which are coding exons. The gene spans over a 3.5 kb region of chromosome 19q13.2. The Apolipoprotein E is a plasma glycoprotein synthesized mainly in the liver, brain (primarily by neurons and astrocytes), and also by other cells such as macrophages and monocytes [38]. ApoE is involved in the mobilization

and redistribution of cholesterol during neuronal growth and after injury [39]; in nerve regeneration, immunoregulation and activation of several lipolytic enzymes [40].

The three major ApoE isoforms (ApoE2, ApoE3 and ApoE4) differ in two sites of the amino acid sequence (residues 112 and 158) and are encoded by a single genetic locus. The frequencies of the ε2, ε3, and ε4 alleles were estimated at 0.11, 0.72, and 0.17, respectively, but vary widely among populations [41]. The ε4 allele, (the ancestral allele), is more frequent in populations such as Pygmies (0.407) and Khoi San (0.370). The E2 allele frequency oscillates with no apparent trend and is absent in Native Americans [42].

Many studies have demonstrated an association between the ε4 allele and late-onset familial and sporadic forms of Alzheimer's disease. This allele represents an increased risk seen across different ethnic groups of 3 fold for heterozygous carriers and up to 15 fold for individuals with the E4E4 genotype, when compared to E3E3 genotypes.

ApoE is known to act in a dose dependent manner in AD: the effect of the ε4 allele in the risk for AD increases from 20 to 90% and the mean age of onset decreases from 84 to 68 years with the increase in the number of ε4 alleles [43]. The ε2 allele has been shown to have an impact on longevity and may confer protection against Alzheimer's disease [44].

Distinct binding properties of the different ApoE isoforms to the Aβ peptide [45] and tau protein [46] have been suggested to underlie the disparities associated with each genotype. In particular, the ApoE4 isoform binds to the Aβ peptide more rapidly than the ApoE3 isoform, forming novel monofibrils that precipitate into dense structures [47]. The fact that ApoE4 does not bind to tau protein in vitro, unlike ApoE2 and ApoE3, has raised the possibility that this interaction between ApoE3 and tau serves as a protection against tau phosphorylation and consequent neurofibrillary tangle formation [46, 48]. Nonetheless, the ε4 allele appears to be a risk factor and not an invariant cause of AD, indicating that other environmental or genetic factors may need to be concurrently acting with this allele in order to cause AD [49]. Additionally, while most of the risk at the *APOE* locus

is likely to be encoded at the protein coding polymorphism, it was always clear that other genetic variability at this locus, probably altering ApoE expression, also contributes to the risk of developing AD [50-53]. It seems more likely that this genetic variability in ApoE expression contributes to disease risk, rather than an independent effect of the adjacent gene *TOMM40* contributes to risk.

During the past two decades there have been many studies searching for genetic risk factors: thousands of positive and negative results have been produced, but none apart from ApoE have produced clear and reproducible associations.

1.1.3 - Different Types of Studies in AD Genetics

Methodologically, two main strategies have governed the field: genetic linkage analysis and case-control association studies. For the first strategy, researchers have used informative families where a clear heritability of the disease is present and no mutations have been found. In case-control association studies, researchers rely on the comparison of the frequencies of a determined allele between a group of AD patients and a group of healthy individuals. I will discuss the currently known genes and risk factors associated with AD, with a focus on genome wide association studies as a methodology, as well as the previously overlooked role of homozygosity and recessive cases in the major complexity of AD genetics.

Linkage studies

Genetic linkage studies aim to identify chromosomal regions associated with disease by measuring the correlated segregation of particular markers with a determined phenotype within a family. This is quantitatively expressed as lod scores (log-likelihood ratio of the existence of linkage versus the inexistence of linkage). Linkage tests may be done using parametric or non-parametric analysis. In the first case a specific inheritance model is assumed, while the second type of analysis is said to be

model-free. Depending on the type of disease and information available on the studied family/families, one of these analysis will be more appropriate [54].

This type of studies usually involves three sequential steps: 1) the identification of the disease causative locus; 2) sequencing the region found in the previous step in a cohort of cases and controls in order to define and characterize the mutation(s) found; 3) uncover the molecular and biological functions of the genes found [55]. Several factors are known to complicate this approach in AD: i) difficulties in getting large, complete and informational multigeneration families; ii) the potential inclusion of phenocopies (individuals with a sporadic indistinguishable form of disease); and iii) genetic heterogeneity, since the fact that the pattern of disease in families is consistent with a major gene component does not necessarily imply that only one gene or factor is involved. Additionally, linkage mapping suffers from limitations such as the low resolution of the results: usually these studies do not identify one gene or one mutation associated with a disease, instead, a chromosomal region (many times, a very large region), is identified; also, the strongest linkage signals tend to come from recessive and highly penetrant, thus very rare, disorders [54].

Nonetheless, linkage mapping has been a very important methodology in the study of AD genetics: the four genes undoubtedly associated with AD have been identified primarily by linkage analysis. In addition to these genes, several other genomic regions have been implicated using this method (Table 1). These loci contain innumerous genes that have been considered candidate genes and consequently have been studied in order to identify the real genetic variation responsible for the development of AD. Until now, no specific genes implicated in AD have been identified in these regions and linkage for risk factors has failed, probably due to low odds ratio associated with AD. Most recently, a study by Butler and colleagues used a genome search meta-analysis method to analyze the pooled linkage results from five independent genome scans. These included the results of 2206 affected individuals and 785 families from Caucasian and Caribbean Hispanic ethnicities. This study was able to identify genome-wide suggestive evidence for linkage on chromosomes 1p13.3-q31.1, 7pter-p21.1 and 8p22-p21.1, together with other seven loci presenting nominally significant

evidence for linkage [56]. Interestingly, the most significant locus identified in this study (8p22-p21.1) includes the CLU gene, the top hit from the largest GWAS performed in AD, as discussed below in this chapter.

Table 1: Ten most interesting AD linkage regions.

Chr	Region	Number of studies	Studies	LOD scores	Relevant genes
1	1p13.3-q23.3	2	[57]; [56]	5.2	GSTM4; GSTM1; GSTM3; CSF1; NGF; HMGCS2, PRKAB2; APH1A; CTSS; THEM5; FAM63A; CHRNB2; LMNA; PMVK; FDPS; APOA1BP; GBA; NTRK1; CRP; NCSTN
	1q23.3-q31.1	3	[57]; [56]; [58]	2.1 – 4.0	F11R; USF1; FCER1G; RGS4; APOA2; RXRG, POU2F1; PRDX6; SOAT1; PTGS2
5	5p13-p15	6	[59]; [60]; [61]; [62]; [58]; [63]	1.4 - 2.8	SLC6A3; PRKAA1
8	8p22-p21.1	2	[56]; [64]	>2.0	NAT1; NAT2; LPL; ADRA1A; CHRNA2; CLU
9	9p21	5	[59]; [60]; [62]; [65]; [66]	>1.0 - 4.6	IFT74
	9q22-q34	8	[59]; [60]; [61]; [62]; [58]; [67]; [63]; [66]	1.6 - 4.2	FBP1; GOLM1; ABCA1; DFNB31; TLR4; NDUFA8; PSMB7; HSPA5; POMT1; DBH; RXRA; TRAF2; ABCA2
10	10q21-q22	7	[60]; [61]; [62]; [58]; [67]; [57]; [66]	1.8 - 4.15	ZWINT; UBE2D1; TFAM; BICC1; ANK3; CDC2; EGR2; CTNNA3; LRRTM3; DNAJC12; SIRT1; SRGN; SUPV3L1; TSPAN15; VPS26A; HK1; TACR2; NEUROG3; SAR1A; SGPL1; PSAP; CHST3; PPP3CB; SEC24C; NDST2; CAMK2G; PLAU; VCL; AP3M1; MYST4; KCNMA1
12	12p11-p13	4	[68]; [60]; [62]; [67]	1.4 - 3.9	TNFRSF1A; CNAP1; GAPDH; GNB3; C1R; APOBEC1; MMP3; A2M; PZP; A2MP; OLR1; LRP6; GRIN2B; GYS2; ABCC9; PKP2P1
19	19q12-q13.33	9	[59]; [60]; [61]; [62]; [69]; [58]; [67]; [66]; [70]; [56]	1.6 - 7.7	LRP3; USF2; GAPDHS; PSENEN; AKT2; TGFB1; LIPE; XRCC1; BCL3; APOE ; PVRL2; TOMM40; APOC1; APOC2; ERCC2; CARD8; GYS1; LHB; CD33; NR1H2
21	21q21-q22	4	[61]; [62]; [58]; [67]	1.6 - 4.5	PRSS7; NCAM2; APP; C21orf63; C21orf55; RUNX1; C21orf55; DYRK1A; KCNJ6; BACE2

These regions were selected either from the top 3 regions resulting from the recent meta-analysis performed by Butler and colleagues, or from the overlapping between four or more whole-genome studies assessing linkage with AD risk, based on the AlzGene database of concordant linkage/association regions observed in full genome screens (http://www.alzforum.org/res/com/gen/alzgene/linkage.asp). The relevant genes for each of the represented regions were obtained from crosschecking with the AlzGene database (www.alzgene.org, [71]). Note that: results by [72] and [73] were not considered as they were significantly extended in the same groups' follow-up genome screen (Myers, et al, 2002 and Lee et al., 2006, respectively). The results by Hamshere, et al., 2007 consist of an amalgamated of three data sets previously studied by Kehoe, et al., 1999; Myers, et al., 2002; Blacker, et al., 2003 and Holmans, et al., 2005. Other studies have been published that performed re-analyses of existing data either by incorporating other variables (such as evidence for AD with and without psychosis [e.g. [74], [75], [76]), or parent of origin effects e.g. [77]. Results by [78] and [79] on Amish families were not considered due to the inclusion of MCI cases.

Gene association studies

The quest for genetic risk factors in clinical genetics has mainly focused on the study of candidate gene (usually focusing only in variants altering the coding sequence of a gene). These types of studies rely on a rather simple principle: to test if a determined allelic or genotypic variant occurs more or less often in a group of people with a particular disease when compared to a similar group of healthy individuals.

The success of this approach relies in an in-depth understanding of the disease and disease pathways, in such a way that the researcher will be able to select, not only the right gene, but also the right variant(s) to be studied – a clear needle in a haystack quest. Additionally, the two studied groups need to be homogeneous, well characterized and large enough to allow a statistically powerful analysis [80].

Most gene association studies in AD have studied one or two variants in one or two genes. The large number of genes and even more variants clearly reduce the chances of true positive findings. Nonetheless, several thousands of positive associations have been reported in the literature. Most of these are certainly false positives resulting mainly from population substructure, poor statistical analysis and publication bias toward positive results [81, 82]. Positive associations have been broadly published, while negative results (unless convincingly refuting previous results) would not be reported. When reported, most of these studies pointed the small sample sizes, or specific genetic population backgrounds as main reasons for the negative results. Other issues, particularly in large epidemiologic studies include case definition or the ability to accurately identify AD versus vascular or other forms of dementia; age of dementia onset estimation; and phenotypic variations in the disease.

In order to address these very large numbers of conflicting reports, a database (the AlzGene database) was created which systematically collects, summarizes and meta-analyzes the results for all the genetic variants studied in association to AD [71]. As of August 31st 2009 the top ten results in

this database included: *ApoE* (E2/3/4), *ACE*, *TFAM*, *IL1B*, *CST3*, *TNK1*, *CHRNB2*, *SORL1*, *SORCS1* and *hCG2039140*. The possibility of some of these genes being associated with AD has been reviewed elsewhere [83]. The second top hit, *ACE*, has repeatedly been reported as associated with AD [84], atherosclerosis [85] and hypertension [86].

Genome wide association studies

The development of platforms able to genotype up to one million SNPs and of powerful analytical frameworks able to distinguish true associations, together with the completion of the International Human HapMap project [87], have provided unprecedented tools to the study of the so called "common disease-common variant" hypothesis. This theory proposes that common polymorphisms (usually defined as having a minor allele frequency of over 1%) may contribute to the overall susceptibility to common diseases [88-90]. This type of analysis addresses one of the major pitfalls of case-control association studies: the coverage of the study. Instead of studying one or two genetic variants we are now able to inquire the majority of common variability in the genome, by means of testing tagging SNPs, i.e., polymorphisms in linkage disequilibrium with each other. This means that if one knows the genotype in one locus, one can predict with a high accuracy [depending mainly on the strength of the linkage disequilibrium (LD) and the allele frequencies] the genotypes occurring at linked loci nearby. However, these new platforms continue to require the study of large numbers of samples and the power of the study remains a major issue. In this way, the general success of any genome wide association studies (GWAS) depends on factors as sample size, frequency of risk alleles and individual effect sizes. The smaller the risk associated with any given common variant, the greater the number of samples needed to identify the variant. The risk allele frequency relates to the effect size: a risk variant that has a high odds ratio but is rare in the population is much more difficult to identify with GWAS than a common variant linked to moderate or mild risk within the population. Additionally, population stratification (when the case and control groups are not well

matched genetically), population admixture (when several distinct, unrecognized sub-populations make part of the same cohort) and non-reported relatedness in the population may also be problems if the study is not correctly designed [91].

GWAS in Alzheimer's disease

Since ApoE has been the only risk factor consistently associated with LOAD and some reports have estimated a heritability of AD between 60 and 80% [92, 93], it is believed that genetic variability plays a critical role in LOAD and that several risk factors are still to be uncovered. Daw and colleagues had estimated the existence of four to seven additional genes contributing to this genetic variability [3]. The first major result from the GWAS in AD showed that there was not another locus conferring the same risk as APOE and it is now known that any additional risk factor(s) in AD exerts a small effect on the risk of disease (odds ratio 1.1-2.0). Clearly this implies that large samples are needed in order to have definite and conclusive results.

Several GWAS in AD have now been published. The first of these studies was performed by Grupe et al, in which the approach used was to first generate a short list of candidate SNPs by analyzing 17 343 SNPs using a Celera cSNP platform in DNA pools from one set of cases and controls. Markers meeting a previously defined significance criteria were then typed in DNA pools from a second set of samples and, again the SNPs satisfying the criteria were then individually genotyped in four sets of cases and controls and finally in a fifth sample. In total, 3870 samples from the United States and the United Kingdom were studied, from which 1808 were LOAD cases and 2062 controls. In this study a large number of candidate SNPs was identified, but only SNPs in LD with APOE were reported to be study-wide significant [94].

Coon et al reported on 1086 individuals (664 neuropathologically confirmed AD cases and 422 healthy subjects) [95]. Reiman et al reported on a total of 1146 individuals in two main series: the discovery set (including 446 neurophatologically confirmed AD cases and 290 healthy controls) and

the replication set (including a neurophatologic sample of 197 AD cases and 114 controls and a clinical sample of 218 LOAD cases and 146 controls) [96]. This study used the same population as the study by Coon et al, enriched with clinical samples. The analysis performed stratified the sample according to ApoE genotype, which allowed the detection of a modifying effect on the risk of AD associated with *GAB2* among ApoE E4 carriers.

Liu et al reported on a total of 103 LOAD patients and 170 first-degree relatives from the Genetic Research in Isolated Populations (GRIP) program carried out in a recently isolated population from The Netherlands. The significant results from this study (4173 SNPs) were then tested in an independent sample of 197 unrelated individuals from the same region [57].

Abraham and colleagues pooled the DNA samples of 1802 LOAD cases and 1239 controls and interrogated 561494 SNPs. They identified a set of 109 SNPs with a significant association with AD and genotyped them individually. In addition to APOE, one SNP close to the lecithin retinol acyltransferase gene was identified [97]. Li et al managed to identify four SNPs: two located in *GOLPH2*, one in chromosome 9 and another one intergenic between *ATP8B4* and *SLC27A2* by studying 753 AD patients and 736 matched controls [98]. The study by Bertram et al identified several associations, using 484522 SNPs in an Affymetrix platform on a sample of 1376 individuals from 410 families [99]. Beecham and colleagues used Illumina's HumanHap550 beadchips to analyze a first cohort of 492 LOAD cases and 498 controls followed by a validation set of 238 cases and 220 controls [100]. Carrasquillo MM et al, using Illumina HumanHap300 BeadChips identified the *PCDH11X* gene in chromosome Xq21.3 as associated with the development of AD, by studying 844 LOAD cases and 1255 controls in stage 1 and evaluating 25 SNPs with the most significant allelic associations in four additional series (1547 cases and 1209 controls) [101]. Feulner TM and colleagues, using the Sentrix HumanHap550 Genotyping Beadchip, generated genome-wide data in a German cohort of 491 AD patients and 479 controls. The results obtained were analyzed only for the genes included in the top results list on the Alzgene database [102]. Additionally to ApoE, nominally

significant associations were found for six of the ten studied genes (*CH25H*, *PGBD1*, *LMNA*, *PCK1*, *MAPT* and *SORL1*).

All the studies were able to pick up ApoE as a major risk factor, what was expected since the estimated odds ratio of AD over the age of 65 years associated with the apoE4 allele, relative to apoE3, is 3.18 (95% CI 2.93–3.45, p< 0.00001) [103]. This association between LOAD and ApoE started to represent a positive control for GWAS in AD. As a result, from these studies several SNPs, genes and genetic regions have been associated with the risk of developing AD (Table 2). Nonetheless, the gold standard for the detection of an authentic association is independent replication in large and phenotypically similar samples [81] and no studies have managed to achieve this gold standard.

Currently, only the data from Reiman et al and by Li et al studies is publicly available (http://www.tgen.org/research/neuro_gab2.cfm and http://www.GSK.com, respectively).

Table 2. Main features and results of genome wide studies performed in AD

Study	Platform	#SNPs	Population	Genotype data publicly available	# Subjects		Replication stage(s)	Significant results
					Cases	Ctrls		
Grupe, 2007 [94]	Celera (cSNPs)	17343	UK and USA	No	380	396	4 replication tiers (UK2, 309 cases+349 controls; UK3, 503 cases+643 controls; WU, 376 cases+344 controls; SD, 240 cases+330 controls) Total number of samples (4 tiers) = 1428 cases + 1666 controls	rs157581 (TOMM40, chr 19q13.32) rs405509 (APOE, chr 19q13.32) rs3745833 (GALP, chr 19q13.42) rs1554948 (TNK1, chr 17p13.1) rs1132899 (APOC2, chr 19q13.32) rs11622883 (14q32.13) rs8192708 (PCK1, chr 20q13.31) rs505058 (LMNA, chr 1q22) rs3800324 (PGBD1, chr 6p22.1) rs6907175 (LOC651924, chr 6q24.1) rs1859849 (7p15.2) rs41310885 (hCV22274641, FAM63A, chr 1q21.2) rs2074877 (MYH13, chr 17p13.1) rs41271951 (hCV15746640, CTSS, chr 1q21.2) rs444013 (UBD, chr 6p22.1) rs9608099 (BCR, chr 22q11.23) rs2882676 (ACAN, chr 15q26.1) rs13022344 (TRAK2, chr 2q33.1) rs11016976 (EBF3, chr 10q26.3)
Coon, 2007 [95]	Affymetrix (500K)	502627	USA, Netherlands; overlaps with Reiman, 2007	No	664	422	No	rs4420638 (APOC1, chr 19q13.32)
Reiman, 2007 [96]	Affymetrix (500K)	312316	USA, Netherlands; overlaps with	Yes	446	290	Neuropathological replication cohort (197 cases + 114 controls); Clinical replication	rs901104; rs1385600; rs1007837; rs2510038; rs4945261; rs7101429; rs10793294;

			Coon, 2007				cohort (218 cases + 146 controls) Total number of samples (2 cohorts) = 415 cases + 260 controls	rs4291702; rs7115850; rs2373115 (GAB2, chr 11q14.1)
Liu, 2007 [57]	Affymetrix (500K)	262000	Netherlands	No	103 LOAD patients and 170 first-degree relatives from a pedigree with 4645 members	-	4173 SNPs in 197 unrelated subjects from the GRIP population	Chr 1q25 (RGSL2, RALGPS2, C1orf49); Chr 3q22-24 (NMNAT3, CLSTN2); Chr 10q22-24 (HTR7, MPHOSPH1, CYP2C); Chr 11q25 (OPCML, HNT)
Li, 2008 [98]	Affymetrix (500K)	469438	Canada and UK	Yes	753	736	120 SNPs in 418 cases + 249 controls	rs10868366; rs7019241 (GOLM1, chr 9q21.33) rs4420638 (APOC1, chr 19q13.32) also significant in Coon, 2007; rs10519262 (chr 15q21.2); rs9886784 (chr 9p24.3)
Abraham, 2008 [97]	Illumina HumanHap300 + Illumina Sentrix HumanHap240S	561494	UK	No	1082 LOAD pooled	1239 controls pooled	114 SNPs were individually genotyped in the cases and controls used in the pools and 1400 controls were added	ApoE; rs727153 (~13 kb from the start of transcription of LRAT, chr 4q32.1), also showed increased significance (10 fold) when additional controls were combined
Bertram, 2008 [99]	Affymetrix 500K	484522	Self-reported European descent from NIMH Genetics Initiative Study sample	No	1376 AD from 410 families	-	4 SNPs in 3 independent AD family samples: NIA- 1040 samples from 329 families; NCRAD- 1108 samples from 331 families; CAG- 483 samples from 215 sibships (Total: 2689 samples 1816 affected and 845 unaffected)	Rs4420638 (located 340 bp 3' of APC1 and probably reflects the effects of apoE E4 allele); rs 11159647 (in predicted gene NT_026437.1360, chr14q31.2), also significant in replication samples and in TGEN's data; rs179943 (in ATXN1, chr 6p22.3), trend for significance in replication samples; rs3826656 (in predicted gene NT_011109.848, chr 19q13.33),

								also significant in replication samples; rs2049161 (in cDNA BC040718, chr 18p11.31)
Beechmam, 2009 [100]	Illumina HumanHap550	532000	USA	No	492 LOAD	496	1 SNP in 238 cases + 220 controls	ApoE; rs11610206 (downstream of FAM113B, chr 12q13), also significant in replication sample 1q42; 4q28; 6q14; 19q13
Carrasquillo, 2009 [101]	Illumina HumanHap300	313504	USA	No	844 LOAD	1255	25 SNPs in 1547 cases + 1209 controls	ApoE; rs5984894, rs2573905 (in PCDH11X, chr Xq21.3), also significant in replication sample
Harold, 2009 [104]	Illumina 610 quadchip (Illumina HumanHap550 and HumanHap300 in some samples)	529205	Europe USA	Yes	3941	7848	2 significant SNPs in 2023 cases + 2340 controls	ApoE; rs11136000 (in CLU, chr 8p21-p12), also significant in replication sample rs3851179 (5' of PICALM, chr 11q14, also significant in replication sample)
Lambert, 2009 [105]	Illumina Human 610-Quad BeadChip	537029	France Europe	No	2032	5328	Significant SNPs in 3978 probable AD cases + 3297 controls from Belgium, Finland, Italy and Spain	rs11136000, rs2279590, rs9331888 (in CLU, chr 8p21-p12), also significant in replication sample rs6656401 (in CR1, chr 1q32), also significant in replication sample

Most recently, two larger studies have uncovered the first genes to achieve genome wide significance in AD. The first of these studies reported the analysis of over 16000 individuals (3941 cases and 7848 controls in stage 1 and 2023 cases and 2340 controls in stage 2), using Illumina 610-quad chips. Two SNPs significantly associated with AD, outside the *APOE* locus, were identified: rs11136000 located in an intron of *CLU* on chromosome 8 and rs3851179 located close to the gene *PICALM* on chromosome 11. Although the risks associated with *CLU* and *PICALM* genes are relatively small (ApoE odds ratio~3; CLU/PICALM~0.86) these are the first genes to reach genome wide significance [104]. Biologically, both of these genes seem to be perfect fits in the already known biological pathways to disease, adding to the inference that these are real AD genes.

Clusterin or apolipoprotein J is, like APOE, a lipoprotein expressed in most mammalian tissues with higher levels present in brain, ovary, testis and liver [106]. CLU interacts with different molecules, including lipids, amyloid proteins, components of the complement membrane attack complex (MAC) and immunoglobulins [107]. Accordingly, it has been proposed to be involved in a number of physiological processes such as ongoing synapse turnover [108], apoptosis [109, 110], cytoprotection at fluid-tissue boundaries, membrane recycling during development and in response to injury and regulation of complement-mediated MAC [107, 111]. Clusterin has also been proposed to be a form of secreted heat-shock protein or chaperone molecule [112, 113].

Several lines of evidence suggest that CLU has a central (either protective or pathogenic) role in the pathway leading to Alzheimer's disease. First, CLU mRNA has been reported to be elevated in AD affected brain areas such as hippocampus, either when brains from AD patients were compared to one Huntington patient [114], or to controls [115]. Likewise, Oda et al reported a statistically significant difference in the Clusterin content of 10 non-AD individuals and 25 AD patients [111]. Second, Clusterin is one of the components of amyloid plaques [116-119]. Third, is able to bind soluble Aβ through a specific, reversible and high-affinity interaction in cerebrospinal fluid [120, 121] to form complexes able to cross the blood-brain barrier by a high affinity receptor mediated process

involving transcytosis [122]. Fourth, reduced levels of ApoE and increased levels of CLU have been correlated with the number of E4 alleles, suggesting a compensatory induction of CLU in the brain of AD individuals with the E4 allele of ApoE presenting low brain levels of ApoE [123]. Moreover, CLU was shown to prevent aggregation and polymerization of synthetic Aβ and to enhance the oxidative stress caused by Aβ in vitro [124, 125], and to facilitate Aβ uptake in cell culture experiments [126]. Clusterin appears to regulate the toxicity and conversion of Aβ into soluble forms [124, 125, 127, 128]. Together with ApoE, suppresses Aβ deposition [129] and may modify Aβ clearance at the blood brain barrier [130].

The second gene, *PICALM*, encodes the phosphatidylinositol-binding clathrin assembly protein, also known as *CALM*: clathrin assembly lymphoid-myeloid leukemia gene. It is also ubiquitously expressed with particular high levels in neurons. This gene has been mainly associated with leukemia, thus its relation to AD may appear not as direct (Figure 1), as the one observed for CLU. Nonetheless, its involvement in clathrin-mediated endocytosis (essential to the intracellular trafficking of proteins and lipids) [131] and in the fusion of synaptic vesicles to the presynaptic membrane by directing the trafficking of VAMP2 [132] have lead Harold et al to propose two interesting hypothesis for the role of PICALM in AD. In this way, genetic variability in PICALM may result on synapse perturbations, possibly through synaptic vesicle cycling, or on alterations of APP processing through endocytic pathways, culminating in changes in Aβ levels [104].

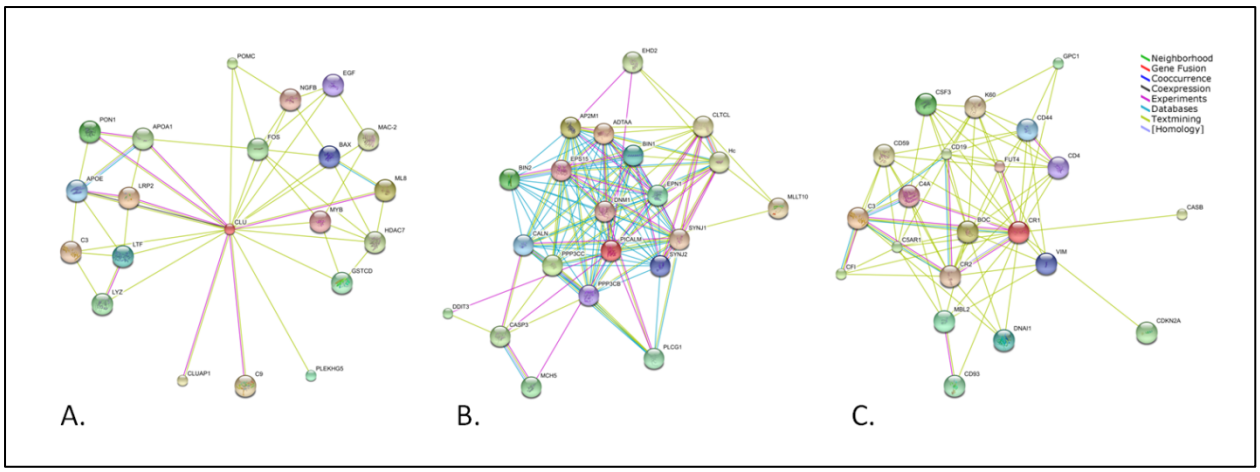


Figure 1. Predicted interactions for Clu (A), Picalm (B) and CR1 (C). The software STRING 8.0 (available at http://bit.ly/bjO27e) was used to establish a network of predicted interactions for Clu (A), Picalm (B) and CR1 (C) proteins. Accessed on August 2009. STRING is a database of known and predicted protein interactions. The interactions include direct (physical) and indirect (functional) associations; they are derived from four sources: genomic context, high-throughput experiments, (conserved) coexpression and previous knowledge. STRING quantitatively integrates interaction data from these sources for a large number of organisms, and transfers information between these organisms where applicable. The database currently covers 2,483,276 proteins from 630 organisms. From the three represented proteins, CLU is the one with a more direct relation with AD: APOE is directly connected to CLU. Interestingly, CLU also interacts with proteins present in CR1 network (like C3). No other genes consistently associated with AD are present in PICALM or CR1 networks, indicating that these proteins may be involved in new pathobiological pathways.

The other large AD GWAS reported on 14635 samples (2032 cases and 5328 controls in the first stage, and 3978 cases and 3297 controls in the replication series). In addition to *APOE*, *CLU* was also associated with AD (OR=0.86, 95%CI0.81-0.90). Additionally, this study identified variation at the *CR1* locus, on chromosome 1q32 to be associated with AD (OR=1.21, 95% CI 1.14-1.29). To our knowledge, no polymorphisms in the *CR1* gene have been studied before in relation to AD [133]. *CR1*, the complement component (3b/4b) receptor 1 (Knops blood group) is a member of the receptors of complement activation family. The gene encodes a monomeric single-pass type I membrane glycoprotein found on erythrocytes, leukocytes, glomerular podocytes, and splenic follicular dendritic cells that mediates cellular binding to particles and immune complexes that have activated complement [134].

Three complement pathways are known: the classical, alternative and lectin-mediated cascades, which have different activation triggers, but all terminate with the production of the membrane

attack complex. High enough concentrations of MAC result in cell lysis. This may lead to tissue damaging when the complement activation tight regulation, (that occurs through the action of several different endogenous complement inhibitor proteins), is deficient [135]. Typically, the complement classical pathway is the one associated with AD [136], mainly due to three facts: 1) C1q, the first protein in this pathway, efficiently binds to aggregated Aβ, activating the pathway and further enhancing Aβ aggregation and fibril formation [137, 138]; 2) early complement activation proteins (C1q, C4 and C3) and the MAC have been found to co-localize with senile plaques, NFTs and dystrophic neurites in AD brains [139-141]; and 3) increased mRNA levels of complement proteins are present in AD brains when compared to controls [142]. More recently, activation of the alternative pathway in AD brains has also been demonstrated: Aβ activates the alternative pathway in vitro [143]; factor B mRNA is present in AD frontal cortex, and factor D cleaved split products of factor B, Bb and Ba, are significantly increased, in AD brains [144]. After the discovery that the complement system can be activated in the brain by several senile plaques and neurofibrillary tangle related components, in the absence of antibodies, and that neurons are the major source of complement proteins in the brain, the involvement of the complement system in AD has been widely accepted [145]. However, if this involvement has a protective or deleterious effect, has been extensively debated. In fact, it has been proposed that binding of C1q to misfolded proteins in early AD, together with C4BP that decreases MAC activation, are favorable and enable clearance of the misfolded material. But, when the system is overwhelmed by amyloid, this protein binds extensively to C1q leading to the full activation of the complement, ultimately leading to detrimental inflammation and neurodegeneration [135].

In summary, as in other aging degenerative diseases, the complement system has an important role in AD, and one may expect that the activation of the classical or the alternative pathways (or both) by Aβ will lead to neurodegeneration in individuals with a genetic predisposition [146]. This will possibly result from an unbalance between the expression of regulator proteins, and one or more cascade proteins (Figure 2). This model is able to explain, at least partially, the presence of

neuropathological changes in the brains of non-demented individuals [147] since the genetic variability in the complement genes may be responsible for different complement reactions to the presence of NFT and senile plaques. Interestingly, as mentioned above, one function attributed to clusterin is in the regulation of complement-mediated membrane attack complex. Together with vitronectin, clusterin binds to the nascent amphiphilic C5b-9 complex, rendering it water soluble and lytically inactive, raising the possibility that the genetic risk conferred by clusterin for the development of AD, may arise from its regulation role in the complement system.

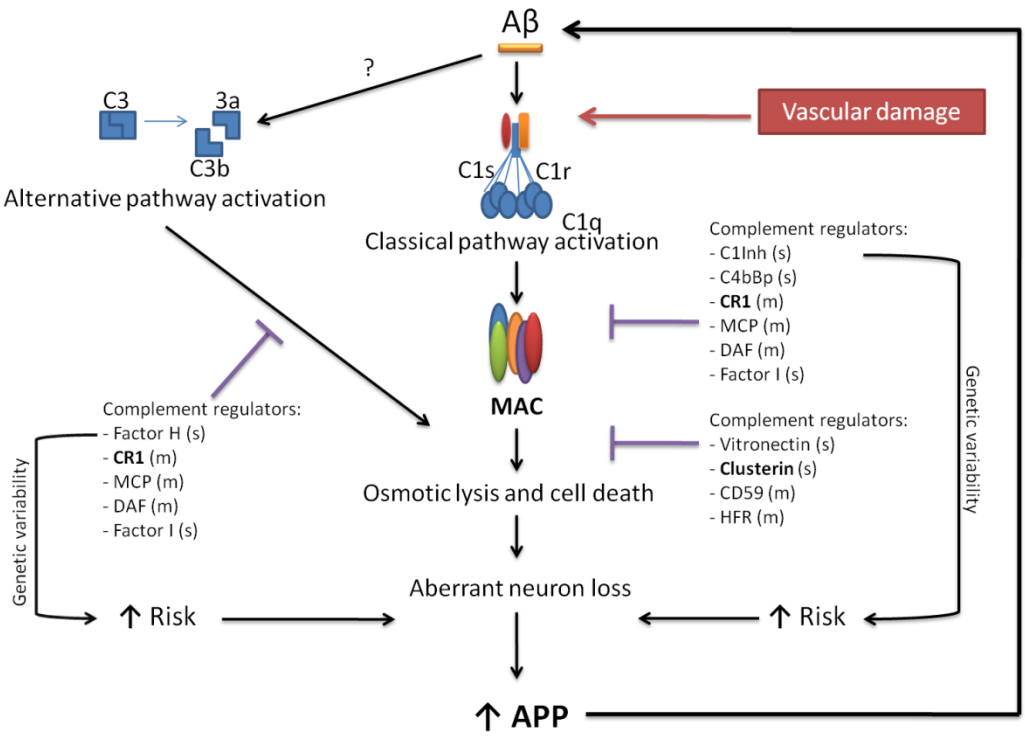


Figure 2. Genetic variability and the role of the complement system in AD. Aβ and probably NFTs are able to activate the classical and alternative complement pathways. These are regulated by several membrane (m) and soluble (s) proteins at different stages. The genetic variability in *CR1* is now known to be associated with the risk of developing AD. Increased numbers of samples are needed to know if the same is true for any other component of these pathways. Drawn from the work of Tenner [148] and McGeer and McGeer [149].

1.1.4 - The Vascular Risk Factors and AD

These genetic findings point to the role of vascular damage in Alzheimer's pathogenesis and vascular risk factors in its aetiology. It is noteworthy that many of the genes now implicated have a direct role at the blood brain interface. As such, these findings are consistent with the epidemiologic

literature, which has consistently reported this association for almost 20 years. The evidence base for the prevention of AD and related brain pathologies is strongest for control of vascular risk factors [150]. These factors include, for example, obesity, hypertension, cardiovascular disease, diabetes, and hypercholesterolemia. The low risk ratios observed in genetic studies attempting to identify new susceptibility genes for AD may, in fact, be due to lack of information on presence and/or severity of this vascular involvement. AD exists against a background continuum of vascular pathology, which may modulate risk for clinically manifest disease.

Hypertension is a risk factor for stroke, ischemic white matter lesions, silent infarcts, general atherosclerosis, myocardial infarction and cardiovascular morbidity and mortality. This risk increases with increasing blood pressure also at apparently healthy blood pressure ranges [151]. Several longitudinal studies have suggested an association between Alzheimer's disease and previous hypertension [152-156].

ACE, which is currently the second top hit on Alzgene database [71], plays a classical role in blood pressure regulation as part of the renin-angiotensin system [157]. This system may play a role in dementia pathogenesis because of its effects on vascular and metabolic homeostasis, as well as amyloid metabolism. The gene encoding for ACE is the only AD susceptibility gene whereby effect modification is observed by vascular phenotype, particularly with population stratification by APOEe4 allele, systolic blood pressure, body mass index, and waist and hip circumferences [158]. Thus, the renin angiotensin system may also provide a link between obesity, hypertension, and vascular syndromes, such as type 2 diabetes, and health of the brain [157, 159] because human brain and adipose tissue express the rennin angiotensin system [160].

Cholesterol is important in AD, not only because of its relationship with cardiovascular disease, but due to its role in amyloid metabolism [161]. APOE and CLU are two proteins involved in lipid transport in the peripheral and central nervous systems [162]. APOE regulates cholesterol homeostasis in astrocytes and microglia, and is related to blood cholesterol levels [163]. In addition,

mutations in the *APP* gene upregulate Abeta40 and Abeta42 production, [164] and Abeta processing is sensitive to cholesterol levels and lipid trafficking. Brain cholesterol levels increase during AD progression [163].

The epidemiology of blood cholesterol levels in relation to AD and other forms of dementia are mixed. High cholesterol levels in mid-life may increase risk for subsequent dementia and AD [152, 165, 166], however, in late-life, low cholesterol levels have been predictive of subsequent dementia [167, 168] or no association has been observed [169, 170]. Nevertheless, even within the mid-life cholesterol literature, results are conflicting, as some studies have not found high cholesterol to predict later dementia [171-173].

While there remain a number of questions regarding the amyloid hypothesis in relationship to AD, the potential link to cholesterol metabolism and vascular damage is noteworthy [174].

1.1.5 – The Next Steps

Two major and clearly non-exclusive pathways have been recently discussed as the future guidelines in the genome wide analysis of complex disorders [175-177]. The first is the extension of the assembly of studies containing larger (tens or even hundreds of thousands) and representative samples in order to identify variants with lower frequencies that may have been missed until now and that could explain the so elusive fractions of the "missing heritability" in AD. The increase in the number of studied samples will inevitably result in the discovery of new variants and probably new genes associated with AD, but the real net value of these variants is highly disputable. Therefore, the actual dilemma now is to know how far one should take these studies in order to keep a positive balance between the resources applied and the gathered genetic returns [175, 178]. Clearly, still to discover rare variants and variants with small effect sizes will be difficult to replicate, due to reduced power and restriction to specific populations, respectively. Nonetheless, in order to identify these

rare variants one may speculate the need to use new chips with a better coverage of rare variants and the resequencing of previously identified regions. One may also predict that, being AD a complex disorder, more emphasis will be put in the study of endophenotypes, as happened in the study by Liu et al that used cognitive function as an endophenotype of AD and identified the *RGSL2*, *RALGPS2* and *C1orf49* genes as potential causative genes located in one of the associated genomic regions [57]. This approach will obviously require precise and clearly defined clinical assessments.

The second line of investigation will rely in the sequencing of whole exomes and whole genomes, expectantly unveiling several new rarer risk variants. These higher risk variants may be related to the commoner lower risk variants found with GWAS. In fact, a classical example comes from the study of Parkinson's disease genetics: rare high risk variants in the α-synuclein gene are the cause of monogenic PD [179], while a common haplotype of this same gene has been established as a moderate risk cause of sporadic disease (Simon-Sanchez et al, unpublished data).

Both lines of research will be followed by large resequencing efforts, in order to identify the real risk variants. Many GWAS have identified regions of the genome associated with AD, but in many cases the real risk/causal changes are not known. Additionally, in the cases where SNPs have been associated with a disease, these SNPs identified may in fact be in LD with the real variants causing the association. Further functional data, (for instance, the effects of a determined variant in gene expression), although not always possible to obtain, will be essential not only to validate the previously identified variants, but also because many variants map to non coding protein sequences, gene deserts, or any genomic region without any functional elements. The integration of these functional studies with information on multiple variants from the same gene in different populations and the effects of epigenetics and epistasis will be vital. The interpretation of these results, however, will be extremely complex and will most likely require the combination of disciplines as integrative genomics and systems biology. Although difficult, this approach will ultimately allow a more profound understanding of the molecular pathways underlying AD and AD risk, as well as the subsequent identification of effective biomarkers and drugs.

1.2 - Genetics of Frontotemporal Lobar Degeneration

Frontotemporal lobar degeneration (FTLD) is a genetically and pathologically heterogeneous degenerative disorder [180, 181]. It is a frequent cause of dementia and accounts for 5 to 10% of all dementia patients and 10 to 20% of patients under 65 years of age [182]. A number of different clinical syndromes fall into the FTLD spectrum: the most common subtype consists of patients who present with behavioral or personality changes (behavioral variant frontotemporal dementia or bvFTD). Less commonly, patients present with language impairment and three different disorders are described (often under the collective term primary progressive aphasia or PPA): semantic dementia (SemD), progressive nonfluent aphasia (PNFA) and the logopenic/phonological variant of PPA (LPA) [183, 184]. There is also an overlap of FTLD with motor disorders: the parkinsonian disorders, corticobasal syndrome (CBS) and progressive supranuclear palsy (PSP) and motor neuron disease (known as FTD-MND when there is an overlap) [185, 186]. Although estimates vary widely, FTLD is commonly familial [187] with a positive familial history being observed in up to 50% of patients, suggesting a strong genetic component to the disease. As expected from the clinical and pathological heterogeneity, the molecular genetics of FTLD is complex. Five genes and one confirmed genetic locus are currently known to cause FTLD of which two genes are relatively common (microtubule-associated protein tau, *MAPT*, and progranulin, *GRN*) and three are rare causes (valosin-containing protein, *VCP*, chromatin modifying protein 2B, *CHMP2B*, and the gene encoding TDP-43, *TARDP*) [181, 188]. *MAPT* mutations are associated with tau-positive inclusions pathologically as are the generally sporadic pathologies of corticobasal degeneration, PSP and Pick's disease. *GRN*, *VCP*, *TARDP* and *CHMP2B* mutations are associated with the other major pathological FTLD type where there are tau-negative inclusions [180, 189]. This second subtype can be further subdivided into those with TDP-43-positive pathology, known as FTLD-TDP (of which there are four subtypes; this group includes GRN, VCP and TARDP mutations) and those without, known as FTLD-UPS (including CHMP2B mutations).

The clinical syndromes comprising FTLD most commonly show severe degeneration of the prefrontal and anterior temporal lobes. However, the pattern of atrophy can vary markedly, involving both hemispheres to equal or differing extents and sometimes also involving the parietal cortex. This pathologic heterogeneity in large part explains the complex range of clinical syndromes that are included under the FTLD grouping and also the difficulty associated with classifying this group of neurodegenerative diseases. Nevertheless, at the molecular pathologic level it is now well recognized that there are only two main subtypes based on the profile of immunohistochemical staining: FTLD with tau-positive pathology (tau-positive FTLD), and FTLD with tau-negative ubiquitin-positive inclusions (FTLD-U) [190].

1.2.1 – Microtubule Associated Protein Tau (*MAPT*)

Tau is a microtubule associated protein that binds to tubulin and works to stabilize microtubules and promote microtubule assembly [191]. Tau may have additional functions, including the regulation of kinesin-dependent transport of vesicles and organelles along the MTs [192]. The *MAPT* gene is located in an atypical genomic region of chromosome 17 and produces six isoforms of the tau protein by alternative splicing of exons 2, 3 and 10. These diverse tau isoforms likely have specific physiologic roles since they are differentially expressed during development and appear to be somewhat differentially distributed in neuronal subpopulations [193, 194]. The interaction of tau with the microtubules occurs through either three (3R) or four (4R) imperfect repeat sequences in the carboxyl terminal of the protein. One of these four domains is encoded by exon 10; hence, the alternative splicing of this exon determines the number of microtubules binding domains. In the adult human brain 3R and 4R tau are present in approximately equal quantities [195-197]. In addition, 3R and 4R tau isoforms were reported to have different effects on the transport rate of cargo along the MTs [198].

Several pathogenic mutations in MAPT have been associated with different clinical phenotypes. These mutations may be roughly divided into two categories, depending on the pathogenic mechanism involved: protein function and splicing regulation mutations. The first group of mutations impairs the ability of tau to interact with microtubules or to promote microtubule assembly; the second corresponds to mutations located in exon 10 (with the exception of those occurring in codon 301) and in flanking intronic regions that affect the alternative splicing of this exon and consequently increase the 4R/3R tau ratio. 4R tau appears to aggregate more readily than 3R tau, thus overproduction of 4R tau may lead to an excess of free 4R and consequently promote tau aggregation [199].

