Universitat Autònoma de Barcelona

The genetic basis of Alzheimer's disease and its related endophenotypes

Sonia Moreno Grau

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons: http://cat.creativecommons.org/?page id=184

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons: http://es.creativecommons.org/blog/licencias/

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license: https://creativecommons.org/licenses/?lang=en

Universitat Autònoma de Barcelona

DOCTORAL THESIS

The genetic basis of Alzheimer's disease and its related endophenotypes

Sonia Moreno Grau

Dr. Agustín Ruiz Laza, Thesis Director

Dr. Joan Xavier Comella Carnicé, Thesis Tutor

Department if Biochemistry and Molecular Biology

Faculty of Bioscience

Autonomous University of Barcelona

2019

Universitat Autònoma de Barcelona

DOCTORAL THESIS

The genetic basis of Alzheimer's disease and its related endophenotypes

Sonia Moreno Grau

Dr. Agustín Ruiz Laza, Thesis Director

Dr. Joan Xavier Comella Carnicé, Thesis Tutor

PhD programme in Biochemistry, Molecular Biology and Biomedicine

Department if Biochemistry and Molecular Biology

Autonomous University of Barcelona

2019

Author signature

Thesis Director signature

Certificate of Direction

Dr. Agustín Ruiz Laza, Research director of Fundació ACE, and Dr. Joan Xavier Comella Carnicé head of the Vall d'Hebron research institute and full professor of the Universitat Autònoma de Barcelona

Certify:

That work entitle "**The genetic basis of Alzheimer's Disease and its related endophenotypes**" presented by Sonia Moreno Grau, candidate of the PhD program in Biochemistry, Molecular Biology and Biomedicine, has been done under our direction and meets all the requirements to be defended in front of the corresponding Thesis Committee.

Dr. Agustín Ruiz Laza,
Thesis Director

Dr. Joan Xavier Comella Carnicé,
Thesis Tutor

Scientific director of the Fundació ACE

Head of the Genomic group of Fundació ACE

Head of the Vall d'Hebrón Research Institute (VHIR)

Head of the Cell signaling and Apoptosis group of VHIR

A Joaquim Moreno, el meu pare

A Maria Luisa Grau, la meua mare

No encuentro una manera mejor de empezar que agradeciéndole a Agustín Ruiz Laza, el director de esta tesis, su confianza, apoyo y dedicación. Gracias por guiarme cuando he estado perdida. Gracias por abrirme puertas, y nunca cerrarlas. Gracias por ayudarme a entender la ciencia desde muchas perspectivas. Pero sobre todo, gracias por hacerme ver que esto a lo que nos dedicamos es algo realmente apasionante. Aprender de ti siempre ha sido, es y será un placer.

A Joan Comella, el tutor de esta tesis, y a Joaquim Soriano, gracias por darme esa primera oportunidad que me ha llevado hasta aquí. Gracias por estar dispuestos a ayudarme.

Gracias a Lluís Tárraga y Mercè Boada, por creer y hacer posible que un trabajo de estas características también tenga su espacio en Fundació ACE. Sin la perspectiva clínica este trabajo no sería, gracias por darme la oportunidad de formar parte del equipo.

A Itziar, Laura y Adela, si hemos sobrevivido a GR@ACE, ya no hay nada que se nos ponga por delante ☺. Itziar, gracias por tus helados, tus galletitas y tu dron! El mundo es más divertido con ellos! Gracias por tu apoyo durante todos estos años. Gracias porque contigo todo ha sido inmensamente fácil, gracias por esa complicidad. Laura, gràcies per estar sempre disposta a donar un cop de mà, i a més fer-ho amb una sonrisa. No oblide aquell divendres oscur a les 20h del vespre al lab, però tot i així, ho vam dur amb humor! Adela, ens va costar arrancar el lab però al final ho vam conseguir. Gràcies pel suport durant tot aquest temps.

Oscar, aunque aún estoy muy lejos de "leer poesía en awk", una gran parte de lo que hoy sé de programación, lo aprendí de ti. Siempre te estaré agradecida por todo el tiempo que invertiste en enseñarme.

Elvira, Sergi, Gemma, Begoña y Marta, gracias por formar parte de esta aventura. Gracias por las palabras de ánimo y por el apoyo. A Pablo, el último en llegar, pero no por ello menos importante. Gracias por los cafetitos de las tres, gracias por el buen rollo.

A toda la familia de la Unidad de Memoria, centro de día y hospital de día, sin vuestro trabajo esta tesis no habría sido posible. Gracias.

María Eugenia Sáez y Antonio González, gracias por enseñarme los intríngulis del mundillo de los GWAS. Gracias por responder SIEMPRE a todas y cada una de mis dudas. Sin vuestro apoyo este trabajo estaría cojo. Muchas gracias.

A todo el equipo del Banco nacional de ADN y el Centro Nacional de Genotipado, muy especialmente a Ángel Carracedo e Inés Quintela. Gracias por hacerlo todo tan, tan fácil siempre. Gracias por permitirme aprender de vosotros, y dejarme formar parte de un equipo humano maravilloso. Gracias a Luís Miguel Real, y a todo el equipo del Hospital Universitario de Valmé, por colaborar en este proyecto.

Gracias a Alfredo Ramírez por darme la oportunidad de visitar su grupo en la Universidad de Bonn y el centro de genotipado Life and Brain. Holger, thank you for making me feel at home during the time I was in Bonn. Steffanie Heilmann, thank you for sharing your knowledge with me.

A Carlos Chruchaga y a Vicky. Gracias por darme la oportunidad de poder disfrutar de una de las experiencias más enriquecedoras a nivel laboral y personal. Ha sido un inmenso privilegio poder formar parte del equipo durante unos meses. A vosotros y a todos los que forman parte

del grupo, gracias por acogerme. Vicky, gràcies per estar sempre disposta a compartir els teus coneixements amb mí, per guiar-me en als primers moments d'aquest treball i també en als últims. Gràcies per donar-me l'oportunitat de poder fer la meua estancia amb tú. Aquí o allà has sigut el més paregut que he tingut a un "post-doc". Gràcies pel suport durant tots aquests anys.

Maria Luisa i Joaquim, els meus pares, si penso en vosaltres trobo un món més càlid, dolç i tendre. Si penso en vosaltres, per molt lluny que em trobí, em sento a casa. Serien infinites totes les coses que deuria agrair-vos, però no sé si fer-ho aquí. Doncs no vull enfadar a la mamà, ja que ella, al seu rol de mare, veu com una gran ofensa que la seua filla li faci agraïments. Això, ja en diu molt. Així doncs, agraeixo a l'univers, al destí, l'atzar per deixar-me a les vostres mans. No podria tenir uns pares millors que vosaltres. Gràcies per ensenyar-me que els objectius van sense límits. Gràcies per un amor tan incondicional.

Als meus iaios, i molt especialment, al meu iaio Joaquim, gràcies per tota una vida d'ensenyances. A Esther, Vicent, Jovi, Maria Jose, Juan, Toni, Pili, Javi, Olga, Celia, i Juanjo, que se que pensen en mi amb amor. A tota la meua família i amics, gràcies per donar-me suport i ànims. Gràcies per creure en mi.

María José, Pepe, Elena e Ismael, gracias por animarme, por pensar en mí, por enviarme positivismo.

A Lucia, Ana i Nerea, gràcies perquè amb vosaltres l'amistat no te límits. Seguim sumant.

Blanca, gracias por compartir conmigo los últimos minutos de este largo parto. Y siempre pensando en las "nosotras" del después. Lo que ha unido una tesis no lo separan ni 8,000 Km.

Ángel, Sergio, Celia, Mariajo, Maria y Silvia gracias porque con vosotros Barcelona dejó de ser una gran ciudad para convertirse en una gran familia. Mariajo, eras mi lugar favorito de Barcelona, y ahora de Salamanca.

A Pablo, gracias por ser mejor de lo que podría desear, gracias por estar siempre. A veces las palabras son insuficientes.

En último lugar, me gustaría agradecer a todos los participantes de este estudio, a los pacientes, a sus familias, y a todas las personas que altruistamente decidieron donar una muestra a la ciencia. Sin ellos este trabajo no habría sido posible.

A todos los que me habéis acompañado durante estos años, este trabajo también es vuestro.

A todos vosotros GRACIAS.