At the present time, 44 different potential pathogenic MAPT mutations have been reported in a total of 132 tauopathy families (FTD mutation database, http://www.molgen.ua.ac.be/ADMutations). With the exception of two mutations in exon 1, all coding mutations are clustered in exons 9 to 13, which encode the microtubule binding domains and its flanking regions. The intronic mutations are all located in the introns flanking the alternatively spliced exon 10. The most frequent mutations are P301L in exon 10 and IVS10+16C>T in intron 10.

1.2.2 – Progranulin (*GRN*)

The physical proximity of *MAPT* and *GRN* genes on chromosome 17q21 delayed the genetic explanation of the disease in several FTLD families.

Presently, 68 different pathogenic *GRN* mutations have been reported in 208 families (FTD mutation database, http://www.molgen.ua.ac.be/FTD mutations). Mutations are found scattered over the gene and include a variety of genetic alterations, with most being small insertions and deletions generating a frameshift. All *GRN* mutations identified thus far create functional null alleles that cause a partial reduction in GRN production or haploinsufficiency [199, 200]. Most *GRN* mutations

introduce a premature termination codon leading to the subsequent degradation of mutant RNA by nonsense-mediated decay. A limited number of *GRN* mutations are missense variants displaying a different disease mechanism and the pathogenicity of these is further discussed in chapter 5. These include splice-site mutations in the splice donor site of the noncoding exon 0 that are expected to lead to the retention of intron 0 in the mRNA and consequent direct nuclear mRNA degradation [201, 202]; mutations within the Met1 translation initiation codon and an unusual exon 1 splice-site mutation that directly destroys the Kozak sequence preventing GRN production [199-201].

The most frequently observed *GRN* mutations are the R493X nonsense mutation in exon 11, and the c.-8+5G>C (IVS0+5G>C) splice-site mutation in intron 0, which was reported in eight independently ascertained branches from a Belgian founder family [199, 201].

GRN encodes a precursor glycoprotein of 68 kD that is composed of 7.5 tandem repeats of a highly conserved motif with 12 cysteines [203]. Proteolytic cleavage by an elastase-like activity results in the generation of a family of 6-kD peptides called granulins. GRN is generally thought to be a secreted mitogenic factor that is expressed in a variety of tissues. GRN appears to regulate cell cycle progression and cell migration in a range of tissue remodeling processes including development, wound repair/inflammation, and tumorigenesis [203, 204]. The function of GRN in the central nervous system is not yet fully understood, but high GRN expression has been observed in a subset of pyramidal neurons and in activated microglial cells [205]. Because all pathogenic *GRN* mutations described to date are expected to reduce the levels of functional GRN and granulins, a role for GRN in maintaining neuronal survival has been suggested [200].

1.2.3 - Charged multivesicular body protein 2B (*CHMP2B*)

In 2004, Skibinski et al. identified a complex mutation in CHMP2B that segregated with the disease in a large dementia family from Denmark [206]. In this family, dementia presented typically in the late fifties, with an insidious change in personality and behavior. The mutation, located in the acceptor

splice site of the most C-terminal exon (exon 6), was shown to produce two aberrant transcripts resulting from the inclusion of intron 5 and the use of a cryptic splice site in exon 6. At the protein level, these transcripts are predicted to result in a 36–amino acid C-terminal truncation of the CHMP2B protein or the replacement of the same region with 29 nonphysiologic amino acids. Subsequent sequencing analyses of CHMP2B in FTLD patients revealed that mutations in this gene are not a common cause of FTLD or FTD-ALS. Few mutations have been identified, including missense mutations with uncertain pathogenicity [207-209] and a nonsense R186X mutation in two unaffected members of an Afrikaner FTD family but not in their affected relatives [210].

Human CHMP2B belongs to the charged multivesicular body protein family. The exact function of this protein is unknown, but overexpression of the two Danish CHMP2B mutant proteins in cell culture was shown to disrupt the CHMP2B localization and resulted in the formation of dysmorphic organelles of the late endosomal pathway [206].

1.2.4 - Valosin-containing protein (*VCP*)

FTD associated with inclusion body myopathy and Paget's disease of the bone (IBMPFD) is a rare multisystemic disorder with autosomal dominant pattern of inheritance. Patients typically present with adult-onset proximal and distal muscle weakness, early-onset Paget's disease of bone, and (in ~ 30% of the patients) early-onset FTD [211]. After the identification of a locus for IBMPFD in chromosome 9p21-p12, missense mutations in VCP were identified in 13 IBMPFD families [212]. To date, thirteen mutations in VCP have been reported in 30 families with a clear hot-spot at codon 155 in exon 5 (http://www.molgen.ua.ac.be/FTDMutations).

VCP is a member of the AAA-ATPase family with a tripartite structure comprising an N-terminal CDC48 domain and two central D1- and D2-domains that bind and hydrolyze adenosine triphosphate [213]. It acts as a molecular chaperone in a range of cellular activities, including ubiquitin-dependent protein degradation, cell-cycle regulation, and apoptosis. Most of the VCP mutations are situated

within or close to the highly conserved CDC48 domain, which is involved in the ubiquitin binding. VCP mutations are likely to cause a loss or alteration of the normal function of the protein, leading to impaired ubiquitin-proteasome system activity and the accumulation of ubiquitinated proteins within cells [212].

CHAPTER 2

OBJECTIVES

CHAPTER 2 - OBJECTIVES

Alzheimer's disease and Frontotemporal lobar degeneration are known to be multifactorial, complex disorders where several factors seem to play concerted roles. Most of AD and FTLD cases are apparently sporadic, with an onset late in life. A very small proportion of cases are known to be caused by genetic mutation. The genetic component of these diseases has been a major factor leading research in the field. When a new gene is found to be associated with determined pathology, new biological venues and new drug targets can be explored.

The molecular analysis of early onset AD and FTD patients can, many times, not only facilitate the disease diagnosis, but also, be a valuable asset in genetic counseling. In this way, chapter 4 of this thesis was designed to allow the characterization at a molecular level, of a large number of FTLD and AD samples. This strategy lead us to the identification of many pathogenic mutations causing disease in different subjects. Additionally, it also resulted in the detection of many variants with unclear pathogenicity. In chapter 5 we try to overcome this issue by sequencing the genes known to harbor mutations associated with AD and FTLD in several healthy individuals and diverse human samples from the Human Genome Diversity Panel.

After the studies performed in chapter 4 the genetic causes for the disease in several patients were revealed. However, a very large subset of these cases (early and late onset, with a clear familial aggregation or sporadic) remained to be explained, from a genetic perspective. In chapter 6 we take advantage of recent methodological developments in order to study genetic variability at the whole genome level and try to find new genes associated with Alzheimer's disease. In this way, the genetic component of late onset sporadic disease was studied in a large genome wide association study and the possible role of homozygosity was analyzed both in late and early onset AD.

CHAPTER 3

METHODS

CHAPTER 3 – METHODS

Several of the studies in this thesis use variations of the same methodologies. In this way, a general description of these common methods is provided in this chapter, while the specific methodologies and samples used for each of the studies are provided in the respective chapters.

3.1 - Polymerase Chain Reaction and Sequencing analyses

The polymerase chain reaction (PCR) was used in several of the studies presented here to amplify particular DNA sequences from different genes, several orders of magnitude, in order to generate thousands to millions of copies of a particular amplicon. The method relies on thermal cycling, consisting of cycles of repeated heating and cooling of the reaction for DNA melting and enzymatic replication of the DNA. Primers containing sequences complementary to the target region along with a DNA polymerase are used to enable selective and repeated amplification. In Table 3 are represented all the sequences of primers and probes used in the different studies of this work. As PCR progresses, the DNA generated is itself used as a template for replication, setting in motion a chain reaction in which the DNA template is exponentially amplified.

Table 3. Sequences of primers and probes used in the different studies.

GENE	EXON/TARGET	SEQUENCE FORWARD	SEQUENCE REVERSE	USE
PSEN1	Exon 3	ACAAAGTTCTGTTTTTCTTTCCC	CAGCATTTCTCAGAGGTGAGG	PCR/Sequencing
PSEN1	Exon 4	CGTTACCTTGATTCTGCTGA	GACATGCTGTAAAGAAAAGCC	PCR/Sequencing
PSEN1	Exon 5	GATTGGTGAGTTGGGGAAAAGTG	ATACCCAACCATAAGAAGAACAGG	PCR/Sequencing
PSEN1	Exon 6	AGGGTTGTGGGACCTGTTAATT	AGTTAATTCTGAAAGACAGACCC	PCR/Sequencing
PSEN1	Exon 7	GGAGCCATCACATTATTCTAAA	AACAAATTATCAGTCTTGGGTTT	PCR/Sequencing
PSEN1	Exon 8	TTACAAGTTTAGCCCATACATTTT	TCAAGTTCCCGATAAATTCTAC	PCR/Sequencing
PSEN1	Exon 9	TGTGTGTCCAGTGCTTACCTG	TGTTAGCTTATAACAGTGACCCTG	PCR/Sequencing
PSEN1	Exon 10	CCAGCTAGTTACAATGACAGC	TCAAAAAGGTTGATAATGTAGCT	PCR/Sequencing
PSEN1	Exon 11	GGTTGAGTAGGGCAGTGATA	TTAAAGGGACTGTGTAATCAAAG	PCR/Sequencing
PSEN1	Exon 12	GTCTTTCCCATCTTCTCCAC	GGGATTCTAACCGCAAATAT	PCR/Sequencing
PSEN2	Exon 3	GTCCTCCACTGCCTTTGTCTCAC	CTTCCCTTCTCCCTCCCGCATCAG	PCR/Sequencing
PSEN2	Exon 4	CAAAAATCCGTGCATTACATGG	CTGCAGGTACAGTGACCAAC	PCR/Sequencing
PSEN2	Exon 5	AGCCTCGAGGAGCAGTCAG	GCAGACGGAGAGAAGCGT	PCR/Sequencing
PSEN2	Exon 6	GGTATCAGTCTCAGGATCATGGG	TGGGGAAGACTGGAGCTCGATG	PCR/Sequencing
PSEN2	Exon 7	GTAAAGAGGGCCAGGTTGGG	GTGCAGCACTGGGGACGATTT	PCR/Sequencing
PSEN2	Exon 8	ACTGGAGAATGAGAATTTGGG	GAAAGCCACGGCCAGGAAG	PCR/Sequencing

PSEN2	Exon 9	ACCGCCTGAGACGTGAACCTT	TCCCTCTGCCCCTCCTGAACT	PCR/Sequencing
PSEN2	Exon 10	CTCTGACCAGCTGTTGTTTC	AGCCTCCACCCTCTGTCT	PCR/Sequencing
PSEN2	Exon 11	TTCCATTCTGTGCACGCCTC	ACCTGCCCCCACCACAATG	PCR/Sequencing
PSEN2	Exon 12	ACACCAGGGATCACCACGCTCAC	TGCCTCCTCCTCACCAAGTAAACA	PCR/Sequencing
APP	Exon 16	CTTCAGGCCTAGAAAGAAGT	GGATGAACCAGAGTTAATAGG	PCR/Sequencing
APP	Exon 17	AACCTCATCCAAATGTCCCC	ATTCCCACTTGGAAACATGC	PCR/Sequencing
MAPT	Exon 1	CAACACTCCTCAGAACTTATC	CAGTGATCTGGGCCTGCTGTG	PCR/Sequencing
MAPT	Exon 9	CGAGTCCTGGCTTCACTCC	CTTCCAGGCACAGCCATACC	PCR/Sequencing
MAPT	Exon 10	GGTGGCGTGTCACTCATCC	GTACGACTCACACCACTTCC	PCR/Sequencing
MAPT	Exon 11	GCTCATTCTCTCTCCTCCTC	GCAGTTCCAGCCTCACCAGG	PCR/Sequencing
MAPT	Exon 12	GTCCTGTCATTGTCTTCTTC	ACCCACTGGATGCTGCTGAG	PCR/Sequencing
MAPT	Exon 13	CTTTCTCTGGCACTTCATCTC	CCTCTCCACAATTATTGACCG	PCR/Sequencing
GRN	Exon 0	CGCCTGCAGGATGGGTTAAGG	GCGTCACTGCAATTACTGCTTCC	PCR/Sequencing
GRN	Exon 1	GGGCTAGGGTACTGAGTGAC	AGTGTTGTGGGCCATTTG	PCR/Sequencing
GRN	Exon 2	TGCCCAGATGGTCAGTTC	GCTGCACCTGATCTTTGG	PCR/Sequencing
GRN	Exon 3	GGCCACTCCTGCATCTTTAC	TGAATGAGGGCACAAGGG	PCR/Sequencing
GRN	Exon 4	TTAGTGTCACCCTCAAACC	ACTGGAAGAGGAGCAAAC	PCR/Sequencing
GRN	Exon 5	GGGCCTCATTGACTCCAAGTGTA	GGTCTTTGTCACTTCCAGGCTCA	PCR/Sequencing
GRN	Exon 6	TCCCTGTGTGCTACTGAG	AAGCAGAGAGGACAGGTC	PCR/Sequencing
GRN	Exon 7	TACCCTCCATCTTCAACAC	TCACAGCACACAGCCTAG	PCR/Sequencing
GRN	Exon 8	ATACCTGCTGCCGTCTAC	GAGGGCAGAAAGCAATAG	PCR/Sequencing
GRN	Exon 9	TGTCCAATCCCAGAGGTATATG	ACGTTGCAGGTGTAGCCAG	PCR/Sequencing
GRN	Exon 10	TGGACTGGAGAAGATGCC	CGATCAGCACAACAGACG	PCR/Sequencing
GRN	Exon 11	CATGATAACCAGACCTGC	AGGGAGAATTTGGTTAGG	PCR/Sequencing
GRN	Exon 12	CGCCTGCAGGATGGGTTAAGG	GCGTCACTGCAATTACTGCTTCC	PCR/Sequencing
APP	Exon 5	TGCCCACTGGCTGAAGAAA	GTCATCCTCCTCCGCATCA	Exon-qPCR
APP	Exon 5	CAGAATCCACATTGTCA	Fluorophore: VIC/MGB	Exon-qPCR
GRN	Exon1	TGGCCAATGGAAACTGAGGTA	CGGGTAGCGCTCAGACTACAG	Exon-qPCR
GRN	Exon1	CGGGTCATCGCGC	Fluorophore: FAM/MGB	Exon-qPCR
PRNP	Exon2	ATGGAGCGCGTGGTTGA	GGCCTGAGATTCCCTCTCGTA	Exon-qPCR
PRNP	Exon2	CAGATGTGTATCACCC	Fluorophore: FAM/MGB	Exon-qPCR
APP	196999	CCTCTCAACTCCTTATTTCAA (FAM)	ACATTCCTCTCTCTGCCTCTA	fm-qPCR
APP	188463	TTAAAGCAATCAATCATGG (FAM)	GGGTTCTGTGAATATGGG	fm-qPCR
PRNP	108991	GGAAGGTTGGACATCTTT (FAM)	GCCTGGCCACAATTTTAC	fm-qPCR
PRNP	452	GAACCTAAAACTCTAAGGAAGCG	AGCCTCCATAACCACATGAA	fm-qPCR
VCP	Exon3	CAAGAACTTGGTCCTGCCTG	GCTTTCTGGTCTAGGGACAGC	PCR/Sequencing
VCP	Exon 5	CCCAGTCCTGACAGTTACCAC	GAGCTTGGCATTTTGACCC	PCR/Sequencing
VCP	Exon 6	CCCAGGATTAGACATTGGGAC	TTTGCACACTAGGTAGTGGAATG	PCR/Sequencing
VCP	Exon 10	GTCTCTAGCCAGTTCCCAGC	GGTCACCCTAGGCCTGTCTC	PCR/Sequencing
CHMP2B	Exon 1	CGGCTTCAAACTCCGTAGTGC	AGAAGCAAGAGGGAGGCAGAGAAGT	PCR/Sequencing
CHMP2B	Exon 2	AGA TGGGGTTTCTCCA TGTTGGT	CACAGAATCCCAATAAAAGTCAGCA	PCR/Sequencing
CHMP2B	Exon 3	TTGCCTTCTTAGATTTCTCTTTTGC	GAAAACGGA TTGTCCTCAAACAA	PCR/Sequencing
CHMP2B	Exon 4	CTTGCCTGCTGTTCAGA TGAGGT	GCCAATGTGTAACAGTGATGTGG	PCR/Sequencing
CHMP2B	Exon 5	CATTATTCCAGATTTAATGGCTCAA	AACGTGCATTAGGAAACATTTGGA	PCR/Sequencing
CHMP2B	Exon 6	ATCATCTTTTCTTAGTTGGAAATAG	ACATGTGTTTTGTAAAAGAAATCTG	PCR/Sequencing
TARDP	Exon 4	CACTGCATCCAGTTGAAACC	GGATTTCATGAACACACCCTG	PCR/Sequencing
TARDP	Exon 6	GGGGTTTAAATGAAATGAGTGTTC	TTTGAAGCAAAATAAACTACTGCC	PCR/Sequencing
FUS	Exon 15	TCGCTGGGTTAGGTAGGAGG	TATTCCAGTTCCTGCTGGGC	PCR/Sequencing

CLU	Exon 1	GTGTCATCCCTCTCTGCCTG	AGGGTCTCTCCTGGAGCC	PCR/Sequencing
CLU	Exon 2	GGCCTGATAGAGAGGCACTG	GACAGCTGAGGGCAGTGAG	PCR/Sequencing
CLU	Exon 3	GTCACCATGGCAACCCAG	CTTGACAGCCCCTGAACTG	PCR/Sequencing
CLU	Exon 4	CAAGCCAGCCAATGCTAATC	TGAACTGGAGCAAGGGTAGG	PCR/Sequencing
CLU	Exon 5	AAATCGAGATGACACCCGC	TGGTCACTTGCTGTGTGACTG	PCR/Sequencing
CLU	Exon 6	ACAGCTCAAAGGCTGCAGAG	ATTGCACCATTGCATTCCAG	PCR/Sequencing
CLU	Exon 7	CTGCCGTGTGATAAATGCTC	TGCATGTGCTAACTTCCATTC	PCR/Sequencing
CLU	Exon 8	CACCAGGCTATAGAGTCTGCAC	GGTTTCCTTGGAGAAGCAGG	PCR/Sequencing
CLU	Exon 9	AGTGAGACTCTGTCTCGGGG	AGCCCAAGAACAAAGCCAG	PCR/Sequencing

In order to proceed to sequencing analysis, the PCR products need to be verified in agarose-TBE gels with ethidium-bromide, and visualized with ultraviolet transillumination. After this verification, and in order to remove unincorporated primers, dNTPs, DNA polymerase and salts used during the PCR amplification that can interfere with downstream applications such as sequencing, Agencourt AMPure PCR Purification system (Agencourt, www.agencourt.com) was utilized. This system uses Agencourt's AMPure solid-phase paramagnetic bead technology for high-throughput amplification of PCR amplicons. In this cleaning reaction, an optimized buffer is used to selectively bind PCR amplicons 100bp and larger to paramagnetic beads.

The Agencourt AMPure procedure can be divided on three main stages and all of them can be performed by a Biomek FX v2.5e (Build: 17) automated work station (Beckman Coulter, www.beckmancoulter.com) (Figure 3):

1) Selective binding of PCR products to paramagnetic beads and separation of the beads with a magnetic field: 25µl of Agencourt AMPure are added to each well of the reaction plate containing the PCR products. After pipette-mixing, the solution is incubated for 3 minutes off the magnet to bind PCR products over 100bp to the magnetic beads. In order to separate the beads of the solution, the reaction plate is placed onto an Agencourt SPRIplate (Solid Phase Reversible Immobilization) for 3-5 minutes. This provides a magnetic field to each well, separating the beads from the solution.

2) Washing the beads to remove contaminants: After these 3-5 minutes, and still with the reaction plate positioned on the Agencourt SPRIplate, the cleared solution is aspirated and discarded from the reaction plate. Two washes of 30 µl of 70% ethanol during 30 seconds are performed afterwards.

3) Eluting the purified PCR products from the magnetic beads: After complete removal of the ethanol and air-drying the plate for 5-10 minutes, 30 µl of elution buffer (molecular grade water) are added to each well of the reaction plate and the solution is then pipette-mixed. Purified samples can be collected and dispensed on a new reaction plate for future usage.

Each specific purified product is now ready to be amplified using forward or reverse primers with Applied Biosystems BigDye terminator v3.1 sequencing chemistry, under the thermocycling conditions shown in Table 4.

Table 4. Thermocycling conditions used for sequencing with Applied Biosystems BigDye terminator v3.1

	Stage 1			Stage 2
	25 Cycles			1 Cycle
Temperature	96°C	50°C	60°C	4°C
Time	30	15	4 min	HOLD

In order to efficiently purify the sequencing products and to have superior quality sequencing data, Agencourt CleanSEQ technology is used. Similar to the procedure used with the Agencourt Ampure, the Agencourt CleanSEQ method can be divided into three main steps, and all of them can be performed by a Biomek FX v2.5e (Build: 17) automated work station (Figure 3):

1) Binding of sequencing extension products to magnetic beads: 5 µl of Agencourt CleanSEQ regardless of the sequencing reaction volume are added to each well of the reaction plate along with 21.4µl of 85% ethanol. After pipette-mixing the solution, the reaction plate is placed onto the Agencourt SPRIPlate for 2-3 minutes so that the magnetic beads are pulled to the side of each well.

2) Washing the beads to remove contaminants: After these 2-3 minutes, and still with the reaction plate positioned on the Agencourt SPRIplate, the cleared solution is aspirated and discarded from the reaction plate. Two washes with 30µl of 85% ethanol during 30 seconds are performed.

3) Eluting the purified PCR products from the magnetic beads: After complete removal of the ethanol and air-drying the plate for 5-10 minutes, 30 µl of elution buffer (molecular grade water) are added to each well of the reaction plate and pipette-mixed. After 5 minutes incubation, purified samples are now collected and dispensed on a new reaction plate for future usage.

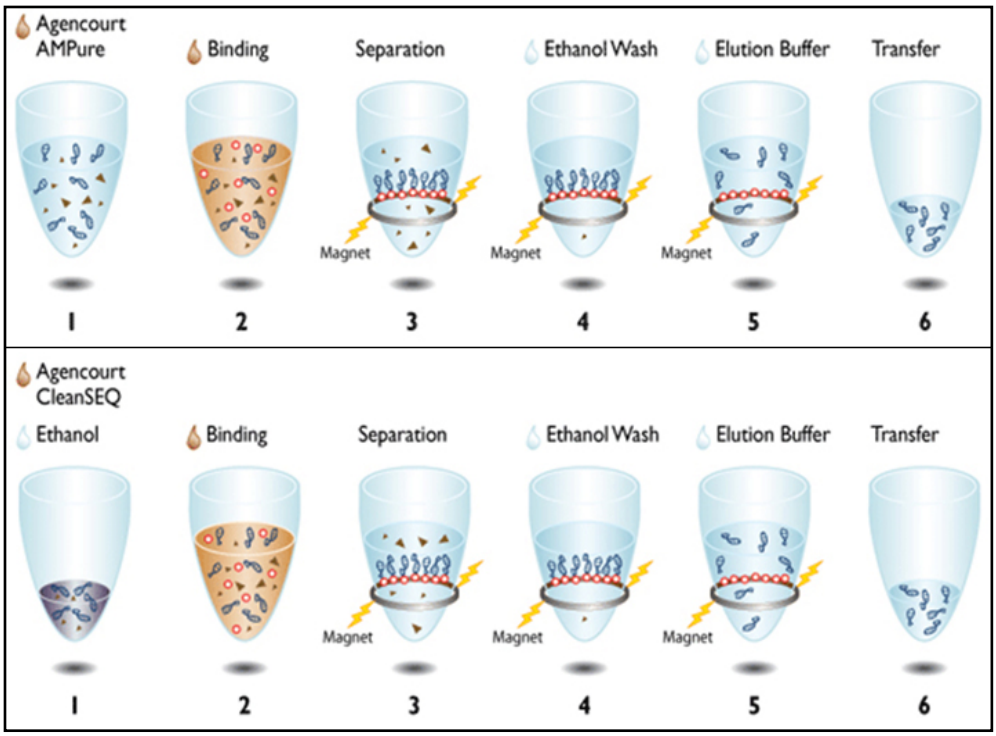


Figure 3. Overview of the PCR (upper panel) and sequencing reactions (lower panel) cleaning processes employing Agencourt technology (from http://bit.ly/1AqzIF and http://bit.ly/3d568p).

Purified sequencing reactions are now ready to be electrophoresed. All the sequencing analyses reported in this thesis were performed on an ABI3730 XL genetic analyzer (Applied Biosystems, www.appliedbiosystems.com) and the results analyzed with Sequencher software v4.1.4 (Gene Codes, www.genecodes.com).

All changes found to deviate from the wild type sequence were also verified by PCR amplification of a fresh DNA aliquot and sequenced in both forward and reverse directions.

3.2 - Whole Genome Genotyping

This methodology was used in the studies from chapter 6 and in the study of structural changes in AD in chapter 4.

3.2.1 - Illumina platform

The platform used for these studies was the Illumina's Infinium genome-wide genotyping beadarray (Illumina, www.illumina.com). This technology is based on 3 μm silica beads that self assemble by Van de Waals forces and hydrostatic interactions with the walls of micro-wells over a planar silica substrate in a slide format. When randomly assembled to this substrate, the beads have a uniform spacing of approximately 5.7 μm. Each bead is covered with hundreds of thousands of copies of a specific 50 mer nucleotide that act as the capture sequences for the fragmented target DNA. These oligos are coupled to the beads through an amine linkage, creating a "bead type" for each bead-oligo combination. "Bead types" are pooled together to form assay panels up to 1,000,000 targets depending on the chip version.

The concept of the Infinium whole-genome genotyping assays is based on direct hybridization of fragmented whole-genome amplified genomic DNA to these 50 mer locus-specific oligos. After hybridization, each SNP is "tagged" by an enzymatic-based extension assay, using labeled nucleotides. These "tags" are visualized by staining with an immunohistochemistry assay that increases the overall sensitivity of the assay (Figure 4).

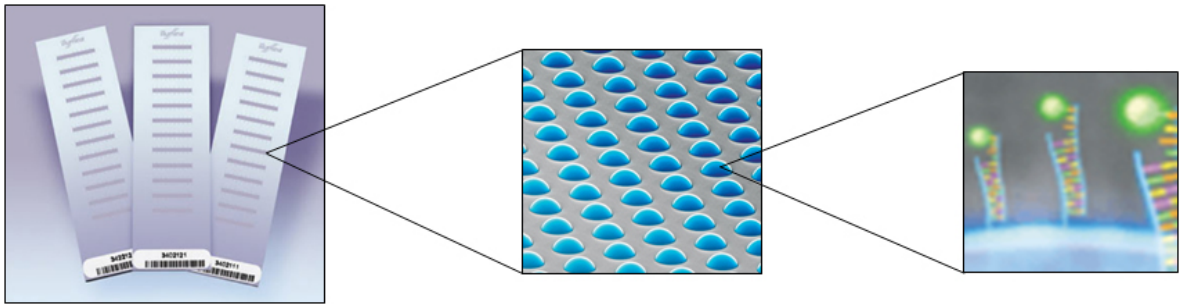


Figure 4. Beadchips. Each bead is covered with thousands of copies of a specific 50 mer capture sequence that hybridize fragmented DNA, making a beadtype. Bead types are pooled together to form assay panels up to 1,000,000 targets depending on the chip version (from http://www.illumina.com/).

Most of the samples subjected to whole genome genotyping were assayed using human610-quad beadchips. This beadchip assays more than 610000 rationally selected tag SNPs and markers per sample and four samples are used in each chip. It offers comprehensive genomic coverage across many populations, captures the majority of known variation in regions of the genome, based on HapMap release 23, and enables the detection of both know and novel CNV regions. Approximately 60,000 additional markers, developed in collaboration with deCODE Genetics, specifically target regions known or likely to contain CNVs, including segmental duplications and regions in the "unSNPable genome" (http://www.illumina.com/downloads/CNVdeCode_TechNote.pdf).

The Human610-quad beadchip assays 309978 markers within 10 kb of a RefSeq gene, 7577 non-synonymous SNPs (based on RefSeq and Ensembl databases), 5728 MHC SNPs, 20293 sex chromosome SNPs and 138 mitochondrial SNPs. It also has 3938 DGV (Toronto Database of Genomic Variants as of February 2008) regions represented. In this way, this product is suited for both whole-genome association studies and loss of heterozygosity/copy number analysis (Figure 5).

Figure 5. Illumina Human610-quad beadchip (from: http://www.illumina.com/).

tagSNPs are *loci* that can serve as proxies for many other SNPs. The use of tagSNPs greatly improves the power of association studies, as the same information and power from a larger number of SNPs can be gathered by genotyping only a subset of *loci*, thus making possible to achieve more statistical power and genomic coverage using fewer SNPs and statistical tests, compared to other strategies using larger numbers of randomly chosen SNPs.

3.2.2 - Data analysis for Infinium assays

The Infinium assay protocols feature single-tube sample preparation without PCR or ligation steps, significantly reducing labor and potential sample handling errors. An enzymatic discrimination step provides high call rates and accuracy with unconstrained locus selection. The Infinium workflow (Figure 6) can be divided on five main stages: 1) sample preparation, 2) sample fragmentation and hybridization, 3) extension and staining, 4) scanning, and 5) data analysis.

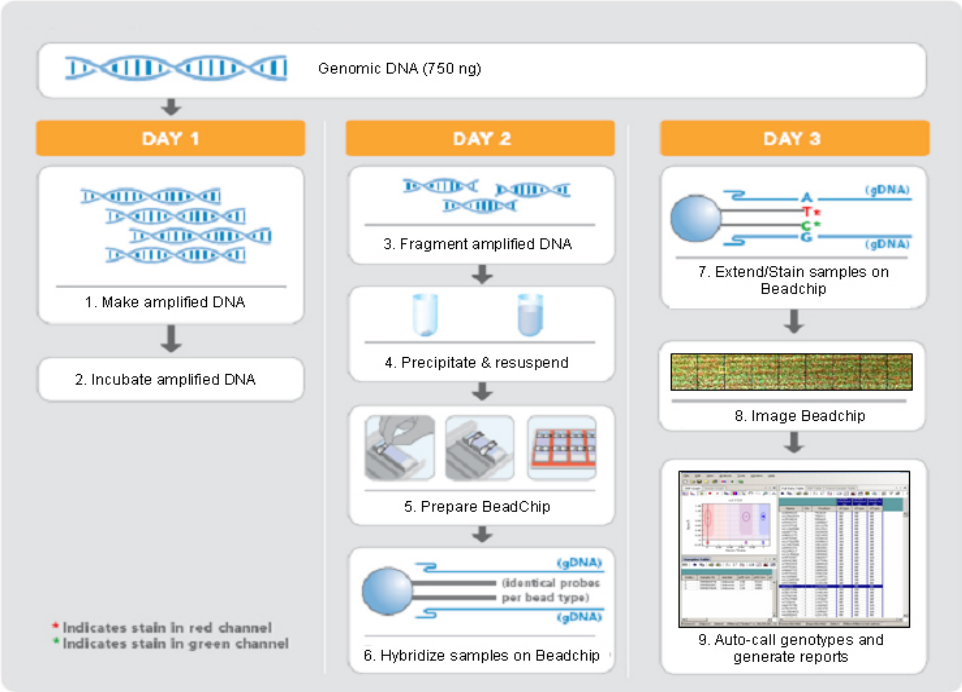


Figure 6. Infinium II assay workflow (from: http://bit.ly/5mRyN).

In order to extract genotypes from the raw image data, a normalization procedure must be undertaken. The BeadStudio software package, normalizes raw data from each BeadChip, generates genotype calls, performs clustering and analyzes data intensity using data generated from Infinium assays. Because the performance of external controls can vary from sample to sample, Illumina's standard normalization is performed by means of a self-normalization algorithm, which draws on information contained in the array itself. The normalization algorithm is designed to remove outliers, adjusts for channel-dependent background and global intensity differences, and also scales the data. During this process, the X and Y color channels (alleles A and B) undergo an affine coordinate transformation to make the data appear as accurate as possible with the homozygotes lying along the transformed X or Y axes.

Five steps are performed: 1) Outlier removal; 2) translation correction in which the asymptotes are fitted to candidate AA or BB homozygotes. Two straight lines are fit into candidate homozygote A and B alleles. The intersection of these two lines is computed and defines the translated origin; 3) rotational correction: the angle between the AA homozygote asymptote and the translated X-axis is

used to define the rotational correction. This angle corresponds to the theta (θ) parameter; 4) shear correction: the angle of the BB homozygote asymptote with respect to the translated and rotated Y-axis is used to define the shear correction. This angle corresponds to the shear parameter; 5) scaling correction: statistical centroids are computed for the candidate AA homozygotes to define an X-axis scaling parameter, and for candidate BB homozygotes to define a Y-axis scaling parameter. The translated, rotated, shear-corrected data are normalized to a scale of ~1 using the scaling parameters.

After normalization, the genotyping data are transformed to a polar coordinate plot of normalized intensity $R = X_{norm} + Y_{norm}$ and allelic intensity ratio $\theta_{subject} = (2/\Pi) * \arctan (Y_{norm}/X_{norm})$, where X_{norm} and Y_{norm} represent transformed normalized signals from alleles A and B for a particular *locus*. The log R ratio of signal intensities is shown as $\log_2 *(R_{subject}/R_{expected})$ and is the base 2 logarithm of the ratio of observed intensity versus expected intensity. $R_{expected}$ is computed from a linear interpolation of the observed allelic ratio ($\theta_{subject}$) with respect to the canonical genotype clusters. The three canonical genotype clusters were generated at one point in time by training on 120 normal samples, and serve as standards for all future experiments. In addition to computing $R_{expected}$, the observed allelic intensity ration ($\theta_{subject}$) is used to estimate a quantitative B allele frequency for the particular SNP in the given sample by using interpolation of known B allele frequencies of the three canonical clusters (0, 0.5 and 1.0).

These two transformed parameters (log R ratio and B allele frequency) are then plotted along the entire genome for all SNPs in the array using the Illumina Genome viewer tool (IGV) within BeadStudio software (Illumina, www.illumina.com). This allows identification of chromosomal aberrations and tracks of contiguous homozygous calls (Figures 7 and 8).

Accordingly, these platforms can be used to perform genome-wide association tests, but also to look at copy number variation and to score homozygosity. This can be done by direct visualization with the IGV tool of log R ratio and B allele frequency.

Because there is considerable redundancy at each *locus* interrogated (each SNP is assayed several times within a beadtype), both alleles of each SNP are assayed multiple times; accordingly, B allele value for an individual SNP in a single sample gives an estimate of the proportion of times an individual allele at each polymorphism was called A or B. This metric is simply viewed as B allele frequency, thus an individual homozygous for A allele would have a score close to 0, an individual homozygous for B allele a score close to 1, and a score approximately 0.5 would indicate a heterozygous A/B genotype. Significant deviation of these values in contiguous SNPs is indicative of an alteration in heterozygosity at a *locus*.

$R_{observed}$ is a measure of the signal intensity for a *locus* and thus, when compared with an average expected value for that locus ($R_{expected}$), the resulting log R ratio provides an indirect measure of copy number. A signal in a sample that is stronger than the expected signal is indicated by a log R ratio above 1.00 and indicates a copy number increase. On the other hand, a weaker signal than that expected for a certain SNP results in a decreased log R ratio and is suggestive of a deletion.

Considering these two metrics, a profile indicative of genomic duplication is an increase in log R ratio in the presence of four allele clusters outside the expected three. These clusters correspond to A/A/A, A/A/B, A/B/B and B/B/B genotypes (Figure 7). On the other hand, a profile indicative of heterozygous deletions would be a decrease in log R ratio in the presence of a lack of heterozygous calls (Figure 7).

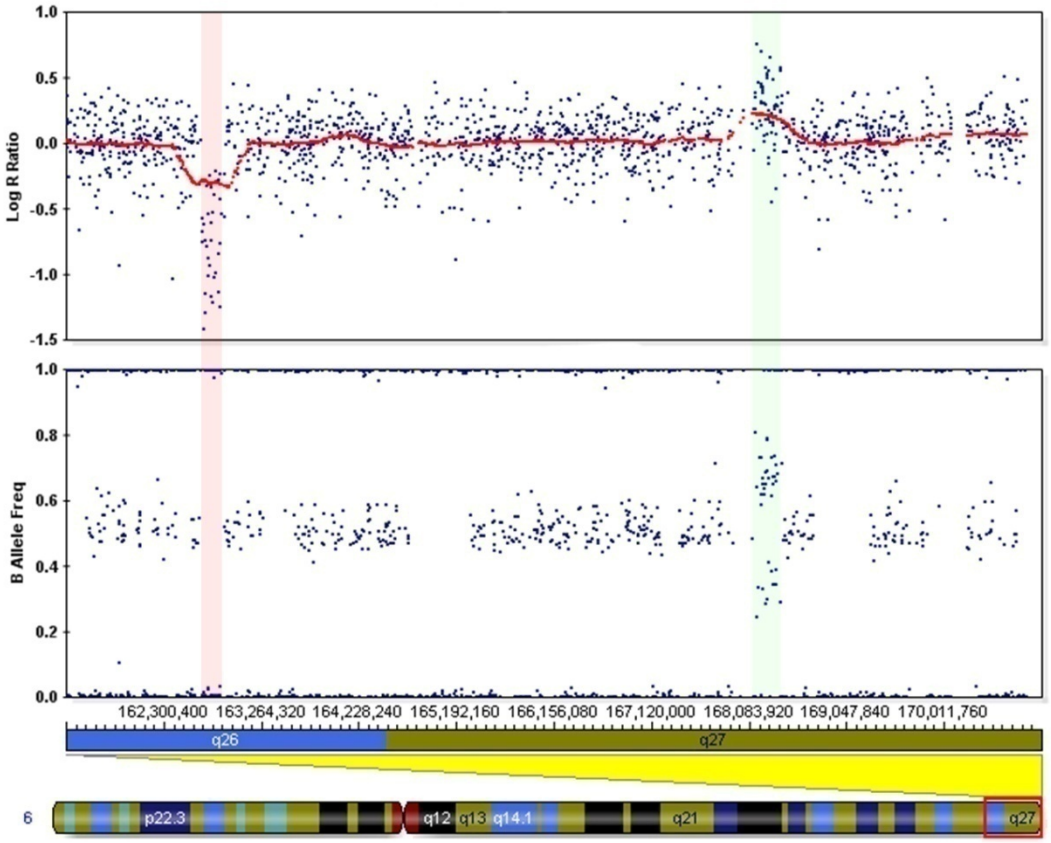


Figure 7. Typical appearance of genomic duplications and heterozygous deletions in BeadStudio IGV. Log R ratio (upper panel) and B allele frequency (lower panel) visualization. A drop in log R ratio along with lack of heterozygous calls in B allele frequency plot shows a heterozygous 0.17 Mb deletion in chromosome 6q26 (shaded in red). An increase of log R ratio coupled with the appearance of four clusters in the B allele frequency plot (corresponding with A/A/A, A/A/B, A/B/B and B/B/B genotypes) indicates a 0.24 Mb heterozygous duplication in chromosome 6q27 of this same sample (shaded in green).

The resolution of this technique is sensitive to SNP density of the assay at any particular genomic *locus*. For all platforms this varies throughout the genome. However, typically we observe copy number variations as small as 50 kb using the Infinium II assays (317,511 SNPs).

Areas of extended homozygosity are defined as a continuous track of homozygous genotypes where the track has to be at least 1 Mb in length and contain at least 10 SNPs. Visual examination will reveal the lack of heterozygous calls in the B allele frequency and a normal Log R ratio (Figure 8).

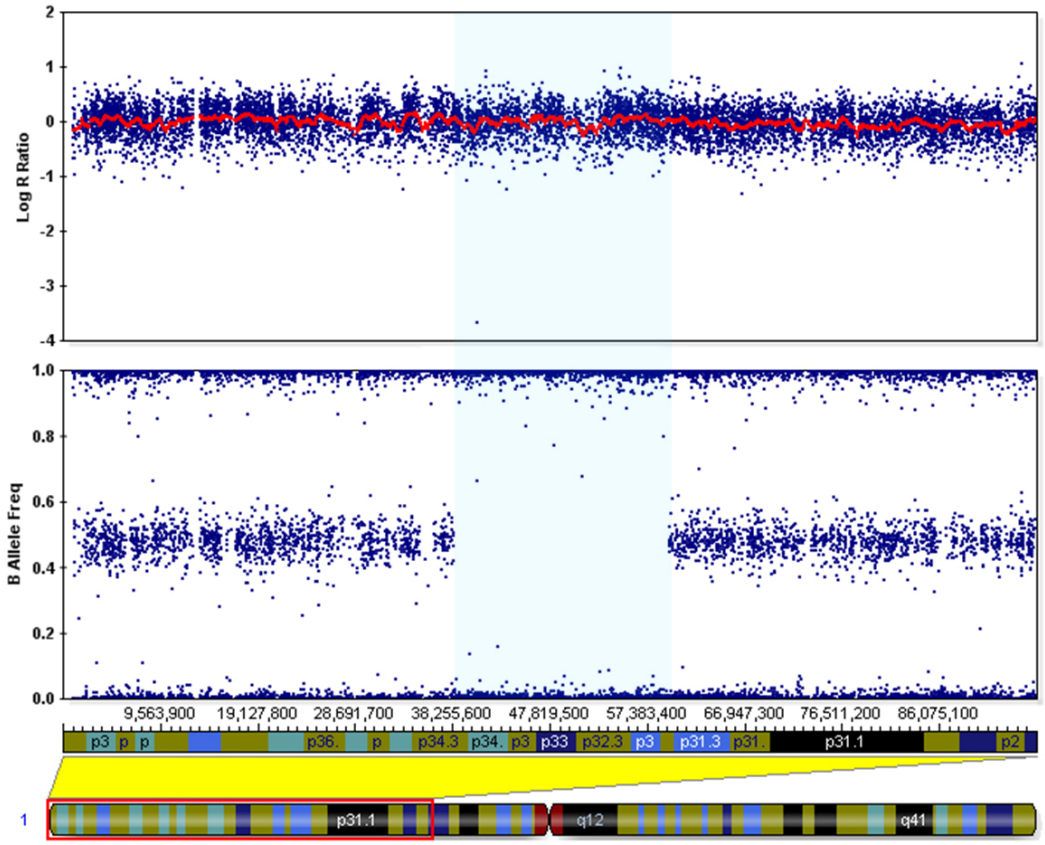


Figure 8. Typical appearance of Log R ratio (upper panel) and B allele frequency (lower panel) for an extended homozygosity track. Lack of heterozygous calls in the B allele frequency, along with no copy variation changes, indicates an extended homozygous track (shaded in blue) spanning more than 19 Mb in chromosome 1.

CHAPTER 4

MOLECULAR GENETICS ANALYSIS OF DEMENTIA CASES

Based on:

1) **Guerreiro RJ**, Baquero M, Blesa R, Boada M, Brás JM, Bullido MJ, Calado A, Crook R, Ferreira C, Frank A, Gómez-Isla T, Hernández I, Lleó A, Machado A, Martínez-Lage P, Masdeu J, Molina-Porcel L, Molinuevo JL, Pastor P, Pérez-Tur J, Relvas R, Oliveira CR, Ribeiro MH, Rogaeva E, Sa A, Samaranch L, Sánchez-Valle R, Santana I, Tàrraga L, Valdivieso F, Singleton A, Hardy J, Clarimón J. Genetic screening of Alzheimer's disease genes in Iberian and African samples yields novel mutations in presenilins and APP. Neurobiol Aging. 2008 Jul 28.

2) McNaughton D*, Knight W*, **Guerreiro R**, Ryan N, Poulter M, Nicholl D, Hardy J, Revesz T, Lowe J, Rossor M, Collinge J, Mead S. Duplications of APP – but not PRNP – are a significant cause of early onset dementia in a large UK referral series (Submitted).

3) Rohrer JD, **Guerreiro R**, Vandrovcova J, Uphill J, Reiman D, Beck J, Isaacs AM, Authier A, Ferrari R, Fox NC, Mackenzie IR, Collinge J, Warren JD, de Silva R, Holton J, Revesz T, Hardy J, Mead S, Rossor MN. The heritability and genetics of frontotemporal lobar degeneration Neurology. 2009 Nov 3;73(18):1451-6.

4) **Guerreiro RJ**, Santana I, Bras JM, Revesz T, Rebelo O, Ribeiro MH, Santiago B, Oliveira CR, Singleton A, Hardy J.Novel progranulin mutation: screening for PGRN mutations in a Portuguese series of FTD/CBS cases. Mov Disord. 2008 Jul 15;23(9):1269-73.