Table of Contents

Abbreviations

A20	Protein A20
AAO	Age at onset
ABCA7	ATP binding cassette subfamily A member 7
ABCG1	ATP binding cassette subfamily G member 1
ABI3	AP2/B3-like transcriptional factor family protein
AC074212.3	AC074212.3 gene
AC099552.4	AC099552.4 gene
ACE	Angiotensin I converting enzyme
ACOT11	Acyl-CoA thioesterase 11
AD	Alzheimer's disease
ADAM10	ADAM metallopeptidase domain 10
ADAMTS1	ADAM metallopeptidase with thrombospondin type 1 motif 1
ADAPTED	Alzheimer's disease Apolipoprotein Pathology for Treatment Elucidation and Development
ADC	Alzheimer's disease centers
ADGC	The Alzheimer's disease Genetics Consortium
ADNI	The Alzheimer's disease neuroimaing initiative
ADRA2B	Adrenoceptor alpha 2B
AFR	Individuals from African ancestry
AgeCode	The German Study on Aging, Cognition and Dementia
AGER	Advanced glycosylation end-product specific receptor
AGPAT1	1-acylglycerol-3-phosphate O-acyltransferase 1
AIBL	Australian biomarkers and Lifestyle Study
AICD	Intracellular domain of APP
ALPK2	Alpha kinase 2
aMCI	Amnestic mild cognitive impairment
AMR	Individuals from American ancestry
ANKRD31	Ankyrin repeat domain 31
AP	Antagonistic pleiotropy
APH1	Anterior pharynx defective 1
APH1B	Aph-1 homolog B, gamma-secretase subunit
APH2	Anterior pharynx defective 2
APOE	Apolipoprotein E
APP	Amyloid precursor protein
APPsα	The α-secretase-generated APP ectodomain fragment
APPsβ	The β-secretase-generated APP ectodomain fragment
APT	Affimetrix power tool
arAD	Autosomal recessive Alzheimer's disease
ASA	American Statistical Association guidelines
ATN	Amyloid, tau and neurodegeneration
ATP5H	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit D
AV45	Florbetapir
AVLT	Auditory-Verbal Learning Test
AVROH	Average lengths of runs of homozygosity per individual

Aβ	Amyloid-beta
BACE1	Beta secretase 1
BACE2	Beta secretase 2
BAG4	BCL2 associated athanogene 4
BCHE	Butyrylcholinesterase
BCKDK	Branched chain keto acid dehydrogenase kinase
BDNF	Brain derived neurotrophic factor
BDRS	Blessed Dementia Rating Scale
BIN1	Bridging integrator 1
BIRC	Baculoviral inhibitors of apoptosis repeat containing
BIRC3	Baculoviral IAP repeat containing 3
BIRC4	X-linked inhibitor of apoptosis
BRAINEAC	The Brain eQTL Almanac
BRCA1	BRCA1 DNA repair associated
BRCA2	BRCA2 DNA repair associated
BZRAP1-AS1	TSPOAP1, SUPT4H1 and RNF43 antisense RNA 1
CAA	Cerebral amyloid angiopathy
CADD Score	Combined Annotation Dependent Depletion
CAMCOG battery	The Cambridge Cognitive Examination
CASS4	Cas scaffold protein family member 4
CCHS	The Copenhagen City Heart Study
CCNE2	Cyclin E2
CCS	The Cache County Study
CD2AP	CD2 associated protein
CD33	CD33 molecule
CD–CV	Common disease–common variant hypothesis
CDR	Clinical Dementia Rating
CeGEN	Spanish national center for genotyping
CELF1	CUGBP Elav-like family member 1
CEU	Utah Residents (CEPH) with Northern and Western European Ancestry
cFLIP	CASP8 and FADD like apoptosis regulator protein
CGPS	The Copenhagen General Population Study
CHARGE	Cohorts for Heart and Aging Research in Genomic Epidemiology
CHS	The Cardiovascular Health Study
CLFAR	CASP8 and FADD like apoptosis regulator
CLU	Clusterin
CNEs	Cognitively normal elders
CNTNAP2	Contactin associated protein like 2
CNV	Copy number variants
COL4A3BP	Collagen Type IV Alpha 3 Binding Protein; also known as *CERT1,* Ceramide transporter I
COMT	Catechol-O-methyltransferase
COR	Concordance rate
CR	Cognitive reserve
CR1	Complement C3b/C4b receptor 1
CSF	Cerebrospinal fluid
CTD-2235C13.3	Lnc RNA ENST00000607296.1

CTF-α/β	The C terminal fragment-α/β
CTSF	Cathepsin F
DEGESCO	The Dementia Genetics Spanish Consortium
DGV	The Database of Genomic Variants
DNA	Deoxyribonucleic acid
DPY19L4	Dpy-19 like 4
DR3	Death receptor 3 protein
DR4	Death receptor 4 protein
DR6	Death receptor 6 protein
DRs	Death receptors
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders, 4th Edition
EADB	European Alzheimer's disease biobank
EADI	The European Alzheimer's disease Initiative
EAS	Individuals from East Asian ancestry
ECHDC3	Enoyl-CoA hydratase domain containing 3
EGA	European genome phenome archive
EMCI	Early mild cognitive impairment
EOAD	Early-Onset Alzheimer's disease
EPHA1	EPH receptor A1
eQTL	Gene expression quantitative trait
ESP	The Exome Sequencing project
ESPR1	Epithelial splicing regulatory protein 1
ESR2	Estrogen receptor 2
EUR	Individuals from European ancestry
ExaC	Exome Aggregation Consortium
EXOC3L2	Exocyst complex component 3 like 2
FACE	Fundacio ACE
FACEHBI	Fundacio ACE healthy brain iniciative
FAD	Familial Alzheimer's disease
FAIM	Fas apoptotic inhibitory molecule protein
FAIM1	Fas apoptotic inhibitory molecule
FAIM2	Fas apoptotic inhibitory molecule 2
FAQ	Functional Activities Questionnaire
FasL	Fas ligand (TNF superfamily, member 6) protein
FASLG	Fas ligand gene
FBB	Florbetaben
FERMT2	Fermitin family member 2
FOS	Framingham Offspring Cohort
FROH	Runs of homozygosity based estimates of the inbreding coefficient
FRY	FRY microtubule binding protein
FST	Follistatin
FUMA	Funtional mapping and annotation
GATK	Genome Analysis Toolkit
GC	Genomic control
GDS	Global deterioration scale
GenADA	Multi-Site Collaborative Study for Genotype-Phenotype Associations in Alzheimer's

	disease
GERAD	Genetic and Environmental Risk in AD
GLIS3	GLIS family zinc finger 3
Gr@ACE	Genome research at Fundacio ACE
GRS	Genetic risk score
GSA	The Infinium global screening array
GTEx	The Gene-Tissue Expression project
GWAS	Genome-wide association studies
GWAX	Genome wide association by proxy
GWS	Genome-wide significance
HBEGF	Heparin binding EGF like growth factor
HC	Healthy controls
HCV	Hippocampal volume
HESX1	HESX homeobox 1
HLA	Human leukocyte antigen
HLA-DRB1	Major histocompatibility complex, class II, DR beta 1
HLA-DRB5	Major histocompatibility complex, class II, DR beta 5
HMGCR	3-hydroxy-3-methylglutaryl-CoA reductase
HRC	Haplotype reference consortium
HS3ST1	Heparan sulfate-glucosamine 3-sulfotransferase 1
HW	Hardy-Weinberg
HWE	Hardy-Weinberg equilibrium
IBD	Identity by descent
IGAP	The International Genomics of Alzheimer's Project
IGAP12	The International Genomics of Alzheimer's Project Stage 1 and 2
IGHV1-67	Immunoglobulin heavy variable 1-67 (pseudogene)
IL1B	Interleukin 1 beta
IL1RAP	Interleukin 1 receptor accessory protein
IMH.Pekin	The Institute of Mental Health of Pekin
INPP5D	Inositol polyphosphate-5-phosphatase D
INTS8	Integrator complex subunit 8
IQCK	IQ motif containing K
KANL	The Knight-ADRC-NIA-LOAD cohort
KAT8	Lysine acetyltransferase 8
KCTD2	Potassium channel tetramerization domain containing 2
LASA	The Longitudinal Aging Stydy Amsterdam
LD	Linkage disequilibrium
LDL	Low-density lipoprotein
LDSC	LD score regression
LMCI	Late mild cognitive impairment
LMDR	Logical memory delayed recognition
LOAD	Late-Onset Alzheimer's disease
LRLD	Long range linkage disequilibrium
MA	Mutation accumulation
MAF	Minor allele frequency
MANS	The Multidimensional Assessment of Neurodegenerative Symptoms questionnaires

MAP	The Memory and Aging Project
MAPT	Microtubule associated protein tau
MCI	Mild cognitive impairment
MCJ	Mayo clinic Jacksonville
MCR	Mayo clinic Rochester
MEF2C	Myocyte enhancer factor 2C
MFE-30	The Spanish version of the Memory Failures in Everyday Life Questionnaire
MIR142	MicroRNA 142
MKX	Mohawk homeobox
MMSE	Mini-Mental State Examination
MRI	Magnetic resonance imaging
MONICA	Monitoring trends and determinants in cardiovascular disease
MS4A4	Membrane spanning 4-domains A4A
MS4A4E	Membrane spanning 4-domains A4E
MS4A6A	Membrane spanning 4-domains A6A
NACC	National Alzheimer's Cordinating center
NBACE	The Fundacio ACE neuropsychological battery
NCT	Nicastrin
NDUFAF6	NADH:ubiquinone oxidoreductase complex assembly factor 6
NECAB1	N-terminal EF-hand calcium binding protein 1
NFT	Neurofibrillary tangles
NGS	Next generation sequencing
NHS	The Nurses Health Study
NIA	The National Institute of Aging
NIA-AA	The National Institute on Aging and Alzheimer's Association's
NINCDS-ADRDA	the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association
NINDS-AIREN	the Neuroepidemiology Branch of the National Institute of Neurological Disorders and Stroke convened an International Workshop with support from the Association Internationale pour la Recherche et l'Enseignement en Neurosciences
NME8	NME/NM23 family member 8
NOTCH4	Notch receptor 4
NPI-Q	Neuropsychiatric Inventory-questionnaire
NROH	Number of runs of homozygosity
NS	The Nun Study
NxC	The Neocodex-Murcia study
OARD1	O-acyl-ADP-ribose deacylase 1
OHI	Open House Iniciative
OPH	Osborne Park Hospital
PAHT	The Personality & Total Health through Life project
PCA	Principal Component Analysis
PCBD1	Pterin-4 alpha-carbinolamine dehydratase 1
PCR	Polymerase chain reaction
PET	Positron emission tomography
PFDN1	Prefoldin subunit 1
PGRS	Polygenic Risk Score
PICALM	Phosphatidylinositol binding clathrin assembly protein

PIHAT	Genome-wide estimates of identity-by-descent
PLCG2	Phospholipase C gamma 2
PLD3	Phospholipase D family member 3
POLK	DNA polymerase kappa
PPARGC1A	PPARG coactivator 1 alpha
Pr AD	Probable Alzheimer's diseae
PSEN1	Presenilin 1
PSEN2	Presenilin 2
Pss AD	Possible Alzheimer's disease
PTK2B	Protein tyrosine kinase 2 beta
PVE	Partial-volume effect
RABEP1	RAB GTPase binding effector protein 1
RELN	Reelin
RIN3	Ras and Rab interactor 3
RNA	Ribonucleic acid
ROH	Runs of homozygosity
RORA	RAR related orphan receptor A
ROS	The Religious Orders Study
ROSMAP	The Religious Orders Study and Memory and Aging Project
RP11-21L19.1	SPON1 antisense RNA 1 lnc RNA
RP11-678B3.2	RP11-678B3.2 processed pseudogene
RRM2B	Ribonucleotide reductase regulatory TP53 inducible subunit M2B
SAD	Sporadic Alzheimer's disease
SAS	Individuals from South Asian ancestry
SCD	Subjective cognitive decline
SCIMP	SLP adaptor and CSK interacting membrane protein
SKAT	SNP-set Kernel Associatio test
SKATO	The optimal unified test
SLC24A4	Solute carrier family 24 member 4
SNP	Single nucleotide polymorphism
SORL1	Sortilin related receptor 1
SPIDR	Scaffold protein involved in DNA repair
SPON1	Spondin 1
SPPL2A	Signal peptide peptidase like 2A
SROH	Sum of the total lenght of runs of homozygosity per individual
SUCLG2	Succinate-CoA ligase GDP-forming beta subunit
SUCLG2P4	Succinate-CoA ligase GDP-forming beta subunit pseudogene 4
SUVR	Standardized uptake values ratios
SVD	Small vessel disease
SYNE2	Spectrin repeat containing nuclear envelope protein 2
TF	TARCC (Texas Alzheimer’s Research & Care ConsortiuM) and FRONTIER (Facing Rural Obstacles to health Now Through Intervention, Education & Research)
TGEN	The translational genomics research Institute Study
TNF	Tumor necrosis factor
TNFAIP3	TNF alpha induced protein 3
TNFR1	TNF receptor superfamily member 1A protein

TNFRSF1A	TNF receptor superfamily member 1A gene
TNFRSF21	TNF receptor superfamily member 21
TNFRSF25	TNF receptor superfamily member 25
TNFSRF10A	TNF receptor superfamily member 10A
TNFSRF12A	TNF receptor superfamily member 12A
TOMM40	Translocase of outer mitochondrial membrane 40
TP53INP1	Tumor protein p53 inducible nuclear protein 1
TREM1	Triggering receptor expressed on myeloid cells 1
TREM2	Triggering receptor expressed on myeloid cells 2
TRIP4	Thyroid hormone receptor interactor 4
TWEAKR	Tumor necrosis factor receptor superfamily member 12A precursor protein
UBR5	Ubiquitin protein ligase E3 component n-recognin 5
UDS	The Uniform dataset protocol
UNC5B	Unc-5 netrin receptor B
UNC5C	Unc-5 netrin receptor C
USP6NL	USP6 N-terminal like
VaD	Vascular dementia
WES	Whole exome sequencing
WGS	Whole genome sequencing
WHO MONICA	World Health Organization Monitoring trends and determinants in cardiovascular disease Project
WMH	White matter hyperintensities
WWOX	WW domain containing oxidoreductase
ZCWPW1	Zinc finger CW-type and PWWP domain containing 1
ZNF282	Zinc finger protein 282
ZNF423	Zinc finger protein 423
5HTT	Solute carrier family 6 member 4

Publication at this thesis

Section 3.1

Sonia Moreno-Grau, O. Rodríguez-Gomez, A. Sanabria, A. Perez-Cordon, D. Sanchez-Ruiz, C. Abdelnour, S. Valero, I. Hernandez, M. Rosende-Roca, A. Mauleon, L. Vargas, A. Lafuente, S. Gil, M. A. Santos-Santos, M. Alegret, A. Espinosa, G. Ortega, M. Guitart, A. Gailhajanet, I. de Rojas, O. Sotolongo-Grau, S. Ruiz, N. Aguilera, J. Papasey, E. Martin, E. Peleja, Alzheimer's Disease Neuroimaging Initiative, F. Lomeña, F. Campos, A. Vivas, M.Gomez-Chiari, M.A. Tejero, J. Giménez, M. Serrano-Ríos, A. Orellana, L. Tarraga, A. Ruiz*, M.Boada, on behalf of the FACEHBI study group. **Exploring *APOE* genotype effects on Alzheimer's disease risk and amyloid β burden in individuals with subjective cognitive decline: TheFundacio ACE Healthy Brain Initiative (FACEHBI) study baseline results.** *Alzheimer's and Dementia*. May, 2018; 14(5), 634-343

Sonia Moreno-Grau, Isabel Hernández, Stefanie Heilmann-Heimbach, Susana Ruiz, Maitée Rosende-Roca, Ana Mauleó, Liliana Vargas, Octavio Rodríguez-Gómez, Montserrat Alegret, Ana Espinosa, Gemma Ortega, Nuria Aguilera, Carla Abdelnour, Alzheimer's Disease Neuroimaging Initiative, GR@ACE consortium, Silvia Gil, Wolfgang Maier, Oscar Sotolongo-Grau, Lluís Tárraga, Alfredo Ramirez, Jesús López-Arrrieta, Carmen Antúnez, Manuel Serrano-Ríos, Mercè Boada and Agustín Ruiz. **Genome-wide significant risk factors in chromosome 19 and the *APOE* locus**. *Oncotarget* May 15; 2018; 9(37): 24590–24600

Section 3.2

Sonia Moreno-Grau, Bruna Barneda, Paulina Carriba, Juan Marín, Oscar Sotolongo-Grau, Isabel Hernández, Maitée Rosende-Roca, Ana Mauleón, Liliana Vargas, Ana Espinosa, Montserrat Alegret, Octavio Rodriguez, Gemma Ortega, Maria Victoria Fernández, Jesús López-Arrieta, Lluís Tárraga, Mercè Boada, Carmen Antúnez, Joaquin López, Agustín Ruiz, and Joan Xavier Comella. **Evaluation of Candidate Genes Related to Neuronal Apoptosis in Late-Onset Alzheimer's Disease**. *Journal of Alzheimer's Disease*, 2015; 45(2), 621-629