CHAPTER 4 – MOLECULAR GENETICS ANALYSIS OF DEMENTIA CASES

4.1 - Alzheimer's Disease

4.1.1 - Mutation screening in Iberian EOAD patients

In this study we have sequenced the presenilins and *APP* genes in a large series of early onset AD patients from Iberia. Additionally, and in order to have a better knowledge of genetic variation within these genes, we have performed the same sequencing analysis in a series of unrelated African individuals from seven different populations obtained from CEPH-HGDP, as well as in Iberian controls. The latter results are presented in chapter 5.

METHODS

Alzheimer series

A total of 231 unrelated patients (61.4% women) were recruited from 9 Iberian centers. All individuals included in this study were Caucasian with apparent Spanish or Portuguese ancestry. Mean age at onset was 52.9 years, ranging from 31 to 64. Seventy-four patients (32%) showed a family history of dementia (defined as at least one affected first degree relative) and 99 (43%) did not report any familial aggregation of disease. In 25% of individuals, no information was available. For all patients, the diagnosis of probable AD was established according to the standard Diagnostic and Statistical Manual, revision 4 (DSM IV) [214] criteria and the National Institute of Neurological Disorders and Stroke, and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) protocol guidelines [215]. Written informed consent was obtained from all participants or surrogates.

DNA Sequencing

Genomic DNA was isolated by standard procedures. The exonic regions of *APP* (exons 16 and 17), *PSEN1* (exons 3-12) and *PSEN2* (exons 3-12) genes, as well as the flanking intronic sequences, were PCR amplified using the respective primers (see Table 3 in the Methods section for primers' sequences) and Roche FastStart PCR Master Mix polymerase (Roche Diagnostics Corp., IN). Each PCR product was sequenced using the same forward and reverse primers with Applied Biosystems BigDye terminator v3.1 sequencing chemistry and run on an ABI3730xl (Applied Biosystems) genetic analyzer as per manufacturer's instructions. The sequences were analyzed with Sequencher software, version 4.2 (Genecodes, VA) [11, 216].

RESULTS

Mutation screening in AD patients

A total of five novel non synonymous mutations were found. Three of them (H214D, c.640C>G; L248R, c.743T>G; and S365A, c.1093T>G) were present in *PSEN1*, one (M174V, c.520A>G) in *PSEN2*, and a change in residue 716 of the *APP* gene (I716F, c.2146A>T). Four previously described mutations were also detected in *PSEN1*: T116N, c.347C>T; M233T, c.697A>C; V272A, c.815T>C; and A260V, c.779C>T.

Additionally, another new mutation in *PSEN2* (R71W, c.211T>C) was detected in a 77 year-old woman with disease onset at 75 years. This patient was not included in the early onset Alzheimer's disease series, but was studied at request of a clinician.

DISCUSSION

Mutations found in Alzheimer cases and their interpretation

APPI716F*: exon 17; transmembrane (TM) domain of the protein. APOE 3/3.*

The novel *APP* I716F variant is associated with the earliest age at onset described for this locus. Clinical symptoms of the proband started at the age of 31, and the patient died two years later. Neuropathological examination revealed the presence of neurofibrillary degeneration (stage VI of Braak and Braak) [217] and amyloid deposits (stage C), thus confirming the clinical diagnosis. The patient's father died at age 41 with clinically diagnosed Alzheimer's disease. Although we were not able to demonstrate segregation of the mutation with the disease in this family, it is likely to be a pathogenic variant, since other mutations have been described in the same residue (I716V and I716T) [218-220].

PSEN1 T116N*: Exon 5, hydrophilic loop I (HL-I), conserved domain in PSEN2 T122, where pathogenic variants (T122P and T122R) have been described. APOE 3/4.*

This mutation was found in a man with onset of symptoms at age 37. It has previously been reported as a pathogenic mutation on three occasions [9, 221, 222], with some segregation information and with an onset age between 30 and 40 years, being associated with an aggressive familial type of AD with a rapid progression. In the same residue, the T116I mutation was also reported [9, 223]. According to the algorithm we propose (Figure 17, on chapter 5) this is a Definitely Pathogenic mutation.

PSEN1 H214D*: Exon 7, HL-IV, conserved in H220 PSEN2 domain where no mutations are described. APOE 3/3.*

This mutation has not been previously reported. It was found in a woman with first clinical symptoms at the age of 55, including atypical signs such as bradykinesia and mild bilateral action tremor. Family history was positive: the father and grandmother of the proband presented late onset dementia. This mutation alters a conserved residue between the two presenilins. The H214Y

mutation was previously described in one family [9]. According to our algorithm (Figure 17, on chapter 5), this mutation is considered to be Possibly Pathogenic.

PSEN1 M233T*: Exon 7, TM-V, residue conserved in PSEN2 M239, where M239V and M239I are described. APOE 3/3.*

This mutation was found in a man who presented the first clinical symptoms at age 35 and died at the age of 42. The proband was first observed in a psychiatric hospital presenting atypical first symptoms with frontal profile (dysexecutive syndrome) and behavioural symptoms (depression/apathy and aggressiveness). He also presented extrapyramidal signs such as dysarthria, left hand apraxia, face and foot dystonia and pyramidal signs (Babinski). Other late signs included myoclonus and tonic-clonic seizures. Neuroimaging studies revealed hippocampal and bi-parietotemporal atrophy. The family history was unclear. This mutation has been reported several times before [9, 224, 225], as well as other mutations in the same residue (M233L, M233V and M233I) [222, 226-228]. It alters a conserved residue between *PSEN1* and *PSEN2* located in the fifth transmembrane domain. Pathogenic mutations have also been reported in the homologous *PSEN2* residue. It produces a greater proportion of Aβ42 [34, 229, 230] and fits the helix rule [33], aligning with other mutations in transmembrane domain 5 (Figure 9). This mutation is Definitely Pathogenic, based on our definition (Figure 17, on chapter 5).

PSEN1 L248R*: Exon 7, TM-VI, residue not conserved in PSEN2 V254, no mutations described. APOE 3/3.*

This mutation has not been reported before. It was found in a man presenting an age at onset of 54 years that died at the age of 65. Neuroimaging studies revealed a prominent atrophy in the lateral fissure together with a less prominent atrophy in parietofrontal regions. Family history was reported as negative. The mutation L248R does not alter a conserved residue between presenilins (V254 in *PSEN2*) and is located in the sixth transmembrane domain of PSEN1. Although it aligns with other mutations found in the same transmembrane domain (Figure 9), the paucity of the genetic data

means that we cannot be certain of the pathogenicity of this mutation. We would assign this mutation as Possibly Pathogenic (Figure 17, on chapter 5).

PSEN1 A260V*: Exon 8, TM-VI, residue conserved in PSEN2 A266 where no mutations are described. APOE 3/3.*

This mutation was found in a woman who presented the first clinical symptoms at the age of 30. Imaging studies revealed hippocampal and parietotemporal atrophy. SPECT revealed bilateral temporoparietal hypoperfusion. Family history was positive: the mother, one aunt and one cousin developed early onset dementia (<40 years). This mutation has been reported several times before [20, 231]. It alters a conserved residue between presenilins (PSEN2 A266) and is located in the sixth transmembrane domain of PSEN1. It aligns with other mutations in this domain (Figure 9) and increases the Aβ 42/Aβ 40 ratio [232]. This mutation is Definitely Pathogenic, based on our definition (Figure 17, on chapter 5).

PSEN1 V272A*: Exon 8, HL-VIa, residue conserved in PSEN2 V278 where no mutations are described. APOE 3/3.*

This mutation was found in a man who at the age of 34 years presented myoclonus and dementia. The patient died at age 42 and had a positive familial history of dementia: the father and three siblings also presented with dementia. These three siblings also carried the V272A mutation. This mutation alters a residue conserved between PSEN1 and PSEN2 but is not in a transmembrane domain (HL-VIa). It was previously reported in a three generations family with four affected subjects, two of which carried the V272A mutation, and associated with an increased Aβ42 levels in plasma [233]. Interpretation of these data as in the scheme presented in Figure 17, on chapter 5, indicates that this mutation is Definitely Pathogenic, based on our definition.

PSEN1 S365A*: Exon 10, HL-VIb, residue not conserved in PSEN2 it does not align with any aminoacid in PSEN2. APOE 4/4.*

The *PSEN1* S365A mutation is a novel variation. It was found in a woman with clinical symptoms beginning at age 55, and was not present in her two healthy siblings. Although the proband's father suffered from dementia, in him, the disease started at the age of 70. Biological specimens were not available and no further segregation studies were possible. A mutation at this residue (S365Y) has been previously reported in a patient, who also carried the *PSEN1* M146V mutation, which is almost certainly pathogenic [222]. This residue is not conserved between *PSEN1* and *PSEN2* and is not in a transmembrane domain (HL-VIb). Clearly, we cannot be certain whether this variant is pathogenic. We suggest this mutation is Possibly Pathogenic (Figure 17, on chapter 5).

PSEN2 R71W*: Exon 4, N-Terminal, residue not conserved in PSEN1 Q65. APOE 2/4.*

This novel variant was found in an elderly sporadic Alzheimer case which was not part of the central screening in this study. Neuroimaging features included chronic periventricular microangiopathic leukoencephalopathy with multiple nucleo-capsular and periventricular white matter lacunar infarctions. This mutation is Possibly Pathogenic.

PSEN2 M174V*: Exon 6, TM-III, residue not conserved in PSEN1 I168. APOE 3/3.*

This novel variant was harbored by a woman who developed Alzheimer's disease at 54 years of age. Neuroimaging features included atrophy in both parietal regions (R>L) and SPECT revealed hypoperfusion in temporoparietal regions (R>L). This variant alters a non conserved residue between presenilins (I168 in PSEN1) located in the third transmembrane domain of PSEN2. Due to the lack of family history and the fact that this is a not-conserved residue between PSEN1 and PSEN2 we assign this mutation as Possibly Pathogenic.

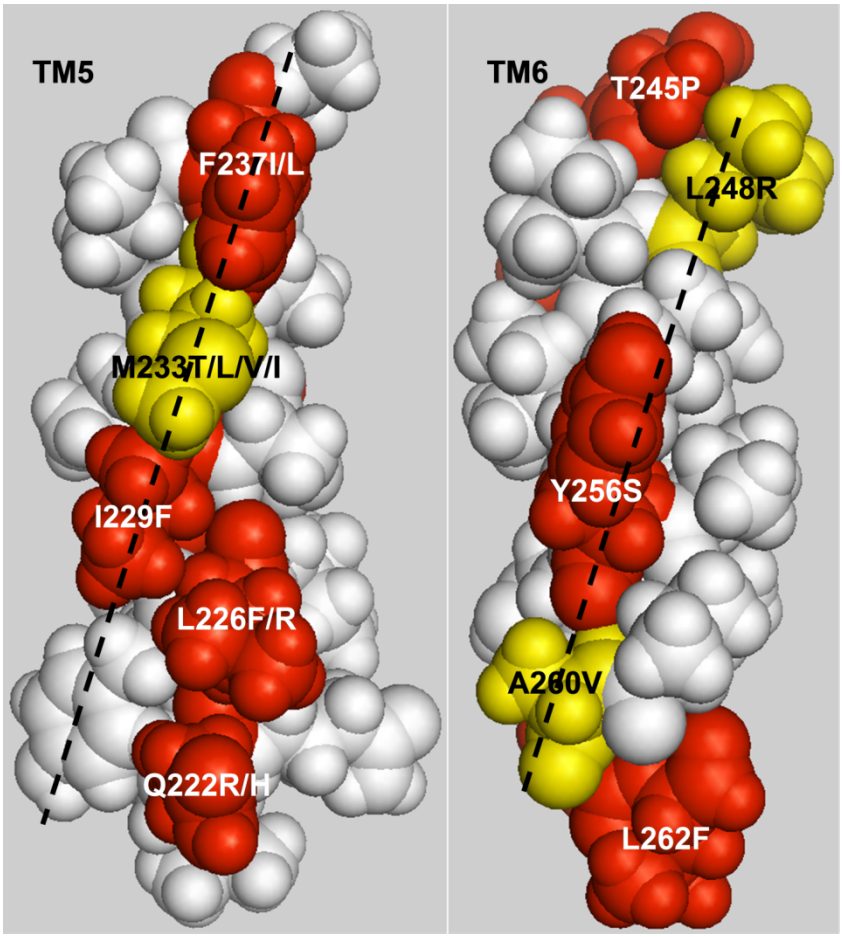


Figure 9. Example of the alignment of the mutations found in this study (in yellow) and the mutations previously described (red) in transmembrane domains 5 and 6 of *PSEN1*.

4.1.2 - Structural changes in Alzheimer's disease

With the development of therapeutics targeting amyloid-β for Alzheimer's disease or prion protein in prion disease, ascertainment and diagnosis of dementia in the earliest stages is becoming increasingly important [234]. Mutations in any of four genes are causative of early onset dementias due to either Alzheimer's disease (*PSEN1*, *PSEN2* and *APP*) or prion disease (prion protein gene, *PRNP*) [235]. Although some mutations in these genes may generate distinct clinicopathological phenotypes, they are often hard to distinguish in the very early stages, typically manifesting with episodic memory deficits and executive dysfunction. Patients presenting before old age with evidence of a familial disease should be screened for mutations in these genes; however, conventional sequencing will not detect gene deletions or duplications. Over recent years it has become recognized that copy number variations, caused by duplication or deletion of parts of chromosomes, are relatively common and occur with a non-random distribution across the genome [236]. The altered regions may include genes that encode for proteins known to be implicated in disease.

The association between AD and Down's syndrome has long been known, providing evidence that a higher gene dosage of *APP* is sufficient to produce an AD phenotype.

Recently a number of duplications of small segments of chromosome 21 which include the *APP* locus (*APP*d) have been reported in five French families [13, 14], confirming a much earlier report of *APP*d in France [237]. The prevalence and phenotypic spectrum of *APP*d are yet to be fully defined although the estimated frequency in the selected Rovelet-Lecruz cohort was 8%, about half the contribution of missense *APP* mutations to early onset, autosomal dominant AD [9]. Finnish cases of autosomal dominant early-onset AD with prominent CAA where no causative mutation was identified [15] have subsequently been screened, and some individuals found to harbor *APP*d [16]. Whilst these studies uncovered an important mutational mechanism at *APP*, several questions

remain about the frequency of these mutations in larger, less selected *APP* patient cohorts in different countries and the associated phenotypic spectrum.

In the *PRNP* gene, point mutations and insertions [238, 239] or deletions [240] within the octapeptide repeat region, have been associated with inherited forms of human prion diseases. In common with AD, prion diseases are caused by protein misfolding. Transgenic models have clearly established that expression of the prion protein gene is an important determinant of incubation time to prion disease [241, 242]. By analogy with AD, duplication of *PRNP* would therefore be an excellent candidate mutation for susceptibility to human prion disease. Although virtually all familial concurrence of prion disease is associated with known mutational mechanisms at *PRNP*, the cause of sporadically occurring Creutzfeldt-Jakob disease remains unexplained. The possibility of *PRNP* copy number variation in human prion disease has not been explored.

In this study we investigated the presence and frequency of copy number variations at *APP* and *PRNP* loci in a large referral cohort of UK patients, and present an optimized methodology for cost-effective high-throughput screening of patients for such structural variation. We identified 5 probands and 11 suspected relatives with *APP*d but no *PRNP* duplication. These data emphasize the importance of *APPd* with respect to *APP* missense mutations in a cohort of FAD families followed-up since the discovery of *APP* as the first AD gene. Clinical histories highlight the prominence of seizures, which may precede dementia. An absence of family history may relate, in part, to a novel mutational mechanism.

METHODS

Patient cohorts

The AD cohort comprised two subgroups. One group was formed from a referral series of 381 patients for a diagnostic test of AD gene mutation, average age was 53 and 47% were male. This series were screened for mutations in *APP*, *PSEN1, PSEN2*, and *PRNP* although not all patients were

screened for all genes as, for example, some patients were referred prior to the identification of the presenilins as causal genes and thus these genes have not been screened. Our practice is to sequence these genes in patients with unexplained early onset dementia, or in older patients where there is additional evidence of a familial disease. This cohort largely comprises probands with familial dementia followed-up at the Dementia Research Centre, UCL Institute of Neurology. A further 492 patients comprise a collection of patients thought to have AD but with insufficient clinical evidence to justify screening of causal genes. The prion disease cohort comprised 231 patients with probable or definite sporadic CJD and a further 427 cases referred to the National Prion Clinic or MRC Prion Unit for *PRNP* gene testing as a result of prion disease being amongst the differential diagnoses, average age was 55 and 50% were male. All patients were *PRNP* mutation negative. All patients gave informed consent for genetic analysis. Ethical approval for the study was given by the University College London Hospitals NHS Trust Local Ethics Committee.

Sample Preparation

Genomic DNA (gDNA) was extracted from peripheral whole blood or post mortem brain material using standard techniques. gDNA quality was assessed using agarose gel electrophoresis. Only those gDNA samples of average molecular weight >10Kb were included in the study. Concentration was assessed using a NanoDrop spectrophotometer (ThermoScientific, NanoDrop Products, Wilmington, DE) and samples were diluted to 20ng/µl in 1X tris-EDTA buffer prior to use.

Real-Time Quantitative PCR (exon-qPCR)

APP alleles were quantified on an ABI 7000 Sequence Detection System (Applied Biosystems, Foster City, CA, USA) using the 5' nuclease assay and duplexed minor groove binding (MGB) probes designed to detect *APP* (exon 5) and *GRN* (exon1). Relative quantification of *APP* using *GRN* as the internal reference was determined using the ΔΔCt method. Similarly, *PRNP* alleles were quantified using duplexed MGB probes designed to detect *PRNP* (exon 2) and *APP* (exon 5), see Table 3 in the Methods section for primers' and probes' sequences. These tests were designed primarily to demonstrate *APP* or *PRNP* copy number variation (CNV), however, the choice of the internal

reference in the screening of individual cohorts allowed for the detection of rare phenocopies by the possible detection of *GRN* CNV in the AD cohort and *APP* CNV in the prion disease cohort respectively. Twenty nanograms of gDNA were amplified in a 25µl reaction volume containing 1X Expression Mix (Applied Biosystems), and the respective primer (900nM) and probe (200nM) sets for both the gene of interest and the internal reference using the universal protocol. Each sample was quantified in triplicate with each replicate having a different plate reference to minimise the affect of interwell variation. One sample with trisomy 21 was included as a positive control on all plates. All samples with a dosage quotient (expressed as a ΔCT) two standard deviations from the mean were considered as potential CNVs.

Fluorescent Microsatellite Quantitative PCR (fm-qPCR)

APP alleles were quantified by the genotyping of two microsatellites located within intron 1 of *APP* (196999; Chr21: 26,460,990 – 26,461,175) and 330kb centromeric from *APP* (188463; Chr21: 25,841,358 – 25,841,530). *PRNP* alleles were similarly quantified by genotyping microsatellites located 30Kb (108991;Chr 20: 4,589,668 - 4,589,950) and 165Kb (452; Chr 20: 4,454,279 - 4,454,435) telomeric of *PRNP* (NCBI Map Viewer Build 36.3), see Table 3 in the Methods section for primers' and probes' sequences. Twenty nanograms of gDNA were amplified by denaturation at 95˚C for 1 minute followed by 24 cycles of 95˚C for 30 seconds, 53˚C for 30 seconds and 72˚C for 1 minute. Amplicons were analysed using an ABI 3130xl automated sequencer and GeneMapper software v.4.0 (ABI). In all individuals heterozygous for a given marker the allelic ratio was assessed by relative peak area. Cases were deemed CNV positive by fm-qPCR if allelic ratios were two standard deviations from the mean and if i) both of a microsatellite pair had altered allelic ratio or ii) one of a microsatellite pair had an altered peak ratio with the other genotyped as homozygous or if the other was further from the gene located on the same side of the chromosome.

Illumina 610 Bead Array

Samples identified as potential CNV by exon-qPCR and/or fm-qPCR were assayed using the Illumina bead array system according to the manufacturers protocol. All samples were assayed using the Illumina Human610-Quad BeadChips (Illumina Inc, San Diego, CA, USA) as per manufacturer's instructions, using 200 ng of genomic DNA. These BeadChips assay more than 610.000 tag SNPs and markers, including 60.000 CNV markers, based on the International Human HapMap project release 23 (www.hapmap.org). All the samples genotyped had a genotype success rate of more than 99%. In order to examine individual chromosomes for structural mutations we used the visualization tool Genome Viewer version 3.2.9 within Beadstudio version 3.1.3.0 (Illumina Inc, San Diego, CA, USA). Data was analyzed using the Humane Genome Build 35 and two metrics were visualized: B allele frequency and log R ratio. The former is the theta value for an individual SNP corrected for cluster position, which gives an estimate of the proportion of times an individual allele at each polymorphism is called A or B. The log R ratio is the log2 ratio of the observed normalized R value for the SNP divided by the expected normalized R value for the SNPs theta value. Expected R is calculated from the values theta and R, where R is the intensity of dye-labeled molecules that have hybridized to the beads on the array and theta is the ratio of signal at each polymorphism for beads recognizing an A allele to beads recognizing a B allele. The expected R value for any individual at any typed SNP is calculated using a large population of typed individuals. Therefore, the ratio of observed R to expected R in any individual at any SNP gives an indirect measure of the binding efficiency of detected alleles for each polymorphism and thus of genomic copy number. An R above 1 is indicative of an increase in copy number, and values below 1 suggest a deletion. We have shown previously that this is a very reliable method for detecting genomic copy number mutations [243-245]. We looked at both log R ratio and B allele frequency plots across the whole genome in each sample, with particular interest in the region containing the *APP* gene, in order to be able to confirm the results previously obtained.

RESULTS

Laboratory findings

Using Exon-qPCR comparison was made between the amplification of *APP* or *PRNP* and an internal reference marker (*APP-GRN* and *PRNP-APP* respectively). 21 samples from the AD cohort and 19 samples from the prion disease were considered as potential CNV's based on the selection criteria of samples with a ΔC_t value greater than 2 standard deviations (2SD) from the mean (Table 5).

Table 5. Performance of screening methods for APP and PRNP CNVs.

***APP* CNV Screen**						
	Total Sample Number	**Double homozygotes**	**Number Screened**	**Positives**	**True Positive**	**False Positive**
Exon-qPCR	1531	N/A	1531	21	5	16
fm-qPCR	873	54	819	42	5	37
Both	873	N/A	873	5	5	0
Illumina Array	10	N/A	10	5	5	N/A

***PRNP* CNV Screen**						
	Total Sample Number	**Double homozygotes**	**Number Screened**	**Positives**	**True Positive**	**False Positive**
Exon-qPCR	1531	N/A	1531	19	0	19
fm-qPCR	658	33	625	11	0	11
Both	658	N/A	632	0	0	0
Illumina Array	N/A	N/A	N/A	N/A	N/A	N/A

Using fm-qPCR two polymorphic microsatellites within intron 1 and 330kB centromeric to *APP* and two microsatellites 30kb and 165Kb telomeric to *PRNP* were amplified using a standard PCR protocol for analysis. A comparison of peak area was made in all heterozygous individuals for either one or both of the microsatellite pair. The samples screened were able to be assessed for CNV's in this way with only 6.1% of individuals being double homozygotes in the AD cohort and 5.0% individuals in the *PRNP* cohort (Table 5). 42 samples from the AD cohort and 11 samples from the *PRNP* cohort were considered as potential CNVs. Three allelic peaks, appeared on the electropherogram of Proband 3,

microsatellite 196999. This was consistent with an unbalanced translocation that does not appear to include Down's syndrome critical regions.

From the prion disease cohort one sample with fm-qPCR appeared to have an abnormal peak area ratio between microsatellite allelic areas on electropherograms, as well as, a high ranking on exon-qPCR although it did not quite fall 2SD from the mean. In addition, the abnormal ratio occurred for marker 452 (microsatellite furthest away from the *PRNP* sequence) and not for marker 108991 (microsatellite closest to the *PRNP* gene) and therefore could not represent a contiguous duplication involving *PRNP*. Consequently, no samples of the *PRNP* cohort were positive for both methods and none were carried forward for verification by Illumina array (Figures 10 and 11). Five samples from the *APP* cohort were deemed positive by both exon-qPCR and fm-qPCR. All these samples were confirmed as heterozygote duplicates by the Illumina array. The complementarity of the two techniques therefore represents a highly specific screening process for heterozygous duplications, using the Illumina array as the gold standard assay. We cannot be certain of high sensitivity as only 10 samples were screened using the Illumina array, however 0/5 samples positive on a single assay were confirmed to have a duplication. The specificity of fm-qPCR may be improved by considering the heterozygous microsatellite allele sizes, in that a large proportion of the variation in allelic ratio was determined by stutter peaks when allele sizes were similar (Figure 12).

The Illumina 610 Bead array was used for the verification of duplication in samples that were identified as potential CNVs from exon-qPCR and fm-qPCR. These data showed duplicated regions of Chr21q, which included the *APP* gene locus in all 5 probands identified as CNV's by both, exon- and fm-qPCR. There was considerable heterogeneity in duplication size (2.77, 6.35, 4.96, 6.49Mb, see Figures 10 and 11) showing that these patients represented separate duplication events. Although Proband 3 had been identified by fm-qPCR as having three microsatellite alleles, the array revealed this patient to have an extended discontinuous duplication of Chr21q. There are no available single nucleotide polymorphisms for the Chr21p, so this could not be assessed. The large duplicated region

is interrupted by a region of unduplicated SNPs (see Figure 10). In conjunction with the microsatellite haplotyping, these data are consistent with an unbalanced translocation of chromosome 21.

Of the five positive individuals of the AD cohort, 4 (1%), (Probands 1,2,3 and 5) originated from the referral series of 381 patients for a diagnostic test of AD gene mutation and screened for *APP*, *PSEN1*, *PSEN2* and 1 (0.2%), (Proband 4) originated from 492 patients comprising a collection of patients thought to have AD but with insufficient clinical evidence to justify screening of causal genes.

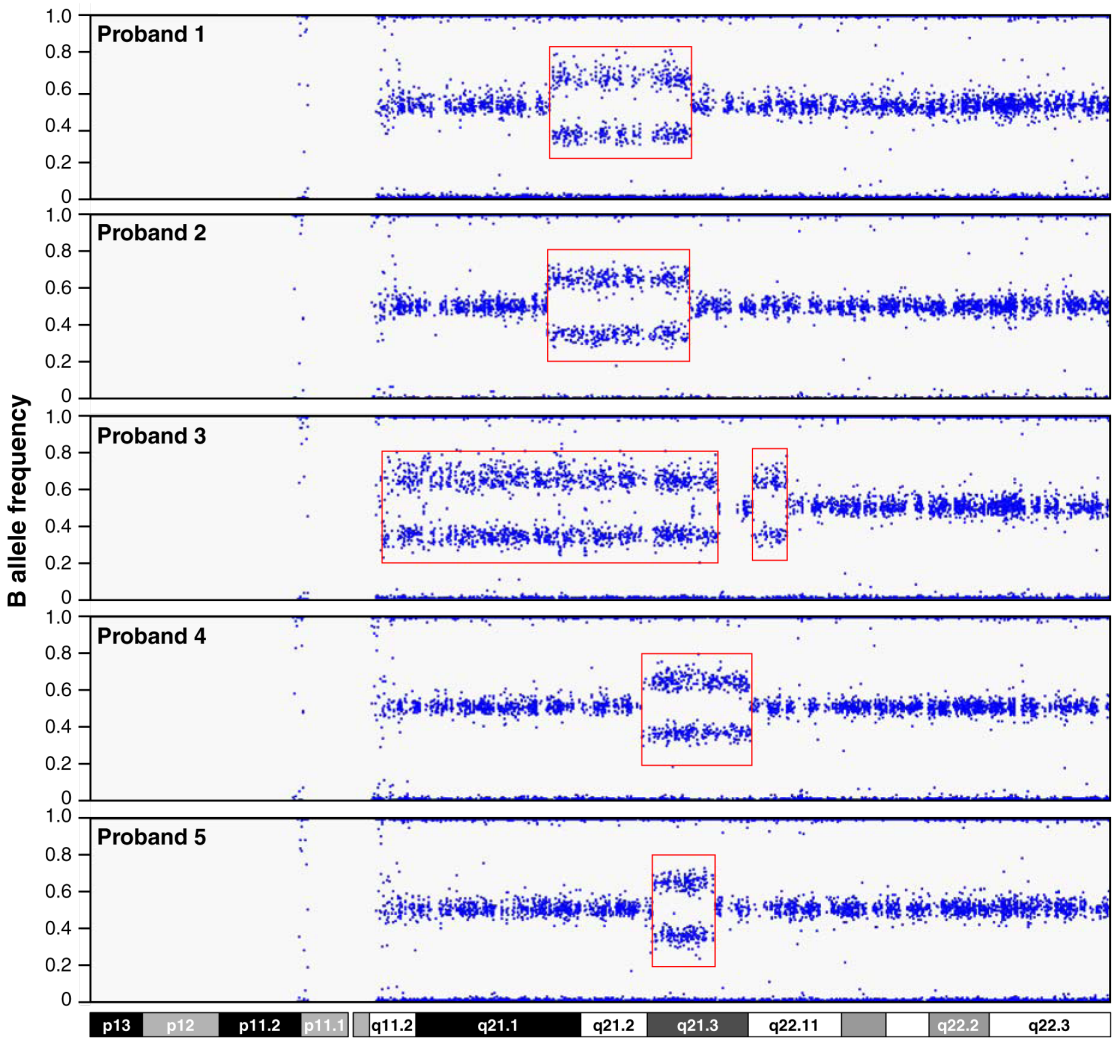


Figure 10. Confirmation of APPdup using Illumina array technology. Heterozygous duplications in 5 probands. Regions of duplication are highlighted in red boxes. In order to examine individual chromosomes for structural mutations the visualization tool Genome Viewer version 3.2.9 within Beadstudio version 3.1.3.0 (Illumina Inc, San Diego, CA, USA) was used. An ideogram of Chromosome 21 and the B allele frequency is illustrated. 1 pixel=56 kb.

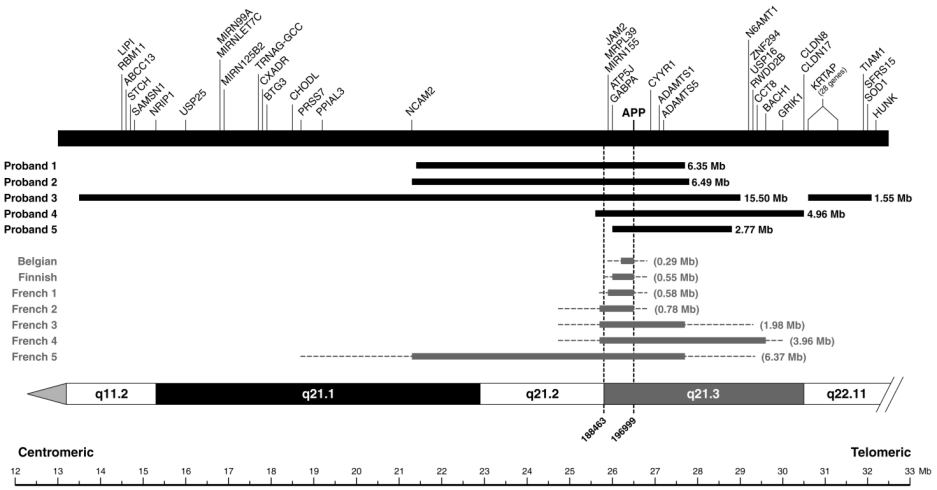


Figure 11. Chr21 diagram showing APPdup regions with respect to genes. *APP* duplications in 5 probands. Black horizontal bars indicate the extent of heterozygous duplications. Minimal sizes (in parenthesis) of previously reported duplications are indicated by grey horizontal bars and the intervals of the duplication boundaries by dotted lines. Gene content excludes pseudogenes and open reading frames. Vertical bars indicate the transcription start site of genes. Proband 1 has a partial duplication of *NCAM2* and Proband 2 has a full duplication of this gene.

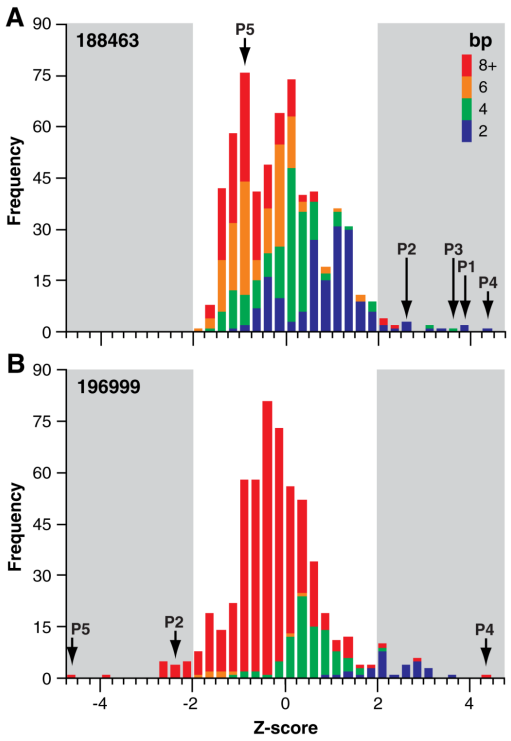


Figure 12. APPdup detection using alteration in the ratio of area under the curve of microsatellite alleles on electropherogram traces. Histograms are shown to illustrate the distribution of allelic peak area ratios. Variation occurring among the ratios of microsatellite allelic areas appeared to be polymodal. This observation was caused by 'stutter' peaks immediately preceding each microsatellite allele. Microsatellite stutter peaks accentuate the peak of the shorter allele to a degree, which is based upon the size differential of the two alleles. Thus, the degree of difference in base pairs between two alleles accounts for part of the variation in allelic ratios. When the two microsatellite alleles are widely separated in length, the ratio is not affected by stutter peaks, whereas when alleles are adjacent the ratio is affected most. This phenomenon accounted for most of the false positives using this method alone. Differences in base number between allelic pairs are illustrated by different colour. Shaded areas highlight samples 2SD from the mean. Microsatellite marker 188463 Proband 5 had a normal allelic peak ratio. Array analysis confirmed that this marker lay outside the duplication boundary for this individual. Microsatellite marker 196999 Proband 3 had three microsatellite peaks. And Proband 1 was uninformative due to homoygosity.

Clinical and investigation findings

The clinical and investigation findings for the five probands/families where APPd were confirmed follow next (Tables 6-8, Figures 13 and 14).

Table 6. Clinical features of 11 individuals with definite or possible APPd in the London series.

Pedigree	Proband	APPdup	Age of onset	Age of Death	Clinical features	Seizures	ICH
1III.1	1	Yes	48	-	Memory loss, apraxia, dysexecutive	Late onset tonic-clonic	Yes
1II.1		-	-	63	Dementia	No	-
1II.2		-	Late 40s	62	Dementia	No	-
1II.3		-	-	49	Dementia	No	-
2III.2	2	Yes	40	-	Cognitive decline, memory loss, myoclonus	Early onset tonic-clonic (from age 11)	No
2III.1		-	50	-	Memory loss, word finding difficulty, altered mood	Generalized and complex partial seizures	No
3II.1	3	Yes	39	-	Memory loss, social withdrawal, cognitive decline	Probably complex partial seizures	No
4II.1	4	Yes	-	-	-	-	No
4I.1		-	-	49	Dementia	-	Yes
5III.1	5	Yes	48	-	Memory loss	Late onset generalized and complex partial	No
5II.1		-	61	66	Memory loss, word finding difficulty	No	No

A further 6 family members had some evidence of the typical clinical syndrome of APPd (see family histories and pedigrees). ICH = intracranial hemorrhage.

Proband 1 (APOE E3, E4)

This 54 year-old man had a progressive impairment of episodic memory, beginning insidiously at the age of 48 years. Impairment of verbal memory was clearly impaired at presentation, although general intellectual functions were preserved. He went on to develop an apperceptive agnosia, limb apraxia and dysexecutive features. He was a former smoker. At 52 years he had a generalized tonic-clonic seizure and elected not to initiate anti-convulsant treatment, although donepezil 10mg was

prescribed around this time. There was an autosomal dominant family history of early onset AD: his father died at 62 years after onset of disease in his late 40s, a paternal uncle died at 63 years and a paternal aunt died at 49 years, both after an undefined dementing illness (Figure 13).

Initial MRI brain scanning showed widespread T2 FLAIR white matter hyperintensities consistent with subcortical vascular disease; multiple small, deep cerebral and cerebellar haemorrhages on T2* imaging and an old left-sided amygdala haemorrhage (Figure 14). Electroencephalography was unremarkable. There was no history of sustained hypertension and cholesterol was 3.5mmmol/L. Tests for Notch 3 (CADASIL), *ITM2B* (Familial British Dementia), *APP, PSEN1*, *PSEN2*, *PRNP* and *MAPT* mutations were all negative. Cerebrospinal fluid (CSF) analysis was unremarkable except for total tau of 555pg/mL and Aβ-42 of 128pg/mL. Normal ranges are yet to be conclusively established for these variables but an upper limit of 445 pg/mL for t-tau and lower limit of 427 pg/mL for Aβ-42 have recently been proposed [246]. See Table 7 for longitudinal neuropsychological assessments.

Table 7. Longitudinal neuropsychometry for proband 1.

Donepezil was initiated between the 2007 and 2008 assessments.

DATE (AGE) / TEST	2006 (52)	2007	Jan 2008	Nov 2008
VIQ	90	87	88	Unknown
PIQ	76	80	78	Unknown
RMTwords (%ile)	CHANCE	<5	<5	<5
GNT (%ile)	25	5-10	10-25	10
MMSE	25	23	Unknown	26

VIQ/PIQ = Verbal and performance subscales of WAIS-R (WECHSLER 1981)
RMTw/RMTf = Recognition memory test for words and faces (*short version) (WARRINGTON EK 1984 & 1996)
GNT = Graded naming test (MCKENNA & WARRINGTON 1980)
MMSE = Mini-mental state examination (FOLSTEIN 1975)

Proband (APOE E3, E4) and family 2

This man belongs to a family with a strong history of both early onset seizures and mid-life cognitive impairment. He developed a variety of seizure types from the age of 11 years with brief absences occurring up to five times per day and generalised tonic-clonic seizures on average once per year. A flurry of seizures at the age of 40 years heralded increasing social withdrawal, apathy and an inability to follow simple commands. His cognition progressively deteriorated, including problems with memory, mood and wandering, as well as prominent myoclonus. From the age of 49 years he had word-finding difficulties, which progressed to the extent that he would barely speak. He became increasingly immobile and urine incontinent.

He had no other past medical history, though he had previously been a heavy smoker. On examination at the age of 50 years he had an MMSE score of 10/30, a paucity of speech, perseveration, poor concentration, moderate limb dyspraxia and visuoperceptual difficulties. General neurological examination revealed a supranuclear gaze palsy affecting vertical movements, a brisk jaw jerk and brisk limb reflexes. His final MRI brain scan at the age of 50 years revealed severe generalised cerebral volume loss and bilateral hippocampal atrophy. There was extensive periventricular white matter signal change, more so on the right than the left. EEG showed an excess of slow wave activity with absent alpha rhythm, while telemetry revealed bifrontal spike activity. CSF was acellular with normal total protein, negative oligoclonal bands and negative Whipple's PCR. Genetic screening for *APP*, *PSEN1*, *PSEN2*, *ITM2B, DRPLA, HD* and common mitochondrial mutations were negative. White cell enzymes were normal. He also underwent skeletal muscle and skin biopsies. The former showed no evidence of mitochondrial myopathy and the latter showed any Lafora bodies. Longitudinal neuropsychological testing was undertaken (Table 8) which demonstrated a relentless, slow, global cognitive decline.

Table 8. Longitudinal Neuropsychometry with corresponding electroencephalographic and imaging profile for proband 2.

YEAR (AGE) / TEST	1998 (44)	2000	2001	2002	2003	2004 (50)
VIQ	71	78	76	75	62	63
PIQ	73	67	63	64	62	57
RMTwords	29/50	34/50	27/50	16/25*	CHANCE	CHANCE
RMTfaces	28/50	27/50	25/50	14/25*	CHANCE	CHANCE
GNT (%ile)	10-25	10-25	10-25	10	5-10	5
MMSE		22		15	10	10
MRI	- Frontal white matter lesions - No regional atrophy	- White matter lesions both hemispheres - No regional atrophy	- White matter lesions both hemispheres - Bilateral hippocampal atrophy	→	- White matter lesions both hemispheres - Generalized atrophic change, greater anteriorly - Bilateral hippocampal atrophy right smaller than left	- Generalized, severe cerebral volume loss - Extensive white matter change more prominent in right hemisphere - Bilateral hippocampal atrophy
EEG	- Frontotemporal slowing	- Traces of alpha rhythm, large amounts of theta in both hemispheres	→		- Widespread slowing - No normal rhythms - No posterior alpha rhythm	

VIQ/PIQ = Verbal and performance subscales of WAIS-R (WECHSLER 1981)
RMTw/RMTf = Recognition memory test for words and faces (*short version) (WARRINGTON EK 1984 & 1996)
GNT = Graded naming test (MCKENNA & WARRINGTON 1980)
MMSE = Mini-mental state examination (FOLSTEIN 1975)

His sister also developed both generalised and absence seizures at the age of 11 years. From the age of 50 years she had a gradual, progressive impairment of episodic memory, with mild word-finding difficulties, low mood and emotional lability. Her MRI brain scan showed, scattered, bilateral, periventricular areas of high signal in the deep white matter, basal ganglia and pons. There was also a high signal area in the right frontal lobe consistent with an old infarct. CSF examination was unremarkable. Early onset seizures also affected his mother, a maternal aunt, one of his sisters, the daughter and granddaughter of one of his unaffected sisters, as well as, the grandson of another of his unaffected sisters although samples are not available for testing of segregation in these individuals. His mother died aged 64 years with a long history of psychiatric problems. Lifelong

generalized epilepsy was complicated by post-ictal paranoid psychosis and, although little information is available on her early life, problems with activities of daily living, self-care and an inability to remember her medication were present at least from the age of 62 years. Histological post-mortem examination of her brain revealed no convincing evidence of a neurodegenerative process. A cavity, consistent with previous haemorrhage or infarct, was identified in the white matter with abundant haemosiderin and surrounding gliosis. Several of the cortical and leptomeningeal vessels had evidence of hyaline degeneration though no amyloid deposition. Scanty diffuse amyloid plaques were present in the striatum although no other amyloid deposition was identified. There were no neurofibrillary tangles.

Proband 3 (APOE E3, E3)

This Greek man presented at 42 years old with a three-year history of mislaying items and forgetting events. Later problems with route-finding were noted with increasing social withdrawal. Episodes of uncontrollable weeping occurred with associated malaise, rolling of the eyes and post-ictal confusion. These episodes were responsive to anti-convulsant therapy. He was an only child of two healthy parents and had no known extended family history of dementia.

MRI brain scans revealed generalized atrophy and EEG revealed an absence of alpha rhythm with disorganised activity and delta discharges. CSF analysis was unremarkable. Screening of *APP*, *PSEN1* and *PRNP* was negative. Neuropsychometry revealed a relatively preserved social façade, marked difficulty with non-verbal reasoning, dysexecutive features and impaired episodic memory. Comprehension was intact, word-retrieval problems and speech production errors (in the form of substitutions, deletions and transpositions) were evident as well as apperceptive agnosia. Reading, writing, calculation, visuoperceptual and visuospatial skills were impaired and limb apraxia was present.

Proband (APOE E3, E4) and family 4

Little clinical information about this proband is available, although it is known that she suffered from

a progressive temporoparietal dementia. Later she developed extrapyramidal features, although the sporadic use of dopamine antagonists may have contributed. Age at onset is unknown but by 57 years she was resident in a nursing home and had a severe dementia with dysphasia, dyspraxia, motor perseveration, utilisation behaviour and wandering. Nothing is known about her only sibling but her mother is known to have died aged 49 years from pathologically confirmed 'cerebral haemorrhage' and 'cystic degeneration of the brain' after a presenile dementing illness.

Proband (APOE E3, E3) and family 5

This 55-year-old woman had progressive memory problems since the age of 48 years when she fell out of bed having had a focal seizure. Seizures consisted of an olfactory aura followed by a focal seizure affecting her right arm. Complex partial and atonic seizures progressed and were refractory to therapy. She developed progressive cognitive symptoms. Aged 53 her MMSE was 16/30 and Addenbrooke's Cognitive Examination 50/100. Aged 54y MMSE was 12/30 and a year later, MMSE was 4/30. Her cognitive profile was typical of AD. Apart from generally brisk reflexes, neurological examination was normal.

MRI imaging showed atrophy of the left temporal lobe compared to the right. HMPAO SPECT scan which showed some irregular diminution of perfusion of the left parieto-occipital region compared to the right. These appearances were felt to represent ischaemic change or AD.