Section 3.3

Sonia Moreno-Grau, Itziar de Rojas, Isabel Hernández, Inés Quintela, Laura Montrreal, Montserrat Alegret, Begoña Hernández-Olasagarre, Laura Madrid, Antonio González-Perez, Olalla Maroña, Maitée Rosende-Roca, Ana Mauleón, Liliana Vargas, Asunción Lafuente, Carla Abdelnour, Octavio Rodríguez-Gómez, Silvia Gil, Miguel Ángel Santos-Santos, Ana Espinosa, Gemma Ortega, Ángela Sanabria, Alba Pérez-Cordón, Pilar Cañabate, Mariola Moreno, Silvia Preckler[a], Susana Ruiz, Nuria Aguilera, Juan Antonio Pineda, Juan Macías, Emilio Alarcón-Martín, Oscar Sotolongo-Grau, GR@ACE/DEGESCO consortium, Alzheimer's Disease Neuroimaging Initiative, Marta Marquié, Gemma Monté-Rubio, Sergi Valero, Alba Benaque, Jordi Clarimón, Maria Jesus Bullido, Guillermo García-Ribas, Pau Pástor, Pascual Sánchez-Juan, Victoria Álvarez, Gerard Piñol-Ripoll, Jose Maria García-Alberca, José Luis Royo, Emilio Franco, Pablo Mir, Miguel Calero, Miguel Medina, Alberto Rábano, Jesús Ávila, Carmen Antúnez, Luis Miguel Real, Adelina Orellana, Ángel Carracedo, María Eugenia Sáez, Lluís Tárraga, Mercé Boada and Agustín Ruiz. **Genome-wide association analysis of dementia and its clinical endophenotypes reveal novel loci associated with Alzheimer's disease and three causality networks: the GR@ACE project.** *Alzheimer's and Dementia* 2019; Aug 13. pii: S1552-5260(19)35117-9

Section 3.4

Sonia Moreno-Grau, and Agustín Ruiz. **Autosomal Recessive Alzheimer's disease (arAD): homozygosity mapping of genomic regions containing arAD loci.** 2019. [In preparation]

Summary

Alzheimer's disease (AD) is the leading cause of dementia worldwide. Although clinical symptoms mainly appear in the elderly, it is suggested that AD is preceded by a long continuum, which includes preclinical and prodromal stages. AD presents a strong genetic component. From a genetic point of view, two forms of the disease have been recognized: the autosomal dominant form for early-onset AD (EOAD, <65 years old); and the polygenic form, mainly described for late-onset AD cases (LOAD, >65 years old). Nearly 40 genetic risk factors have been associated with LOAD. Of these, the Ɛ4 allele of the *APOE* gene was the first to be identified, and remains the major genetic risk factor for AD. In spite of extensive genetic research, the disease heritability remains largely unexplained, and the disease mechanisms incomprehensible.

This thesis aims at understanding the role of certain known AD loci and identifying new genetic risk factors and related pathways for AD.

First, we analyzed the role of a major AD loci, the *APOE Ɛ4* genetic risk factor, in preclinical stages of AD, and its relationship with the cerebral amyloid beta (Aβ) burden in the AD continuum. Our results supported the hypothesis that *APOE* genotypes can be used instrumentally to enrich at risk populations of AD. Furthermore, *APOE* genotypes explained 11% of brain amyloid variance, which supported the role of additional genetic and/or epigenetic factors.

Next, considering the emergence of controversial AD genetic findings, e.g. the rs3865444-*CD33* marker, in the close vicinity of *APOE* gene, we decided to analyze whether long-range linkage disequilibrium (LRLD) patterns in this important AD region could be disturbing additional genetic discoveries. We detected a weak and non-universal LRLD between the *APOE Ɛ2* allele and rs3865444 in *CD33* locus, but forces causing this LRLD remain elusive.

To uncover new AD genetic loci, we first applied a candidate gene approach, testing whether genes related with neuronal apoptosis were associated with AD. Our results did not show evidence of a genetic relationship between tested markers and AD.

After that, we moved to a hypothesis-free strategy. We generated the largest Spanish genome-wide dataset for dementia, the GR@ACE Stage I, and explored the impact of phenotypic heterogeneity on genetic findings and biological pathways to dementia. We detected three gene categories operating differently across AD subgroups of patients. After combining genomic and co-expression data, we identified vasculature regulation as a primary event in the causative mechanism for purer forms of AD. Meta-analysis of GR@ACE with additional genetic datasets

revealed two novel AD loci: *ANKRD31*-rs4704171 (in the *HMGCR* genomic region) and *NDUFAF6*-rs10098778, and confirmed *SCIMP*-rs7225151 and *CD33*-rs3865444 to be genome-wide significant. Our results indicate that, genetic discoveries can be strongly disturbed by the presence of clinical subgroups of AD patients.

Finally, to investigate genetic heterogeneity in AD, we looked for recessive variants in runs of homozygosity (ROHs). We first detected that increased homozygosity was present in AD cases respect to controls. Next, we efficiently captured inbred AD cases from an outbred population, which supposes a refined method to analyze inbreeding in AD. The use of complementary sequencing data led us to select ROHs harboring potential recessive variants. Of these, the *SPON1* gene appears to be a novel candidate AD locus with a potential recessive component, as it is a biological partner of *APP*. Potential recessive variants were found in the *CNTNAP2* and *TP53INP1/NDUFAF6* genetic regions previously associated with AD. Overall, our results further suggest that recessive effects may explain a portion of AD heritability.

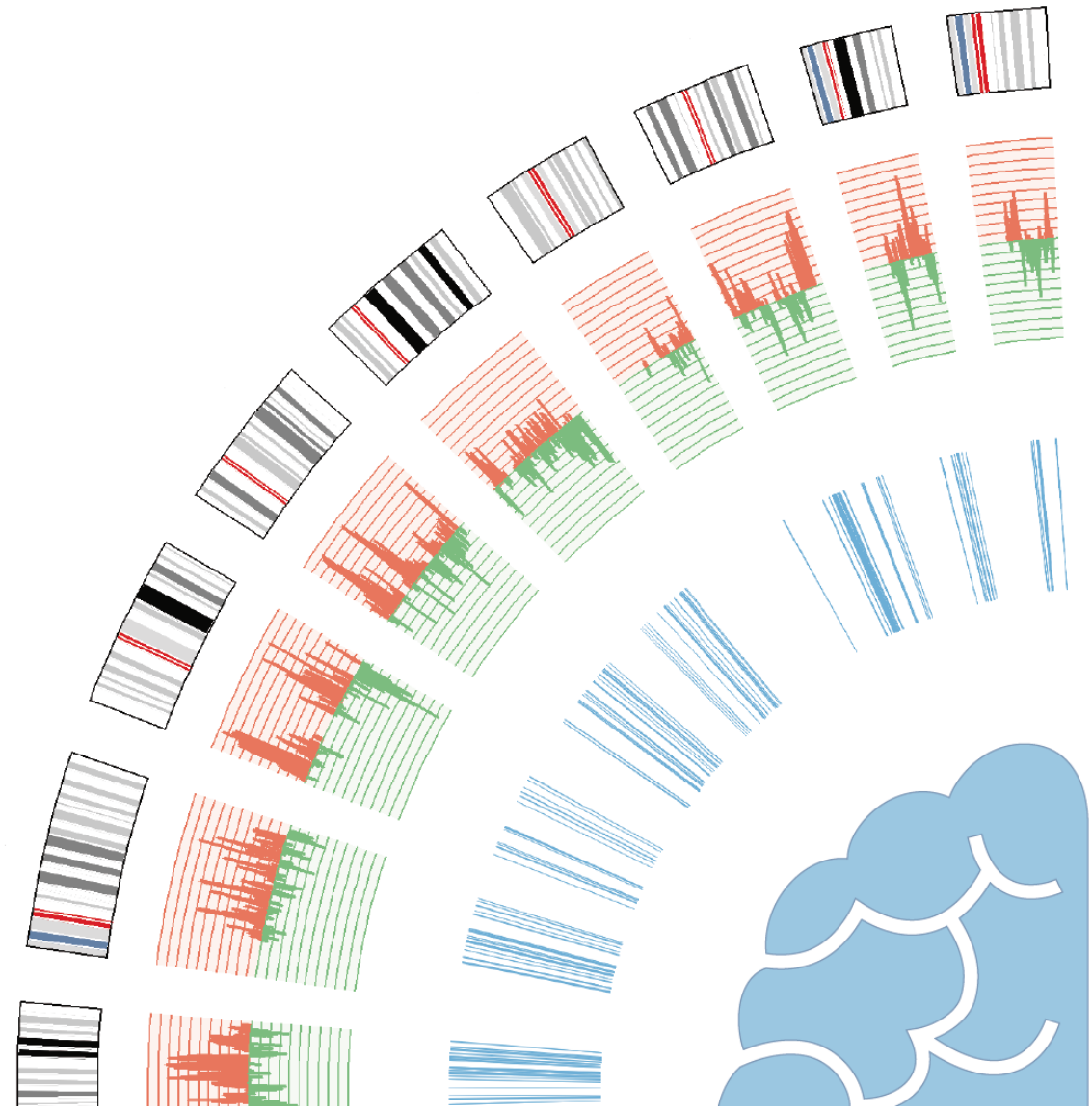
CHAPTER 1

INTRODUCTION

Modified from Moreno-Grau S, and Ruiz, A

The Genetic landscape of Alzheimer's Disease

Recent Research in Alzheimer's Disease. Volume 1. Page 1-16



The deep transformation of human societies during the 20th century has strongly affected the life expectancy of human populations. It is estimated that nearly 2 billion people will be over 60 years old in 2050 (1). Consequently, the prevalence of age-related neurodegenerative diseases will increase dramatically.

Dementia is a syndrome that mainly affects elderly people. It causes a disruption of a person's daily activities, and is ultimately fatal. The characteristic symptoms of dementia are difficulties with memory, language, problem-solving, and other cognitive skills, which differ from the symptoms of age-related cognitive decline (2). In 2015, more than 47 million people worldwide had dementia (3). These numbers have increased gradually, establishing this syndrome as a public health priority (4). Therefore, providing strategies to prevent, diagnose, and treat dementia seems essential.

1.1. Alzheimer's disease: the commonest form of dementia

Dementia has numerous causes. Alzheimer's disease (AD) is the most common one among them, accounting for 60 to 80% of all cases (5). In its typical presentation, AD is characterized by memory loss, i.e. forgetting recently learned information, but additional cognitive functions may become impaired first (6). Most AD patients begin with clinical symptoms after 65 years of age, a form of AD called late-onset AD (LOAD). A small portion (<10%) develop AD before 65 years of age, a type known as early-onset AD (EOAD) (7).

AD brains are pathologically defined by the accumulation of amyloid-beta (Aβ) peptides in the extraneuronal space (senile plaques) and in the walls of the cerebral vessels (cerebral amyloid angiopathy, CAA); by the intracellular accumulation of hyperphosphorylated tau protein (neurofibrillary tangles, NFTs); and by neuronal loss (8). Despite that, brain autopsies have demonstrated that most AD cases co-occur with other types of neuropathology, producing what is named mixed dementia (9), (10). In that sense, roughly 80% of clinical AD patients also present additional brain vascular pathology (11). Cerebrovascular pathology is suggested as a major risk factor for AD (12). The control of cardiovascular risk factors, as well as improvements in education level, has been shown to decrease AD incidence (13), (14), although its prevalence is steadily increasing (15). Despite that, there is a large controversy about the role of vascular factors in AD. I address this specific topic in **Section 3.3.**

The natural history of the disease is preceded by a long continuum (**Figure 1**). Several studies suggest that AD brain changes begin silently several years before clinical onset (2). Brain

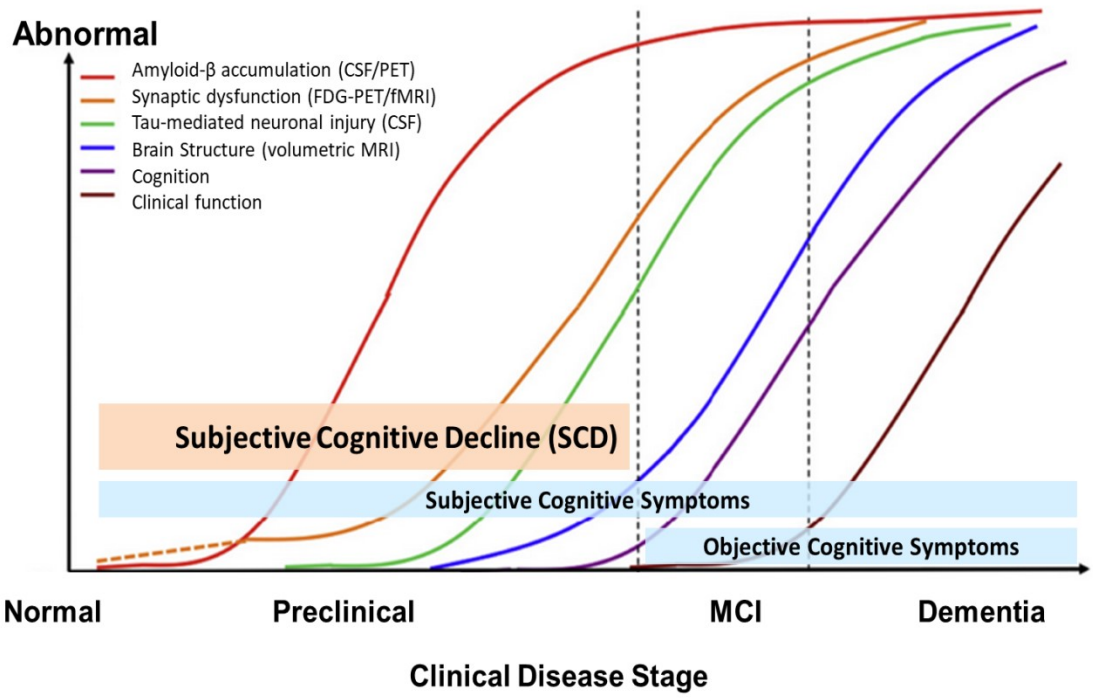


FIGURE 1. Alzheimer's disease continuum
MCI: mild cognitive decline. This figure has been modified from Jack et al. (16)

levels of Aβ peptides and tau protein are altered before the symptoms, as well as glucose metabolism, commonly used as a proxy of neurodegeneration. These brain alterations are used as disease endophenotypes or biomarkers, which can be measured with molecular neuroimaging techniques, e.g. positron emission tomography (PET), or quantified in cerebrospinal fluid (CSF).

Current research identifies three stages of Alzheimer's disease: preclinical AD, mild cognitive impairment (MCI), and dementia due to AD. Preclinical AD is seen in those individuals with measurable brain changes, but no evident cognitive decline. In that context, subjective cognitive decline (SCD), i.e. individuals with subjective memory concerns, have also been suggested as a cognitively healthy subpopulation at risk of suffering AD (17). Conversely, MCI individuals present objective cognitive decline without dementia (18) (**Figure 1**). Further description of SCD and MCI phenotypes is provided in **Appendix 1.** There is an increasing interest in the study of the clinical and molecular profiles of these subpopulations, as it could establish the basis for identifying converters to AD before the clinical onset.

There is still no effective treatment to cure AD or to slow AD progression, nor are there effective preventative interventions. In this scenario, the elucidation of the biological

mechanisms underlying AD is a priority, as it should lead to the identification of relevant drug targets and, ultimately, to the design of an effective therapy.

1.2. The role of genetics in Alzheimer's disease

Following advanced age, family history is the most important risk factor for AD. The risk of suffering from AD is greater for first-degree relatives of affected individuals (19). In fact, considering familial clustering, two forms of the disease are found in both EAOD and LOAD. Sporadic AD (SAD) is detected in an isolated case in a family, or in cases separated by more than three degrees of relationship, and familial AD (FAD) occurs in ≥2 individuals of a family with at least a third-degree relative (20). A third form, affecting a small percentage of familial EAOD cases (10–15%) (21),(22), occurs in ≥3 individuals in two or more generations, of whom two out of three are first-degree relatives (20), which supports an autosomal dominant mode of inheritance.