The patient's mother died aged 66 years of cancer but had progressive cognitive problems for 5 years including word finding difficulties and neologisms.

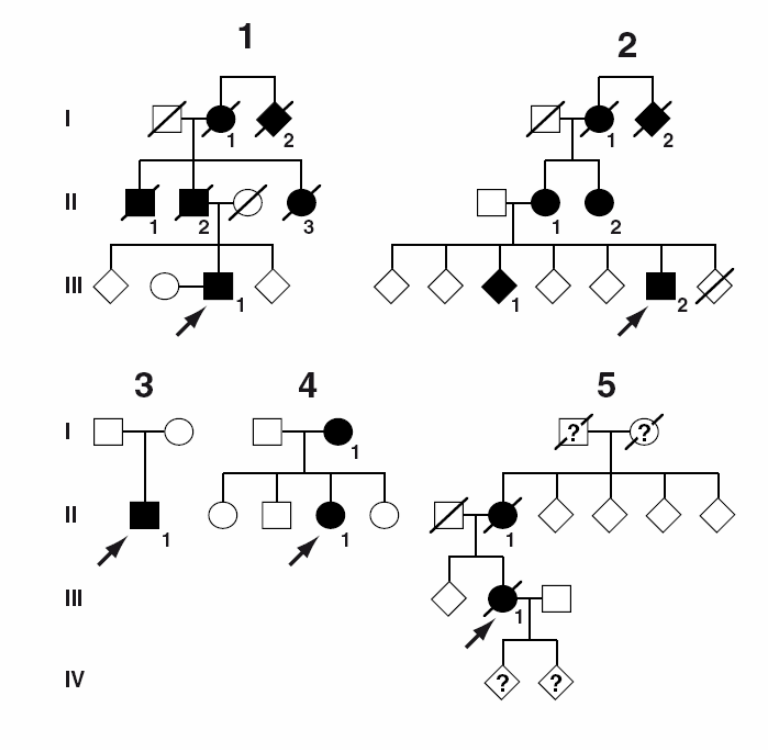


Figure 13. Pedigrees of the five families where a duplication of the *APP* gene was found.

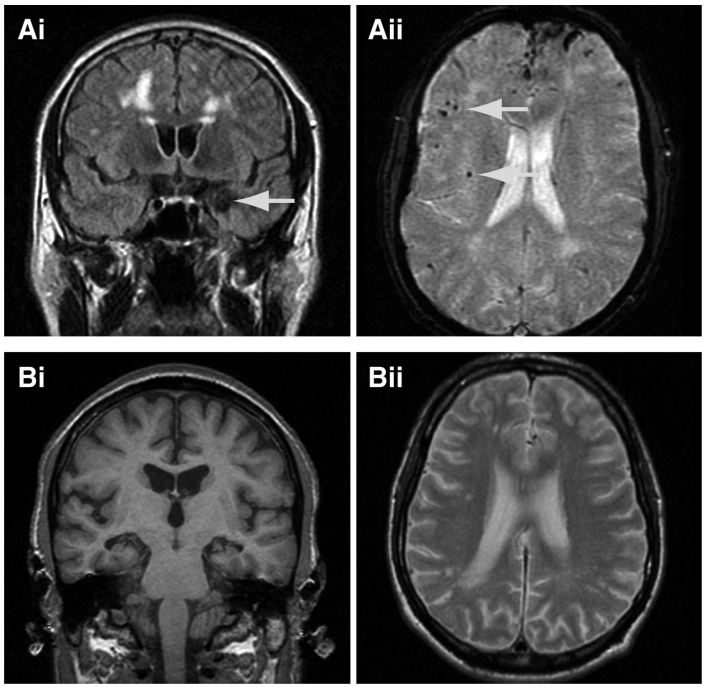


Figure 14. MRI scans from APPd patients. (A) proband 1 showing extensive white matter abnormalities (Aii) and an amygdala hemorrhage (Ai, arrow) (B) proband 2 (aged 47) showing white matter abnormalities on T2 weighted scans (Bii) and hippocampal atrophy (Bi).

DISCUSSION

We describe the screening of a large and heterogenous referral series to identify CNVs causing early onset dementia. Five probands with *APP* duplication were identified in the screen, four with evidence of familial dementia and one with a sporadic disease. The overall frequency of mutation in our series was 0.57% (95% CI 0.19-1.33). Although we confirm *APP*d to be relatively rare, the frequency in this series was comparable to that of missense mutations (5 different APP missense mutations in the series). The previously reported mutation frequency in familial AD was 8% (95% CI 2.6-17.1) in a French series [14], 2.7% (95% CI 0.32-9.3) in a Dutch series of familial AD [17] and 0% in a Swedish and Finnish cohort of EOAD (95% CI 0-2.58) [247]. Our data are not directly comparable with more selected studies, as two from the five probands were not identified as having a familial disease until after mutations were identified and this aspect of the history was reconsidered. Only one from the five probands was considered likely to have cerebral amyloid angiopathy because of a scan finding of a small intra-cerebral hemorrhage.

In one proband (3) we detected a complex discontinuous *APP*d mutation associated with sporadic EOAD. This individual had three microsatellite alleles at one locus and a double allele dose at the second locus tested. The *APP*d patchily involved almost the entire Chr21q, but critically did not include the Down's syndrome Critical Region around 21q22. The largest continuous region of duplication was 15.5Mb. These data are consistent with a different mutational mechanism from other *APP*d, an unbalanced translocation. In four probands, continuous duplications of 1.6-6.5 Mb including APP were detected, one mutation was slightly larger than any previously reported. Although the clinical phenotype of *APP* duplication may vary within families, it does not appear to be influenced by the size of the duplication itself [16], a finding supported by our results. Documented ages at onset range from 40-59 years, which is broadly in keeping with our data (range 39-61) [13, 15]. No clinical features of Down's syndrome have yet been observed although dementia and CAA seem universal with a quarter also suffering intra-cerebral haemorrhage (ICH). There was also a high incidence of seizures (57%) in published series. Our data corroborate these reports with a high

proportion of patients having seizures in the clinical course. One family (B) has an autosomal dominant family history of partial seizures, with partial seizures from adolescence in both patients subsequently affected by a cognitive disorder. Whether these early seizures are coincidental, are caused by a linked or unlinked genetic abnormality, or a very early manifestation of cerebral amyloid angiopathy is unknown. Clinically it would appear that the *APP* duplication-related Alzheimer's disease falls somewhere between the canonical AD phenotype observed in most *APP* mutations and the frequent, CAA-associated ICH seen in the Dutch *APP* mutation [248].

The methods used, successfully identified 5 patients where duplication had occurred and five further samples, which showed borderline abnormalities on exon-qPCR, were all normal using the Illumina array. Each detection method has its own merits and neither can be used exclusively for the detection of CNV. Although not 100% specific exon-qPCR cannot be confounded by genomic admixture due to accidental contamination. In contrast the specificity of fm-qPCR may supercede that of exon-qPCR when one considers the effect of allelic stutter. However fm-qPCR is susceptible to false positive results because of sample contamination and also requires heterozygosity of the marker. The latter issue is largely resolved by typing two markers rather than one. In tandem the techniques are complementary and 100% specific for detecting duplications in our large series. As a further 5 samples, positive on only one assay, were negative when tested by the Illumina array, it is likely that sensitivity is also very high, although we can not be certain as only a small number of samples were tested with a gold standard assay.

The absence of *PRNP* duplication in our large series can be interpreted several ways. First, that *PRNP* duplication is an extremely rare cause of dementia, sufficient for no patients to be sampled by a national referral center in 15 years. Second, that *PRNP* duplication is not causal of human prion disease. Duplication would be expected to increase PrP expression by up to 1.5x; transgenic mice engineered to greatly overexpress *Prnp* (>8X) do not develop clinically evident spontaneous prion disease in their normal lifespan; however, these mice have very short incubation times when provided with a "seed" of abnormal PrP. The high prevalence of abeta deposition in the population

would suggest that an exogenous “seed” is not required to initiate AD, whereas this may be a cause of at least some apparently sporadic prion disease. Third, as duplication is likely to occur involving *PRNP* on chromosome 20 it is possible that these mutations cause a severe developmental phenotype in humans preventing the ascertainment of later onset prion disease.

4.2 - Frontotemporal lobar degeneration

4.2.1 - Mutation screening in a large series of FTLD patients from the UK

The five genes (*MAPT*, *GRN*, *TARDP*, *VCP* and *CHMP2B*) known to underlie some cases of FTLD, do not account for all familial cases of this group of disorders, suggesting that there are other disease-causing genes yet to be discovered. This study was designed to look at the heritability of FTLD in each of the different clinical syndromes and to investigate to what extent the known genes account for this heritability. We also looked at the patients with strong family histories of FTLD but without known mutations and the underlying pathological causes in this group.

METHODS

Sample collection

Blood was collected for DNA extraction from patients attending the Specialist Cognitive Disorders Clinic at the National Hospital for Neurology and Neurosurgery, Queen Square, London and also from patients involved in studies of FTLD at the Dementia Research Centre, Institute of Neurology, Queen Square, London. Samples were taken from 225 patients who had been diagnosed by clinical, behavioral and neuropsychologic assessments with a diagnosis within the FTLD spectrum according to consensus criteria [184, 249-251]. Although samples were collected from patients with and without a family history, there is likely to be some ascertainment bias towards familial cases.

Ethical approval for the study was obtained from the National Hospital for Neurology and Neurosurgery Local Research Ethics Committee. Written research consent was obtained from all patients participating in the study.

Analysis of family history

All patients were given a "modified Goldman score" between 1 and 4 as per Goldman et al, 2005 [252, 253] where 1 is an autosomal dominant family history of FTLD, MND, CBS or PSP, defined as the presence of at least three affected people in two generations with one person being a first-degree relative of the other two, 2 is familial aggregation of three of more family members with dementia but not meeting criteria for 1, 3 is one other affected family member with dementia (modified to give a score of 3 only if there is a history of young-onset dementia within the family i.e. less than 65, and 3.5 if onset above 65) and 4 is no or unknown family history. All patients had had a structured clinical interview, which had included a detailed family tree. This had been discussed with the patient and family members (a minimum of one other person). The data for this study was ascertained from a review of all of the clinical notes: data was available on 222 of the cases with only 3 patients scoring 4 because of an unknown family history (2 with bvFTD and 1 with FTD-MND).

Genetic analysis

All 225 patients were screened for mutations in *MAPT* and *GRN* detecting 39 pathogenic mutations. Of the remaining 186 mutation negative patients, sequencing was obtained for *VCP* exons 3,5,6 and 10 in 160 patients, *TARDBP* exons 4 and 6 in 179 patients, and *CHMP2B* in 92 patients. We also sequenced exon 15 of the FUS gene in 183 patients, which has previously been shown to be causative of motor neurone disease, although currently no mutations have been found in FTLD [254, 255]. PCR amplicons were generated using primers at 500 µM in MegaMix Blue PCR cocktail (Microzone). See Table 3 in the Methods section for primers' sequences. Amplicons cleaned with Microclean (Microzone) were sequenced using Applied Biosytems BigDye v1.1 cycle sequence chemistry with 1 µl BigDye, 5 µl BetterBuffer (Microzone), 0.75 µl sequencing primer at 5 pmol/µl, 2.5 ng of cleaned amplicon and ddH2O to a final volume of 15 µl. Following 25 thermal cycles, sequencing products were precipitated and cleaned used standard EtOH/EDTA methodology and electrophoresed on an ABI 3130xl automated sequencer. Data were analyzed using ABI Seqscape

software v2.5. For more details of *GRN* and *MAPT* sequencing see Beck et al, 2008 and Janssen et al, 2002 [252, 256].

RESULTS

Demographic and family history data (Table 9)

Almost half of the 225 patients had bvFTD as their initial clinical syndrome (44.4%) with the next most common disorders being SemD (16.0%) and PNFA (13.3%). Smaller numbers had CBS or a CBS/PNFA overlap syndrome, PSP or LPA. Average age of onset for the different groups was between 54.8 years (bvFTD) and 62.0 years (PNFA) with a total mean of 57.3 years. In total 58.2% of the patients were male with more male patients in each of the groups apart from the CBS and CBS/PNFA overlap groups. 10.2% of patients had an autosomal dominant inheritance (as defined by a modified Goldman score of 1) but the heritability of FTLD was substantially higher (41.8%) when a family history was defined by a modified Goldman score of 1, 2, 3 or 3.5. The bvFTD group had the largest proportion of cases with a family history (58.0% with modified Goldman score 1 to 3.5 and an average modified Goldman score of 2.9) and the least familial of the disorders were FTD-MND (10.0%, 3.8), LPA (20.0%, 3.6) and SemD (22.2%, 3.8) (Table 9).

Table 9. Demographic and family history data for the cohort of 225 FTLD patients (n = number of cases, AAO = age at onset of symptoms)

Initial clinical syndrome	n	% of total cases	Average AAO	% male	Modified Goldman score (% of cases)					% of cases with score 1-3.5	Average score
					1	2	3	3.5	4		
bvFTD	100	44.4	54.8	64.0	20	17.0	11.0	10	42.0	58.0	2.9
FTD-MND	10	4.4	56.7	70	0	10	0	0	90	10	3.8
SemD	36	16.0	57.9	52.8	0	5.6	5.6	11.1	77.8	22.2	3.8
PNFA	30	13.3	62.0	63.3	0	13.3	0	16.7	70	30	3.7
PNFA/CBS	8	3.6	61.3	12.5	12.5	0	12.5	12.5	62.5	37.5	3.4
CBS	17	7.6	57.6	41.2	5.9	11.8	17.6	17.6	47.1	52.9	3.3
PSP	9	4.0	58.9	55.6	0	33.3	0	0	66.7	33.3	3.3
LPA	15	6.7	60.7	60	6.7	6.7	0	6.7	80	20	3.6
Total	**225**		**57.3**	**58.2**	**10.2**	**13.3**	**7.6**	**10.7**	**58.2**	**41.8**	**3.5**

Genetic analysis

Mutations were found in the *MAPT* and *GRN* genes but no mutations were found in the *CHMP2B, VCP*, *TARDBP* or *FUS* genes. In total 20 patients (8.9%) had mutations in *MAPT* (15 probands) and 19 patients (8.4%) had mutations in *GRN* (13 probands). Of the *MAPT* mutations, thirteen (from eight families) had an intronic 10+16 mutation of which seven families have previously been described [256]. The other previously described mutations were an intronic 10+19 mutation as well as deltaK280, L284R, N296N [257], S320F and G389R mutations. A novel *MAPT* variant was also found, N286N, a synonymous change similar to the N296N which is thought to be pathogenic via its effect on the splicing of exon 10 [257]. Most patients presented with bvFTD (although many developed semantic impairment as the disease progressed) apart from the N296N (CBS) and the L284R (PSP) mutations. The *GRN* mutations were ten C31fs mutations (from four families) and two 385_388delAGTC mutations (from two families) [252]. Other mutations were 603_603insC, 1494_1498delAGTGG, Q300X, L469F, A199V mutations (all previously described in Beck et al, 2008 [252]) as well as 1048_1049insG [258] and R493X mutations. Patients were diagnosed with bvFTD (C31fs, 385_388delAGTC, Q300X), PPA (PNFA or LPA: C31fs, L469F, R493X, 603_603insC) or CBS

(1048_1049insG, A199V, C31fs, 1494_1498delAGTGG) with two patients having a PNFA/CBS overlap (C31fs, 1494_1498delAGTGG).

Of the 186 patients without a known mutation we re-examined the number of cases with a family history and in particular we looked at whether any of the "familial" cases had post-mortem confirmation of FTLD pathology (Table 10). The majority of these cases (125) had a modified Goldman score of 4 but four cases with an autosomal dominant family history (modified Goldman score of 1) were still without a known mutation (all with bvFTD) and in total 61 cases with a modified Goldman score of 1,2,3, or 3.5 were not known to have a mutation. We looked particularly at those cases with a score of 1, 2 or 3 (38 in total) as a score of 3.5 may well represent another family member with old-age onset dementia and therefore less likely to be a true familial history of FTLD. Of these 38 cases seven had pathological confirmation of disease (6 bvFTD and 1 FTD-MND). All of these cases had tau-negative FTLD pathology: 1 case with bvFTD was known to have ubiquitin-positive, TDP-43 negative pathology (FTLD-UPS) without intranuclear inclusions similar to the pathology found in the Danish family with a mutation in *CHMP2B* (but in this case without a *CHMP2B* mutation) but the other six all had FTLD-TDP (i.e. TDP-43 pathology) with all having type 3 pathology according to consensus criteria [180, 189].

Table 10. Number of cases without mutations stratified according to their family history.

	Modified Goldman score (n cases)					**TOTAL (n)**
	1	**2**	**3**	**3.5**	**4**	
bvFTD	4	15	7	9	39	74
FTD-MND	0	1	0	0	9	10
SemD	0	2	2	4	27	35
PNFA	0	2	0	5	20	27
PNFA/CBS	0	0	0	1	5	6
CBS	0	1	2	3	8	14
PSP	0	2	0	0	6	8
LPA	0	0	0	1	11	12
Total	**4**	**23**	**11**	**23**	**125**	**186**

DISCUSSION

Our results confirm previous findings that FTLD is a highly heritable degenerative disorder. Heritability varies between the different clinical syndromes with SemD and LPA having a much lower percentage of cases with a family history compared with bvFTD. SemD is typically a sporadic disease and is associated with type 1 FTLD-TDP [259] although patients with *MAPT* mutations will often develop semantic impairment later in the disease (with behavioral symptoms initially). There are few reports of the associations of LPA with family history although some studies suggest that the majority of cases of LPA actually have Alzheimer-type pathology rather than an FTLD pathology that would account for the lower rate of reported family history in this group [183, 260]. A previous study suggested that FTD-MND was the most heritable of the FTLD syndromes [253] but in our series it was the least heritable, albeit with low numbers in our cohort. Inconsistent results with other series may reflect ethnogeographic clustering of particular causal mutations. Numbers were also low in the PSP group limiting the ability to interpret this data.

Mutations in *MAPT* and *GRN* are relatively common and have a similar prevalence in our series. However, there is variability in the prevalence geographically across different reported series with *MAPT* mutation frequency between 3 and 14% [261-264] and *GRN* mutation frequency between 1 and 16% [199, 201, 202, 265-267], with some series reporting vastly different frequencies within the same country e.g. two studies of FTLD patients in Italy found *GRN* mutation frequencies of 1.6% and 15.2% [265, 266]. A number of recent series have compared the frequency of *MAPT* and *GRN* mutations in FTLD populations [199, 201, 202, 267] (Table 11): some countries have families with a founder effect causing higher *GRN* mutation prevalence than *MAPT* e.g. in Belgium [201]. In the UK, the 10+16 *MAPT* mutation is common with a known founder effect [268] which is likely to account for the higher frequency of *MAPT* mutations in some series [267], although we have also previously shown that patients in our series with the *GRN* C31fs mutation are part of the same family [252],

which is likely to at least partly account for the similar prevalence of *GRN* and *MAPT* mutations in our series.

Table 11. Previously reported series comparing frequencies of MAPT and GRN mutations in an FTLD spectrum population

Series	Geographical area	Number in series	% MAPT mutations	% GRN mutations
This series	UK	225	8.9	8.4
Cruts et al, 2006	Belgium	103	1.9	10.7
Gass et al, 2006	USA	167	4.4	4.8
Le Ber et al, 2007	France	210	2.9	4.8
Pickering-Brown et al, 2008	UK	223	7.6	5.8

We found no mutations in *VCP, CHMP2B* or *TARDP* consistent with previous series suggesting that these are rare causes of FTLD [181]. It remains possible that causal mutations in these genes are present in unscreened exons of these 3 genes, although our sequencing strategy covered all known mutations in these genes to date. We also found no mutations in *FUS* suggesting that mutations in this gene are either not causative or are a very rare cause of FTLD.

Taking into account the known mutations many patients were still found to have a strong family history suggesting that there are still unknown genes that cause FTLD [269]. One locus (on chromosome 9) is known for patients with a clinical phenotype of FTD-MND [270], however this is associated with type 2 FTLD-TDP. Analysis of the pathological cases within this subgroup suggests that there are at least two other groups of patients with a family history without a known mutation: those with FTLD-UPS and those with type 3 FTLD-TDP without a GRN mutation (some of whom have a clinical phenotype of FTD-MND). A number of series have now been reported with FTLD-UPS (ubiquitin-positive, TDP-43 negative) pathology although these have all been reported as sporadic cases [271, 272]. The family reported here is pathologically distinct from these patients lacking the intranuclear inclusions seen in these cases. Although similar pathologically to cases with *CHMP2B* mutations, the single case in our series was negative for mutations in this gene. Studies of type 3

FTLD-TDP suggest that between 30 and 60% of such patients have *GRN* mutations but some patients negative for *GRN* mutations still have a family history [273-275]. There are no known loci associated with such patients suggesting further work needs to be done to clarify the genetic cause in this group as well as patients with familial FTLD-UPS.

4.2.2 - Screening for GRN mutations in a Portuguese series of FTD/CBS cases

Truncating and missense mutations altering splicing in the granulin gene have recently been identified as a cause of FTLD with ubiquitin inclusions linked to chromosome 17q21 [199-201, 276]. It is now apparent that mutations in this gene account for 4 to 26% of FTLD cases in different populations, making it a more common cause of FTLD than *MAPT* mutations [199, 201, 277, 278]. The vast majority of *GRN* mutations identified are nonsense, frameshift and splice-site mutations that would be predicted to cause premature termination of the coding sequence, creating null alleles with the mutant RNA being degraded by nonsense mediated decay [279].

In this study we report the results of *GRN* screening in a Portuguese series of consecutive FTLD patients and describe one family with a novel mutation.

METHODS

Case Series

DNA samples were obtained from forty-six patients (28 females and 18 males) attending the Dementia Clinic at the Coimbra University Hospital. Forty of these patients were diagnosed with FTD, according to the Lund and Manchester criteria, and six with possible FTD (Table 12). All subjects were of apparent Portuguese ancestry. Mean age at onset was 58 years of age (range, 40 – 73 years). Family history was positive in 21 cases, 22 cases were sporadic and family status was unknown for 3 samples. The 40 definite FTD cases included 20 with positive family history, 17 sporadic and 3 unknown. The 20 cases with family history were from 16 different families, with 4 samples coming from one extended pedigree (PT1) and 2 samples from PT2. Additionally, one of the possible FTD cases also presented a positive family history. For the purpose of calculating the frequency of mutations in the FTD population, only those patients and families with definite FTD were considered.

Table 12. Characteristics of the studied sample, stratified by diagnosis

Diagnostic	N	Sex		Mean age	Mean age at onset	Family history		
		F	M			Positive	Negative	Unknown
CBS	2	1	1	56	54	2	0	0
SemD	2	1	1	56	54	1	1	0
FTD-MND	3	2	1	58	63	2	1	0
bvFTD	31	18	13	60	58	15	13	3
Possible FTD	6	5	1	59	56	1	5	0
PD with dementia	2	1	1	63	61	0	2	0

CBS - Corticobasal syndrome; SemD – Semantic dementia; FTD-MND – Frontotemporal dementia associated with motor neuron disease; bvFTD – behavioral frontotemporal dementia; PD with dementia – Parkinson's disease with dementia.

The research protocols for this study were approved by the Ethics board of the University Hospital of Coimbra and all the patients, or surrogates, provided an informed consent.

DNA Sequencing

DNA was extracted using the DNA Isolation Kit for Mammalian Blood (Roche Diagnostics Corp., IN). The 12 coding exons and the non-coding exon 0 of *GRN*, as well as the respective flanking intronic sequences, were PCR amplified using primers previously reported (see Table 3 in the Methods section for primers' sequences) and sequenced according to standard procedures [199, 280].

Histological staining

Seven micron thick tissue sections from the cortex of the proband with the mutation (Family PT2, case III.1) were stained with haematoxylin and eosin and used for immunohistochemistry with antibodies to TDP43 (Abnova; 1:80; pretreatment: pressure cooking), P62 (BD Transduction Laboratories; 1:100; pretreatment: pressure cooking), ubiquitin (Dako; 1:200; pretreatment: pressure cooking), tau (AT8; AutogenBioclear; 1:600; pretreatment: pressure cooking), amyloid-beta peptide (Dako; 1:100; pretreatment: pressure cooking and formic acid) and alpha-synuclein (Novocastra;1:50; pretreatment: pressure cooking and formic acid). Immunohistochemical staining was performed using a standard avidin-biotin method.

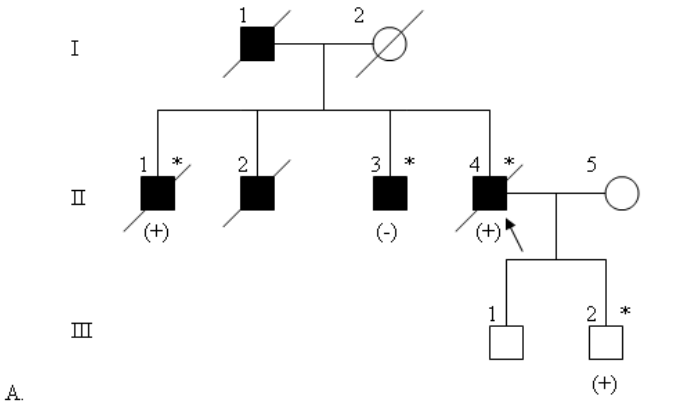
RESULTS

Variants in the *GRN* gene were found in 5 of the 46 Portuguese patients.

DNA sequencing

A heterozygous novel 2bp insertion was found in exon 9 (g.2265_2266insGT) in two affected siblings from family PT2, who presented with corticobasal degeneration in the fifth decade of life. This mutation introduced a premature stop at codon 361 and would be predicted to truncate the remaining 233 residues of the protein (p.Gln300GlnfsX61). However, it is likely that this mRNA will be removed by nonsense mediated decay resulting directly in haploinsuffiency.

A previously identified single nucleotide change, c.764C>T causing a predicted threonine to methionine substitution at residue 182 (p.T182M) was identified in a male patient from family PT1 who had developed cognitive impairment at the age of 37 years. Screening of additional family members demonstrated that this mutation was also present in another affected sibling, but was absent in the third affected sibling (Figure 15A). Thus, it is unlikely that this mutation is pathogenic: a conclusion that others have also suggested [277, 280]



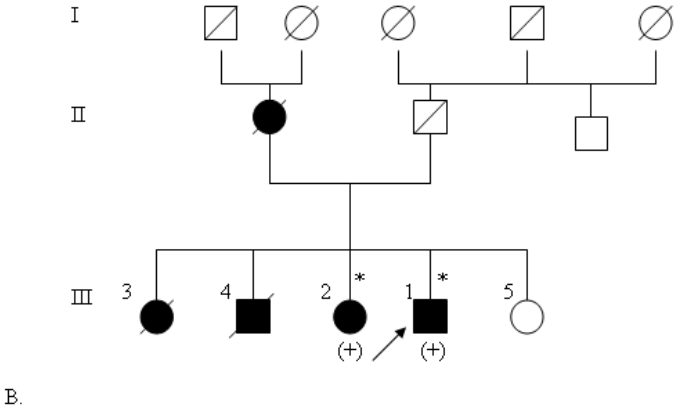


Figure 15. Pedigrees of families PT1 (A) and PT2 (B). Affected persons= black. Square=male; circle=female; diagonal line=death. Probands are illustrated by the arrows. * indicates a DNA sample was available (+) positive for the respective mutation (-) negative for the respective mutation.

Two synonymous *GRN* variants were also observed. This included the novel L14L and the known D128D in two patients with possible FTLD. rs9897526, rs34424835, rs25646 and rs850713 polymorphisms were also observed (Table 13). These variants are unlikely to be pathogenic.

Table 13. Frequencies of the polymorphisms in the sample series

Polymorphism	Location	Alteration	Frequencies
Rs9897526	Intron 3	A/G	AG=0.33 GG=0.67
Rs34424835	Intron 4	delGTCA	- =0.52 GTCA=0.48
Rs25646	Exon 5	A/G	AG=0.11 AA=0.89
Rs850713	Intron 5	C/T	CC=0.41 CT=0.52 TT=0.07

Clinical characteristics of family PT2

Family PT2 is an eight members pedigree spanning two generations (Figure 15B). The proband (III.1) developed bradykinesia, loss of initiative, decreased speech output and dysarthria at 50 years of age. Over the next few years, his symptoms progressed to include memory disturbance, spatial disorientation, gait difficulty and Parkinsonism with central and extremity rigidity without tremor. He developed prominent frontal symptoms including disinhibition, apathy, mutism, hyperphagia and rituals. At the age of 53, the patient had obvious features of corticobasal degeneration including unilateral parkinsonian, alien limb of the left arm and myoclonus. SPECT at that time revealed bilateral frontal and temporal hypoperfusion, more marked on the right side. MRI showed asymmetrical cortical and sub-cortical atrophy on the right side involving caudate nucleus and right hippocampus. Right frontal lobe biopsy at this time revealed cortical neuronal loss without distinctive pathology more marked on superficial layers together with cortical and subcortical gliosis. The patient died at the age of 59.

The proband's sister (III.2) presented with apathy, loss of initiative, decreased voluntary speech, memory problems and urine incontinence by the age of 58 years. Neuropsychological testing showed bradyphrenia with frontal executive dysfunction. Her symptoms progressed to include parkinsonism (bradykinesia, masked facies, rigidity) and a frontal syndrome (disinhibition, apathy,

mutism and dysexecutive syndrome). Similar to the proband, she ultimately developed features of corticobasal degeneration including lateralized parkinsonian signs (right side, including the face), and myoclonus of the right arm. SPECT evaluation showed a bilateral frontal and temporal hypoperfusion more marked on the left side. MRI analysis revealed an asymmetric cortical and subcortical atrophy, which was more marked on the left side.

The eldest sister of the proband (III.3) presented at the age of 50 with progressive dementia and lentification and died a few years later. The family states that her symptoms were similar to those of sister III.2.

An older brother of the proband (III.4) presented with progressive dementia and parkinsonism at the age of 50 years. He died seven years later. The family states that his symptoms were similar to those of the proband.

At 50 years old, subject III.5 is healthy.

The proband's mother (II.1) developed dementia by the age of 70 and died at 72 years old. The father died of heart disease at age 77 without neurological dysfunction.

Neuropathology of the proband

Analysis of brain tissue from case III.1 revealed a relatively small amount of subcortical white matter with a good preservation of the cortical hexalaminar architecture. Astrogliosis was present in the subpial cortical region and Aβ immunohistochemistry was negative. Neither cortical Lewy bodies nor neurofibrillary tangles were found on alpha-synuclein and tau immunohistochemistry. TDP43, P62 and ubiquitin immunohistochemistry demonstrated numerous neurites together with neuronal intracytoplasmic inclusions, more common in the superficial cortical laminae, particularly in layer II, than in the deeper cortical layers. Occasional neuronal intranuclear inclusions were also noted (Figure 16). Numerous TDP43 positive neurites were present in the subcortical white matter where structures resembling coiled bodies were also noted.

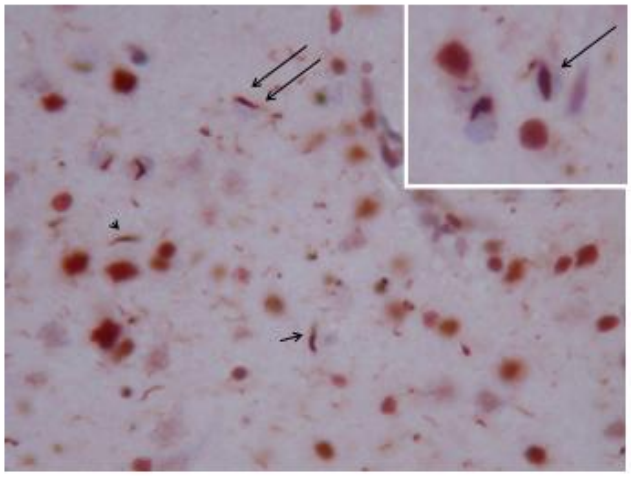


Figure 16. TDP-43 immunohistochemistry of brain tissue from case III.1, Family PT2. TDP-43 stainning revealed TDP-43-positive neurites (arrowhead), neuronal intracytoplasmic (double arrow) and neuronal intranuclear 'cat eye' inclusions (arrow).

DISCUSSION

In this study, we screened a series of 46 Portuguese FTLD patients for mutations in *GRN*, which has recently been associated with FTLD. Pathogenic variations in this gene were found in two siblings (family PT2). This represents 5% of the patients in our series (2/40), including 1/36 (3%) of the probands and 1/16 (6%) of families with definite FTLD. These frequencies are lower than those presented in previous studies [199, 201], but are in accordance, for instance, with the study from Bronner et al. reporting *PGRN* mutations as a cause for 4% of the independent familial cases studied. This disparity may be explained by differences in the patient recruitment process and by differences inherent to diverse populations characteristics. Thus, we conclude that *PGRN* mutations are not a major cause of FTLD in the Portuguese population.

We found a novel 2 bp insertion in exon 9 resulting in a frameshift at codon 300 (p.Gln300GlnfsX61) in two siblings from the same Portuguese family with corticobasal syndrome. It was not possible to further demonstrate segregation of this insertion with the disease, as additional DNA samples were not available. However, the expected effect on the mRNA would be nonsense mediated decay. If the message was to be translated it would result in a protein truncated in the last 233 amino acids,

which is likely to be unstable. In either way haploinsufficiency would lead to disease in this family. Previous studies established that frameshift, as well as nonsense FTLD-associated progranulin mutations, induce degradation of mutant RNA by a nonsense-mediated decay process.[199-201] In the same region of the *GRN* gene, a similar mutation (W304LfsX58) was reported in one patient with an age at onset of 60 years old.[199] In this same codon, a single nucleotide deletion and a point mutation causing a frameshift and a premature termination codon have also been reported [199, 281, 282].

Several other studies have associated *GRN* mutations with CBS. Masellis and colleagues identified a novel splice donor site mutation in the *GRN* gene (IVS7+1G→A) that segregated with corticobasal syndrome in a Canadian family of Chinese origin [283]. Benussi L, et al. described two Italian families presenting highly variable clinical phenotypes that included CBS [284]. Additionally, the study by Gass J, et al. included one American patient with CBS presenting the p.Val279GlyfsX5 mutation [199].

In the brain biopsy of the proband from family PT2 (subject III.1) there were numerous ubiquitin, P62 and TDP43 positive neuronal intracytoplasmic inclusions and occasional neuronal intranuclear inclusions together with numerous neurites with a similar immunohistochemical profile. The inclusions were, however, tau and alpha-synuclein negative. These appearances are in keeping with the diagnosis of 'frontotemporal lobar degeneration with ubiquitin-positive, tau-negative inclusions' (FTLD-U) [274, 285] and correspond to type 3 in the scheme proposed by Sampathu et al. [286] or type 1 according to the diagnostic classification of FTLD-U proposed by Mackenzie et al. [287]. Features of classical CBS, characterized by neuronal and glial tau pathologies, including neurofibrillary tangles, oligodendroglial coiled bodies, astrocytic plaques and threads [288] were absent in the case studied neuropathologically. It is of interest that FTLD-U has also been reported to present clinically as CBS or progressive supranuclear palsy [289, 290].

In family PT1, a point mutation (p.T182M) was found in the *GRN* gene. The mean age at onset of the disease in this family is 42 years old. This mutation was found in two deceased, affected siblings but was absent in the third living affected sibling, suggesting that this variant is not the genetic cause of dementia in the PT1 family, or at least not the sole cause. The majority of pathogenic mutations found in *PGRN* are frameshift or nonsense mutations causing null alleles. Even so, van der Zee and colleagues found three missense mutations and have suggested that these may cause functional haploinsufficency based on *in silico* modeling of protein structure [278]. Nevertheless, the point mutations, which are strongly suspected to be pathogenic, are those that disrupt the protein structure and probably results in an unstable protein. It is unlikely that this is the case for this T182M variant.

In summary, we identified one novel frameshift mutation as being the probable cause of disease in two of the FTD patients, representing 5% of our series. Although this is a small series and definite conclusions cannot be made, these data suggest that *PGRN* mutations are a rare cause of FTLD in the Portuguese population and confirm that corticobasal syndrome is part of the phenotype of *GRN* mutations.

CHAPTER 5

WHAT IS NORMAL IN THE GENETICS OF DEMENTIA

Based on:

1) **Guerreiro RJ**, Baquero M, Blesa R, Boada M, Brás JM, Bullido MJ, Calado A, Crook R, Ferreira C, Frank A, Gómez-Isla T, Hernández I, Lleó A, Machado A, Martínez-Lage P, Masdeu J, Molina-Porcel L, Molinuevo JL, Pastor P, Pérez-Tur J, Relvas R, Oliveira CR, Ribeiro MH, Rogaeva E, Sa A, Samaranch L, Sánchez-Valle R, Santana I, Tàrraga L, Valdivieso F, Singleton A, Hardy J, Clarimón J. Genetic screening of Alzheimer's disease genes in Iberian and African samples yields novel mutations in presenilins and APP. Neurobiol Aging. 2008 Jul 28

2) **Guerreiro RJ**, Washecka N, Hardy J, Singleton A. A thorough assessment of benign genetic variability in GRN and MAPT. Hum Mutat. 2009 Dec 17.

CHAPTER 5 – WHAT IS NORMAL IN THE GENETICS OF DEMENTIA

5.1 - Alzheimer's disease

A general problem in clinical genetics, is that when a locus for a disease is found by positional cloning, and subsequent point mutations are discovered, the gene is sequenced in others with the disease and novel DNA changes are described whose pathogenicity is then not assessed by either linkage or association [291]. Usually, these mutations are simply screened for in a number of controls and pathogenicity is assumed if they are not found. In addition, most sequencing is done in highly studied populations, such as Caucasians or East Asians and little is done in other populations. The result of this strategy is that variants are reported and assumed to be pathogenic. This is damaging both from a basic scientific perspective, because it misleads research on basic mechanisms and from a clinical genetic perspective, because it could lead to incorrect information being given, especially to those from under investigated communities. Examples of the precedent of misassignment of pathogenicity are the mutations E318G and InsR352 within *PSEN1*. Both variants were initially reported to be pathogenic and were later found to either be a normal coding variant (E318G) or a rare non-pathogenic mutation (InsR352) [292, 293]. With this background, we have embarked on a systematic reassessment of pathogenicity of the genes involved in AD. First, we sequenced the presenilin and *APP* genes in a large series of early onset AD patients from Iberia (results presented in chapter 4). Then, in order to have a better knowledge of genetic variation within these genes, we have performed the same sequencing analysis in a series of unrelated African individuals from seven different populations obtained from CEPH-HGDP, as well as in Iberian controls. We chose to study this African series because genetic diversity is greatest in Africans and this population has been little studied. Therefore our chance of finding variants was highest in this population.

METHODS

Sequencing of control individuals (Portuguese and CEPH-HGDP)

Written informed consent was obtained from 121 neurologically normal and age-matched control subjects from the Iberian peninsular (mean age at collection 67.4). These were either patient's spouses or unrelated caregivers. Coding exons 3 to 12 of *PSEN1* and *PSEN2*, and exons 16 and 17 of the *APP* gene were sequenced in this series and in 130 African samples obtained from the CEPH-HGDP [294]. These samples originated from 7 different African populations: 29 Biaka Pygmy, 13 Mbuti Pygmy, 23 Mandenka, 22 Yoruba, 2 San, 17 Bantu, and 24 Mozabite.

RESULTS

Screening in Iberian controls

In the 121 controls screened for mutations in *PSEN1*, *PSEN2* and *APP* we found no non-synonymous changes.

Screening in African individuals

Two non-synonymous changes were found in the *PSEN1* gene in the African series: the previously reported mutation R35Q (c.104G>A) [9, 222] which was present in one Mozabite individual and a new variant (V191A, c.572T>C), found in a San subject.

In the *PSEN2* gene we found three new non-synonymous changes (R29H, c.86C>T; L143H, c.428T>A; and A252T, c.754G>A). Strikingly, twenty individuals presented the R62H (c.185G>A) variation, which had been previously described as a variant with an unclear pathogenic role in AD [216].

In the *APP* gene, a new non-synonymous variation (H733P, c.2198A>C) was found in an individual from the Mandenka population.

DISCUSSION

The next paragraphs summarize the main features of the variations found in the African population.

APP H733P*: found in a Mandenka sample (Exon 17, C-Terminal).*

This variant was found in exon 17 of the *APP* gene. It alters a residue located in the C-terminal of the protein. No variants have been reported at this residue or close to this residue before.

PSEN1 R35Q*: found in a Mozabite sample (Exon 4, N-Terminal), residue not conserved in PSEN2 G37.*

This variant has been described before in 2 families with AD, but in neither case did it segregate with disease and it was interpreted as non-pathogenic [9, 222]. The data we present here suggests this interpretation is correct. The residue is not conserved in PSEN2 and no studies of the effect of this mutation on Aβ metabolism have been reported.

PSEN1 V191A*: found in a San sample (Exon 7, HL-III, residue conserved in PSEN2 V197).*

This variant has not been reported before. It was found in exon 7 of *PSEN1* and the residue where this mutation occurs is conserved between PSEN1 and PSEN2.

PSEN2 R29H*: found in a Mandenka Sample (Exon 3, N-Terminal), residue conserved in PSEN1 R27.*

This variation has not been previously reported. It is located in exon 3 of the *PSEN2* gene and alters a residue that is conserved between PSEN1 and PSEN2.

PSEN2 R62H *found in 20 African samples (allele frequency 9%) (Exon 4, N-Terminal), residue conserved in PSEN1 R60.*

This variant had previously been reported in a sporadic Alzheimer case. These data indicate it is a relatively common polymorphism in African populations. This residue, located in the N-terminal of the protein, is conserved between PSEN1 and PSEN2 (R60 in PSEN1) and studies of the effect of the variant on Aβ metabolism showed no alteration in the production of Aβ42 [295].

PSEN2 L143H *found in a Bantu sample (Exon 5, TM-II, residue not conserved in PSEN1 A137).*

This variant occurs in exon 5 of *PSEN2*. It alters a residue that is not conserved between PSEN1 and PSEN2 (A137 in PSEN1), located in the second transmembrane domain of the protein. No mutations have been reported in the correspondent residue of *PSEN1* and it does not fit with the helix rule for pathogenic mutations.

PSEN2 A252T*: found in a Mandenka and in a Yoruba sample (Exon 7, TM-VI, residue conserved in PSEN1 A246)*

This variant, found in exon 7 of the *PSEN2* gene modifies a residue conserved between PSEN1 and PSEN2 (A246 of *PSEN1*) that localizes in the sixth transmembrane domain of PSEN2. Two different pathogenic mutations have been reported in the corresponding *PSEN1* residue.

With this background, we herein propose a scale for grading mutations as not pathogenic, possibly pathogenic, probably pathogenic and definitely pathogenic (Figure 17). We recognize that this scale is merely pragmatic and subject to improvement, not least because it is possible that some variants may increase the risk of disease rather than being truly pathogenic.

1) Segregation: Has the mutation been shown to segregate with the disease? This is clearly the strongest evidence. We would suggest that if a mutation has been shown to segregate with the disease in three or more cases in a family it should be regarded as Definitely Pathogenic and if it has shown segregation in two cases, Probably Pathogenic.

2) Association: Has the mutation been found in one case and not in controls? Evidence of association with the disease is fundamentally weaker than segregation because it is not clear how many controls from different populations have been sequenced. Thus, one does not know, for any individual mutation, what the denominator is. However, we suggest that if a mutation has been found in at least three early onset non related Alzheimer cases and in no controls; and more than 100 controls have been sequenced, that it be regarded as Probably

Pathogenic. If less than three have been found, then we would suggest the designation as Possibly Pathogenic.

3) Residue and Aß levels: Have other pathogenic mutations been described in that residue before? If it is a presenilin mutation, does the mutation alter a residue conserved between PSEN1 and PSEN2 and, if the residue is in a transmembrane domain does it follow the helix rule? Does the mutation alter APP processing such that a greater proportion of Aβ42/Aβ40 ratio is produced? We would suggest that if two of the answers to any of these questions is "yes", then this should allow the "promotion" of a mutation.

4) Obviously, finding the mutation in controls is strong evidence that it is not simply pathogenic (although small changes in risk will remain a possibility).