Thus, AD presents a strong genetic component. From a genetic point of view, two mechanisms for the disease are recognized, which do not fully correlate with age at onset: the autosomal dominant mechanism of EOAD and the polygenic mechanism of LOAD. Autosomal dominant AD is an infrequent Mendelian disorder, with an estimated prevalence of <1% (23). It is caused by highly penetrant mutations in three genes (see **Section 1.2.1**). However, familial EOAD cases most often present negative screening results for known Mendelian mutations, and roughly 40% are sporadic EOAD cases. The genetic basis underlying these cases remains elusive, but the presence of hidden familial cases, incompletely penetrant dominant mutations (e.g. mutations in *SORL1* (24),(25) and *ABCA7* (26),(27),(28) loci), or recessive genes have been proposed (29). The role of recessiveness in AD is assessed with further detail in **Section 1.2.5**. Either way, EOAD is almost exclusively a genetic disorder. Heritability estimations fall in a range of 92 to 100% (29). That is, nearly 100% of the factors causing the disease are genetic.

Conversely, the sporadic form of LOAD is often a complex disease, with a polygenic background (29), which represents roughly 75% of all AD cases. Although environmental factors might increase the risk of LOAD, studies in monozygotic and dizygotic pairs of twins have suggested genetics have a large influence (30),(31) (see **Section 1.2.3**). Nowadays, heritability estimations for LOAD vary from 13 to 70% (30),(31),(32).

1.2.1. Autosomal dominant mutations for familial EOAD: *APP*, *PSEN1*, and *PSEN2*

Initial genetic studies in familial EOAD were strongly influenced by neuropathological findings. In 1984, Aβ peptide was purified from cerebrovascular amyloid aggregates and senile plaques of AD brains (33),(34),(35). It was subsequently demonstrated that this peptide was encoded by the *APP* gene, mapped on chromosome 21 (36),(37),(38),(39). Some years later, an autosomal dominant mutation was detected in the *APP* locus, segregating in EOAD families (40),(41),(42),(43). The negative screening results for *APP* mutations in several families (44),(45) also suggested additional AD genes played a role. In fact, linkage studies pointed to two new genomic regions, 14q24.3 (46),(47),(48) and 1q31–q42 (49),(50). This enabled the identification of two new genes, *PSEN1* gene in 14q24.3 and *PSEN2* (51) in 1q31–q42.

To date, September 2019, more than 300 mutations have been identified in these three genes: 67 in *APP* (including gene duplications), 230 in *PSEN1*, and 39 in *PSEN2* (http://www.molgen.ua.be/ADMutations) (52). *PSEN1* mutations are the most frequent cause of familial EOAD, whereas *PSEN2* and *APP* are relatively infrequent. With the exception of two potential recessive mutation reported for *APP* (A673V and E693Δ) (53),(54), all of them presented an autosomal dominant mode of inheritance.

1.2.2. Understanding the amyloid hypothesis of AD

The *APP* gene encodes a ubiquitously expressed type 1 transmembrane protein. Its function is poorly understood. APP undergoes complex proteolytic processes. Two major canonical pathways are described: the non-amyloidogenic path and the amyloidogenic path (55). In the competing and physiologically predominant non-amyloidogenic pathway, APP cleavage generates the α-secretase-generated APP ectodomain fragment (APPsα), p3-related fragments, the C terminal fragment-α (CTFα), and the intracellular domain of APP (AICD). This process is mediated by α-secretases, e.g. ADAM10, the major α-secretase in the brain (56), and the γ-secretase complex, which includes five proteins: nicastrin (NCT), anterior pharynx defective 1 and 2 (APH1 and APH2), PSEN1, and PSEN2. By contrast, processing along the amyloidogenic pathway generates the Aβ peptides, the β-secretase-generated APP ectodomain fragment (APPsβ), CTFβ, and AICD through β-secretase (BACE1 and BACE2) and γ-secretase cleavage (**Figure 2**). Several species of Aβ oligomers have been described, with different lengths and biochemical properties. $A\beta_{1-40}$ has been identified mainly in brain blood vessels and $A\beta_{1-42}$ in senile plaques of AD patients. Mutations causing AD are mostly located near the β-proteolytic

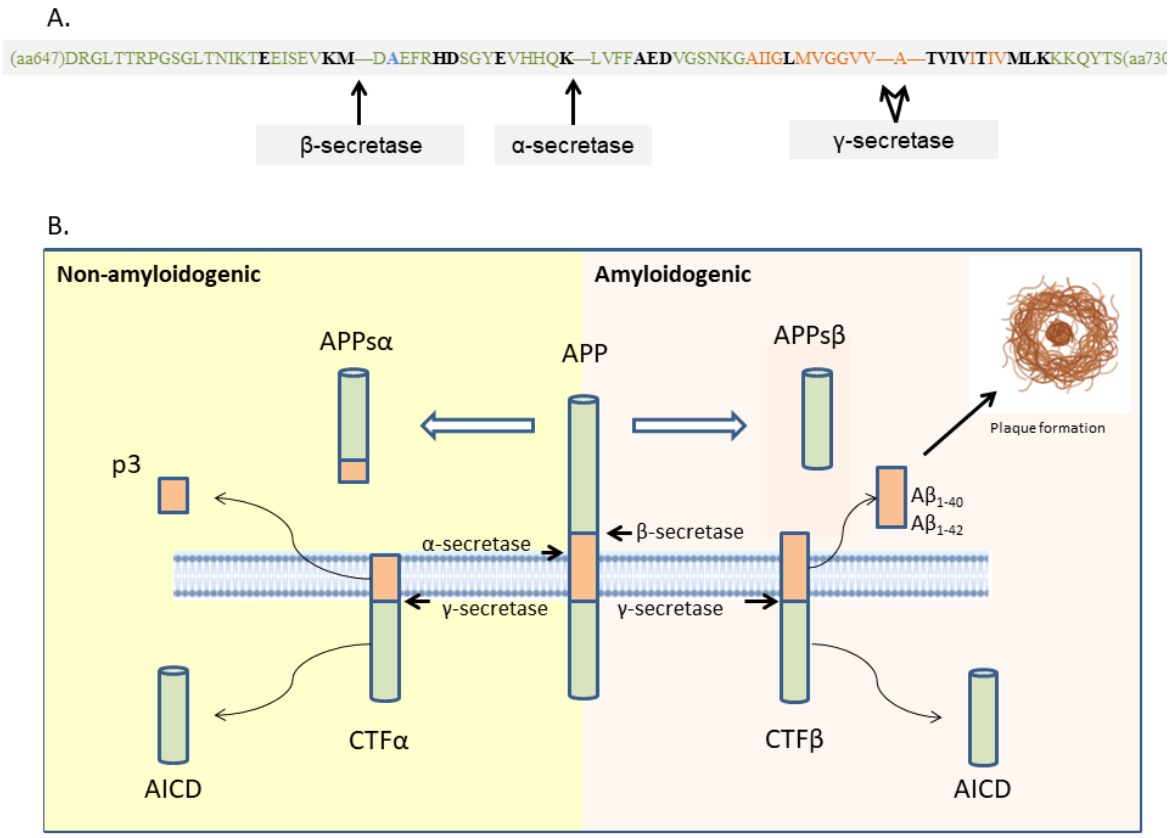


FIGURE 2. A. Point mutations for the APP protein. B. Proteolytic processes for APP
The green sequence represents the extracellular and intracellular domains; the orange sequence represents the transmembrane domain. Pathogenic dominant point mutations are highlighted in black and recessive ones in blue. Not included are deletions and duplications. This figure has been adapted from Müller et al. (55) and Cacace et. al (57)

cleavage site of APP or in presenilins, which strongly disrupt the physiological mechanism of APP processing, mainly increasing Aβ production (58) (**Figure 2**).

The amyloid hypothesis of AD claims that the accumulation of Aβ in the brain is the primary event in AD pathology (59). Glenner and Wong were the first to propose this hypothesis; they thought that the Aβ peptide had a central role in brain vasculature and suggested that AD is a form of cerebrovascular amyloidosis (34). Later, this hypothesis was reinterpreted, pointing to the accumulation of Aβ in senile plaques as main driver of the formation of neurofibrillary tangles, neuronal loss, vascular damage, and finally dementia (59).