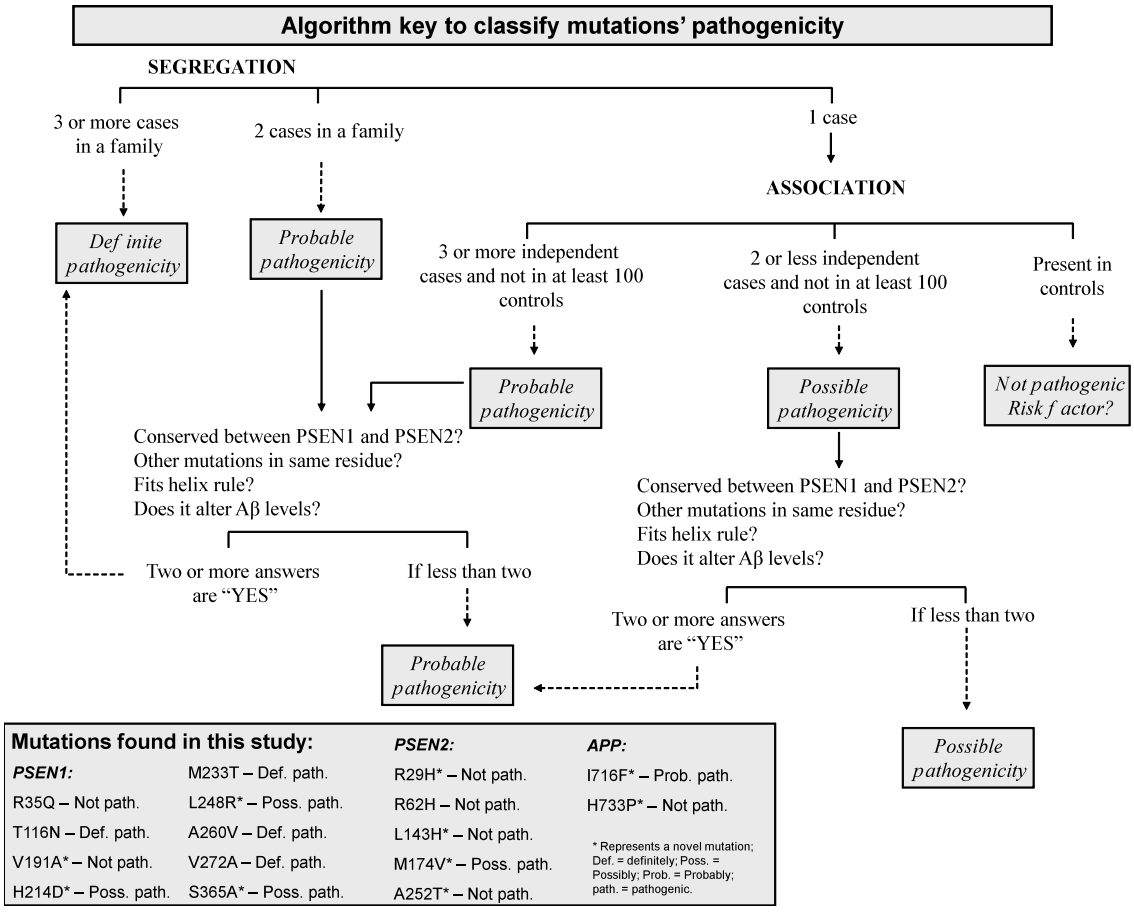


Figure 17. Algorithm to evaluate the pathogenicity of PSENs variants, classifying them as not pathogenic, possibly pathogenic, probably pathogenic, or definitely pathogenic.

5.2 - Frontotemporal lobar degeneration

Hundreds of pathogenic mutations have been described in *APP*, *PSEN1*, *PSEN2*, *MAPT* and *GRN*, leading to early onset Alzheimer's disease or FTLD. However, it is often unclear wheter identified variants are true disease causing mutations or rare benign changes. The dementia genes mentioned above were identified by positional cloning in different family based studies [11, 31, 200, 201, 296, 297]. These genes have subsequently been sequenced in other AD and FTLD patients and several novel mutations described. Frequently, pathogenicity has been assumed following screening for the variant in a relatively small number of control subjects. Most of the screening efforts have been done in an *ad hoc* manner in highly studied populations, such as Caucasians or East Asians and little has been done in other populations. The main result of this strategy is that variants have been reported and assumed to be pathogenic, sometimes with little supporting evidence. This is particularly problematic from a clinical genetic perspective, because it can lead to incorrect information being given to family members, but also from a fundamental scientific perspective, because it has the potential to mislead research on the basic mechanisms iunderlying the pathological process [298].

In the previous section of this chapter we described a systematic reassessment of the pathogenicity of variants present in genes involved in AD. In that study, in addition to screening the presenilin and *APP* genes in a large series of early onset AD patients from Iberia, we performed sequencing analysis in a series of unrelated African individuals from seven different populations obtained from the Centre d'Etude du Polymorphisme Humain - Human Genome Diversity Cell Line Panel (CEPH-HGDP), as well as in Iberian controls. The study of these 130 African samples allowed us to identify five new variants in the screened genes [298].

In order to extend these results to other genes frequently involved in dementia, we have performed a thorough assessment of common benign genetic variability in the *GRN* and *MAPT* loci in 232 samples from different regions of the world, including 87 African samples. Together with our

previous report, we were able to generate a list of variants that may otherwise have been classified as pathogenic mutations.

METHODS

Samples

GRN and *MAPT* were sequenced in a series of 282 samples from different regions of the world (106 Asian, 84 European, 5 Middle Eastern and 87 African), obtained from the CEPH-HGDP [294]. The number of samples from each population, geographic origin and region are detailed in Table 14.

Table 14. Number of samples studied per region, geographic origin and population

Region	Number of samples	Geographic origin	Number of samples	Population	Number of samples
Asia	106	Cambodia	10	Cambodian	10
		China	48	Dai	2
				Daur	2
				Han	15
				Hezhen	3
				Lahu	2
				Miaozu	4
				Mongola	4
				Naxi	3
				Orogen	1
				She	3
				Tu	2
				Tujia	1
				Xibo	4
				Yizo	2
		Japan	10	Japanese	10
		Pakistan	32	Balochi	15
				Pathan	17
		Siberia	6	Yakut	6
Europe	84	France	30	French	15
				French Basque	15
		Italy	15	Sardinian	15
		Italy (Bergamo)	3	North Italian	3
		Orkney Islands	6	Orcadian	6
		Russia	15	Russian	15
		Russia Caucasus	15	Adygei	15
Middle East	5	Israel (Carmel)	5	Druze	5

North Africa	10	Algeria (Mzab)	10	Mozabite	10
Subsaharan Africa	77	Democratic Republic of Congo	10	Mbuti Pygmy	10
		Kenya	11	Nantu N.E.	11
		Namibia	5	San	5
		Nigeria	22	Yoruba	22
		Senegal	21	Mandenka	21
		South Africa	8	Bantu S.E. Pedi	1
				Bantu S.E. S. Sotho	1
				Bantu S.E. Tswana	2
				Bantu S.E. Zulu	1
				Bantu S.W. Herero	2
				Bantu S.W. Ovambo	1

DNA Sequencing

The exonic regions of *GRN* (exons 1-12, plus noncoding exon 0; NM_002087.2) and *MAPT* (exons 1,9-13; NM_005910.4), as well as the flanking intronic sequences, were PCR amplified using oligonucleotide primers (see Table 3 in the Methods section for primers' sequences) and Roche FastStart PCR Master Mix polymerase (Roche Diagnostics Corp., IN). Each PCR product was sequenced using the same forward and reverse primers with Applied Biosystems BigDye terminator v3.1 sequencing chemistry and run on an ABI3730xl genetic analyzer as per manufacturer's instructions (Applied Biosystems). The sequences were analyzed with Sequencher software, version 4.2 (Genecodes, VA).

RESULTS

The sequencing of *MAPT* and *GRN* in 282 control samples from the human genome diversity re-sequencing panel allowed us to identify sixteen different non-synonymous changes. Of these, three were new non-synonymous changes in *MAPT* (c.50C>T, p.T17M; c.88A>G, p.T30A; c.898G>A, p.V300I) and thirteen were found in the coding region of *GRN*. From these latter, eight are new changes (c.163A>T, p.R55W; c.205G>A, p.A69T; c.355_357del, p.N119del; c.662G>C, p.C221S; c.824G>A, p.P275L; c.1126C>T, p.D376N; c.1193G>A, p.S398L; and c.1691, p.R564H); four are already reported variants, until now classified with unclear pathogenicity (c.359G>T, p.S120Y; c.545G>A, p.T182M; c.1298G>A, p.R433Q; and c.1544G>C, p.G515A); and one is already known as a non-pathogenic change (c.55C>T, p.R19W).

All coding non-synonymous variants were observed in African and Asians individuals and, as expected, a greater percentage of African individuals presented coding changes (twenty nine out of thirty six subjects presenting coding non-synonymous variants) when compared to other populations (Table 15).

Table 15. Non-synonymous variants found in the coding region of GRN and MAPT and in exon 0 of GRN

Gene	Variant		Frequency N (%)	Frequency by population (N)	Frequency by geographic origin (N)	Frequency by region (N)
	DNA	Predicted protein				
GRN	g.96073G>A	EX0+49*	1 (0.4%)	Adygei (1)	Russia Caucasus (1)	1 Europe
GRN	g.96144A>G	EX0+120*	1 (0.4%)	Russian (1)	Russia (1)	1 Europe
GRN	g.96172G>T	EX0+148*	5 (1.8%)	Bantu N.E. (1) Mandenka (3) Yoruba (1)	Kenya (1) Senegal (3) Nigeria (1)	Subsaharan Africa (5)
GRN	g.96199C>G	EX0+175*	1 (0.4%)	French (1)	France (1)	Europe (1)
GRN	g.96206T>C	EX0+182*	2 (0.7%)	Mandenka (2)	Senegal (2)	Subsaharan Africa (2)
GRN	g.96232C>G	EX0+208*	1 (0.4%)	Adygei (1)	Russia Caucasus (1)	Europe (1)
GRN	c.55C>T	p.R19W	8; 1 homozygous (2.8%)	Mandenka (1) Yoruba (3) Mbuti Pygmy (2) Mozabite (2)	Senegal (1) Nigeria (3) Democratic Republic of Congo (2)	Sub-Saharan Africa (6) North Africa (2)

					Algeria Mzab (1)	
GRN	c.163A>T	p. R55W	1 (0.4%)	Cambodia (1)	Cambodia (1)	Asia (1)
GRN	c.205G>A	p.A69T	1 (0.4%)	Xibo (1)	China (1)	Asia (1)
GRN	c.355_357del	p.N119del	1 (0.4%)	Mongola (1)	China (1)	Asia (1)
GRN	c.359G>T	p.S120Y	2 (0.7%)	Balochi (2)	Pakistan (2)	Asia (2)
GRN	c.545G>A	p.T182M	8 (2.8%)	Bantu N.E. (1) Yoruba (1) Mbuti Pygmy (6)	Kenya (1) Nigeria (1) Democratic Republic of Congo (6)	Sub-Saharan Africa (8)
GRN	c.662G>C	p.C221S	1 (0.4%)	Daur (1)	China (1)	Asia (1)
GRN	c.824G>A	p.P275L	1 (0.4%)	Yoruba (1)	Nigeria (1)	Sub-Saharan Africa (1)
GRN	c.1126C>T	p.D376N	2 (0.7%)	Bantu N.E. (1) Mbuti Pygmy (1)	Kenya (1) Democratic Republic of Congo (1)	Sub-Saharan Africa (2)
GRN	c.1193G>A	p.S398L	1 (0.4%)	Pathan (1)	Pakistan (1)	Asia (1)
GRN	c.1298G>A	p.R433Q	2 (0.7%)	Mandenka (1) Yoruba (1)	Senegal (1) Nigeria (1)	Sub-Saharan Africa (2)
GRN	c.1544G>C	p.G515A	4 (1.4%)	Bantu S.E. Tswana (1) Bantu N.E. (1) Yoruba (1) Mozabite (1)	South Africa (1) Kenya (1) Nigeria (1) Algeria Mzab (1)	Sub-Saharan Africa (3) North Africa (1)
GRN	c.1691G>A	p.R564H	1 (0.4%)	Bantu N.E. (1)	Kenya (1)	Sub-Saharan Africa (1)
MAPT	c.50C>T	p.T17M	1 (0.4%)	Yoruba (1)	Nigeria (1)	Sub-Saharan Africa (1)
MAPT	c.88A>G	p.T30A	1 (0.4%)	Bantu S.E. Tswana (1)	South Africa (1)	Sub-Saharan Africa (1)
MAPT	c.898G>A	p.V300I	1 (0.4%)	Mozabite (1)	Algeria Mzab (1)	North Africa (1)

Nucleotide numbering ("c.") reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the reference sequences (GRN: NM_002087.2; MAPT: NM_005910.4), according to journal guidelines (www.hgvs.org/mutnomen). The initiation codon is codon 1. Genomic numbering ("g.") refers to sequence GRN: NG_007886.1. Protein numbering ("p.") refers to sequences GRN: NP_002078.1 and MAPT: NP_058519.2. *Refers to the nomenclature employed by van der Zee and colleagues for an easier comparison.

A Mbuti Pygmy sample harbored two non-synonymous changes in *GRN* (p.R19W and p.T182M). Two samples harbored a *GRN* change and a *MAPT* change: a Mozabite individual presented the *GRN* p.R19W and the *MAPT* p.V300I variants, and a Bantu SE Tswana presented the p.G515A change in *GRN* and the p.T30A in *MAPT*.

Six different changes were found in exon 0 of *GRN*: +49G>A, +120A>G, +148G>T, +175C>G, +182T>C and +208C>G (nomenclature employed by van der Zee and colleages is used here for easier

comparison; for genomic nomenclature please refer to Table 15). From these, only the EX0+148G>T has been previously described in a Belgian FTLD patient [278].

Several synonymous changes were found in the *GRN* gene, including nine new variants (c.228C>T, p.T76T; c.267C>T, p.A89A; c.402G>A, p.P134P; c.507C>G, p.A169A; c.546G>A, p.T182T; c.786C>T, p.S262S; c.1227G>A, p.T409T; c.1515C>T, p.A505A; and c.1554C>T, p.D518D) and two variants until now classified with unclear pathogenicity (c.42G>A, p.L14L; and c.1695C>T, p.C565C). In the *MAPT* gene, seven synonymous changes were found, from which three were new (c.42C>T, p.H14H; c.600C>T, p.P200P; and c.798G>A, p.L266L). Although the vast majority of the synonymous changes were observed in African individuals, this type of variations was found in samples from all the regions (Table 16).

Table 16. *GRN* and *MAPT* coding synonymous variants and respective frequencies

Gene	Variant	Frequency N (%)	Frequency by population (N)	Frequency by geographic origin (N)	Frequency by region (N)
GRN	c.42G>A, p.L14L	15; 5 homozygous (5.3%)	Bantu S.E. Tswana (1) Bantu N.E. (2) Mandenka (1) Mbuti Pygmy (8) San (3)	South Africa (1) Kenya (2) Senegal (1) Democratic Republic of Congo (8) Namibia (3)	Subsaharan Africa (15)
GRN	c.99C>T, p.D33D	3 (1.1%)	Russia (2) Xibo (1)	Russia (2) China (1)	Europe (2); Asia (1)
GRN	c.228C>T, p.T76T	3 (1.1%)	Bantu S.W. Ovambo (1) Bantu N.E. (1) Bantu S.E. Pedi (1)	South Africa (2) Kenya (1)	Subsaharan Africa (3)
GRN	c.267C>T, p.A89A	1 (0.4%)	Balochi (1)	Pakistan (1)	Asia (1)
GRN	c.324C>T, p.D108D	1 (0.4%)	Mozabite (1)	Algeria Mzab (1)	North Africa (1)
GRN	c.384T>C, p.D128D	38; 4 homozygous (13.5%)			Europe (3); Asia (35)
GRN	c.402G>A, p.P134P	2; 1 homozygous (0.7%)	Mbuti Pygmy (1) San (1)	Democratic Republic of Congo (1) Namibia (1)	Subsaharan Africa (2)
GRN	c.507C>G, p.A169A	5 (1.8%)	Bantu S.E. Tswana (1) Bantu N.E. (1) San (3)	South Africa (1) Kenya (1) Namibia (3)	Subsaharan Africa (5)
GRN	c.546G>A, p.T182T	2 (0.7%)	Mandenka (1) Yoruba (1)	Senegal (1) Nigeria (1)	Subsaharan Africa (2)
GRN	c.786C>T, p.S262S	1 (0.4%)	Xibo (1)	China (1)	Asia (1)
GRN	c.1227G>A, p.T409T	2 (0.7%)	Mandenka (1) Yoruba (1)	Senegal (1) Nigeria (1)	Subsaharan Africa (2)
GRN	c.1515C>T,	1 (0.7%)	Adygei (1)	Russia Caucasus (1)	Europe (1)

	p.A505A				
GRN	c.1554C>T, p.D518D	1 (0.7%)	Russian (1)	Russia (1)	Europe (1)
GRN	c.1695C>T, p.C565C	1 (0.7%)	French (1)	France (1)	Europe (1)
MAPT	c.42C>T, p.H14H	1 (0.7%)	Yoruba (1)	Nigeria (1)	Subsaharan Africa (1)
MAPT	c.54C>T, p.Y18Y	1 (0.7%)	Oroqen (1)	China (1)	Asia (1)
MAPT	c.600C>T, p.P200P	1 (0.7%)	Naxi (1)	China (1)	Asia (1)
MAPT	c.681A>G, p.A227A	42 (14.9%)			Asia (7) Europe (29) Middle East (1) North Africa (1) Subsaharan Africa (4)
MAPT	c.765T>C, p.N255N	42 (14.9%)			Asia (7) Europe (29) Middle East (1) North Africa (1) Subsaharan Africa (4)
MAPT	c.798G>A, p.L266L	1 (0.7%)	1 Balochi (1)	1 Pakistan (1)	1 Asia (1)
MAPT	c.810G>A, p.P270P	26 (9.2%)			Asia (8) Europe (12) North Africa (2) Subsaharan Africa (4)

Nucleotide numbering ("c.") reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the reference sequences (*GRN*: NM_002087.2; *MAPT*: NM_005910.4), according to journal guidelines (www.hgvs.org/mutnomen). The initiation codon is codon 1. Protein numbering ("p.") refers to sequences *GRN*: NP_002078.1 and *MAPT*: NP_058519.2.

In order to predict the impact in protein function of the non-synonymous variants found in *GRN* and *MAPT*, we performed an *in silico* analysis using two software - SIFT [299] and PolyPhen [300, 301] (Table 17). From the three variants found in *MAPT*, one was considered to possibly affect the protein function (p.T17M) while the other two were predicted to be benign. However, the confidence in the prediction for the effect of p.T17M was considered to be low, by the software (SIFT), since the sequences used were not diverse enough.

From the twelve missense variations found in *GRN*, three were predicted to be possibly damaging and two to be probably damaging by the Polyphen software. The SIFT program predicted four of these twelve changes to affect protein function. Both programs predicted p.T182M, p.C221S and p.P275L to be possibly damaging to the protein.

Table 17. *GRN* and *MAPT* coding variants altering the amino acid sequence and respective PolyPhen and SIFT predictions of the impact of each variant

Gene	VARIANT					N	PolyPhen		SIFT	
	Location in the gene	DNA change	Location in the protein	Predicted protein change	New variant?		PSIC score	Result	Score	Result
GRN	Ex 1	c.55C>T	ParaGran	p.R19W	Already described	5	0.447	Benign	0.03	Affect protein function*
GRN	Ex 2	c.163A>T	InterParaGran	p.R55W	New	1	1.622	Possibly damaging	0.14	Tolerated
GRN	Ex 2	c.205G>A	Gran G	p.A69T	New	1	0.781	Benign	0.30	Tolerated
GRN	Ex 4	c.355_357del	InterGF	p.N119del	New	1	n.a.	n.a.	n.a.	n.a.
GRN	Ex 4	c.359G>T	InterGF	p.S120Y	Already described	2	1.627	Possibly damaging	0.09	Tolerated
GRN	Ex 5	c.545G>A	InterFB	p.T182M	Already described	8	1.556	Possibly damaging	0.04	Affect protein function
GRN	Ex 6	c.662G>C	Gran B	p.C221S	New	1	3.713	Probably damaging	0.01	Affect protein function
GRN	Ex 7	c.824G>A	InterBA	p.P275L	New	1	3.146	Probably damaging	0.01	Affect protein function
GRN	Ex 9	c.1126C>T	Gran C	p.D376N	New	2	0.32	Benign	1.00	Tolerated
GRN	Ex 10	c.1193G>A	Gran C	p.S398L	New	1	0.895	Benign	0.14	Tolerated
GRN	Ex 10	c.1298G>A	InterCD	p.R433Q	Already described	2	0.395	Benign	0.43	Tolerated
GRN	Ex 11	c.1544G>C	InterDE	p.G515A	Already described	4	1.12	Benign	0.47	Tolerated
GRN	Ex 12	c.1691G>A	Gran E	p.R564H	New	1	0.592	Benign	0.31	Tolerated
MAPT	Ex 1	c.50C>T	N-terminal	p.T17M	New	1	1.557	Possibly damaging	0.02	Affect protein function*
MAPT	Ex 1	c.88A>G	N-terminal	p.T30A	New	1	0.512	Benign	0.13	Tolerated
MAPT	Ex 10	c.898G>A	R2	p.V300I	New	1	0.114	Benign	0.07	Tolerated

* This is a low confidence prediction (this substitution may have been predicted to affect function just because the sequences used were not evolutionarily diverse enough - for specific details please see http://sift.jcvi.org/); N – Number of individuals.

Nucleotide numbering ("c.") reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the reference sequences (*GRN*: NM_002087.2; *MAPT*: NM_005910.4), according to journal guidelines (www.hgvs.org/mutnomen). The initiation codon is codon 1. Protein numbering ("p.") refers to sequences *GRN*: NP_002078.1 and *MAPT*: NP_058519.2.

Considering all the gene variants present in the AD/FTD database (http://www.molgen.ua.ac.be/ADMutations and http://www.molgen.ua.ac.be/FTDMutations), *GRN* contains a higher percentage of non-pathogenic variants and variants of unclear pathogenicity (53%). For the *MAPT* gene, the pathogenicity of a variant seems to be closely related to the location of that variant: until now all the coding variants found in the three last exons of the gene are pathogenic (Figures 18 and 19).

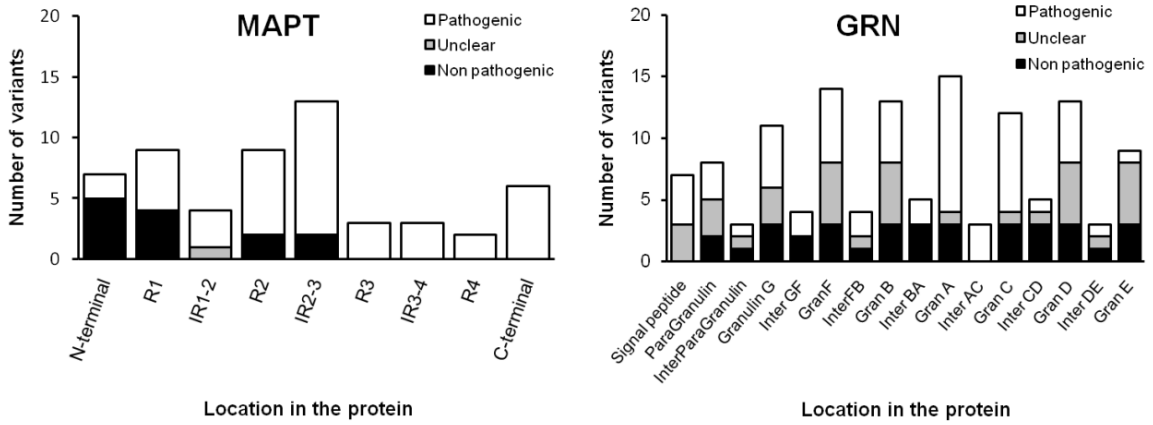


Figure 18. Location of non pathogenic, with unclear pathogenicity and pathogenic variants in functional protein domains found in *GRN* and *MAPT* according to the AD/FTD database (assessed on March 15th 2009) and including the variants found in the present study.

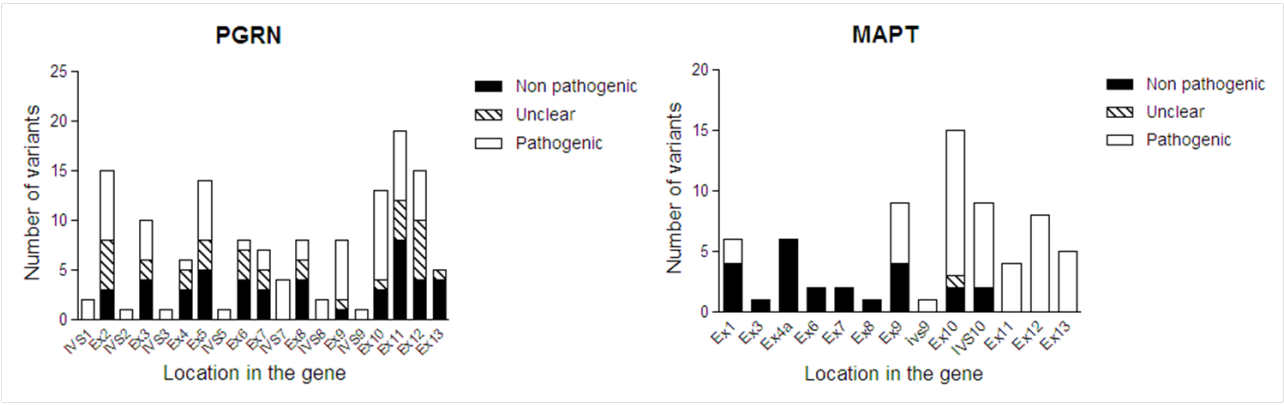


Figure 19. Gene location of non pathogenic, with unclear pathogenicity and pathogenic variants found in *GRN* and *MAPT* according to the AD/FTD database (assessed on March 15th 2009) and including the variants found in the present study.

The number of non-pathogenic variants and changes with an unclear pathogenicity is high in the *GRN* gene. The vast majority of these changes are missense variants. Of the 72 exonic point mutations described in *GRN*, only 5 (7%) are classified as pathogenic mutations (Figures 20 and 21).

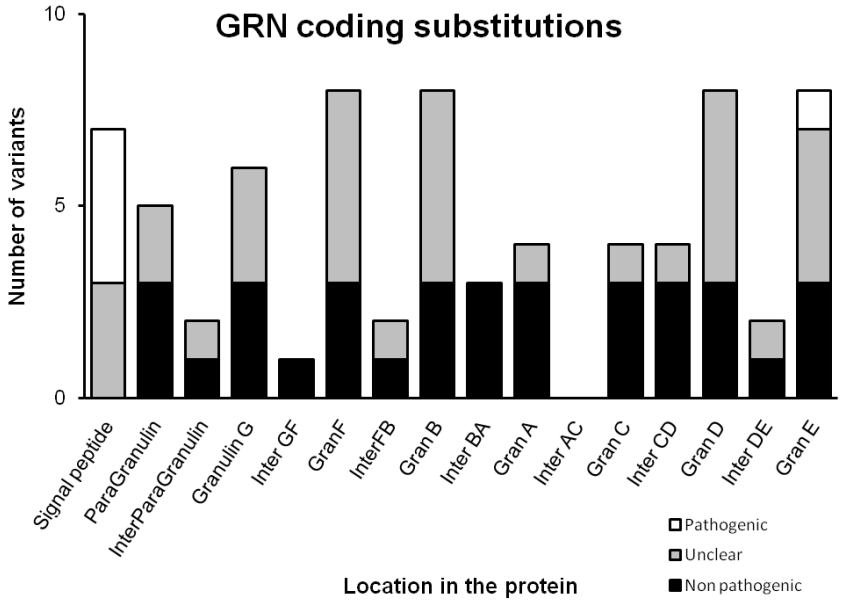


Figure 20. Location of coding substitutions in functional protein domains, classified as non pathogenic, unclear pathogenicity and pathogenic.

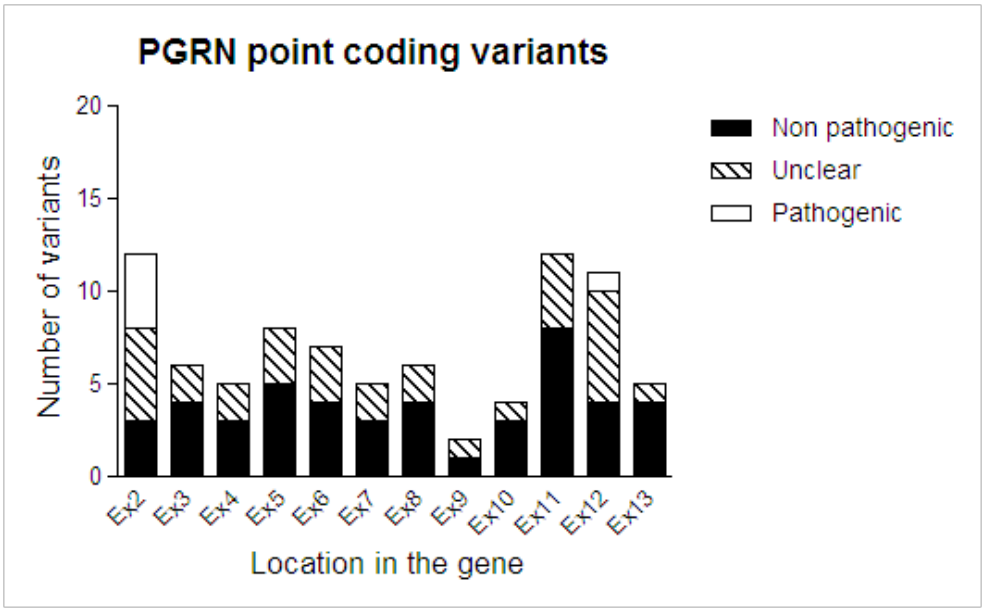


Figure 21. Gene location of coding substitutions in functional protein domains, classified as non pathogenic, unclear pathogenicity and pathogenic.

DISCUSSION

The question of how to distinguish between sequence variants that have an effect in the stability or function of a protein, ultimately leading to a disease and a neutral variation with no influence over the phenotype, is very difficult to answer. Several programs (the most widely used: PolyPhen and SIFT) have been developed in order to address this problem and the *in silico* interpretation of sequence variants is becoming more and more popular. However, some studies have shown that these *in silico* analysis do not account for all the variables involved (such as supramolecular interactions with other molecules or regions involved in control or signaling, and the existence of linkage disequilibrium) and remain flawed [302, 303]. In this way, although these are consistent and easy analysis to perform, they should not replace functional assays or unequivocal linkage/association. We tested all the variants found in *GRN* and *MAPT* genes using SIFT and PolyPhen, expecting that all variants would be predicted to be well tolerated. This did not happen for several variants with discordant results in three of the fifteen predictions.

The variant found in exon 10 of MAPT gene is the first missense change in a microtubule binding domain to be considered as non pathogenic. Although we cannot completely exclude the fact that this Mozabite individual will come to develop dementia in the future, this is unlikely to occur. The p.V300I change is located in the second microtubule binding domain of the protein. Although this is a highly conserved residue between different species, it is only partially conserved among the four microtubule binding domains of the protein. This alteration in the encoded amino acid sequence (valine to isoleucine) is a relatively conservative change, since these are both non polar, neutral amino acids with similar hydropathy indexes (4,2 and 4,5, respectively). Additionally, software used to interpret the functional effects of sequence variants, predicted this to be a benign and well-tolerated variant.

Frameshift and nonsense *GRN* mutations have been clearly demonstrated to induce degradation of the mutant RNA by a nonsense mediated decay process. Conversely, the role of missense mutations in this gene remains unclear. From the 83 missense synonymous and non-synonymous variants in

the *GRN* gene, present in the AD/FTD database and found in this study, only five (6%) remain classified as pathogenic. Three of these are point mutations in the translation initiation codon predicted to result in reduced mRNA levels [p.M1 (g.1A>G, g.2T>C, g.3G>A)] [199-201, 304]. One, (p.A9D) has been reported several times [199, 305, 306] as a pathogenic mutation affecting the GRN signal peptide. This mutation is believed to prevent N-glycosylation and subsequently cause the mislocation of the protein with accumulation of the mutant protein in the Golgi apparatus and prevention of its secretion [307, 308]. Additionally this variant greatly reduces GRN expression due to inefficient translation or degradation [308]. And the last one (p.C521Y) was found in a Spanish family presenting with progressive nonfluent aphasia [309]. While there are strong evidence for the pathogenicity of the first four mutations, this is not the case for p.C521Y. The carriers of this variant showed significant lower scores than controls in word generation and word learning tests, indicating that they could be at a prodromal stage of the disease. This genetic change was found in seven out of 10 family members: two of these presented dementia, one presented primary progressive aphasia and four were non-affected members. This can indicate, as the authors suggest, that p.C521Y may cause disease with reduced penetrance, but a more conservative and perhaps more likely interpretation is that this variant is a benign change not related to disease.

Several other point mutations have been proposed to be pathogenic, including p.P248R, p.R432C [278], p.C139R and p.P451L [310] (Figure 22).

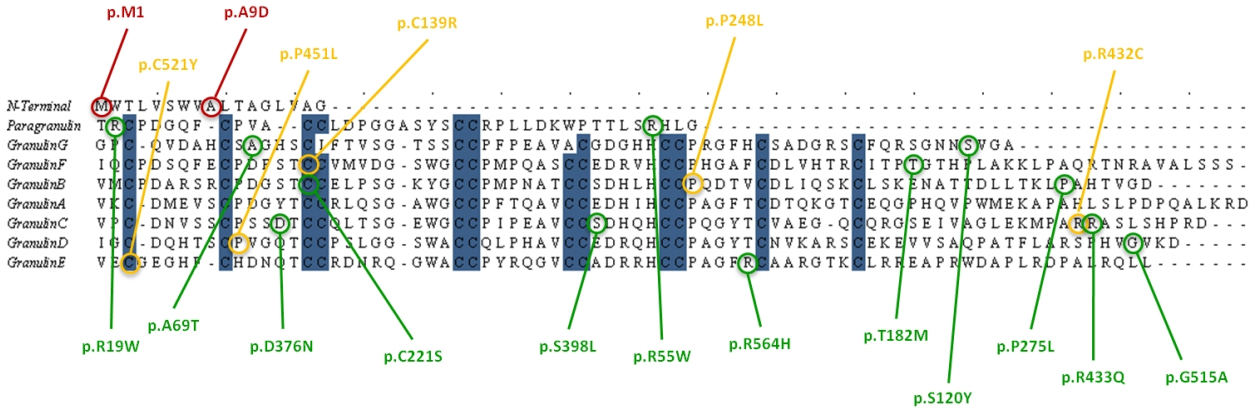


Figure 22. Progranulin sequence alignment of individual granulin domains with conserved cysteine residues highlighted in blue. Pathogenic missense mutations are represented in red, mutations reported as pathogenic [278, 310] or considered as pathogenic in the AD/FTD database [309] are represented in yellow and the variants found in this study are represented in green.

No segregation data is available for any of these variants and the assumption of pathogenicity has mainly relied on *in silico* analysis (SIFT, protein sequence conservation and protein modeling). From these, only p.Pro248Leu and p.Arg432Cys have been studied in vitro. These are significantly expressed but cause deficient transport of the protein through the secretory pathway, resulting in degradation and reduction of GRN secretion (p.P248L presented a reduced secretion by about 70% and p.R432C by about 45%) [308]. The p.P248L variant occurs in a residue adjacent to a cysteine in granulin B and p.R432C is located in the linker region between granulins C and D. In our study we have found three variants predicted by SIFT and Polyphen to affect protein function (p.T182M, p.C221S and p.P275L). The p.T182M variant was found in 8 individuals and is clearly a benign variant, while p.C221S and p.P275L were found in one subject, each. The p.C221S change alters a cysteine residue in granulin B that is 100% conserved between all granulins (as does p.C139R and p.C521Y), while p.P275L is located in the linker region between granulins B and A. Thus the variants found in the present study are dificult to distinguish from thise previously described as pathogenic and represented in yellow in Figure 22.

Although it remains possible that FTLD cases may exist within the HGDP samples, this is unlikely and if there were FTLD cases among HGDP samples we might expect to see some nonsense *GRN* mutations. Instead, we found a large number of probably benign missense variants. Our data supports the hypothesis that *GRN* point variants, with rare exceptions, are not pathogenic and should not be considered as definitely pathogenic unless segregation data is available. Additionally, all clearly pathogenic mutations act by abolishing the expression of the protein, more causing deficient GRN secretion, while linked to a plausible mechanism, should still be viewed with caution. It is important to discriminate between a genetically normal variant and a pathogenic mutation. Examples of the precedent of misassignment of pathogenicity are the mutations p.E318G and p.InsR352 within *PSEN1*. Both variants were initially reported to be pathogenic and were later found to either be normal coding variants (p.E318G) or a rare non-pathogenic mutation (p.InsR352) [292, 293]. Such errors may lead to incorrect and potentially damaging information being given to family members. Several variants found in this study have been previously described in patients. The p.S120Y change in *GRN* has been found before in an ALS-FTLD patient [311]. The p.T182M variant was reported in one FTLD patient [277], in a single case of limb onset sporadic ALS [311] and in two of three affected siblings presenting early onset FTLD [312]. Although this variant has never been found in controls before, its pathogenicity has always been considered unclear. The p.R433Q change was previously found in an African American patient with FTLD. The pathogenicity of this variant was reported as unclear since no family history of FTLD was reported for the affected individual and no other DNA samples were available for segregation analysis. Additionally, another variant at the same codon (p.R433W) had been previously reported in a neurologically normal individual. The p.G515A variant has been previously found in two AD patients [310], in four FTLD patients and two healthy control individuals [199, 278].

In order to be able to evaluate the pathogenicity of variants found in GRN and MAPT we developed two algorithms similar to the one used in the evaluation of PSENs and APP variants. Once again,

these are pragmatic scales for grading mutations as not pathogenic, possibly pathogenic, probably pathogenic and definitely pathogenic (Figures 23 and 24).

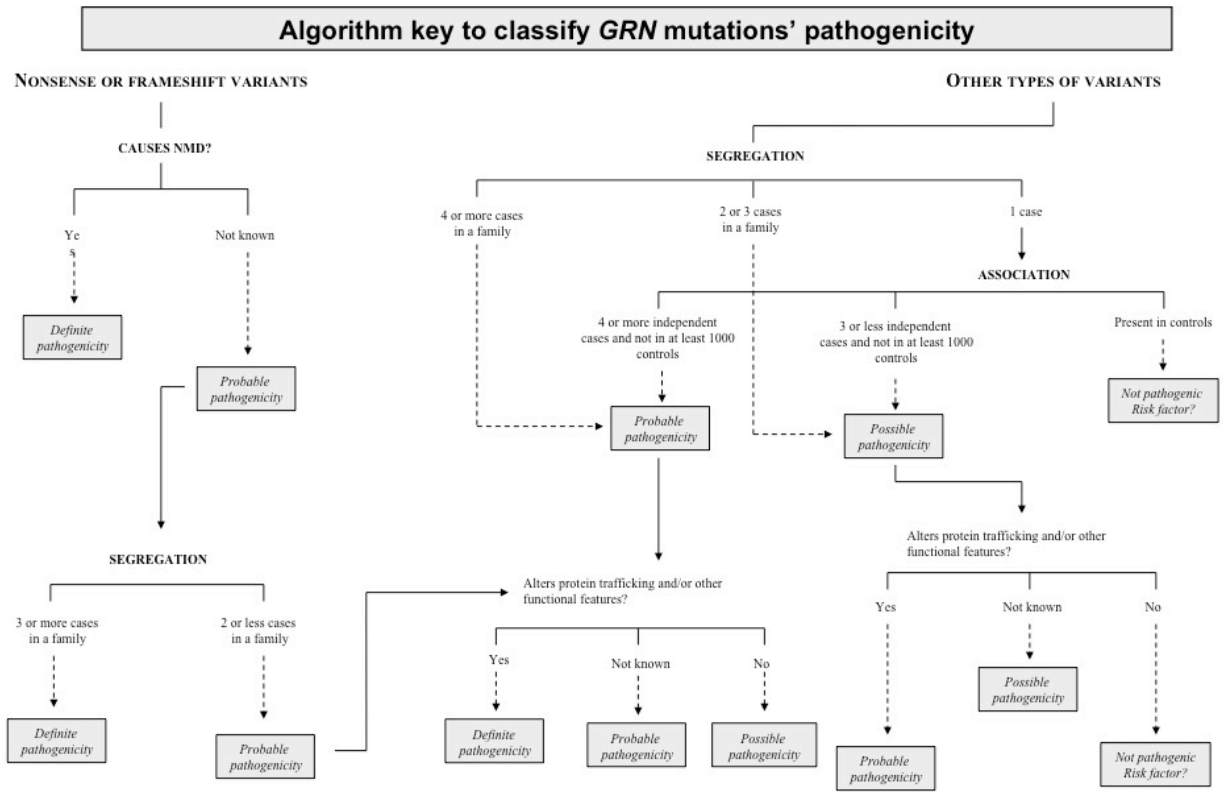


Figure 23. Algorithm to evaluate the pathogenicity of *GRN* variants, classifying them as not pathogenic, possibly pathogenic, probably pathogenic, or definitely pathogenic.

Our study adds to the classification of variants within *MAPT* and *GRN*, and identifies novel variants that could easily be mistaken as pathogenic mutations in future studies. Overall, our results stress the need for a coordinated effort to resequence the genes involved in neurological disorders in large worldwide samples.

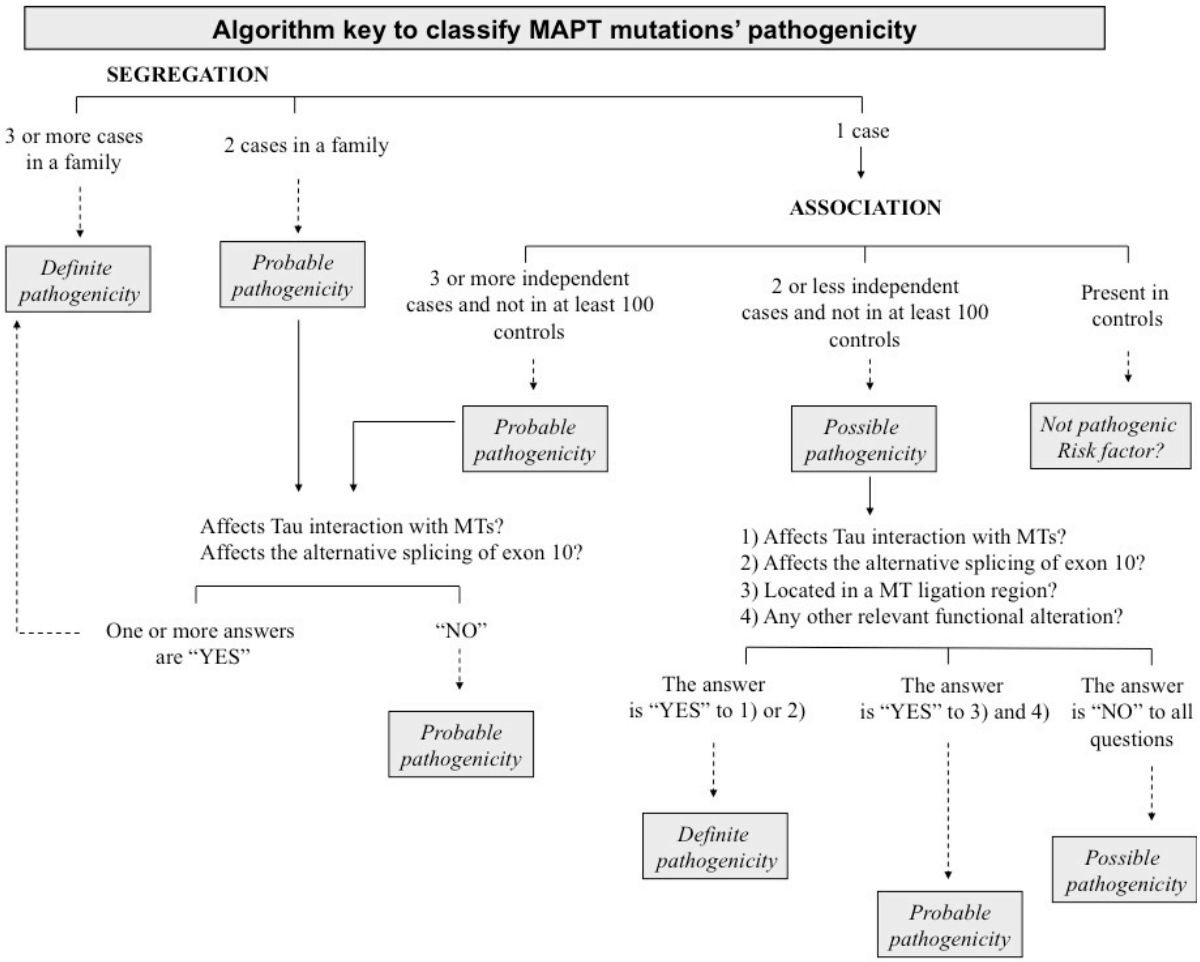


Figure 24. Algorithm to evaluate the pathogenicity of *MAPT* variants, classifying them as not pathogenic, possibly pathogenic, probably pathogenic, or definitely pathogenic.

CHAPTER 6

NEW GENES IN ALZHEIMER'S DISEASE

Based on:

1) Extended tracts of homozygosity identify novel candidate genes associated with late-onset Alzheimer's disease.
Nalls MA*, **Guerreiro RJ***, Simon-Sanchez J, Bras JT, Traynor BJ, Gibbs JR, Launer L, Hardy J, Singleton AB.
Neurogenetics. 2009 Jul;10(3):183-90. Epub 2009 Mar 7

2) Whole genome analysis in a consanguineous family with early onset Alzheimer's disease.
Clarimón J, Djaldetti R, Lleó A, **Guerreiro RJ**, Molinuevo JL, Paisán-Ruiz C, Gómez-Isla T, Blesa R, Singleton A, Hardy J.
Neurobiol Aging. 2008 Apr 1

3) Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease.
Harold D, et al.
Nat Genet. 2009 Oct;41(10):1088-93.

4) Genetic variability in CLU and its association with Alzheimer's Disease.
Guerreiro R, Beck J, Gibbs JR, Santana I, Nalls M, Ribeiro MH, Santiago B, Oliveira CR, Collinge J, Mead S, Singleton AB, Hardy J
PLoS One. 2010 Mar 3;5(3):e9510

CHAPTER 6 – NEW GENES IN ALZHEIMER DISEASE

6.1 - The possible role of homozygosity in Alzheimer's disease

6.1.1 - Late Onset Alzheimer's Disease

Recent research has noted that the human genome contains a higher frequency of extended tracts of contiguous homozygous single nucleotide polymorphisms than previously expected [313-315]. The occurrence of these long homozygous tracts in uninterrupted sequences may represent deletion polymorphisms, loss of heterozygosity, segmental uniparental disomy, low minor allele frequency, or autozygosity. Recent data has suggested that these tracts most likely represent extended homozygosity by virtue of parental descent from a common ancestor [314, 316-318]. The longest tracts are expected to occur in more recently inbred populations with subsequent recombination events interrupting long chromosomal segments [313]. Our group has reported the unexpected high degree of apparent parental consanguinity in control individuals from North America (9.5% of studied individuals harboring homozygous tracts larger than 5Mb) [319]. Similar numbers (6.6%) were presented by Li L et al when studying an outbred population of unrelated Han Chinese [314] and by Gibson et al who reported that 1393 tracts exceeding 1Mb in length were observed in the 209 unrelated HapMap individuals studied [313]. The same conclusion was obtained by McQuillan and colleagues in a recent manuscript. Using pedigrees from two isolated and two more cosmopolitan populations of European origin, they demonstrated that runs of homozygosity up to 4 Mb are common in out bred individuals [315].

The development of platforms able to genotype over one million SNPs has provided an unparalleled opportunity to study tracts of extended homozygosity. One obvious application of this technology is the study of recessive families. Recent research our group has reported a whole genome analysis of a consanguineous family with early onset Alzheimer's disease. This study presented the first catalog of extended homozygosity in EOAD and identified several regions as candidate loci for a recessive

genetic lesion in AD [320]. This same technology may also be used to perform homozygosity mapping for disease in populations with unknown pedigree structures from ostensibly out bred populations [321].

In order to evaluate the role of runs of extended homozygosity in LOAD, we analyzed the association between Alzheimer's disease and quantitative measures of extended homozygosity. The underlying hypothesis of this work is that an excess of large homozygous tracts in AD patients versus controls would support the notion that a recessive component exists for the disease. This led us to perform a statistical comparison of the extended tracts of homozygosity in a series of 837 LOAD patients and 550 neurological normal controls [96].

METHODS

Data characterization

The data used for this analysis is publicly available at http://www.tgen.org/research/neuro_gab2.cfm and was generated by a genome wide analysis of unrelated LOAD cases and healthy controls [322]. Detailed quality control data for this publicly available dataset is described in Reiman et al., 2007 [96]. Due to the necessity of relatively complete genotyping coverage in analyses of extended homozygosity, we increased the genotyping success threshold to 97% successful genotypic calls per individual (allowing for an average call rate per individual of ~98.5%), but causing the exclusion of 24 participants from the initial dataset. The total population in our analyses after exclusions were made (total n=1387) was taken from three cohorts: the neuropathological discovery cohort (n= 722); the neuropathological replication cohort (n= 305), and the clinical replication cohort (n= 360). All participants included in the initial study were of self-reported European ancestry and population structure was evaluated using STRUCTURE [323]. Only participants identified as genetically most similar to European HapMap samples were included in the

analytic dataset, with 14 outliers removed prior to data analysis [96, 324]. A summary of the information available from each cohort used in the analyses is presented in Table 18. Neither individual ages nor sex were available for public access.

Table 18. Summary of the information available online for the data used in this analysis

	Neuropathological discovery cohort		**Neuropathological replication cohort**		**Clinical replication cohort**	
N	722		305		360	
Cases / Controls	LOAD cases	Controls	LOAD cases	Controls	LOAD cases	Controls
	432[a]	290[b]	191[a]	114[b]	214[c]	146[d]
ApoE4 carriers	291 (67%)	61 (21%)	109 (57%)	27 (24%)	113 (53%)	29 (20%)
	Neuropathological cohorts				**Clinical replication cohort**	
	LOAD cases		Controls		LOAD cases	Controls
Mean age* +/- standard deviation	73.5+-6.2 at death		75.8+-7.5 at death		78.9+-7.8 at last clinical visit	81.7+-6.6 at last clinical visit

a- Brain donor cases satisfied clinical and neuropathological criteria for LOAD.
b- Brain donor controls did not have significant cognitive impairment or significant neuropathological features of AD.
c- Clinical cases satisfied criteria for probable AD (living subjects who were at least 65 years old at the time of their death or last clinical assessment and who were independently assessed for their APOE genotype).
d- Clinical controls did not have clinically significant cognitive impairment or significant neuropathological features of AD.
* denotes taken from Reiman et al., 2007 published data

Identification of runs of extended homozygosity

Runs of extended homozygosity were identified using the PLINKv1.02 software package [http://pngu.mgh.harvard.edu/~purcell/plink/contact.shtml - cite]. We utilized robust criteria for inclusion of genomic regions into runs of homozygosity as a means of reducing confounding by copy number. We identified only large runs of homozygosity (ROH), the primary criteria for genomic inclusion in a ROH being at least 1 Mb of consecutive homozygous genotypic calls at adjacent SNP loci. The minimum SNP density coverage was at least 50 SNPs per Mb to be included in a homozygous run, allowing for centromeric and SNP-poor regions to be algorithmically excluded from any analyses. The genome was scanned for ROHs using a sliding window of 50 SNPs, and allowed at

most 2 missing genotypes and 1 heterozygote call per ROH. These large ROHs identified are suggestive of a recent origin, as opposed to shorter runs suggestive of an ancient ROH origin.

Total length of the genome comprised in homozygous runs (expressed in Mb), average length of each homozygous run (in Mb per ROH) and total number of homozygous runs per participant were calculated. The total length of the genome comprised of homozygous runs is the sum of the distance of each individual run per participant. The average length of homozygous runs was generated by dividing the total genomic length of the homozygous runs by the total number of homozygous runs per participant.

Homozygosity mapping

From the initial structural analysis, 1090 consensus regions from overlapping ROHs were defined. These consensus regions were comprised of the loci shared by all overlapping segments of extended homozygosity in a particular genomic region and were used as positional loci for identification of common ROHs in our mapping efforts. Each consensus region was comprised of at least 3 SNPs and 100,00 base pairs and was found in no less than 10 participants. The largest consensus regions were 1,782,642 bp and contained 207 SNPs. The average consensus region defined was 351,744 bp in length and contained ~36 snps.

These consensus ROHs were analyzed using the maxT permutation test algorithm for case/control studies in PLINKv.1.02. This algorithm, which tests the frequency of segmental occurrence in cases compared to that in controls, is identical to that used in copy number variant analyses of disease association although it has been adapted to incorporate the larger ROH segments. 50,000 permutations of the maxT permutation testing algorithm were used to generate empirical p-values and association coefficients to evaluate risk of LOAD attributable to individual runs of homozygosity. These empirical p-values were then corrected for multiple testing incorporating the possibility of false positive results in each of the possible permutations per ROH tested during the label-swapping function.

Additional statistical analyses

We tested hypotheses of increased summary measures of extended homozygosity (i.e., number of ROHs, total and average ROH lengths) being associated with LOAD using simple two tailed T-tests to compare means between cases and controls. These measures were relatively normally distributed.

RESULTS

Identification of risk homozygous regions

In order to localize specific homozygous regions significantly associated with LOAD, empirical p-values for 50,000 permutations of homozygous consensus overlapping regions were determined using a modified CNV algorithm from PLINK toolset, adapted to larger segments (Figure 25). One homozygous consensus region in chromosome 8 was found to be significantly overrepresented in cases when compared to controls (Table 19). This region is not concurrent with homozygous regions previously reported to have a frequency above 25% in healthy individuals [325] nor was it found as an overlapping long contiguous stretch of homozygosity in three different populations (Han Chinese, Taiwan aborigines and Caucasians) [314]. The other top regions most significantly overrepresented in LOAD cases include additional loci on chromosomes 8, 9 and 10 and are associated with empirical p values between 0.001 and 0.01 (0.02 and 0.1 after adjustment for multiple testing corrections).

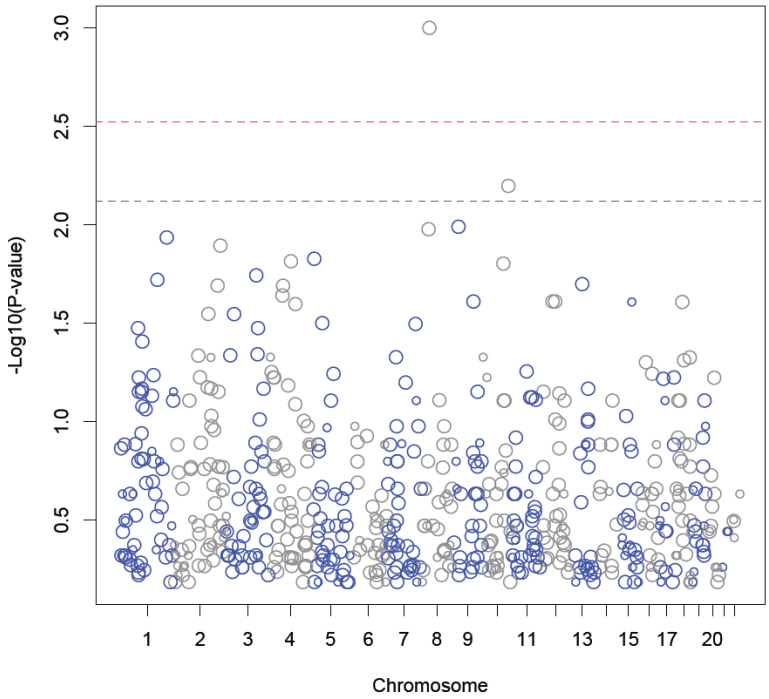


Figure 25. Empirical p-values for runs of homozygosity between cases and controls. The red line denotes significance based on permutation-based adjustments for multiple testing phenomena.

Genes included in the risk homozygous regions

The genes present within the top 3% homozygous consensus regions overrepresented in cases when compared with controls are shown in Table 19. From the seven genes located in the homozygous region in chromosome 8 found to be most significantly different between cases and controls (*RAB11FIP1*; *MGC33309*; *ADRB3*; *EIF4EBP1*; *ASH2L*; *STAR* and *LSM1*), only the *ADRB3* gene (GeneID: 155), coding for the beta-3-adrenergic receptor is represented in the AlzGene database [71]. One SNP in this gene (rs4998) was found to be associated with AD in ApoE4 negative samples by Hamilton G. et al, in a case-control association study performed for candidate insulin signaling genes [326]. Additionally, the steroidogenic acute regulatory protein (StAR), coded by the *STAR* gene (GeneID: 6770), was found to be markedly increased in the cytoplasm of hippocampal pyramidal neurons and in the cytoplasm of other non-neuronal cell types from AD brains, when compared with age-matched controls. Also, the levels of phosphorylated eukaryotic translation initiation factor 4E

binding protein 1, encoded by the gene *EIF4EBP1* (GeneID: 1978), have been previously found to be dramatically increased in AD and positively significantly correlated with total tau and p-tau [327].

Table 19. Top 3% homozygous consensus regions overrepresented in cases when compared to controls and genes present within these regions

Chr	Start of region (bp)	End of region (bp)	Region length (bp)	Empirical p-value	Permutation p-value after adjustment for multiple testing	Candidate genes	Number of cases with homozygous runs	Number of controls with homozygous runs	OR
8	37835460	38143780	308320	0.00099998	0.0170397	RAB11FIP1; MGC33309; ADRB3; EIF4EBP1; ASH2L; STAR; LSM1	40	9	3.02
10	116826992	117301088	474096	0.00633987	0.0908982	ATRNL1	45	14	2.18
9	26314013	26648658	334645	0.0102198	0.129957	-	9	0	n.a.
8	34251617	34375482	123865	0.0105198	0.151257	LOC137107	109	49	1.53
1	213279861	213715720	435859	0.0115798	0.152477	ESRRG	32	9	2.39
2	214386585	214511938	125353	0.0127397	0.155837	PF20	34	10	2.29
5	15390663	15634707	244044	0.0148597	0.161017	LOC391741; FBXL7	27	7	2.59
4	98911102	99056376	145274	0.0152997	0.198896	MGC46496	162	81	1.39
10	93788860	94018536	229676	0.0157197	0.192296	CPEB3	63	25	1.71
3	137687127	137949289	262162	0.0180396	0.213336	STAG1	149	74	1.39
1	169919601	170135865	216264	0.0189996	0.190176	LOC343153	14	2	4.66
13	56558450	56726080	167630	0.0199796	0.241755	FLJ40296	58	23	1.71
2	201456660	201882057	425397	0.0203196	0.211676	AOX2; BZW1; LOC391472; CLK1; PPIL3; NIF3L1; ORC2L; MGC39518; NDUFB3; CFLAR; CASP10	28	8	2.34
4	62126310	62256398	130088	0.0203596	0.196436	LPHN3	11	1	7.31
4	59420217	59599269	179052	0.0228195	0.235695	-	21	5	2.81
9	95407641	95524041	116400	0.0244795	0.240655	PSMA7P	16	3	3.55
12	51809479	51949002	139523	0.0244795	0.240655	LOC400039; CSAD; LOC283337; ITGB7; RARG; MGC11308; ESPL1	16	3	3.55
12	63691379	64048325	356946	0.0244795	0.240655	WIF1; MAN1;	16	3	3.55

						LOC253827			
15	65557360	66059106	501746	0.0246195	0.251595	FLJ12476; LOC145853; MAP2K5; LOC390598; LOC388129	23	6	2.56
18	31690766	31818000	127234	0.0246195	0.251595	C18orf21	23	6	2.56
4	119462492	119635887	173395	0.0252395	0.278194	NDST3; PRSS12	53	21	1.70
2	158898786	159035950	137164	0.0283194	0.272655	LOC130940	18	4	3.00
3	35117703	35239453	121750	0.0283794	0.257595	LOC442078	13	2	4.32
5	52730874	53125379	394505	0.0315194	0.268275	FST; NDUFS4	10	1	6.64
7	132838685	133208570	369885	0.0318394	0.343753	SEC8L1	63	27	1.58
1	80639887	80811557	171670	0.0334193	0.300274	LOC441889; COX6A1P	22	6	2.45
3	145678190	145958179	279989	0.0334193	0.300274	-	22	6	2.45
1	100201372	100553949	352577	0.0390992	0.376492	HIAT1; DKFZp761A078; FLJ10287; MGC14816; DBT; RTCD1; CDC14A	49	20	1.65
3	143407729	143828507	420778	0.0453791	0.422592	MGC40579; XRN1; ATR; PLS1	35	13	1.80
3	17477922	17773653	295731	0.0459391	0.448171	TBC1D5	84	40	1.42
2	112487241	112641378	154137	0.0460991	0.385732	MERTK; FLJ14681; FLJ37440	27	9	2.00
7	42294711	42463020	168309	0.0469591	0.329253	-	6	0	n.a.

Chr – Chromosome; bp – Base Pairs; OR – Odds Ratio; n.a. – not available.

Analysis of extended homozygosity (Table 20)

The AD cases studied presented a slightly higher degree of extended homozygosity when compared with the control group. The most variability observed was associated with the total number of homozygous runs and total run length with suggestive p-values of 0.062 and 0.052 respectively, while the average run length difference between these two groups was not statistically significant ($p>0.10$).

Table 20. Descriptive statistics for measures of extended homozygosity

	N	Measure	Mean	SD
Controls	550	Number Runs	51.35	16.43
		Total Run Length	73.02	28.13
		Average Run Length	1.41	0.22
Cases	837	Number Runs	52.94	15.24
		Total Run Length	76.15	30.48
		Average Run Length	1.42	0.28
Total	1387	Number Runs	52.31	15.73
		Total Run Length	74.91	29.60
		Average Run Length	1.42	0.26

DISCUSSION

In order to identify risk loci significantly different between cases and controls that might contain genetic lesions responsible for AD, empirical p-values for consensus homozygous regions were determined in both cases and controls groups. One consensus ROH in chromosome 8 was found to be significantly overrepresented in cases when compared to controls after correction for multiple testing. This region contains seven genes, three of which have previously been studied in relation to AD (Table 19).

The *ADRB3* gene product, beta-3-adrenergic receptor, is one of three beta-adrenergic receptor subtypes. The product is located mainly in adipose tissue and is involved in the regulation of

adipocyte lipolysis, thermogenesis and oxygen consumption [328]. There is evidence that a disturbance in the insulin signal transduction pathway may be a central and early pathophysiologic event in sporadic AD [329]. Patients with diabetes and insulin resistance have an increased risk of impaired cognition or dementia [330]. Hamilton et al performed a two stage association study to investigate the role of insulin signaling genes in the risk of developing AD showing one SNP in *ADRB3* (rs4998) was studied and found to be associated with AD in ApoE4 negative samples [331].

The steroidogenic acute regulatory protein (StAR), coded by the *STAR* gene, plays a key role in the acute regulation of steroid hormone synthesis by enhancing the conversion of cholesterol into pregnenolone [332]. To test the hypothesis that hormones of the hypothalamic-pituitary-gonadal axis have a role in AD pathogenesis, K.M. Webber et al studied the expression levels of StAR in AD and control brains. The authors found an increase of this protein in AD hippocampal neurons as well as other non-neuronal cells compared to aged matched controls. These results, together with the finding that StAR is present in the same brain regions as the luteinizing hormone (LH) receptors, suggest that steroidogenic pathways regulated by LH may play a role in AD [326].

The *EIF4EBP1* gene encodes one member of a family of translation repressor proteins, which interacts with eukaryotic translation initiation factor 4E (eIF4E). This is a limiting component of the multi-subunit complex that recruits 40S ribosomal subunits to the 5' end of mRNAs. Interaction of this protein with eIF4E inhibits the complex assembly and represses translation. The 4E-BP1protein is phosphorylated in response to several signals, including insulin signaling and together with several other components of the translational machinery is regulated through signaling events that require the mammalian target of rapamycin (mTOR) [333]. Neurofibrillary tangles, composed mainly of hyperphosphorylated tau protein, are one of the neuropathological hallmarks of AD. Tau mRNA levels have been shown not to be altered in sporadic AD brains. Nevertheless, Li X, et al studied the possibility that tau mRNAs in AD brains could be abnormally regulated by investigating the levels of various translation control elements including 4E-BP1, in the brains of AD and controls subjects.

Together with increased levels of p-mTOR, they found increased levels of phosphorylated 4E-BP1 in AD and a positive significant correlation with total tau and phosphorylated tau [327].

Genetic variation in one or more of these genes may account for the development of different phenotypes related to LOAD and it is possible that this form of genetic heterogeneity coexists with the multifactorial, common-disease/common-variant mode of inheritance that is generally studied in whole-genome association [325].

From a methodological perspective, using ROHs to identify possible candidate loci for small effects size or rare recessive variants in unrelated individuals may be useful. This method uses relatively few statistical tests compared to conventional genome-wide association studies to probabilistically model recessive genome-wide associations with disease, increasing power to some degree. The method itself likely trades specificity of results for sensitivity, by seeking to identify large regions of homozygosity harboring as-of-yet unknown recessive disease components. In particular, homozygosity mapping is limited by fine mapping within the risk ROH not being feasible in the discovery population. To identify a more focused risk region in replication of a homozygosity mapping finding, a population of a different ethnic background or an admixed population would have to be studied, as boundaries of the ROHs would likely be different from the discovery cohort. The summary measures of genome-wide extended homozygosity may also prove useful in quantifying distant consanguinity in a single unrelated population, for which F_{st} calculations may not be applicable to gauge genetic distance from a founder group.

Using public whole genome SNP analysis data on late onset Alzheimer's disease and neurologically normal controls we identified an average of 52.1 homozygous runs larger than 1 Mb containing more than 50 consecutive SNPs in 1387 samples. In order to determine if Alzheimer’s disease populations are more consanguineous than healthy controls we statistically compared the total number of homozygous runs, average length of homozygous runs and total length of genome contained in runs of homozygosity between cases and controls. For this analysis we used a stringent

method, assuming that the presence of large tracts of contiguous homozygous SNPs represent a direct association with increased levels of relatively recent consanguinity [96, 319]. Only homozygous runs with sizes over 1 Mb were included in the analysis in order to rule out the effect of copy number variation, which could confound results if smaller minimum run sizes were used. Our borderline significant results suggest that the AD cases in this study may be characterized by a higher degree of extended homozygosity than the control group, although this generalization would need to be tested in a larger sample size.

As the individual ages were not available to the public, it was not possible to establish a relation between age and homozygosity. Still, our group has recently reported the first genetic proof that younger individuals present less homozygosity than older ones, presumably due to populations increased mobility in the last generations. Linear predictive models of homozygosity and current age, showed that younger individuals have smaller percentage of their genome contained in homozygous runs; which are significantly shorter in length [334]. Similar results would be expected for the present populations.

The E4 allele of *APOE* has been the only genetic risk factor consistently associated with familial and sporadic forms of AD [335]. In the present study no association between homozygosity at the APOE gene and AD was found. This is explained by the size of the statistical window used (1Mb), as the methodology used was designed to pick up variants with recently created ROHs that have not been broken up by recombination. High recombination rates in this genomic region in the European American population, prevent detection of any signal at the APOE gene using homozygosity mapping [336].

In summary, we conclude that extended runs of homozygosity are common in out bred populations and present statistical data that show specific large tracts of homozygosity are a risk factor for Alzheimer's disease. The present results, together with the previously reported cases of consanguineous families with AD [320] and the speculation that recessive genes for AD are

responsible for the high AD prevalence in the Wadi Ara [337], demonstrate that particular regions of homozygosity have a significant role in AD genetics. It is possible that future sequencing and additional follow-up analyses will allow the identification of small-effect size recessive risk variant(s) within the described ROHs.

6.1.2 - Early Onset Alzheimer's Disease

Only in 1-2% of all Alzheimer's disease cases, the disease appears before the fifth decade of life (<60-65 years of age). In the majority, but not all, EOAD patients, the disease aggregates within families, and ~10% of these families show an autosomal dominant pattern of inheritance. The three genes unequivocally related to these familial forms of EOAD account for 15-50% of the cases (*PSEN1*), 5% of cases (*APP*), and less than 1% (*PSEN2*) (for review see [1]). However, despite the great effort that has been made to find other genetic causes, no other Mendelian gene has been found during the last 11 years. In summary, although genetic causes for EOAD have been found, a substantial proportion of sporadic and familial EOAD cases are of unknown genetic etiology [216, 225]. Surprisingly, there has been no formal assessment of the possibility that a proportion of cases have a simple recessive aetiology, although there have been two studies of isolated populations which have had a high incidence of disease, although even in these, the primary analyses have been using dominant or additive modes of inheritance [57, 338].

In this study we describe a consanguineous family in which two siblings, sons of a first-cousin, neurologically healthy parents, suffered from EOAD. We catalog the homozygous genomic regions identical by descent in both patients.

METHODS

Patients

We analyzed a Jewish Israeli family originating from Morocco (Figure 26). The family was composed of seven sibs born to first-degree consanguineous parents who died after the age of 90 years without any sign of cognitive impairment. Two sibs suffered from EOAD and agreed to participate in the genetic study. The others refused to give any blood sample. Written informed consent was obtained from both patients.

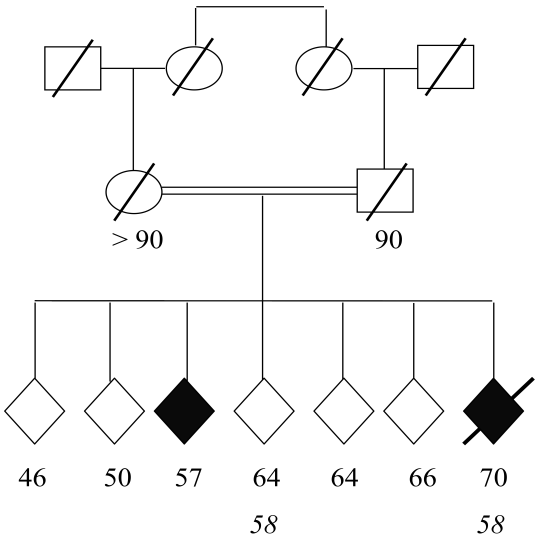


Figure 26. Pedigree of the Israeli family. Squares represent men, circles represent women: diamonds are used in the present generation to obscure identity. Black filling indicates early-onset Alzheimer's disease. Actual ages (and age of death of one of the siblings) are shown underneath. Ages at onset are indicated in Italics.

Genetic Analyses

After informed consent was obtained, blood sample was collected and DNA extraction was performed by standard methods. Mutations in *PSEN1* (exons 3 to 12), *PSEN2* (exons 3 to 12), *APP* (exons 16 and 17), *MAPT*, *GRN*, and *PRNP* were discarded by direct sequencing.

Genome wide SNP typing using the Illumina HumanHap240S Genotyping was performed for the two siblings. A total of 241,848 SNPs (236,852 located in autosomes) were genotyped for each sample. The experiments were carried out as per the manufacturer's instructions. Genotype success rate greater than 95% was obtained for both assays. Visualization of homozygosity tracks and gene copy variation was performed through Genome Viewer tool within Beadstudio v2.2.22 (Illumina Inc., San Diego, CA). B allele frequency and log R ratio were analyzed. Briefly, B allele gives an estimate of the proportion of times an individual allele is called A or B; thus an individual homozygous for the B allele would have a score close to 1, an individual homozygous for the A allele a score close to 0, and a score of 0.5 would indicate a heterozygous genotype. Value of the log R ratio is the log (base 2) ratio of the observed normalized R value for the individual SNP divided by the expected normalized R value for the SNPs theta value. The ratio of observed R to expected R for any SNP gives an indirect

measure of the binding efficiency of detected alleles for each polymorphism and therefore, gives an estimate of genomic copy number. An R above 1 is indicative of an increase in copy number (duplication or triplication), and values below 1 suggest a deletion.

RESULTS

Family description and clinical characteristics

The family consisted of seven siblings from a first-cousin marriage (Figure 26). At the time of the clinical study the mother was 90 years old with mild gait difficulties but otherwise healthy, with no cognitive loss. We are aware that she died a few years later. The father had died at age 90 due to cerebral stroke. The proband is 64-years-old who was born in Morocco and migrated to Israel in 1956. At the age of 58 he presented with a 9-month history of insidious, progressive impairment of memory first noticed by the family. Early symptoms were repeated questioning, forgetting names and loss of interest in his family. Other features included an inability to plan or order tasks, for which he was fired from his work in a tire factory. During the neurological examination, no involuntary movements, pyramidal, extrapyramidal, or cerebellar signs were found. Primitive frontal release signs were absent. Two months after the initial visit the Mini-Mental State Examination score was 9/30. Cognitive examination revealed diffuse cognitive impairment, most notably of short and long-term memory, verbal and visual recall, and visuospatial abilities. Calculation, attention tasks, and information processing speed were markedly impaired. In addition, he had moderate dressing apraxia and right-left disorientation. Other neuropsychological examinations (i.e., Wechsler test, word naming and general knowledge) were difficult to perform due to advanced dementia. According to the NINCDS-ADRDA criteria the clinical examination and neuropsychological profile was consistent with EOAD [215]. The diagnostic workup included brain MRI that showed generalized, symmetrical cerebral atrophy, and an EEG with diffuse mild slowness at the range of 5-7 Hz. The lumbar puncture showed normal protein and glucose levels and PCR from the CSF was negative for

tuberculosis. One of his siblings started with clinical symptoms at the age of 58, when he was laid-off from work due to cognitive decline. One year later, he was institutionalized in a psychiatric institution with a clinical diagnosis of EOAD. He died at the age of 70 and did not present any extrapyramidal sign during the course of disease. The actual ages of the other sibs are 46, 55, 57, 64 and 66. All of them are neurologically healthy. APOE genotyping revealed that both affected siblings carried an APOEε3ε4 genotype.

Whole genome genotyping

We looked at both log R ratio and B allele frequency plots across the whole genome in both siblings with EOAD. This high-throughput genotyping technology allowed a rapid production of high-quality, ultra-dense genotypes and provided direct visualization of extended tracks of homozygosity [322]. Information of all the homozygozity genomic regions greater than 1MB that are shared by both siblings is presented in Table 21. Also, a diagram showing the homozygous regions of the entire genome of both siblings is depicted in Figure 27. There were no regions of homozygosity identical by descent greater than 1MB on chromosomes 7, 12, 15, 18, 19, 20, 21, and 22. Also, regions such as 14q24.3 (*PSEN1*), 1q31-q42 (*PSEN2*), 21q21 (*APP*), 17q21.1 (*MAPT*), 10q23-q25 (*IDE*), 3q21-q27 (*MME*), 1p36.1 (*ECE*), or 11q23.2-24.2 (*SORL1*) did not show shared homozygosity. The total length of genomic homozygosity that was shared by both siblings was of ~250MB, which represents ~8.32% of the genome, with the longest track of homozygosity in chromosome 11 (with 59.1MB). This is considerably longer than would be expected (~2%) and suggests, unsurprisingly, that there has been additional consanguinity in previous generations. There were no duplications or triplications in either of the genomes and all genomic deletions found have been previously reported as polymorphic copy number variations of the human genome [339].

Table 21. Lengths and positions of homozygosity tracks shared by both affected siblings

Chr	From (SNP ID)	Position	To (SNP ID)	Position	Total homozygosity lengh (Mb)	Genes in AlzGene located in these regions
1	rs1390484	64,603,980	rs6681765	76,549,653	11.95	
1	rs2994809	120,153,500	rs661678	145,445,019	25.29	PRKAB2
1	rs2584315	220,273,510	rs851170	221,416,553	1.14	
2	rs13423995	89,769	rs3910069	7,472,223	7.38	
2	rs10189659	193,944,304	rs2123738	195,479,306	1.54	
3	rs1517520	16,482,329	rs17006228	19,761,798	3.28	RFTN1
3	rs9863064	61,590,243	rs704267	71,784,695	10.19	
3	rs1546304	80,618,614	rs893518	81,898,253	1.28	
3	rs12630403	103,152,590	rs1107889	116,721,177	13.57	
4	rs7672923	110,173,872	rs4975181	129,466,845	19.29	CASP6; FABP2
5	rs6870997	34,587,622	rs7711836	55,344,388	20.76	PRKAA1; HMGCS1; NDUFS4
5	rs17237118	67,441,983	rs7732832	90,645,624	23.20	HMGCR; CRHBP
6	rs9460326	11,382,125	rs13202453	21,091,421	9.71	
8	rs10503942	33,300,800	rs16882274	34,447,332	1.15	
8	rs6470874	132,053,392	rs7017860	135,220,187	3.17	
9	rs10814410	36,587	rs1378324	3,834,391	3.80	VLDLR
10	rs10823828	73,129,862	rs607483	78,895,276	5.77	PSAP; CHST3; PPP3CB; SEC24C; NDST2; CAMK2G; PLAU; VCL; AP3M1; MYST4; KCNMA1
10	rs12267798	106,125,107	rs4918515	112,181,982	6.06	SORCS3; SORCS1; hCG2039140
11	rs1503502	19,215,941	rs611962	78,319,725	59.10	BDNF; CAT; MAPK8IP1; TCN1; CNTF; CHRM1; GSTP1; GAL; FADD; INPPL1; UCP2; GAB2
13	rs12430069	59,745,163	rs7999738	61,243,897	1.50	
13	rs1316712	84,627,554	rs2148065	94,432,314	9.80	
14	rs11845833	73,772,002	rs17103130	74,844,099	1.07	NPC2; DLST; FOS
16	rs4081759	65,505,284	rs10431924	67,396,803	1.89	HSD11B2
17	rs16968682	34,640,425	rs16939964	40,779,673	6.14	GRB7; PNMT; PYY

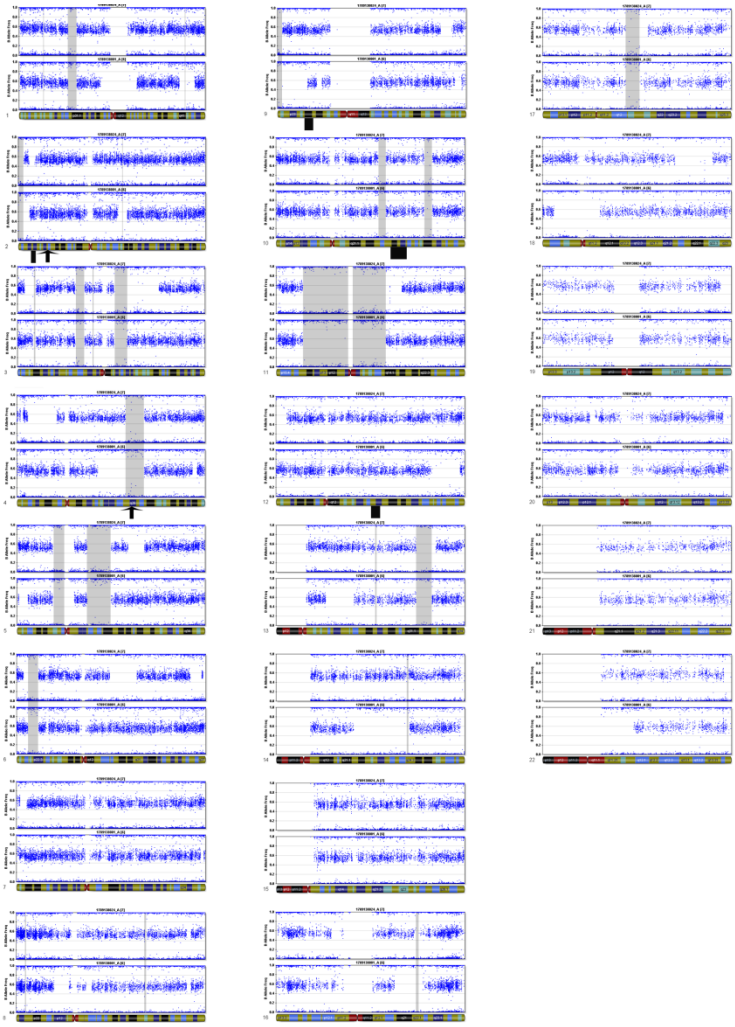


Figure 27. Data from HumanHap240S beadchips revealing the extension of homozygous regions shared by both siblings. The plot shows the B allele frequency across the chromosome. Genotypes for BB have an allele frequency of 1; heterozygous (AB) genotypes have an allele frequency of 0.5; and AA genotypes have an allele frequency of 0.

DISCUSSION

As far as we know, this is the first study describing a family with a segregation pattern of EOAD suggesting a recessive mode of inheritance. We have used a high throughput genotyping technology to accurately define the length of homozygosity that is identical by descent in both siblings with the disease. Since the HumanHap240S beadchip average spacing between SNPs (MAF ≥ 0.05) is 5.5Kb for the CEU population, genomic rearrangements equal or greater than this length could have been detected. Although we are aware that this is an approximate approach and that this is not a powerful sample to limit the loci of interest, our aim was to stress that an autosomal recessive mode of inheritance is a possible cause of dementia and more importantly to provide an initial catalog of candidate loci for us and others to follow-up on. In contrast to autosomal dominant cases, estimates of the incidence of sporadic EOAD have not been accurately described in the literature. However, patients with EOAD and no family history of disease are relatively common. We suggest that recessive segregation of disease could be playing a role in, at least, some forms of AD, such as the present family, and could also account for a proportion of early onset "sporadic" cases. In the present study, it is unlikely that the presence of one single APOE-ε4 allele could lower the age of onset to 58 years, as is the case of the two patients analyzed.

We have narrowed the homozygosity region shared by both siblings to ~250MB, thus discarding ~92% of the genome to contain any hypothetical recessive gene (at least in the present family). In Table 21, we outline this region and note all the loci from Alzgene (www.alzgene.org: [71]). Although this is beyond present sequencing capacity, sequencing less than 10% of the human genome is going to be a reasonable and realistic endeavor in the near future. This makes the present data a significant start for subsequent studies in which large scale genome sequencing will be performed in both familial and sporadic EOAD patients not carrying mutations in any of the known mendelian genes linked to the disease.

6.2 - Genome wide association studies in Alzheimer's disease

Alzheimer's disease is the most common form of dementia, is highly heritable (heritability of up to 76%) but genetically complex [340]. Aiming to identify novel AD loci, several genome-wide association studies (GWAS) have been conducted prior to the study in which we participated. All have identified strong evidence for association to *APOE*, but less convincing evidence implicating other genes [94-101]. This outcome is consistent with the majority of findings from GWAS of other common phenotypes, where susceptibility alleles typically have effect sizes with odds ratios of 1.5 or less, rather than of the magnitude for *APOE* (OR~3). Detecting such modest effects requires much larger samples than those that have been applied in the GWAS of AD to date [341], which have all included fewer than 1,100 cases. Based upon the hypothesis that risk alleles for AD are likely to confer ORs in the range seen in other common diseases, we participated in a collaborative consortium from Europe and the USA. This collaboration was able to get a combined sample of up to 19,000 subjects (before quality control) and conduct a two-stage study. In Stage 1, 14,639 subjects were genotyped on Illumina platforms. 5,715 samples were genotyped for the present study using the Illumina 610-quadchip; genotypes for the remaining subjects were either made available to us from population control datasets or through collaboration and were genotyped on the Illumina HumanHap550 or the HumanHap300 BeadChips. Prior to association analysis, all samples and genotypes underwent stringent quality control, which resulted in the elimination of 53,383 autosomal SNPs and 2,850 subjects. Thus, in Stage 1, 529,205 autosomal SNPs were tested for association in up to 11,789 subjects (3,941 AD cases, 7,848 controls of which 2,078 were elderly screened controls.

In addition to the known association with the *APOE* locus, GWA analysis identified two novel loci at a genome-wide level of significance. Table 22 shows SNPs that were genome-wide significant (GWS) in stage 1. Thirteen GWS SNPS map within or close to the *APOE* locus on chromosome 19 ($p< 3x10^{-8}$ - $2x10^{-157}$) and the top five are shown in Table 22.

Table 22. SNPs showing genome-wide significant association with AD in stage 1 of the GWAS.

SNP	Chr	Closest RefSeq Gene	Location Relative to Gene	MAF	GWAS: 3941 cases 7848 controls P-value (two-tailed)	GWAS OR (95% CI)	Extension: 2023 cases 2340 controls P-value (one-tailed)	Extension OR (95% CI)	Combined: 5964 cases 10188 controls P-value (two-tailed)	Combined OR (95% CI)	Population Attributable Risk (%)
rs2075650 *	19	TOMM40	Intron	0.15	$1.8x10^{-157}$	2.53 (2.37-2.71)					18.7%
rs157580	19	TOMM40	Intron	0.39	$9.6x10^{-54}$	0.63 (0.59-0.66)					
rs6859	19	PVRL2	3' UTR	0.43	$6.9x10^{-41}$	1.46 (1.38-1.54)					
rs8106922	19	TOMM40	Intron	0.40	$5.4x10^{-39}$	0.68 (0.64-0.72)					
rs405509	19	APOE	5’	0.52	$4.9x10^{-37}$	0.70 (0.66-0.74)					
rs11136000	**8**	**CLU**	**Intron**	**0.40**	**$1.4x10^{-9}$**	**0.84 (0.79-0.89)**	**0.017**	**0.91 (0.83-0.99)**	**$8.5x10^{-10}$**	**0.86 (0.82-0.90)**	**8.9%**
rs3851179	**11**	**PICALM**	**5’**	**0.37**	**$1.9x10^{-8}$**	**0.85 (0.80-0.90)**	**0.014**	**0.90 (0.82-0.99)**	**$1.3x10^{-9}$**	**0.86 (0.82-0.90)**	**9.3%**

5 of the13 genome-wide significant SNPs at the *APOE* locus are shown. P-values in the extension sample and the combined sample are also shown for the two SNPs unlinked to the *APOE* locus (highlighted in bold).

Chr = Chromosome; Mb = position in megabases; MAF = minor allele frequency in controls; OR = odds ratio for the minor allele; 95% CI = 95% confidence interval; UTR= untranslated region. * rs2075650 is in linkage disequilibrium with rs429358, the *APOE* ε4 SNP ($r^2 = 0.48$).

Interestingly, *CLU* was also reported to be associated with AD by another GWAS [105], revealing rs11136000 in *CLU* on chromosome 8 as the most significant hit [104, 105].

Clusterin or apolipoprotein J is a lipoprotein expressed in most mammalian tissues with higher levels in brain, ovary, testis and liver [106]. It has been suggested to be involved in a variety of physiological processes, such as, ongoing synapse turnover [108], apoptosis [109, 110], cytoprotection at fluid-tissue boundaries, membrane recycling during development and in response to injury and regulation of complement-mediated membrane attack complex [107, 111]. Accordingly, CLU interacts with several different molecules, including lipids, amyloid proteins, components of the complement membrane attack complex and immunoglobulins [107]. From a biological point of view, it seems to have a potential central role in the pathogenesis of Alzheimer's disease. When control brains are compared with AD brains, CLU mRNA is found to be significantly elevated in AD affected brain areas [115] [111]. Clusterin is one of the proteins found to be part of amyloid plaques, the main neuropathological hallmarks of AD [116-119]. Clusterin specifically binds soluble Aβ in cerebrospinal fluid [120, 121] to form complexes able to cross the blood-brain barrier [122]. It has also been noted by Bertrand and colleagues, that reduced levels of ApoE and increased levels of CLU are correlated with the number of E4 alleles, suggesting a compensatory induction of CLU in the brain of AD individuals with the E4 allele of ApoE presenting low brain levels of ApoE [123].

Because of the genetic and biological evidence suggesting a role for CLU in AD, we sought to characterize the common genetic coding variability in this gene. To accomplish this, we compared the genetic variation found in *CLU*, between a group of 892 AD cases and a group of 684 neurologically healthy controls. Additionally, and since it is now clear that common risk alleles and rare mutations in the same gene may contribute to common sporadic and familial forms of the same disease, respectively; we intended to search for rare mutations that could be related to familial AD

cases in our sample. In this way, we sequenced the entire coding region of *CLU* in a total of 495 AD cases and 330 controls.

The absence of common coding variants able to explain the association with disease observed by Harold et al and Lambert et al, suggested a potential implication of differential gene expression or splicing changes of the gene [104, 105]. In order to determine if common variants within these regions effect expression of nearby (*cis*) mRNA transcripts we conducted an expression quantitative loci (eQTL) analysis for *CLU* genomic region, for gene expression in cortical tissues from a group of 174 neurologically normal controls (GEO Series GSE8919, Myers et al 2007).

METHODS

Sequencing analysis

Portuguese series

After obtaining written informed consent from all participants or surrogates, a blood sample was collected and DNA extracted from a total of 403 AD patients (62% of cases were women). All these were Caucasian with apparent Portuguese ancestry. The mean age of AD patients was 70 years old and mean age at onset was 67 years, ranging from 42 to 86 years. The genes associated with EOAD (*PSEN1*, *PSEN2* and *APP*) have been previously sequenced in all early onset cases (88 cases, with an age at onset before 65 years old) and no mutations were found [32]. Seventy-two patients (18%) showed a family history of dementia (defined as at least one affected first degree relative) and 103 (26%) did not report any familial aggregation of disease. In 56% of individuals, no information was available.

Written informed consent was obtained from 235 neurologically normal and aged control subjects from Portugal (mean age at collection 68 years, 58% women). All controls were subjected to a neurological examination and found free of any symptoms suggestive of cognitive decline.