1.2.3 The genetic etiology of late-onset Alzheimer's disease

1.2.3.1. A major genetic risk factor for late-onset Alzheimer's disease: the APOE locus

The Ɛ4 allele of the *APOE* gene is the strongest genetic risk factor of AD (60), (61), (62). In humans, the *APOE* gene is located on chromosome 19q.13.2 and encodes three major isoforms: APOE2, APOE3, and APOE4. These isoforms correspond to the *APOE* Ɛ2, Ɛ3, and Ɛ4 alleles, which arise from the variation in two nonsynonymous nucleotide polymorphisms in exon 4, rs429358 (C/T) and rs7412 (C/T) (63). The specific variation of the above-mentioned polymorphisms promotes two cysteine-to-arginine substitutions at positions 112 and 158 of the protein sequence, respectively. These polymorphisms have a major impact on structure and function (64). The *APOE* Ɛ2 and Ɛ3 alleles are derived from the ancestral allele Ɛ4 (65). Reported frequencies of *APOE* Ɛ2, Ɛ3, and Ɛ4 alleles in individuals of European ancestry are 0.11, 0.72, and 0.17 (66), respectively, but they are highly variable across human populations (67),(68).

Unlike the mutations in *APP*, *PSEN1*, and *PSEN2*, the *APOE* Ɛ4 allele is considered a genetic risk factor, due to it being neither necessary nor sufficient for provoking the AD phenotype. *APOE* acts in a dose-dependent manner in AD (69). Heterozygotes, carrying one copy of the *APOE* ε4 allele, have an increased AD risk (∼2–5 fold), while homozygotes, carrying two copies, have an approximately 12- to 15-fold increase in risk compared to non-carriers (70). Approximately 30% of the homozygote carriers will develop AD by the age of 75, and >50% by the age of 85 (70). Furthermore, *APOE* Ɛ4 is associated with an earlier age at onset (AAO) of the disease (71). In contrast, the Ɛ2 allele of *APOE* exerts a protective effect on AD (72), and has been linked with familial type III hyperlipoproteinemia (73).

Additional loci have been associated with AD in the vicinity of *APOE* on chromosome 19, e.g. *TOMM40* (74), *EXOC3L2*(75), *CD33* (76), and *PLD3* (77). Most of them have shown inconsistent replication results, and remain controversial (78),(79),(80). In that sense, this thesis has contributed to the analysis of the potential role of genes in the vicinity of the *APOE* locus (see **Section 3.1.2**). Our group also led the European project IMI2-ADAPTED (Alzheimer's disease Apolipoprotein Pathology for Treatment Elucidation and Development), which focused on elucidating the biological mechanism for the *APOE* locus in AD. The work in this thesis contributed in part to the project (unpublished results). Currently, we are testing whether ethnic differences modulate the risk conferred by *APOE* Ɛ4 allele to AD**.**

1.2.3.2. Candidate Gene Approach in Alzheimer's disease

Candidate gene approaches evaluate genetic variations in genes which carry out related functions together. More than 1000 scientific studies have applied this approach to AD, although most of them have failed or shown inconclusive results. A database containing all genes tested for association with AD was created in 2007 (http://www.alzgene.org/) (81). The *SORL1* (82), together with the *ACE* (83),(84), *ADAM10* (85) and *HLA-DRB1* (86), (87) genomic regions, could be considered the only successful findings emerging from the candidate-gene approach to this moment, as they have been confirmed using genome-wide strategies (See Section 1.1.3.4)

In this work, I also applied a candidate gene approach to test the involvement of genes related with neuronal apoptosis in AD (see **Section 3.2.1**).

1.2.3.3. Addressing the common disease–common variant hypothesis

The recognition of common genetic contributors to complex disease has been predicted by the common disease-common variant (CD-CV) hypothesis (88). This theory postulates that the genetic architecture of a complex disease can be explained by the combined effect of common risk variants (with minor allele frequency [MAF] > 0.05 in a population) acting under co-dominant genetic models, i.e. additive and multiplicative. In that scenario, each risk variant confers a small degree of risk, being neither necessary nor sufficient for causing the disease. This theory emerged in the light of quantitative genetics (89),(90),(91),(92), and has guided population genetics over the past decades.

Together with this theoretical framework, technological advances have made possible to launch large-scale genome-wide association studies (GWAS) (93). These technologies include the establishment of patterns of genome-wide variation and linkage disequilibrium across populations (94), the availability of dense genotyping chips, and the availability of large and well-characterized clinical samples. GWAS enable us to interrogate millions of markers (single nucleotide polymorphism, SNPs) for genetic associations with a specific phenotype. Nowadays, they efficiently capture common and low-frequency variants [MAF > 5%–1%]. A particular characteristic of GWAS is that they are hypothesis-free or agnostic, leading to novel discoveries, and not limited by incomplete understanding of disease pathology. Despite that, GWAS also have pitfalls. The genetic effects of common alleles are small, and the detection of genetic signals requires careful cleaning of raw data and very large sample sizes. Therefore, GWAS meta-analyses, also named meta-GWAS, have become a popular way to increase power and reduce false-positive findings (95). Furthermore, dealing with multiple comparisons

requires that stringent statistical significance thresholds are applied in detecting noteworthy SNPs, e.g. $P \leq 5\times10^{-8}$, which is considered genome-wide significance (GWS). This roughly corresponds to a simple Bonferroni correction for 1 million comparisons (96),(97). The conventional suggestive GWAS threshold was established between $5\times10^{-6} \leq P > 5\times10^{-8}$. It defines candidates consistent with the null hypothesis of no association, but with borderline GWS, which should be considered for additional studies (98).

1.2.3.4. Common genetic variation in late-onset Alzheimer's disease

Twenty years after the discovery of *APOE* Ɛ4, two large GWAS (N > 10,000) led to the identification of three additional LOAD loci: *PICALM*, *CLU*, and *CR1* (99),(100). In 2010, the first meta-GWAS for LOAD, integrating imputation procedures, was conducted (N∼30,000) (75). Imputation is a technique in which non-directly genotyped variants are inferred based on patterns of linkage disequilibrium (LD) from a specific reference population (101). It allows for the combination of studies genotyped with different genotyping platforms (95). Thus, *BIN1* genomic variants were associated with LOAD, and previously reported signals were replicated. Subsequently, the combination of large independent projects led to the doubling of the sample size (N∼60,000), increasing the power to detect small effects. In fact, five new genomic regions reached GWS (*ABCA7*, *MS4A4*/*MS4A4E, CD2AP, EPHA1*, and *CD33)* (76),(102). At the same time as these studies, *MS4A* gene cluster was confirmed in a cohort of Spanish individuals by our group (103).

In 2013, a large-scale international collaborative effort was consolidated. The International Genomic Alzheimer's Project (IGAP) mega-consortium comprises four genetic consortia: the Genetic and Environmental Risk in AD (GERAD), the European AD Initiative (EADI), Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE), and the AD Genetics Consortium (ADGC). We, as part of CHARGE, contribute to IGAP genetic research. The first work of the IGAP consortium included 74,046 individuals, and prompted the identification of 10 new GWS signals (*CASS4, SLC24A4-RIN3, FERMT2, HLA-DRB5/HLA-DRB1, INPP5D, MEF2C, PTK2B, CELF1, NME8*, and *ZCWPW1*) (104). All the previous LOAD loci were confirmed (*ABCA7, BIN1, CLU, CR1, CD2AP, EPHA1, MS4A6A-MS4A4E*, and *PICALM*), with the exception of *CD33*, which failed replication. This thesis has contributed to elucidation of the role of *CD33* markers in AD (see **Section 3.1.2** and **Section 3.3**). In addition, markers at the *SORL1* gene, previously identified by candidate gene approach, were confirmed.

Further re-analyses of IGAP data led to the identification of additional signals, i.e. *ATP5H* (105), *TRIP4* (106), and *HS3ST1* (107) loci. *TP53INP1*, *IGHV1-67*, *PPARGC1A, RORA*, and

ZNF423 loci (108)(109) were also detected, but using a gene-based approach for IGAP genetic results. A subsequent GWAS for *APOE* subgroups let the association of the *MAPT* genomic region with LOAD in *APOE Ɛ4* non-carriers (110). The *MAPT* locus encodes the tau protein, which tends to aggregate in neurofibrillary tangles in AD brains. It has also been associated with Parkinson disease and frontotemporal dementia, progressive supranuclear palsy, and corticobasal dementia (110). This finding was replicated by the Dementia Genetics Spanish Consortium (DEGESCO) (111), which is integrated with our group and others. DEGESCO represents an effort to disentangle the genetic architecture of neurodegenerative disease in the Spanish population.

In 2017, the application of trans-ethnic analytic approach promoted the identification of new genomic regions associated with the disease: *PFDN1/HBEGF, USP6NL/ECHDC3,* and *BZRAP1-AS1* (112). Finally, the use of a genome-wide association-by-proxy (GWAX) strategy, adopted by several independent groups, identified *HBEGF* (104), *ECHDC3* (104), *SPPL2A* (104), *SCIMP* (104), *ADAM10* (113), *BCKDK/KAT8* (113), *ACE* (113), *HESX1* (114), *CNTNAP2* (114), *APH1B* (114), *ALPK2* (114), and *AC074212.3* (114) loci associated with AD. Most of the new findings need independent replication.