UK series

Blood samples were collected from 489 AD patients (59% of cases were women) with informed, written consent for research purposes. DNA was extracted using established protocols. A total of 422 cases were of UK origin, 5 were of other European origin, 2 were of non-European origin and 60 were of unknown origin. Patients were referred from the Dementia Research Centre, National Hospital for Neurology, Queen Square, UK, or other specialist cognitive teams in the UK, with either a clinical diagnosis of AD, or alternatively with cognitive impairment and referred for *APP*, *PSEN1* or *PSEN2* gene analysis. The mean age of onset was 54 years with a range of 21-78 years. One or more of the genes causally associated with EOAD (*PSEN1*, *PSEN2* and *APP*) and with prion disease (*PRNP*) were screened and found to be negative for mutations in 207 cases. *PRNP* alone was found to be negative for mutation in a further 88 cases, and the remaining 194 cases were not screened. Fifty-two patients (11%) had a family history of dementia (defined as at least one affected first degree relative) and 70 (14%) did not report any familial concurrence of disease. In 75% no information on family history was available.

A total of 440 healthy control DNAs of Caucasian origin were obtained from ECCAC (European Collection of Cell Cultures) and 192 unrelated individuals from the CEPH (Centre d'Etude du Polymorphisme Humain). Appropriate consent was available for all control individuals.

DNA sequencing and data analysis

Sequencing analysis of *CLU* (GeneID: 1191; isoform 1: NM_001831.2, which is the longest transcript of the gene, encoding 9 exons) was carried out using genomic DNA of a total of 638 Portuguese samples (403 AD cases and 235 neurologically normal controls) and 187 UK samples [92 AD patients

and 95 UK healthy controls (ECCAC)]. All the coding exons plus the flanking intron-exon boundaries of *CLU* were PCR amplified using primers (see Table 3 in the Methods section for primers' sequences) designed using ExonPrimer software (http://ihg2.helmholtz-muenchen.de/ihg/ExonPrimer.html) and Roche FastStart PCR MasterMix polymerase (Roche Diagnostics Corp., IN). Each purified PCR product was sequenced using Applied Biosystems BigDye terminator v3.1 sequencing chemistry and run on an ABI3730xl (Applied Biosystems, CA) genetic analyzer as per manufacturer's instructions. The sequences were analyzed with Sequencher software, version 4.2 (Genecodes, VA).

Further 354 AD patients, 345 UK healthy controls (ECCAC) and 192 unrelated CEPH individuals were sequenced for exon 5 of *CLU* only. Genotypes for rs7982 were confirmed by realtime PCR using MGB allelic discrimination probes in all UK cases described above and an additional 43 definite AD cases.

In order to predict the impact in protein function of the non-synonymous variants found in *CLU* we performed an *in silico* analysis using PolyPhen [300, 301, 342].

SNPs were tested for association with Alzheimer's disease using logistic regression, assuming an additive model. These tests were performed using PLINK v1.07 [343].

eQTL Analysis

Since no gene expression data was available from the populations in which the sequencing analysis was performed, we used genotype and gene expression data from 174 neurologically normal controls from the United States, obtained from a previously published eQTL analysis [344]. Gene expression data is available at NCBI's Gene Expression Omnibus (GEO Series Accession, GSE8919) and genotype data is available from TGen (http://www.tgen.org/research/neuro_gab2.cfm). Based on genotype data samples were checked for quality of genotyping based upon genotype call rate per

sample, population outliers using Structure [323, 345] and Plink and for cryptic relatedness and gender discrepancies with Plink [343].

Genotype data were from the Affymetrix GeneChip Human Mapping 500K Array Set, which assays 502,627 SNPs (Affymetrix, Santa Clara, CA). Genotypes were imputed based on HapMap data using the IMPUTE program resulting in genotype call for 1741175 SNPs after QC of imputed data. Based upon SNP selection criteria of Hardy-Weinberg Equilibrium p-value > 0.001 and the presence of at least 3 minor homozygotes, 322 SNPs were selected for *cis* regional eQTL analysis near *CLU*, being the *cis* region defined as centered on the AD associated locus +/- 250Kb. Gene expression data were from the Illumina HumanRef-8 v1 Expression BeadChip that assays 24,357 RefSeq mRNA transcripts (Illumina Inc., San Diego, CA). Expression data were rank invariant normalized [346-348] using Illumina's BeadStudio Gene Expression module. Expression mRNA transcripts were selected for analysis based upon whether they reliable detected within 95% of samples, are within the *cis* genomic regions of interest and where the probes designed to assay the mRNA transcript did not contain a common polymorphism that have been typed within the HapMap CEU population. This resulted in 6 mRNA transcripts being selected for analysis, including one CLU transcript. Prior to eQTL analysis each of the normalized transcript expression profiles was $\log_2$ transformed and adjusted for available covariates (age at death, gender, cortical region, post mortem interval, tissue bank source and RNA preparation/hybridization batch) by linear regression, the residuals of these regressions were then used as the quantitative trait for eQTL analysis.

The eQTL analysis was performed using Plink's [343] *assoc* function for quantitative traits, which correlates allele dosage with changes in the trait. To correct for the number of tests performed, a Bonferroni correction was applied to the asymptotic p-values based 322 SNPs tested against 6 transcripts in the *CLU* region. This Bonferroni correction for 1932 tests resulted in a significance threshold of 2.587992e-05 for asymptotic p-values.

RESULTS

Sequencing analysis

A total of twenty-four variants were found in both cohorts. From these, eleven were observed in only one subject (Table 22).

Table 23. Changes (synonymous and non-synonymous) found in *CLU* and observed in only one subject (AD case or control) in both series.

Variants seen in only one subject					Sample					
Location in the gene	**DNA change**	**Protein change**	**Present in dbSNP**	**PolyPhen prediction of**	**Diagnosis**	**Sex**	**Age**	**AAO**	**Fam hist**	**Origin**
5'UTR	c.290G>A	N.A.	no	-	Ctrl	M	-	N.A.	N.A.	UK
1	c.40G>T	p.E14X	No	-	Ctrl	M	69	N.A.	N.A.	PT
1	c.126C>G	p.T42T	No	-	Ctrl	F	70	N.A.	N.A.	PT
2	c.132G>A	p.A44A	No	-	AD	M	76	73	Pos	PT
3	c.279C>T	p.Y93Y	rs9331898	-	MCI	F	78	-	-	PT
3	c.284A>G	p.N95S	No	Benign	Ctrl	M	59	N.A.	N.A.	PT
3	c.348C>T	p.N116N	No	-	AD	F	59	59	Neg	PT
4	c.438G>A	p.K146K	No	-	Ctrl	F	69	N.A.	N.A.	PT
5	c.812_814delTCT	p.F272del	No	-	Ctrl	F	28	N.A.	N.A.	UK
5	c.879C>T	p.H293H	No	-	AD	F	73	72	-	PT
6	c.1013G>A	p.R338Q	No	Benign	AD	F	51	-	-	PT

Nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the reference sequences (CLU: NM_001831.2). The initiation codon is codon 1. Genomic numbering refers to reference sequence NC_000008.9. Protein numbering refers to sequence NP_001822.2. AAO: age at onset; Fam hist: Family history; N.A.: not applicable; -: information not available. Prediction of pathogenicity was performed in silico using the PolyPhen software [300, 301, 342].

These eleven variants include synonymous changes (c.126C>G, p.T42T; c.132G>A, p.A44A; c.279C>T, p.Y93Y; c.348C>T, p.N116N; c.438G>A, p.K146K; and c.879C>T, p.H293H), suggesting no functional change for the protein; non-synonymous variants observed in controls and AD patients (c.284A>G, p.N95S; c.764C>T, p.T255I and c.1013G>A, p.R338Q); a nonsense mutation identified in a control individual (c.40G>T, p.E14X); and an in-frame deletion (c.812_814delTCT, p.F272del) present in a control individual.

Ten variants were found in more than one individual in the Portuguese series and seven in the UK series. The genotypic frequencies for these variants were compared between cases and controls (Tables 24 and 25). From these, four variants were found in both cohorts (p.T255I, p.H315H, rs3216167 and p. D380D).

Only the rs3216167 polymorphism located in intron 6 presented a statistically significant difference (p=0.04) between cases and controls and only in the UK cohort. No statistical significant differences were observed for any of the other variants.

In order to predict the impact in protein function of the non-synonymous variants found in *CLU*, we performed an *in silico* analysis using PolyPhen (Tables 23-25). From the eight coding, missense and non-synonymous changes found in *CLU*, three were considered to possibly affect the protein function (p.R234H, p.P322L and p.N369H) while the other five were predicted to be benign. Nonetheless, these three variants are most likely non-pathogenic, since they have been found both in cases and healthy controls (Tables 24 and 25).

Table 24. Variants found in CLU observed in more than one individual in the Portuguese series.

Location in the gene	DNA change	Protein change	Present in dbSNP	Frequency affected	Frequency unaffected	P	OR	PolyPhen prediction of pathogenicity (PSIC score)
Exon 1	c.48C>A	p.S16R	No	0.01425	0.01878	0.5459	0.7517	Benign
Exon 2	c.240C>T	p.D80D	rs9331892	0.005181	0.007653	0.6074	0.6736	-
Exon 5	c.945T>C	p.H315H	rs7982	0.3575	0.3716	0.6296	0.9429	-
IVS6	g.27517690 -/A	NA	rs3216167	0.2642	0.2269	0.1449	1.233	-
Exon 7	c.1105A>C	p.N369H	rs9331936	0.008174	0.002358	0.2469	3.507	Possibly damaging (1.557)
Exon 7	c.1110C>G	p.P370P	rs9331937	0.001362	0.002358	0.6975	0.5765	-
Exon 7	c.1138G>A	p.D380N	rs9331938	0.002725	0.002358	0.9059	1.156	Benign (0.322)
Exon 7	c.1140C>T	p.D380D	rs9331939	0.001362	0.004717	0.3092	0.2869	-
3'UTR	g.27511359 T/C	NA	rs3087554	0.1425	0.1302	0.5646	1.112	-
3'UTR	g.27511354 C/T	NA	no	0.007634	0.005208	0.5001	1.723	-

Nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the reference sequences (CLU: NM_001831.2). The initiation codon is codon 1. Genomic numbering refers to reference sequence NC_000008.9. Protein numbering refers to sequence NP_001822.2. Prediction of pathogenicity was performed in silico using the PolyPhen software [300, 301, 342].

Table 25. Variants found in *CLU* observed in more than one individual in the UK series.

Location in the gene	DNA change	Protein change	Present in dbSNP	Frequency affected	Frequency unaffected	P	OR	PolyPhen prediction of pathogenicity (PSIC score)
5'UTR	c.-229G>C	NA	no	0.34	0.27	0.11	1.50	-
Exon 5	c.701G>A	p.R234H	no	0.001	0.002	0.51	0.47	Possibly damaging (1.641)
Exon 5	c.764C>T	p.T255I	rs41276297	0.003	0.006	0.35	0.53	Benign (0.310)
Exon 5	c.945T>C	p.H315H	rs7982	0.37	0.40	0.41	0.93	-
Exon 5	c.965T>C	p.P322L	no	0.002	0.002	0.73	1.42	Possibly damaging (1.963)
IVS6	g.27517690 -/A	NA	rs3216167	0.34	0.25	0.04	1.68	-
Exon 7	c.1140C>T	p.D380D	rs9331939	0.01	0.02	0.68	0.68	-

Nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the reference sequences (CLU: NM_001831.2). The initiation codon is codon 1. Genomic numbering refers to reference sequence NC_000008.9. Protein numbering refers to sequence NP_001822.2. p values presented here are uncorrected for multiple testing (none is statistically significant after Bonferronni correction).Prediction of pathogenicity was performed in silico using the PolyPhen software [300, 301, 342].

Exons 5 (where the common H315H polymorphism is located) and 6 of *CLU* (where a non-synonymous change, p.R338Q, was found in a patient and was absent from controls) were screened in an additional set of 200 control individuals from Portugal. p.R338Q was not found in these additional 200 control samples.

eQTL Analysis

eQTL analysis did not reveal any statistically significant results within the immediate region of AD associated loci for *CLU* (Figure 28).

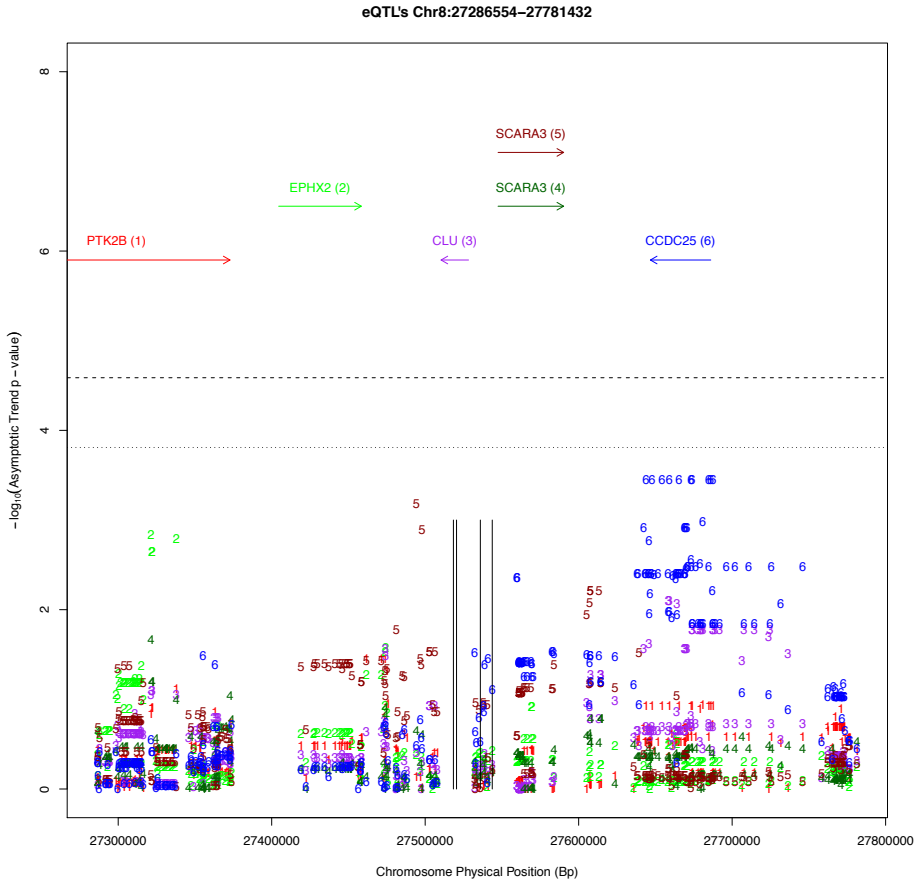


Figure 28. Manhattan plots for eQTL p-values +/- 250Kb of the previously AD associated regions near CLU. For this plot the x-axis represents the physical region of the chromosome and the y-axis is the –log10 of the asymptotic p-values from the eQTL analyses. The horizontal black dashed line represent the statistically significant threshold based on a Bonferonni correction for the number of SNPs and mRNA transcripts tested, while the black dotted represent the suggestive threshold based on the average number of SNPs tested per transcript. The relative positions of previously disease associated SNPs are denoted by vertical black lines. The individual p-values points for the SNP/Transcript tests are indicated by color and number so that the most significant values are the dark red '5's for SCARA3 and the blue '6's for CCDC25. In the plots the transcripts depicted are not all mRNA transcripts within the region but are instead the transcripts within the region with a probe present on the expression array and was well detected within the sample series.

DISCUSSION

We sequenced *CLU* in 495 AD patients and 330 controls. This approach allowed us to find several new variants and, for some changes, compare the genotypic and allelic frequencies between cases and controls.

Our results, despite non-significant, represent an association in the same direction of that found for rs7982 (p.H315H) by Harold et al. The minor allele frequencies (MAF) for this variant observed in the Portuguese and UK control series did not differ significantly from those described by Harold et al (MAF in Portuguese controls: 0.37; MAF in UK controls: 0.40; MAF in Harold et al: 0.40). The inexistence of statistical significance is most likely due to the small number of samples studied here, considering the marginal odds ratios associated with this variant (meta-analysis of the combined sample OR=0.86, based on partially imputed genotypes in Harold et al) we only achieved a 18 and 23% power to detect an effect of this magnitude with the Portuguese and UK cohorts, respectively [104].

From the 14 variants found in more than one individual only one (rs3216167) was found to be significantly associated with the development of AD in the UK cohort (p=0.04, OR=1.68, Table 25). A meta-analysis between the two cohorts also revealed a statistically significant p value (p=0.03 for random-effects meta-analysis, data not shown). rs3216167 is a common SNP located in intron 6 that has previously been reported by Miwa Y. and colleagues to be the only of 11 *CLU* SNPs to be associated with serum levels of total and LDL cholesterol, intima-media complex thickness of the common carotid artery and plaque prevalence in Japanese hypertensive females [349].

The sequencing of the gene allowed us to search for mutations that could be the cause of the disease in EOAD cases. This is clearly of importance, since in many cases, common genetic variation associated with late onset common disease has been found in the same genes where mutations are

known to cause early onset familial forms of the disease [350, 351]. A total of twenty-four variants was found, from which four were non-synonymous coding changes present in only one subject. Three of these changes (p.E14X, p.N95S and p.F272del) were found in controls and one (p.R338Q) was found in an early onset female AD patient. The screening of this variant in 200 additional controls revealed a total of 435 healthy age matched individuals not harboring the p.R338Q change. Nonetheless, the lack of segregation data, the nonexistence of functional data for this variant and the *in silico* prediction by PolyPhen of non-pathogenicity, support a benign outcome for this genetic change in AD pathogenicity.

From the eight coding, non-synonymous changes found, three were predicted by PolyPhen to be possibly damaging (p.R234H, p.P322L and p.N369H). All of these were present in cases and controls, indicating that most probably these are also non-pathogenic variants. The finding of variants in controls, specially the finding of a nonsense mutation in a 69 years old healthy individual, clearly adds to the requirement of an extended evaluation of the pathogenicity of each variant.

The sequencing approach used in this study, allowed us to compare not only one of the variants previously associated with AD (p.H315H) [104], but also, all the coding variants, the changes located in the exon-intron boundaries and in the 5' and 3' untranslated regions of *CLU*. Furthermore, since in our series no clear pathogenicity could be attributed to the variants found, we speculated if this variability could be affecting the expression of the gene. In fact, many GWAS in different disorders have resulted in the identification of non-coding variants, many times located in gene deserts. In these cases, a still poorly understood putative role in gene expression is usually pointed as the possible pathogenic mechanism [178]. In order to dissect this hypothesis, we performed an eQTL analysis of genotype and gene expression. Since no gene expression data was available from the populations in which the sequencing analysis was performed and the results from Harold et al suggest that CLU is associated with AD across populations; and also because this common variation would be present in controls, the eQTL analysis was performed on data from 174 neurologically

normal controls from the United States. No significant differences were found for *cis* SNPs with expression levels. A previous *cis* and *trans* analysis of these cortical signals also did not reveal significant eQTL signals [344], which is consistent with lack of eQTL signals for immediate region in human liver [352] and HapMap lymphoblast [353, 354].

These findings show an absence of common coding variants in *CLU* able to explain the previously established association with AD and the nonexistence of clear pathogenic mutations underlying the disease in early onset familial cases, in our sample. Also, the eQTL analysis of genotype and gene expression from neurological normal controls revealed no significant eQTL signals.

CHAPTER 7

CONCLUSIONS

CHAPTER 7 - CONCLUSIONS

In the first part of this work, the molecular characterization of different AD and FTLD samples lead us to the identification of several new and already known mutations.

- In AD, the screening for mutations in the coding regions of *PSENs*, as well as exons 16 and 17 of *APP* in EOAD patients from the Iberian Peninsula identified three novel mutations in *PSEN1*, one novel mutation in *PSEN2*, and a novel mutation in *APP*. Four previously described mutations in *PSEN1* were also found.
- The screening of a large dementia cohort for *APP* duplications showed that five individuals harbored *APP* duplications, being this, a similar prevalence to the one observed for *APP* missense mutations. *APP* duplications were found to be a significant cause of early onset dementia in the UK. Also, based on the clinical findings, the recognized phenotype was proposed to be expanded in order to include early seizures and apparently sporadic disease.
- In FTLD the molecular screening of *MAPT* and *GRN* in a large cohort from the UK allowed the study of heritability of the different clinical subtypes of this disorder, revealing that FTLD is a highly heritable disorder but that heritability varies between the different syndromes with the behavioral variant being the most heritable and FTD-MND and the language syndromes (particularly semantic dementia) the least heritable. Mutations were found in *MAPT* (8.9% of the cohort) and *GRN* (8.4%) and of the remaining patients without mutations but with a strong family history; seven had pathological confirmation, falling into two groups: type 3 FTLD-TDP without *GRN* mutations (six) and FTLD-UPS (one). *MAPT* and *GRN* mutations were shown to account for a substantial proportion of familial cases; nonetheless, other genes remain to be discovered, particularly in patients with type 3 FTLD-TDP without a *GRN* mutation.
- The sequencing of *GRN* in a consecutive series of 46 FTLD/CBS Portuguese patients lead us to the identification of two mutations: a novel pathogenic insertion (p.Gln300GlnfsX61)

found in a family presenting with corticobasal syndrome, and a previously described point variant (p.T182M) of unclear pathogenicity. Pathogenic mutations in the *GRN* gene were found in one of the 36 probands studied (3% of the probands in our series), indicating that in the Portuguese population, mutations in this gene are not a major cause of FTLD.

In order to evaluate the pathogenicity of several of the variants found to be associated with AD and FTLD in the studies of the first part of this work (chapter4), the same type of analysis was carried in control individuals.

- To determine the extent of normal variability in *APP*, *PSEN1* and *PSEN2*, these genes were sequenced in controls from the Iberian peninsular, and in individuals from seven African populations. In the latter series, we found five new non-synonymous changes in all three genes and a presenilin 2 variant (R62H) that had been previously related to AD.
- For some of the mutations found in AD cases the pathologic consequence continued uncertain and to address this issue we proposed and used a systematic algorithm to classify the putative pathology of AD mutations.
- The sequencing of *GRN* and *MAPT* in samples from the Centre d'Etude du Polymorphisme Humain- Human Genome Diversity Cell Line Panel, identified sixteen different non-synonymous changes, eleven of which were novel variants that could easily be mistaken as pathogenic mutations in future studies.

Overall, these results stressed the need for a coordinated effort to resequence the genes involved in neurological disorders in large worldwide samples.

In the past decade, a major problem in the genetics of AD, has been the inability to find new genes associated either with the early onset form of the disease, or with late onset AD.

To address this problem and in order to identify new loci and genes involved in the pathogenesis of AD, we assessed the possible role of homozygosity in late and early onset AD and participated in the most powerful GWAS in AD to date.

- The comparison of measures of extended homozygosity in a population of 837 late-onset Alzheimer's disease cases and 550 controls lead us to the conclusion that specific large tracts of homozygosity are a risk factor for Alzheimer's disease. We identified one homozygous region on chromosome 8 significantly associated with LOAD containing seven genes from which the most biologically plausible candidates to be involved in AD pathogenesis are *STAR*, *EIF4EBP1*, and *ADRB3*. These results clearly suggested a recessive component to the etiology of LOAD.
- In EOAD we studied the clinical characteristics of an Israeli family comprising two affected siblings born to neurologically healthy parents who were first cousins. As the disease in this family was consistent with an autosomal recessive mode of inheritance, we identified all homozygous regions identical by descent in both siblings, by high-density SNP genotyping and consequently provided the first catalog of autozygosity in EOAD. These results suggested that the regions identified are excellent candidate loci for a recessive genetic lesion causing this disease.

The results from these two studies, together with the speculation that recessive genes for AD are responsible for the high AD prevalence in the Wadi Ara [338], demonstrated that particular regions of homozygosity have a significant role in AD genetics. It is possible that future sequencing and additional follow-up analyses will allow the identification of small effect size recessive risk variant(s) within the described regions of homozygosity.

- The largest AD GWAS to date identified the first SNPs outside the *APOE* locus to reach genome wide significance. SNPs at two loci not previously associated with the disease: at the *CLU* gene (rs11136000, P = 1.4 ×10−9) and 5' to the PICALM gene (rs3851179, P = 1.9 × 10−8)

were observed to be associated with the development of AD. These associations were replicated in stage 2, producing compelling evidence for association with Alzheimer's disease in the combined dataset (rs11136000, P = 8.5 × 10−10, odds ratio = 0.86; rs3851179, P = 1.3 × 10−9, odds ratio = 0.86). To further characterize the strongest association found we sequenced the coding region of *CLU* in a total of 495 AD cases and 330 healthy controls. Coding variants were found in both controls and cases (including a nonsense mutation in a healthy subject), indicating that the pathogenicity of variants found in this gene must be carefully evaluated. Additionally, no significant eQTL associations were observed for the SNPs previously associated with AD, confirming that common variants at the CLU locus do not effect expression of nearby (cis) mRNA transcripts.

In summary, this work allowed: 1) the molecular characterization of AD and FTLD in several subjects and families; 2) the recognition of how crucial is to critically evaluate all the genetic variants found in the dementia related genes; and 3) the identification of putative loci associated with late and early onset AD, by assessing the role of homozygosity in the disease and, by collaborating in a large consortium, the identification of the first two genes significantly associated with sporadic LOAD, after *APOE*.

CHAPTER 8

REFERENCES

CHAPTER 8 - REFERENCES

1. Rademakers, R., M. Cruts, and C. Van Broeckhoven, *Genetics of early-onset Alzheimer dementia.* ScientificWorldJournal, 2003. **3**: p. 497-519.
2. Rogaeva, E., *The solved and unsolved mysteries of the genetics of early-onset Alzheimer's disease.* Neuromolecular Med, 2002. **2**(1): p. 1-10.
3. Daw, E.W., et al., *The number of trait loci in late-onset Alzheimer disease.* Am J Hum Genet, 2000. **66**(1): p. 196-204.
4. Tanzi, R.E., et al., *Amyloid beta protein gene: cDNA, mRNA distribution, and genetic linkage near the Alzheimer locus.* Science, 1987. **235**(4791): p. 880-4.
5. Yoshikai, S., et al., *Genomic organization of the human amyloid beta-protein precursor gene.* Gene, 1990. **87**(2): p. 257-63.
6. Nikolaev, A., et al., *APP binds DR6 to trigger axon pruning and neuron death via distinct caspases.* Nature, 2009. **457**(7232): p. 981-9.
7. Hardy, J. and D. Allsop, *Amyloid deposition as the central event in the aetiology of Alzheimer's disease.* Trends Pharmacol Sci, 1991. **12**(10): p. 383-8.
8. Hardy, J., *Amyloid, the presenilins and Alzheimer's disease.* Trends Neurosci, 1997. **20**(4): p. 154-9.
9. Raux, G., et al., *Molecular diagnosis of autosomal dominant early onset Alzheimer's disease: an update.* J Med Genet, 2005. **42**(10): p. 793-5.
10. Levy, E., et al., *Mutation of the Alzheimer's disease amyloid gene in hereditary cerebral hemorrhage, Dutch type.* Science, 1990. **248**(4959): p. 1124-6.
11. Goate, A., et al., *Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease.* Nature, 1991. **349**(6311): p. 704-6.
12. Di Fede, G., et al., *A recessive mutation in the APP gene with dominant-negative effect on amyloidogenesis.* Science, 2009. **323**(5920): p. 1473-7.
13. Cabrejo, L., et al., *Phenotype associated with APP duplication in five families.* Brain, 2006. **129**(Pt 11): p. 2966-76.
14. Rovelet-Lecrux, A., et al., *APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy.* Nat Genet, 2006. **38**(1): p. 24-6.
15. Remes, A.M., et al., *Hereditary dementia with intracerebral hemorrhages and cerebral amyloid angiopathy.* Neurology, 2004. **63**(2): p. 234-40.
16. Rovelet-Lecrux, A., et al., *APP locus duplication in a Finnish family with dementia and intracerebral haemorrhage.* J Neurol Neurosurg Psychiatry, 2007. **78**(10): p. 1158-9.
17. Sleegers, K., et al., *APP duplication is sufficient to cause early onset Alzheimer's dementia with cerebral amyloid angiopathy.* Brain, 2006. **129**(Pt 11): p. 2977-83.
18. Rogaev, E.I., et al., *Analysis of the 5' sequence, genomic structure, and alternative splicing of the presenilin-1 gene (PSEN1) associated with early onset Alzheimer disease.* Genomics, 1997. **40**(3): p. 415-24.
19. Levy-Lahad, E., et al., *Genomic structure and expression of STM2, the chromosome 1 familial Alzheimer disease gene.* Genomics, 1996. **34**(2): p. 198-204.
20. Rogaev, E.I., et al., *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, 1995. **376**(6543): p. 775-8.
21. Henricson, A., L. Kall, and E.L. Sonnhammer, *A novel transmembrane topology of presenilin based on reconciling experimental and computational evidence.* Febs J, 2005. **272**(11): p. 2727-33.
22. Laudon, H., et al., *A nine-transmembrane domain topology for presenilin 1.* J Biol Chem, 2005. **280**(42): p. 35352-60.

23. Kovacs, D.M., et al., *Alzheimer-associated presenilins 1 and 2: neuronal expression in brain and localization to intracellular membranes in mammalian cells.* Nat Med, 1996. **2**(2): p. 224-9.
24. De Strooper, B., *Aph-1, Pen-2, and Nicastrin with Presenilin generate an active gamma-Secretase complex.* Neuron, 2003. **38**(1): p. 9-12.
25. De Strooper, B., et al., *A presenilin-1-dependent gamma-secretase-like protease mediates release of Notch intracellular domain.* Nature, 1999. **398**(6727): p. 518-22.
26. Zhang, Z., et al., *Destabilization of beta-catenin by mutations in presenilin-1 potentiates neuronal apoptosis.* Nature, 1998. **395**(6703): p. 698-702.
27. Baki, L., et al., *Presenilin-1 binds cytoplasmic epithelial cadherin, inhibits cadherin/p120 association, and regulates stability and function of the cadherin/catenin adhesion complex.* Proc Natl Acad Sci U S A, 2001. **98**(5): p. 2381-6.
28. Smine, A., et al., *Regulation of brain G-protein go by Alzheimer's disease gene presenilin-1.* J Biol Chem, 1998. **273**(26): p. 16281-8.
29. Katayama, T., et al., *Presenilin-1 mutations downregulate the signalling pathway of the unfolded-protein response.* Nat Cell Biol, 1999. **1**(8): p. 479-85.
30. Niwa, M., et al., *A role for presenilin-1 in nuclear accumulation of Ire1 fragments and induction of the mammalian unfolded protein response.* Cell, 1999. **99**(7): p. 691-702.
31. Sherrington, R., et al., *Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease.* Nature, 1995. **375**(6534): p. 754-60.
32. Guerreiro, R.J., et al., *Genetic screening of Alzheimer's disease genes in Iberian and African samples yields novel mutations in presenilins and APP.* Neurobiol Aging, 2008: p. in press.
33. Hardy, J. and R. Crook, *Presenilin mutations line up along transmembrane alpha-helices.* Neurosci Lett, 2001. **306**(3): p. 203-5.
34. Scheuner, D., et al., *Secreted amyloid beta-protein similar to that in the senile plaques of Alzheimer's disease is increased in vivo by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease.* Nat Med, 1996. **2**(8): p. 864-70.
35. Lemere, C.A., et al., *The E280A presenilin 1 Alzheimer mutation produces increased A beta 42 deposition and severe cerebellar pathology.* Nat Med, 1996. **2**(10): p. 1146-50.
36. Duff, K., et al., *Increased amyloid-beta42(43) in brains of mice expressing mutant presenilin 1.* Nature, 1996. **383**(6602): p. 710-3.
37. Kumar-Singh, S., et al., *Mean age-of-onset of familial alzheimer disease caused by presenilin mutations correlates with both increased Abeta42 and decreased Abeta40.* Hum Mutat, 2006. **27**(7): p. 686-95.
38. Siest, G., et al., *Apolipoprotein E: an important gene and protein to follow in laboratory medicine.* Clin Chem, 1995. **41**(8 Pt 1): p. 1068-86.
39. Mahley, R.W., *Apolipoprotein E: cholesterol transport protein with expanding role in cell biology.* Science, 1988. **240**(4852): p. 622-30.
40. Mahley, R.W. and S.C. Rall, Jr., *Apolipoprotein E: far more than a lipid transport protein.* Annu Rev Genomics Hum Genet, 2000. **1**: p. 507-37.
41. Zannis, V.I., P.W. Just, and J.L. Breslow, *Human apolipoprotein E isoprotein subclasses are genetically determined.* Am J Hum Genet, 1981. **33**(1): p. 11-24.
42. Corbo, R.M. and R. Scacchi, *Apolipoprotein E (APOE) allele distribution in the world. Is APOE*4 a 'thrifty' allele?* Ann Hum Genet, 1999. **63**(Pt 4): p. 301-10.
43. Corder, E.H., et al., *Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families.* Science, 1993. **261**(5123): p. 921-3.
44. Corder, E.H., et al., *Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease.* Nat Genet, 1994. **7**(2): p. 180-4.
45. Strittmatter, W.J., et al., *Binding of human apolipoprotein E to synthetic amyloid beta peptide: isoform-specific effects and implications for late-onset Alzheimer disease.* Proc Natl Acad Sci U S A, 1993. **90**(17): p. 8098-102.

46. Strittmatter, W.J., et al., *Hypothesis: microtubule instability and paired helical filament formation in the Alzheimer disease brain are related to apolipoprotein E genotype.* Exp Neurol, 1994. **125**(2): p. 163-71; discussion 172-4.
47. Sanan, D.A., et al., *Apolipoprotein E associates with beta amyloid peptide of Alzheimer's disease to form novel monofibrils. Isoform apoE4 associates more efficiently than apoE3.* J Clin Invest, 1994. **94**(2): p. 860-9.
48. Weisgraber, K.H., *Apolipoprotein E: structure-function relationships.* Adv Protein Chem, 1994. **45**: p. 249-302.
49. Hyman, B.T., et al., *Apolipoprotein E epsilon4/4 in a neuropathologically normal very elderly individual.* Arch Neurol, 1996. **53**(3): p. 215.
50. Bekris, L.M., et al., *APOE mRNA and protein expression in postmortem brain are modulated by an extended haplotype structure.* Am J Med Genet B Neuropsychiatr Genet, 2009.
51. Chartier-Harlin, M.C., et al., *Apolipoprotein E, epsilon 4 allele as a major risk factor for sporadic early and late-onset forms of Alzheimer's disease: analysis of the 19q13.2 chromosomal region.* Hum Mol Genet, 1994. **3**(4): p. 569-74.
52. Lambert, J.C., et al., *Contribution of APOE promoter polymorphisms to Alzheimer's disease risk.* Neurology, 2002. **59**(1): p. 59-66.
53. Lambert, J.C., et al., *Distortion of allelic expression of apolipoprotein E in Alzheimer's disease.* Hum Mol Genet, 1997. **6**(12): p. 2151-4.
54. Dawn Teare, M. and J.H. Barrett, *Genetic linkage studies.* Lancet, 2005. **366**(9490): p. 1036-44.
55. Altshuler, D., M.J. Daly, and E.S. Lander, *Genetic mapping in human disease.* Science, 2008. **322**(5903): p. 881-8.
56. Butler, A.W., et al., *Meta-analysis of linkage studies for Alzheimer's disease--a web resource.* Neurobiol Aging, 2009. **30**(7): p. 1037-47.
57. Liu, F., et al., *A genomewide screen for late-onset Alzheimer disease in a genetically isolated Dutch population.* Am J Hum Genet, 2007. **81**(1): p. 17-31.
58. Blacker, D., et al., *Results of a high-resolution genome screen of 437 Alzheimer's disease families.* Hum Mol Genet, 2003. **12**(1): p. 23-32.
59. Pericak-Vance, M.A., et al., *Identification of novel genes in late-onset Alzheimer's disease.* Exp Gerontol, 2000. **35**(9-10): p. 1343-52.
60. Curtis, D., B.V. North, and P.C. Sham, *A novel method of two-locus linkage analysis applied to a genome scan for late onset Alzheimer's disease.* Ann Hum Genet, 2001. **65**(Pt 5): p. 473-81.
61. Olson, J.M., K.A. Goddard, and D.M. Dudek, *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, 2002. **71**(1): p. 154-61.
62. Myers, A., et al., *Full genome screen for Alzheimer disease: stage II analysis.* Am J Med Genet, 2002. **114**(2): p. 235-44.
63. Lee, J.H., et al., *Expanded genomewide scan implicates a novel locus at 3q28 among Caribbean hispanics with familial Alzheimer disease.* Arch Neurol, 2006. **63**(11): p. 1591-8.
64. Lee, J.H., et al., *Analyses of the National Institute on Aging Late-Onset Alzheimer's Disease Family Study: implication of additional loci.* Arch Neurol, 2008. **65**(11): p. 1518-26.
65. Scott, W.K., et al., *Ordered-subsets linkage analysis detects novel Alzheimer disease loci on chromosomes 2q34 and 15q22.* Am J Hum Genet, 2003. **73**(5): p. 1041-51.
66. Hamshere, M.L., et al., *Genome-wide linkage analysis of 723 affected relative pairs with late-onset Alzheimer's disease.* Hum Mol Genet, 2007. **16**(22): p. 2703-12.
67. Holmans, P., et al., *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, 2005. **135B**(1): p. 24-32.
68. Pericak-Vance, M.A., et al., *Complete genomic screen in late-onset familial Alzheimer disease. Evidence for a new locus on chromosome 12.* Jama, 1997. **278**(15): p. 1237-41.

69. Li, Y.J., et al., *Age at onset in two common neurodegenerative diseases is genetically controlled.* Am J Hum Genet, 2002. **70**(4): p. 985-93.
70. Sillen, A., et al., *Expanded high-resolution genetic study of 109 Swedish families with Alzheimer's disease.* Eur J Hum Genet, 2008. **16**(2): p. 202-8.
71. Bertram, L., et al., *Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database.* Nat Genet, 2007. **39**(1): p. 17-23.
72. Kehoe, P., et al., *A full genome scan for late onset Alzheimer's disease.* Hum Mol Genet, 1999. **8**(2): p. 237-45.
73. Lee, J.H., et al., *Fine mapping of 10q and 18q for familial Alzheimer's disease in Caribbean Hispanics.* Mol Psychiatry, 2004. **9**(11): p. 1042-51.
74. Bacanu, S.A., et al., *Linkage analysis of Alzheimer disease with psychosis.* Neurology, 2002. **59**(1): p. 118-20.
75. Avramopoulos, D., M.D. Fallin, and S.S. Bassett, *Linkage to chromosome 14q in Alzheimer's disease (AD) patients without psychotic symptoms.* Am J Med Genet B Neuropsychiatr Genet, 2005. **132B**(1): p. 9-13.
76. Hollingworth, P., et al., *Increased familial risk and genomewide significant linkage for Alzheimer's disease with psychosis.* Am J Med Genet B Neuropsychiatr Genet, 2007. **144B**(7): p. 841-8.
77. Bassett, S.S., D. Avramopoulos, and D. Fallin, *Evidence for parent of origin effect in late-onset Alzheimer disease.* Am J Med Genet, 2002. **114**(6): p. 679-86.
78. Ashley-Koch, A.E., et al., *An autosomal genomic screen for dementia in an extended Amish family.* Neurosci Lett, 2005. **379**(3): p. 199-204.
79. Hahs, D.W., et al., *A genome-wide linkage analysis of dementia in the Amish.* Am J Med Genet B Neuropsychiatr Genet, 2006. **141B**(2): p. 160-6.
80. Hattersley, A.T. and M.I. McCarthy, *What makes a good genetic association study?* Lancet, 2005. **366**(9493): p. 1315-23.
81. Ioannidis, J.P., et al., *Replication validity of genetic association studies.* Nat Genet, 2001. **29**(3): p. 306-9.
82. Choudhry, S., et al., *Population stratification confounds genetic association studies among Latinos.* Hum Genet, 2006. **118**(5): p. 652-64.
83. Bertram, L. and R.E. Tanzi, *Thirty years of Alzheimer's disease genetics: the implications of systematic meta-analyses.* Nat Rev Neurosci, 2008. **9**(10): p. 768-78.
84. Webster J, et al., *Whole genome association analysis shows that ACE is a risk factor for Alzheimer's disease and fails to replicate most candidates from Meta-analysis.* Int J Mol Epidemiol Genet 2009: p. (in press).
85. Sayed-Tabatabaei, F.A., et al., *Angiotensin-converting enzyme gene polymorphism and carotid artery wall thickness: a meta-analysis.* Stroke, 2003. **34**(7): p. 1634-9.
86. Staessen, J.A., et al., *The deletion/insertion polymorphism of the angiotensin converting enzyme gene and cardiovascular-renal risk.* J Hypertens, 1997. **15**(12 Pt 2): p. 1579-92.
87. Consortium, I.H., *A haplotype map of the human genome.* Nature, 2005. **437**(7063): p. 1299-320.
88. Chakravarti, A., *Population genetics--making sense out of sequence.* Nat Genet, 1999. **21**(1 Suppl): p. 56-60.
89. Cargill, M., et al., *Characterization of single-nucleotide polymorphisms in coding regions of human genes.* Nat Genet, 1999. **22**(3): p. 231-8.
90. Lander, E.S., *The new genomics: global views of biology.* Science, 1996. **274**(5287): p. 536-9.
91. Simon-Sanchez, J. and A. Singleton, *Genome-wide association studies in neurological disorders.* Lancet Neurol, 2008. **7**(11): p. 1067-72.
92. Gatz, M., et al., *Role of genes and environments for explaining Alzheimer disease.* Arch Gen Psychiatry, 2006. **63**(2): p. 168-74.

93. Pedersen, N.L., S.F. Posner, and M. Gatz, *Multiple-threshold models for genetic influences on age of onset for Alzheimer disease: findings in Swedish twins.* Am J Med Genet, 2001. **105**(8): p. 724-8.
94. Grupe, A., et al., *Evidence for novel susceptibility genes for late-onset Alzheimer's disease from a genome-wide association study of putative functional variants.* Hum Mol Genet, 2007. **16**(8): p. 865-73.
95. Coon, K.D., et al., *A high-density whole-genome association study reveals that APOE is the major susceptibility gene for sporadic late-onset Alzheimer's disease.* J Clin Psychiatry, 2007. **68**(4): p. 613-8.
96. Reiman, E.M., et al., *GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers.* Neuron, 2007. **54**(5): p. 713-20.
97. Abraham, R., et al., *A genome-wide association study for late-onset Alzheimer's disease using DNA pooling.* BMC Med Genomics, 2008. **1**: p. 44.
98. Li, H., et al., *Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease.* Arch Neurol, 2008. **65**(1): p. 45-53.
99. Bertram, L., et al., *Genome-wide association analysis reveals putative Alzheimer's disease susceptibility loci in addition to APOE.* Am J Hum Genet, 2008. **83**(5): p. 623-32.
100. Beecham, G.W., et al., *Genome-wide association study implicates a chromosome 12 risk locus for late-onset Alzheimer disease.* Am J Hum Genet, 2009. **84**(1): p. 35-43.
101. Carrasquillo, M.M., et al., *Genetic variation in PCDH11X is associated with susceptibility to late-onset Alzheimer's disease.* Nat Genet, 2009. **41**(2): p. 192-8.
102. Feulner, T.M., et al., *Examination of the current top candidate genes for AD in a genome-wide association study.* Mol Psychiatry, 2009.
103. Rubinsztein, D.C. and D.F. Easton, *Apolipoprotein E genetic variation and Alzheimer's disease. a meta-analysis.* Dement Geriatr Cogn Disord, 1999. **10**(3): p. 199-209.
104. Harold, D., et al., *Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease.* Nat Genet, 2009. **41**(10): p. 1088-93.
105. Lambert, J.C., et al., *Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease.* Nat Genet, 2009. **41**(10): p. 1094-9.
106. de Silva, H.V., et al., *Apolipoprotein J: structure and tissue distribution.* Biochemistry, 1990. **29**(22): p. 5380-9.
107. Jones, S.E. and C. Jomary, *Clusterin.* Int J Biochem Cell Biol, 2002. **34**(5): p. 427-31.
108. Danik, M., et al., *Localization of sulfated glycoprotein-2/clusterin mRNA in the rat brain by in situ hybridization.* J Comp Neurol, 1993. **334**(2): p. 209-27.
109. Wong, P., et al., *Genomic organization and expression of the rat TRPM-2 (clusterin) gene, a gene implicated in apoptosis.* J Biol Chem, 1993. **268**(7): p. 5021-31.
110. Jenne, D.E. and J. Tschopp, *Clusterin: the intriguing guises of a widely expressed glycoprotein.* Trends Biochem Sci, 1992. **17**(4): p. 154-9.
111. Oda, T., et al., *Purification and characterization of brain clusterin.* Biochem Biophys Res Commun, 1994. **204**(3): p. 1131-6.
112. Wilson, M.R. and S.B. Easterbrook-Smith, *Clusterin is a secreted mammalian chaperone.* Trends Biochem Sci, 2000. **25**(3): p. 95-8.
113. Michel, D., et al., *Stress-induced transcription of the clusterin/apoJ gene.* Biochem J, 1997. **328 (Pt 1)**: p. 45-50.
114. Duguid, J.R., et al., *Changes in brain gene expression shared by scrapie and Alzheimer disease.* Proc Natl Acad Sci U S A, 1989. **86**(18): p. 7260-4.
115. May, P.C., et al., *Dynamics of gene expression for a hippocampal glycoprotein elevated in Alzheimer's disease and in response to experimental lesions in rat.* Neuron, 1990. **5**(6): p. 831-9.
116. McGeer, P.L., T. Kawamata, and D.G. Walker, *Distribution of clusterin in Alzheimer brain tissue.* Brain Res, 1992. **579**(2): p. 337-41.

117. Takamaru, Y., *[Analysis of the 80 kDa antigen in senile plaque amyloid].* Hokkaido Igaku Zasshi, 1994. **69**(2): p. 191-201.
118. McGeer, P.L., et al., *Pathological proteins in senile plaques.* Tohoku J Exp Med, 1994. **174**(3): p. 269-77.
119. Kida, E., N.H. Choi-Miura, and K.E. Wisniewski, *Deposition of apolipoproteins E and J in senile plaques is topographically determined in both Alzheimer's disease and Down's syndrome brain.* Brain Res, 1995. **685**(1-2): p. 211-6.
120. Ghiso, J., et al., *The cerebrospinal-fluid soluble form of Alzheimer's amyloid beta is complexed to SP-40,40 (apolipoprotein J), an inhibitor of the complement membrane-attack complex.* Biochem J, 1993. **293 (Pt 1)**: p. 27-30.
121. Golabek, A., et al., *Amyloid beta binding proteins in vitro and in normal human cerebrospinal fluid.* Neurosci Lett, 1995. **191**(1-2): p. 79-82.
122. Zlokovic, B.V., *Cerebrovascular transport of Alzheimer's amyloid beta and apolipoproteins J and E: possible anti-amyloidogenic role of the blood-brain barrier.* Life Sci, 1996. **59**(18): p. 1483-97.
123. Bertrand, P., et al., *Association of apolipoprotein E genotype with brain levels of apolipoprotein E and apolipoprotein J (clusterin) in Alzheimer disease.* Brain Res Mol Brain Res, 1995. **33**(1): p. 174-8.
124. Oda, T., et al., *Clusterin (apoJ) alters the aggregation of amyloid beta-peptide (A beta 1-42) and forms slowly sedimenting A beta complexes that cause oxidative stress.* Exp Neurol, 1995. **136**(1): p. 22-31.
125. Matsubara, E., et al., *Apolipoprotein J and Alzheimer's amyloid beta solubility.* Biochem J, 1996. **316 (Pt 2)**: p. 671-9.
126. Hammad, S.M., et al., *Interaction of apolipoprotein J-amyloid beta-peptide complex with low density lipoprotein receptor-related protein-2/megalin. A mechanism to prevent pathological accumulation of amyloid beta-peptide.* J Biol Chem, 1997. **272**(30): p. 18644-9.
127. Boggs, L.N., et al., *Clusterin (Apo J) protects against in vitro amyloid-beta (1-40) neurotoxicity.* J Neurochem, 1996. **67**(3): p. 1324-7.
128. DeMattos, R.B., et al., *Clusterin promotes amyloid plaque formation and is critical for neuritic toxicity in a mouse model of Alzheimer's disease.* Proc Natl Acad Sci U S A, 2002. **99**(16): p. 10843-8.
129. DeMattos, R.B., et al., *ApoE and clusterin cooperatively suppress Abeta levels and deposition: evidence that ApoE regulates extracellular Abeta metabolism in vivo.* Neuron, 2004. **41**(2): p. 193-202.
130. Bell, R.D., et al., *Transport pathways for clearance of human Alzheimer's amyloid beta-peptide and apolipoproteins E and J in the mouse central nervous system.* J Cereb Blood Flow Metab, 2007. **27**(5): p. 909-18.
131. Kim, J.A. and H.L. Kim, *Cell-free expression and functional reconstitution of CALM in clathrin assembly.* Exp Mol Med, 2001. **33**(2): p. 89-94.
132. Harel, A., et al., *Evidence for CALM in directing VAMP2 trafficking.* Traffic, 2008. **9**(3): p. 417-29.
133. Lambert, J.C., et al., *Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease.* Nat Genet, 2009.
134. Ahearn, J.M. and D.T. Fearon, *Structure and function of the complement receptors, CR1 (CD35) and CR2 (CD21).* Adv Immunol, 1989. **46**: p. 183-219.
135. Sjoberg, A.P., L.A. Trouw, and A.M. Blom, *Complement activation and inhibition: a delicate balance.* Trends Immunol, 2009. **30**(2): p. 83-90.
136. Akiyama, H., et al., *Inflammation and Alzheimer's disease.* Neurobiol Aging, 2000. **21**(3): p. 383-421.
137. Rogers, J., et al., *Complement activation by beta-amyloid in Alzheimer disease.* Proc Natl Acad Sci U S A, 1992. **89**(21): p. 10016-20.

138. Webster, S., C. Glabe, and J. Rogers, *Multivalent binding of complement protein C1Q to the amyloid beta-peptide (A beta) promotes the nucleation phase of A beta aggregation.* Biochem Biophys Res Commun, 1995. **217**(3): p. 869-75.
139. McGeer, P.L., et al., *Activation of the classical complement pathway in brain tissue of Alzheimer patients.* Neurosci Lett, 1989. **107**(1-3): p. 341-6.
140. Veerhuis, R., et al., *Early complement components in Alzheimer's disease brains.* Acta Neuropathol, 1996. **91**(1): p. 53-60.
141. Webster, S., et al., *Molecular and cellular characterization of the membrane attack complex, C5b-9, in Alzheimer's disease.* Neurobiol Aging, 1997. **18**(4): p. 415-21.
142. Walker, D.G. and P.L. McGeer, *Complement gene expression in human brain: comparison between normal and Alzheimer disease cases.* Brain Res Mol Brain Res, 1992. **14**(1-2): p. 109-16.
143. Bradt, B.M., W.P. Kolb, and N.R. Cooper, *Complement-dependent proinflammatory properties of the Alzheimer's disease beta-peptide.* J Exp Med, 1998. **188**(3): p. 431-8.
144. Strohmeyer, R., Y. Shen, and J. Rogers, *Detection of complement alternative pathway mRNA and proteins in the Alzheimer's disease brain.* Brain Res Mol Brain Res, 2000. **81**(1-2): p. 7-18.
145. McGeer, E.G. and P.L. McGeer, *Innate immunity in Alzheimer's disease: a model for local inflammatory reactions.* Mol Interv, 2001. **1**(1): p. 22-9.
146. Zipfel, P.F., *Complement and Immune Defense: From Innate Immunity to Human Diseases.* Immunol Lett, 2009.
147. Hof, P.R., P. Glannakopoulos, and C. Bouras, *The neuropathological changes associated with normal brain aging.* Histol Histopathol, 1996. **11**(4): p. 1075-88.
148. Tenner, A.J., *Complement in Alzheimer's disease: opportunities for modulating protective and pathogenic events.* Neurobiol Aging, 2001. **22**(6): p. 849-61.
149. McGeer, P.L. and E.G. McGeer, *The possible role of complement activation in Alzheimer disease.* Trends Mol Med, 2002. **8**(11): p. 519-23.
150. Gustafson D and S. I, *Control of vascular risk factors*, in *Vascular dementia in clinical practice*. 2009, Cambridge University Press: L-O Wahlund TE, S Gauthier ed.
151. Kannel, W.B., *Risk stratification in hypertension: new insights from the Framingham Study.* Am J Hypertens, 2000. **13**(1 Pt 2): p. 3S-10S.
152. Kivipelto, M., et al., *Hypertension and hypercholesterolaemia as risk factors for Alzheimer's disease: potential for pharmacological intervention.* CNS Drugs, 2002. **16**(7): p. 435-44.
153. Launer, L.J., et al., *Midlife blood pressure and dementia: the Honolulu-Asia aging study.* Neurobiol Aging, 2000. **21**(1): p. 49-55.
154. Qiu, C., B. Winblad, and L. Fratiglioni, *The age-dependent relation of blood pressure to cognitive function and dementia.* Lancet Neurol, 2005. **4**(8): p. 487-99.
155. Skoog, I., et al., *15-year longitudinal study of blood pressure and dementia.* Lancet, 1996. **347**(9009): p. 1141-5.
156. Stewart, R., et al., *Change in blood pressure and incident dementia: a 32-year prospective study.* Hypertension, 2009. **54**(2): p. 233-40.
157. Goossens, G.H., E.E. Blaak, and M.A. van Baak, *Possible involvement of the adipose tissue renin-angiotensin system in the pathophysiology of obesity and obesity-related disorders.* Obes Rev, 2003. **4**(1): p. 43-55.
158. Gustafson, D.R., et al., *The ACE Insertion Deletion polymorphism relates to dementia by metabolic phenotype, APOEvarepsilon4, and age of dementia onset.* Neurobiol Aging, 2008.
159. Katzov, H., et al., *A cladistic model of ACE sequence variation with implications for myocardial infarction, Alzheimer disease and obesity.* Hum Mol Genet, 2004. **13**(21): p. 2647-57.
160. Strazzullo, P., et al., *Genetic variation in the renin-angiotensin system and abdominal adiposity in men: the Olivetti Prospective Heart Study.* Ann Intern Med, 2003. **138**(1): p. 17-23.

161. Sparks, D.L., et al., *Induction of Alzheimer-like beta-amyloid immunoreactivity in the brains of rabbits with dietary cholesterol.* Exp Neurol, 1994. **126**(1): p. 88-94.
162. Nuutinen, T., et al., *Clusterin: A forgotten player in Alzheimer's disease.* Brain Res Rev, 2009.
163. Hoglund, K. and K. Blennow, *Effect of HMG-CoA reductase inhibitors on beta-amyloid peptide levels: implications for Alzheimer's disease.* CNS Drugs, 2007. **21**(6): p. 449-62.
164. Pangalos, M.N., S.J. Jacobsen, and P.H. Reinhart, *Disease modifying strategies for the treatment of Alzheimer's disease targeted at modulating levels of the beta-amyloid peptide.* Biochem Soc Trans, 2005. **33**(Pt 4): p. 553-8.
165. Notkola, I.L., et al., *Serum total cholesterol, apolipoprotein E epsilon 4 allele, and Alzheimer's disease.* Neuroepidemiology, 1998. **17**(1): p. 14-20.
166. Whitmer, R.A., et al., *Midlife cardiovascular risk factors and risk of dementia in late life.* Neurology, 2005. **64**(2): p. 277-81.
167. Mielke, M.M., et al., *High total cholesterol levels in late life associated with a reduced risk of dementia.* Neurology, 2005. **64**(10): p. 1689-95.
168. Reitz, C., et al., *Relation of plasma lipids to Alzheimer disease and vascular dementia.* Arch Neurol, 2004. **61**(5): p. 705-14.
169. Li, G., et al., *Serum cholesterol and risk of Alzheimer disease: a community-based cohort study.* Neurology, 2005. **65**(7): p. 1045-50.
170. Yoshitake, T., et al., *Incidence and risk factors of vascular dementia and Alzheimer's disease in a defined elderly Japanese population: the Hisayama Study.* Neurology, 1995. **45**(6): p. 1161-8.
171. Kalmijn, S., et al., *Metabolic cardiovascular syndrome and risk of dementia in Japanese-American elderly men. The Honolulu-Asia aging study.* Arterioscler Thromb Vasc Biol, 2000. **20**(10): p. 2255-60.
172. Stewart, R., et al., *Twenty-six-year change in total cholesterol levels and incident dementia: the Honolulu-Asia Aging Study.* Arch Neurol, 2007. **64**(1): p. 103-7.
173. Tan, Z.S., et al., *Plasma total cholesterol level as a risk factor for Alzheimer disease: the Framingham Study.* Arch Intern Med, 2003. **163**(9): p. 1053-7.
174. Hardy, J., *The amyloid hypothesis for Alzheimer's disease: a critical reappraisal.* J Neurochem, 2009. **110**(4): p. 1129-34.
175. Goldstein, D.B., *Common genetic variation and human traits.* N Engl J Med, 2009. **360**(17): p. 1696-8.
176. Hirschhorn, J.N., *Genomewide association studies--illuminating biologic pathways.* N Engl J Med, 2009. **360**(17): p. 1699-701.
177. Kraft, P. and D.J. Hunter, *Genetic risk prediction--are we there yet?* N Engl J Med, 2009. **360**(17): p. 1701-3.
178. Hardy, J. and A. Singleton, *Genomewide association studies and human disease.* N Engl J Med, 2009. **360**(17): p. 1759-68.
179. Polymeropoulos, M.H., et al., *Mutation in the alpha-synuclein gene identified in families with Parkinson's disease.* Science, 1997. **276**(5321): p. 2045-7.
180. Cairns, N.J., et al., *Neuropathologic diagnostic and nosologic criteria for frontotemporal lobar degeneration: consensus of the Consortium for Frontotemporal Lobar Degeneration.* Acta Neuropathol, 2007. **114**(1): p. 5-22.
181. Mackenzie, I.R. and R. Rademakers, *The molecular genetics and neuropathology of frontotemporal lobar degeneration: recent developments.* Neurogenetics, 2007. **8**(4): p. 237-48.
182. Graff-Radford, N.R. and B.K. Woodruff, *Frontotemporal dementia.* Semin Neurol, 2007. **27**(1): p. 48-57.
183. Gorno-Tempini, M.L., et al., *The logopenic/phonological variant of primary progressive aphasia.* Neurology, 2008. **71**(16): p. 1227-34.

184. Gorno-Tempini, M.L., et al., *Cognition and anatomy in three variants of primary progressive aphasia.* Ann Neurol, 2004. **55**(3): p. 335-46.
185. Kertesz, A., W. Davidson, and D.G. Munoz, *Clinical and pathological overlap between frontotemporal dementia, primary progressive aphasia and corticobasal degeneration: the Pick complex.* Dement Geriatr Cogn Disord, 1999. **10 Suppl 1**: p. 46-9.
186. Lomen-Hoerth, C., T. Anderson, and B. Miller, *The overlap of amyotrophic lateral sclerosis and frontotemporal dementia.* Neurology, 2002. **59**(7): p. 1077-9.
187. Chow, T.W., et al., *Inheritance of frontotemporal dementia.* Arch Neurol, 1999. **56**(7): p. 817-22.
188. Benajiba, L., et al., *TARDBP mutations in motoneuron disease with frontotemporal lobar degeneration.* Ann Neurol, 2009. **65**(4): p. 470-3.
189. Mackenzie, I.R., et al., *Nomenclature for neuropathologic subtypes of frontotemporal lobar degeneration: consensus recommendations.* Acta Neuropathol, 2009. **117**(1): p. 15-8.
190. Morris, H.R., et al., *The genetic and pathological classification of familial frontotemporal dementia.* Arch Neurol, 2001. **58**(11): p. 1813-6.
191. Hirokawa, N., *Microtubule organization and dynamics dependent on microtubule-associated proteins.* Curr Opin Cell Biol, 1994. **6**(1): p. 74-81.
192. Sato-Harada, R., et al., *Microtubule-associated proteins regulate microtubule function as the track for intracellular membrane organelle transports.* Cell Struct Funct, 1996. **21**(5): p. 283-95.
193. Goedert, M., et al., *Cloning and sequencing of the cDNA encoding an isoform of microtubule-associated protein tau containing four tandem repeats: differential expression of tau protein mRNAs in human brain.* Embo J, 1989. **8**(2): p. 393-9.
194. Kosik, K.S., et al., *Developmentally regulated expression of specific tau sequences.* Neuron, 1989. **2**(4): p. 1389-97.
195. Gustke, N., et al., *Domains of tau protein and interactions with microtubules.* Biochemistry, 1994. **33**(32): p. 9511-22.
196. Goedert, M. and R. Jakes, *Expression of separate isoforms of human tau protein: correlation with the tau pattern in brain and effects on tubulin polymerization.* Embo J, 1990. **9**(13): p. 4225-30.
197. King, M.E., et al., *Differential assembly of human tau isoforms in the presence of arachidonic acid.* J Neurochem, 2000. **74**(4): p. 1749-57.
198. Vershinin, M., et al., *Multiple-motor based transport and its regulation by Tau.* Proc Natl Acad Sci U S A, 2007. **104**(1): p. 87-92.
199. Gass, J., et al., *Mutations in progranulin are a major cause of ubiquitin-positive frontotemporal lobar degeneration.* Hum Mol Genet, 2006. **15**(20): p. 2988-3001.
200. Baker, M., et al., *Mutations in progranulin cause tau-negative frontotemporal dementia linked to chromosome 17.* Nature, 2006. **442**(7105): p. 916-9.
201. Cruts, M., et al., *Null mutations in progranulin cause ubiquitin-positive frontotemporal dementia linked to chromosome 17q21.* Nature, 2006. **442**(7105): p. 920-4.
202. Le Ber, I., et al., *Progranulin null mutations in both sporadic and familial frontotemporal dementia.* Hum Mutat, 2007. **28**(9): p. 846-55.
203. Bateman, A. and H.P. Bennett, *Granulins: the structure and function of an emerging family of growth factors.* J Endocrinol, 1998. **158**(2): p. 145-51.
204. He, Z., et al., *Progranulin is a mediator of the wound response.* Nat Med, 2003. **9**(2): p. 225-9.
205. Daniel, R., et al., *Cellular localization of gene expression for progranulin.* J Histochem Cytochem, 2000. **48**(7): p. 999-1009.
206. Skibinski, G., et al., *Mutations in the endosomal ESCRTIII-complex subunit CHMP2B in frontotemporal dementia.* Nat Genet, 2005. **37**(8): p. 806-8.

207. Cannon, A., et al., *CHMP2B mutations are not a common cause of frontotemporal lobar degeneration.* Neurosci Lett, 2006. **398**(1-2): p. 83-4.
208. Parkinson, N., et al., *ALS phenotypes with mutations in CHMP2B (charged multivesicular body protein 2B).* Neurology, 2006. **67**(6): p. 1074-7.
209. Rizzu, P., et al., *CHMP2B mutations are not a cause of dementia in Dutch patients with familial and sporadic frontotemporal dementia.* Am J Med Genet B Neuropsychiatr Genet, 2006. **141B**(8): p. 944-6.
210. Momeni, P., et al., *Genetic variability in CHMP2B and frontotemporal dementia.* Neurodegener Dis, 2006. **3**(3): p. 129-33.
211. Kovach, M.J., et al., *Clinical delineation and localization to chromosome 9p13.3-p12 of a unique dominant disorder in four families: hereditary inclusion body myopathy, Paget disease of bone, and frontotemporal dementia.* Mol Genet Metab, 2001. **74**(4): p. 458-75.
212. Watts, G.D., et al., *Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein.* Nat Genet, 2004. **36**(4): p. 377-81.
213. Woodman, P.G., *p97, a protein coping with multiple identities.* J Cell Sci, 2003. **116**(Pt 21): p. 4283-90.
214. AmericanPsychiatricAssociation, *Diagnostic and Statistical Manual of Mental Disorders*. 4th revised ed. ed, ed. A.P. Association. 1994, Washington DC, USA.
215. McKhann, G., et al., *Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease.* Neurology, 1984. **34**(7): p. 939-44.
216. Cruts, M., et al., *Estimation of the genetic contribution of presenilin-1 and -2 mutations in a population-based study of presenile Alzheimer disease.* Hum Mol Genet, 1998. **7**(1): p. 43-51.
217. Braak, H. and E. Braak, *Neuropathological stageing of Alzheimer-related changes.* Acta Neuropathol, 1991. **82**(4): p. 239-59.
218. Eckman, C.B., et al., *A new pathogenic mutation in the APP gene (I716V) increases the relative proportion of A beta 42(43).* Hum Mol Genet, 1997. **6**(12): p. 2087-9.
219. De Jonghe, C., et al., *Pathogenic APP mutations near the gamma-secretase cleavage site differentially affect Abeta secretion and APP C-terminal fragment stability.* Hum Mol Genet, 2001. **10**(16): p. 1665-71.
220. Terreni L, et al., *Novel pathogenic mutation in an Italian patient with familial Alzheimer's disease detected in APP gene.* Neurobiol Aging, 2002. **23**(1S): p. S319.
221. Romero, I., et al., *A presenilin-1 Thr116Asn substitution in a family with early-onset Alzheimer's disease.* Neuroreport, 1999. **10**(11): p. 2255-60.
222. Rogaeva, E.A., et al., *Screening for PS1 mutations in a referral-based series of AD cases: 21 novel mutations.* Neurology, 2001. **57**(4): p. 621-5.
223. La Bella, V., et al., *A novel mutation (Thr116Ile) in the presenilin 1 gene in a patient with early-onset Alzheimer's disease.* Eur J Neurol, 2004. **11**(8): p. 521-4.
224. Kwok, J.B., et al., *Two novel (M233T and R278T) presenilin-1 mutations in early-onset Alzheimer's disease pedigrees and preliminary evidence for association of presenilin-1 mutations with a novel phenotype.* Neuroreport, 1997. **8**(6): p. 1537-42.
225. Campion, D., et al., *Early-onset autosomal dominant Alzheimer disease: prevalence, genetic heterogeneity, and mutation spectrum.* Am J Hum Genet, 1999. **65**(3): p. 664-70.
226. Aldudo, J., M.J. Bullido, and F. Valdivieso, *DGGE method for the mutational analysis of the coding and proximal promoter regions of the Alzheimer's disease presenilin-1 gene: two novel mutations.* Hum Mutat, 1999. **14**(5): p. 433-9.
227. Houlden, H., et al., *A novel presenilin mutation (M233V) causing very early onset Alzheimer's disease with Lewy bodies.* Neurosci Lett, 2001. **313**(1-2): p. 93-5.
228. Mendez, M.F. and A. McMurtray, *Frontotemporal dementia-like phenotypes associated with presenilin-1 mutations.* Am J Alzheimers Dis Other Demen, 2006. **21**(4): p. 281-6.

229. Qi-Takahara, Y., et al., *Longer forms of amyloid beta protein: implications for the mechanism of intramembrane cleavage by gamma-secretase.* J Neurosci, 2005. **25**(2): p. 436-45.
230. Qi, Y., et al., *Distinct mechanisms by mutant presenilin 1 and 2 leading to increased intracellular levels of amyloid beta-protein 42 in Chinese hamster ovary cells.* Biochemistry, 2003. **42**(4): p. 1042-52.
231. Ikeda, M., et al., *The clinical phenotype of two missense mutations in the presenilin I gene in Japanese patients.* Ann Neurol, 1996. **40**(6): p. 912-7.
232. Kametani, F., et al., *Human wild presenilin-1 mimics the effect of the mutant presenilin-1 on the processing of Alzheimer's amyloid precursor protein in PC12D cells.* J Neurol Sci, 2001. **188**(1-2): p. 27-31.
233. Jimenez-Escrig, A., et al., *New V272A presenilin 1 mutation with very early onset subcortical dementia and parkinsonism.* Eur J Neurol, 2004. **11**(10): p. 663-9.
234. Burns, A. and S. Iliffe, *Alzheimer's disease.* Bmj, 2009. **338**: p. b158.
235. Hardy, J. and H. Orr, *The genetics of neurodegenerative diseases.* J Neurochem, 2006. **97**(6): p. 1690-9.
236. McCarroll, S.A. and D.M. Altshuler, *Copy-number variation and association studies of human disease.* Nat Genet, 2007. **39**(7 Suppl): p. S37-42.
237. Delabar, J.M., et al., *Beta amyloid gene duplication in Alzheimer's disease and karyotypically normal Down syndrome.* Science, 1987. **235**(4794): p. 1390-2.
238. Collinge, J., *Prion diseases of humans and animals: their causes and molecular basis.* Annu Rev Neurosci, 2001. **24**: p. 519-50.
239. Owen, F., et al., *Insertion in prion protein gene in familial Creutzfeldt-Jakob disease.* Lancet, 1989. **1**(8628): p. 51-2.
240. Beck, J.A., et al., *Two-octapeptide repeat deletion of prion protein associated with rapidly progressive dementia.* Neurology, 2001. **57**(2): p. 354-6.
241. Asante, E.A., et al., *Absence of spontaneous disease and comparative prion susceptibility of transgenic mice expressing mutant human prion proteins.* J Gen Virol, 2009. **90**(Pt 3): p. 546-58.
242. Bueler, H., et al., *Mice devoid of PrP are resistant to scrapie.* Cell, 1993. **73**(7): p. 1339-47.
243. Knight, M.A., et al., *A duplication at chromosome 11q12.2-11q12.3 is associated with spinocerebellar ataxia type 20.* Hum Mol Genet, 2008. **17**(24): p. 3847-53.
244. Simon-Sanchez, J., et al., *Genomewide SNP assay reveals mutations underlying Parkinson disease.* Hum Mutat, 2008. **29**(2): p. 315-22.
245. van de Leemput, J., et al., *Deletion at ITPR1 underlies ataxia in mice and spinocerebellar ataxia 15 in humans.* PLoS Genet, 2007. **3**(6): p. e108.
246. Wallin, A.K., et al., *CSF biomarkers for Alzheimer's Disease: levels of beta-amyloid, tau, phosphorylated tau relate to clinical symptoms and survival.* Dement Geriatr Cogn Disord, 2006. **21**(3): p. 131-8.
247. Blom, E.S., et al., *Low prevalence of APP duplications in Swedish and Finnish patients with early-onset Alzheimer's disease.* Eur J Hum Genet, 2008. **16**(2): p. 171-5.
248. Hendriks, L., et al., *Presenile dementia and cerebral haemorrhage linked to a mutation at codon 692 of the beta-amyloid precursor protein gene.* Nat Genet, 1992. **1**(3): p. 218-21.
249. Boeve, B.F., A.E. Lang, and I. Litvan, *Corticobasal degeneration and its relationship to progressive supranuclear palsy and frontotemporal dementia.* Ann Neurol, 2003. **54 Suppl 5**: p. S15-9.
250. Litvan, I., et al., *Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): report of the NINDS-SPSP international workshop.* Neurology, 1996. **47**(1): p. 1-9.
251. Neary, D., et al., *Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria.* Neurology, 1998. **51**(6): p. 1546-54.

252. Beck, J., et al., *A distinct clinical, neuropsychological and radiological phenotype is associated with progranulin gene mutations in a large UK series.* Brain, 2008. **131**(Pt 3): p. 706-20.
253. Goldman, J.S., et al., *Comparison of family histories in FTLD subtypes and related tauopathies.* Neurology, 2005. **65**(11): p. 1817-9.
254. Kwiatkowski, T.J., Jr., et al., *Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis.* Science, 2009. **323**(5918): p. 1205-8.
255. Vance, C., et al., *Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6.* Science, 2009. **323**(5918): p. 1208-11.
256. Janssen, J.C., et al., *Clinical features of frontotemporal dementia due to the intronic tau 10(+16) mutation.* Neurology, 2002. **58**(8): p. 1161-8.
257. Spillantini, M.G., et al., *A novel tau mutation (N296N) in familial dementia with swollen achromatic neurons and corticobasal inclusion bodies.* Ann Neurol, 2000. **48**(6): p. 939-43.
258. Rohrer JD, et al., *Corticobasal syndrome associated with a novel 1048_1049insG progranulin mutation.* JNNP: p. (in press).
259. Snowden, J., D. Neary, and D. Mann, *Frontotemporal lobar degeneration: clinical and pathological relationships.* Acta Neuropathol, 2007. **114**(1): p. 31-8.
260. Mesulam, M., et al., *Alzheimer and frontotemporal pathology in subsets of primary progressive aphasia.* Ann Neurol, 2008. **63**(6): p. 709-19.
261. Houlden, H., et al., *Frequency of tau mutations in three series of non-Alzheimer's degenerative dementia.* Ann Neurol, 1999. **46**(2): p. 243-8.
262. Rosso, S.M., et al., *Frontotemporal dementia in The Netherlands: patient characteristics and prevalence estimates from a population-based study.* Brain, 2003. **126**(Pt 9): p. 2016-22.
263. Signorini, S., et al., *Prevalence of pathogenic mutations in an Italian clinical series of patients with familial dementia.* Curr Alzheimer Res, 2004. **1**(3): p. 215-8.
264. Stanford, P.M., et al., *Frequency of tau mutations in familial and sporadic frontotemporal dementia and other tauopathies.* J Neurol, 2004. **251**(9): p. 1098-104.
265. Benussi, L., et al., *Progranulin Leu271LeufsX10 is one of the most common FTLD and CBS associated mutations worldwide.* Neurobiol Dis, 2009. **33**(3): p. 379-85.
266. Borroni, B., et al., *Progranulin genetic variations in frontotemporal lobar degeneration: evidence for low mutation frequency in an Italian clinical series.* Neurogenetics, 2008. **9**(3): p. 197-205.
267. Pickering-Brown, S.M., et al., *Frequency and clinical characteristics of progranulin mutation carriers in the Manchester frontotemporal lobar degeneration cohort: comparison with patients with MAPT and no known mutations.* Brain, 2008. **131**(Pt 3): p. 721-31.
268. Pickering-Brown, S., et al., *Evidence of a founder effect in families with frontotemporal dementia that harbor the tau +16 splice mutation.* Am J Med Genet B Neuropsychiatr Genet, 2004. **125B**(1): p. 79-82.
269. Seelaar, H., et al., *Distinct genetic forms of frontotemporal dementia.* Neurology, 2008. **71**(16): p. 1220-6.
270. Le Ber, I., et al., *Chromosome 9p-linked families with frontotemporal dementia associated with motor neuron disease.* Neurology, 2009. **72**(19): p. 1669-76.
271. Mackenzie, I.R., et al., *Atypical frontotemporal lobar degeneration with ubiquitin-positive, TDP-43-negative neuronal inclusions.* Brain, 2008. **131**(Pt 5): p. 1282-93.
272. Roeber, S., et al., *TDP-43-negative FTLD-U is a significant new clinico-pathological subtype of FTLD.* Acta Neuropathol, 2008. **116**(2): p. 147-57.
273. Geser, F., et al., *Clinical and pathological continuum of multisystem TDP-43 proteinopathies.* Arch Neurol, 2009. **66**(2): p. 180-9.
274. Josephs, K.A., et al., *Neuropathologic features of frontotemporal lobar degeneration with ubiquitin-positive inclusions with progranulin gene (PGRN) mutations.* J Neuropathol Exp Neurol, 2007. **66**(2): p. 142-51.

275. Josephs, K.A., et al., *Evaluation of subcortical pathology and clinical correlations in FTLD-U subtypes.* Acta Neuropathol, 2009. **118**(3): p. 349-58.
276. Cruts, M., S. Kumar-Singh, and C. Van Broeckhoven, *Progranulin mutations in ubiquitin-positive frontotemporal dementia linked to chromosome 17q21.* Curr Alzheimer Res, 2006. **3**(5): p. 485-91.
277. Bronner, I.F., et al., *Progranulin mutations in Dutch familial frontotemporal lobar degeneration.* Eur J Hum Genet, 2007. **15**(3): p. 369-74.
278. van der Zee, J., et al., *Mutations other than null mutations producing a pathogenic loss of progranulin in frontotemporal dementia.* Hum Mutat, 2007. **28**(4): p. 416.
279. Goedert, M. and M.G. Spillantini, *Frontotemporal lobar degeneration through loss of progranulin function.* Brain, 2006. **129**(Pt 11): p. 2808-10.
280. Schymick, J., et al., *Progranulin mutations and ALS or ALS-FTD phenotypes.* J Neurol Neurosurg Psychiatry, 2007.
281. Llado, A., et al., *Late-onset frontotemporal dementia associated with a novel PGRN mutation.* J Neural Transm, 2007.
282. Van Deerlin, V.M., et al., *Clinical, genetic, and pathologic characteristics of patients with frontotemporal dementia and progranulin mutations.* Arch Neurol, 2007. **64**(8): p. 1148-53.
283. Masellis, M., et al., *Novel splicing mutation in the progranulin gene causing familial corticobasal syndrome.* Brain, 2006. **129**(Pt 11): p. 3115-23.
284. Benussi, L., et al., *A novel deletion in progranulin gene is associated with FTDP-17 and CBS.* Neurobiol Aging, 2006.
285. Josephs, K.A., et al., *Frontotemporal lobar degeneration and ubiquitin immunohistochemistry.* Neuropathol Appl Neurobiol, 2004. **30**(4): p. 369-73.
286. Sampathu, D.M., et al., *Pathological heterogeneity of frontotemporal lobar degeneration with ubiquitin-positive inclusions delineated by ubiquitin immunohistochemistry and novel monoclonal antibodies.* Am J Pathol, 2006. **169**(4): p. 1343-52.
287. Mackenzie, I.R., et al., *Heterogeneity of ubiquitin pathology in frontotemporal lobar degeneration: classification and relation to clinical phenotype.* Acta Neuropathol (Berl), 2006. **112**(5): p. 539-49.
288. Ince, P.G., et al., *Disorders of movement and system degenerations*, in *Greenfield's Neuropathology*, D.W. Ellison, Louis,D.N. and Love,S. Eds., Editor. 2007: London.
289. Grimes, D.A., C.B. Bergeron, and A.E. Lang, *Motor neuron disease-inclusion dementia presenting as cortical-basal ganglionic degeneration.* Mov Disord, 1999. **14**(4): p. 674-80.
290. Paviour, D.C., et al., *Frontotemporal lobar degeneration with ubiquitin-only-immunoreactive neuronal changes: broadening the clinical picture to include progressive supranuclear palsy.* Brain, 2004. **127**(Pt 11): p. 2441-51.
291. Hardy, J. and A. Singleton, *Reporting and interpretation of genetic variants in cases and controls.* Neurology, 2007. **69**(1): p. 111-2.
292. Mattila, K.M., et al., *The Glu318Gly mutation of the presenilin-1 gene does not necessarily cause Alzheimer's disease.* Ann Neurol, 1998. **44**(6): p. 965-7.
293. Boeve, B.F., et al., *Frontotemporal dementia and parkinsonism associated with the IVS1+1G->A mutation in progranulin: a clinicopathologic study.* Brain, 2006. **129**(Pt 11): p. 3103-14.
294. Cann, H.M., et al., *A human genome diversity cell line panel.* Science, 2002. **296**(5566): p. 261-2.
295. Walker, E.S., et al., *Presenilin 2 familial Alzheimer's disease mutations result in partial loss of function and dramatic changes in Abeta 42/40 ratios.* J Neurochem, 2005. **92**(2): p. 294-301.
296. Levy-Lahad, E., et al., *Candidate gene for the chromosome 1 familial Alzheimer's disease locus.* Science, 1995. **269**(5226): p. 973-7.
297. Hutton, M., et al., *Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17.* Nature, 1998. **393**(6686): p. 702-5.

298. Guerreiro, R.J., et al., *Genetic screening of Alzheimer's disease genes in Iberian and African samples yields novel mutations in presenilins and APP.* Neurobiol Aging, 2008.
299. Ng, P.C. and S. Henikoff, *SIFT: Predicting amino acid changes that affect protein function.* Nucleic Acids Res, 2003. **31**(13): p. 3812-4.
300. Sunyaev, S., et al., *Prediction of deleterious human alleles.* Hum Mol Genet, 2001. **10**(6): p. 591-7.
301. Ramensky, V., P. Bork, and S. Sunyaev, *Human non-synonymous SNPs: server and survey.* Nucleic Acids Res, 2002. **30**(17): p. 3894-900.
302. Tchernitchko, D., M. Goossens, and H. Wajcman, *In silico prediction of the deleterious effect of a mutation: proceed with caution in clinical genetics.* Clin Chem, 2004. **50**(11): p. 1974-8.
303. Chan, P.A., et al., *Interpreting missense variants: comparing computational methods in human disease genes CDKN2A, MLH1, MSH2, MECP2, and tyrosinase (TYR).* Hum Mutat, 2007. **28**(7): p. 683-93.
304. Le Ber, I., et al., *Phenotype variability in progranulin mutation carriers: a clinical, neuropsychological, imaging and genetic study.* Brain, 2008. **131**(Pt 3): p. 732-46.
305. Mukherjee, O., et al., *HDDD2 is a familial frontotemporal lobar degeneration with ubiquitin-positive, tau-negative inclusions caused by a missense mutation in the signal peptide of progranulin.* Ann Neurol, 2006. **60**(3): p. 314-22.
306. Spina, S., et al., *Corticobasal syndrome associated with the A9D Progranulin mutation.* J Neuropathol Exp Neurol, 2007. **66**(10): p. 892-900.
307. Mukherjee, O., et al., *Molecular characterization of novel progranulin (GRN) mutations in frontotemporal dementia.* Hum Mutat, 2008. **29**(4): p. 512-21.
308. Shankaran, S.S., et al., *Missense mutations in the progranulin gene linked to frontotemporal lobar degeneration with ubiquitin-immunoreactive inclusions reduce progranulin production and secretion.* J Biol Chem, 2008. **283**(3): p. 1744-53.
309. Cruchaga, C., et al., *Cortical Atrophy and Language Network Reorganization Associated with a Novel Progranulin Mutation.* Cereb Cortex, 2008.
310. Brouwers, N., et al., *Genetic variability in progranulin contributes to risk for clinically diagnosed Alzheimer disease.* Neurology, 2008. **71**(9): p. 656-64.
311. Schymick, J.C., et al., *Progranulin mutations and amyotrophic lateral sclerosis or amyotrophic lateral sclerosis-frontotemporal dementia phenotypes.* J Neurol Neurosurg Psychiatry, 2007. **78**(7): p. 754-6.
312. Guerreiro, R.J., et al., *Novel progranulin mutation: screening for PGRN mutations in a Portuguese series of FTD/CBS cases.* Mov Disord, 2008. **23**(9): p. 1269-73.
313. Gibson, J., N.E. Morton, and A. Collins, *Extended tracts of homozygosity in outbred human populations.* Hum Mol Genet, 2006. **15**(5): p. 789-95.
314. Li, L.H., et al., *Long contiguous stretches of homozygosity in the human genome.* Hum Mutat, 2006. **27**(11): p. 1115-21.
315. McQuillan, R., et al., *Runs of homozygosity in European populations.* Am J Hum Genet, 2008. **83**(3): p. 359-72.
316. Devilee, P., A.M. Cleton-Jansen, and C.J. Cornelisse, *Ever since Knudson.* Trends Genet, 2001. **17**(10): p. 569-73.
317. Raghavan, M., et al., *Genome-wide single nucleotide polymorphism analysis reveals frequent partial uniparental disomy due to somatic recombination in acute myeloid leukemias.* Cancer Res, 2005. **65**(2): p. 375-8.
318. Woods, C.G., et al., *A new method for autozygosity mapping using single nucleotide polymorphisms (SNPs) and EXCLUDEAR.* J Med Genet, 2004. **41**(8): p. e101.
319. Simon-Sanchez, J., et al., *Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals.* Hum Mol Genet, 2007. **16**(1): p. 1-14.

320. Clarimon, J., et al., *Whole genome analysis in a consanguineous family with early onset Alzheimer's disease.* Neurobiol Aging, 2008.
321. Leutenegger, A.L., et al., *Estimation of the inbreeding coefficient through use of genomic data.* Am J Hum Genet, 2003. **73**(3): p. 516-23.
322. Gibbs, J.R. and A. Singleton, *Application of genome-wide single nucleotide polymorphism typing: simple association and beyond.* PLoS Genet, 2006. **2**(10): p. e150.
323. Pritchard, J.K., M. Stephens, and P. Donnelly, *Inference of population structure using multilocus genotype data.* Genetics, 2000. **155**(2): p. 945-59.
324. *A haplotype map of the human genome.* Nature, 2005. **437**(7063): p. 1299-320.
325. Lencz, T., et al., *Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia.* Proc Natl Acad Sci U S A, 2007. **104**(50): p. 19942-7.
326. Webber, K.M., et al., *Steroidogenic acute regulatory protein (StAR): evidence of gonadotropin-induced steroidogenesis in Alzheimer disease.* Mol Neurodegener, 2006. **1**: p. 14.
327. Li, X., et al., *Levels of mTOR and its downstream targets 4E-BP1, eEF2, and eEF2 kinase in relationships with tau in Alzheimer's disease brain.* Febs J, 2005. **272**(16): p. 4211-20.
328. Meyers, D.S., et al., *Beta 3-adrenergic receptor-mediated lipolysis and oxygen consumption in brown adipocytes from cynomolgus monkeys.* J Clin Endocrinol Metab, 1997. **82**(2): p. 395-401.
329. Hoyer, S., *Causes and consequences of disturbances of cerebral glucose metabolism in sporadic Alzheimer disease: therapeutic implications.* Adv Exp Med Biol, 2004. **541**: p. 135-52.
330. Craft, S. and G.S. Watson, *Insulin and neurodegenerative disease: shared and specific mechanisms.* Lancet Neurol, 2004. **3**(3): p. 169-78.
331. Hamilton, G., et al., *Candidate gene association study of insulin signaling genes and Alzheimer's disease: evidence for SOS2, PCK1, and PPARgamma as susceptibility loci.* Am J Med Genet B Neuropsychiatr Genet, 2007. **144B**(4): p. 508-16.
332. Sugawara, T., et al., *Steroidogenic factor 1-dependent promoter activity of the human steroidogenic acute regulatory protein (StAR) gene.* Biochemistry, 1996. **35**(28): p. 9052-9.
333. Proud, C.G., *Regulation of mammalian translation factors by nutrients.* Eur J Biochem, 2002. **269**(22): p. 5338-49.
334. Nalls, M.A., et al., *Measures of autozygosity in decline: globalization, urbanization, and its implications for medical genetics.* PLoS Genet, 2009. **5**(3): p. e1000415.
335. Saunders, A.M., et al., *Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease.* Neurology, 1993. **43**(8): p. 1467-72.
336. Yu, C.E., et al., *Comprehensive analysis of APOE and selected proximate markers for late-onset Alzheimer's disease: patterns of linkage disequilibrium and disease/marker association.* Genomics, 2007. **89**(6): p. 655-65.
337. Bowirrat, A., et al., *The very high prevalence of AD in an Arab population is not explained by APOE epsilon4 allele frequency.* Neurology, 2000. **55**(5): p. 731.
338. Farrer, L.A., et al., *Identification of multiple loci for Alzheimer disease in a consanguineous Israeli-Arab community.* Hum Mol Genet, 2003. **12**(4): p. 415-22.
339. Iafrate, A.J., et al., *Detection of large-scale variation in the human genome.* Nat Genet, 2004. **36**(9): p. 949-51.
340. Avramopoulos, D., *Genetics of Alzheimer's disease: recent advances.* Genome Med, 2009. **1**(3): p. 34.
341. WTCCC, *Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.* Nature, 2007. **447**(7145): p. 661-78.
342. Sunyaev, S., V. Ramensky, and P. Bork, *Towards a structural basis of human non-synonymous single nucleotide polymorphisms.* Trends Genet, 2000. **16**(5): p. 198-200.

343. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses.* Am J Hum Genet, 2007. **81**(3): p. 559-75.
344. Myers, A.J., et al., *A survey of genetic human cortical gene expression.* Nat Genet, 2007. **39**(12): p. 1494-9.
345. Falush, D., M. Stephens, and J.K. Pritchard, *Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies.* Genetics, 2003. **164**(4): p. 1567-87.
346. Workman, C., et al., *A new non-linear normalization method for reducing variability in DNA microarray experiments.* Genome Biol, 2002. **3**(9): p. research0048.
347. Schadt, E.E., et al., *Feature extraction and normalization algorithms for high-density oligonucleotide gene expression array data.* J Cell Biochem Suppl, 2001. **Suppl 37**: p. 120-5.
348. Tseng, G.C., et al., *Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects.* Nucleic Acids Res, 2001. **29**(12): p. 2549-57.
349. Miwa, Y., et al., *Insertion/deletion polymorphism in clusterin gene influences serum lipid levels and carotid intima-media thickness in hypertensive Japanese females.* Biochem Biophys Res Commun, 2005. **331**(4): p. 1587-93.
350. Swaroop, A., et al., *Unraveling a multifactorial late-onset disease: from genetic susceptibility to disease mechanisms for age-related macular degeneration.* Annu Rev Genomics Hum Genet, 2009. **10**: p. 19-43.
351. Ross, O.A., et al., *Familial genes in sporadic disease: common variants of alpha-synuclein gene associate with Parkinson's disease.* Mech Ageing Dev, 2007. **128**(5-6): p. 378-82.
352. Schadt, E.E., et al., *Mapping the genetic architecture of gene expression in human liver.* PLoS Biol, 2008. **6**(5): p. e107.
353. Stranger, B.E., et al., *Population genomics of human gene expression.* Nat Genet, 2007. **39**(10): p. 1217-24.
354. Veyrieras, J.B., et al., *High-resolution mapping of expression-QTLs yields insight into human gene regulation.* PLoS Genet, 2008. **4**(10): p. e1000214.