Recently, we and others, as part of IGAP consortium, conducted the largest case-control meta-GWAS for LOAD published to date, which comprised 35,274 cases and 59,163 controls (N = 94,437). It confirmed 20 previously reported LOAD loci, identified five new genome-wide significant signals (*IQCK, ACE, ADAMTS1, WWOX*, and *OARD1*) (115), and replicated markers in *ADAM10* and *ECHDC3*, which were previously detected by alternative GWAS strategies (112),(116). *NDUFAF6*, *MIR142/BZRAP1-AS1*, *SPPL2, FST*, and *SUCLG2P4* variants remained as suggestive signals ($5\times10^{-6} \geq P > 5\times10^{-8}$). This thesis contributed significantly to the validation of several IGAP results. See Kunkle et al. (115). The most relevant genetic findings for AD to date are shown in **Figure 3.** Significant findings regarding the *SCIMP* and *NDUFAF6* genomic regions were made during the research for this thesis (**Section 3.3**).

1.1.3.5. Rare genetic variation in late-onset Alzheimer's disease

Although the genetic architecture of complex diseases has been predicted by the CD-CV model, strong evidence suggests that rare (MAF < 1%) large-effect alleles might be important contributors to such diseases (117). In recent years, the improvement in next-generation sequencing (NGS) technology, i.e. whole-exome or whole-genome sequencing (WES or WGS),

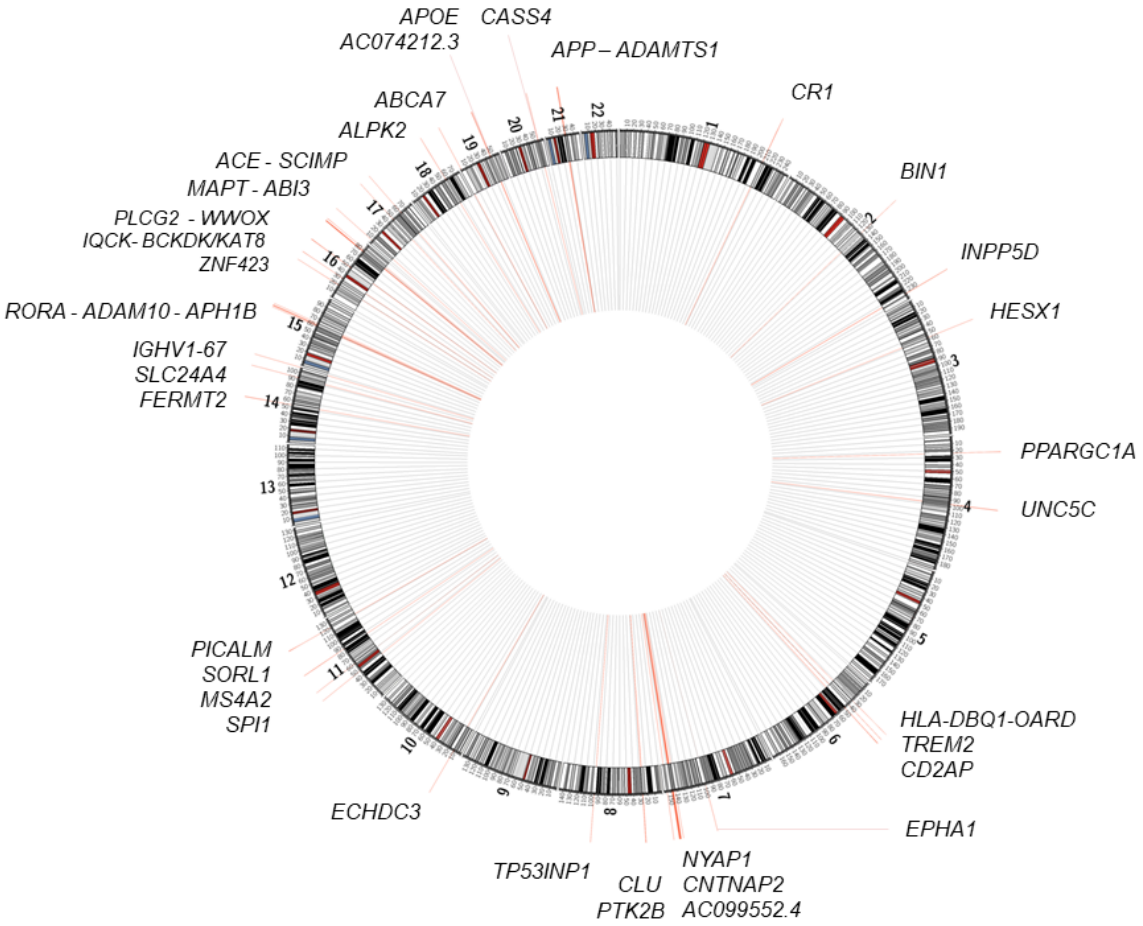


FIGURE 3. Loci associated with Late Onset Alzheimer's disease
The circular ideogram was created using Circos (118). Previously genome-wide loci not reaching significance in the lastest IGAP or follow-up studies are not included.

together with the construction of large haplotype references (119), has been providing high accuracy in tests for association at ever lower minor allele frequencies.

With this technology, new genetic variants near to or in *TREM2* (120), (121) ,(122), the strongest genetic risk factor after *APOE Ɛ4, ABI3* (122), *PLCG2* (122), *APP* (123), *UNC5C* (124) , *PLD3* (77), *AC099552.4* (125), *ABCA7* and *SORL1*(25), (26), (126), (127), (128) have been associated with AD (**Figure 3**). Although some of them are still under debate, as inconsistent results were shown for the association of the *PLD3* variant with sporadic LOAD (78), (79), (80), (129); others has been widely replicated. The Spanish consortium DEGESCO successfully replicated rare variants in the *TREM2* (130) and *PLCG2* (131) loci.

1.2.4. From genomics to biological pathways

The ultimate goal of gene discovery is the identification of biological pathways and critical processes underlying diseases and traits (117), (132). When the amyloid cascade hypothesis was

formulated for the first time (133), the genetic landscape of the most common AD phenotype, sporadic LOAD, was mostly unknown. Failures of clinical trials that were focused on Aβ pathology suggested that the amyloid hypothesis might be specific to Mendelian forms, or even that it might need a deeper reformulation. In that scenario, the application of pathway analysis to large GWAS results has brought new biological insights. The roles of immune and cholesterol pathways (134), (135), regulation of endocytosis and protein ubiquitination (135), or, more recently, the APP metabolism and tau-binding protein pathways, have been highlighted (115).

It should be noted that the power of pathway analysis is highly dependent on the quality and specificity of annotation, which can be questionable. In that context, it is suggested that the implementation of a more integrative or holistic approach might promote better comprehension of the disease mechanisms. In this thesis, I describe the role of a new biological pathway for AD using an integrative approach, i.e. also including data from gene co-regulatory expression networks, and considering clinical subgroups of AD patients (see **Section 3.3**).

1.2.5. Runs of homozygosity: A footprint of recessive inheritance in Alzheimer's disease

Mutations in *APP*, *PSEN1*, and *PSEN2* explained less than 10% of EOAD cases. However, EOAD has an estimated heritability close to 100% (29). In that context, it is suggested that autosomal recessive loci might account for most AD cases (~90%) (29), but nowadays, only two recessive mutations in the *APP* gene have been described for EOAD (53),(54). In the same way, current genetic findings for LOAD, discovered using an additive mode of inheritance, explained a limited part of disease heritability (~31%) (136). Several reasons for the missing heritability can be put forward: the lack of statistical power to detect uncharacterized variants with small effects, and the presence of substantial genetic heterogeneity. The genetic architecture of AD, as with most human diseases, is intricate, but the majority of studies have overlooked the role of recessive components underpinning the disease.

Long runs of homozygosity (ROHs), i.e. long stretches of consecutive homozygous genotypes (>1 Mb), are a recognized signature of recessive inheritance and consanguinity. Thus far, it is well known that inbreeding increases the incidence of recessive diseases (137). It increases the chance of harboring the same mutation as a recent common ancestor, arriving at the index case from both parents simultaneously. These recessive alleles are usually embedded in long segments of homozygosity, and, traditionally, have been identified using a homozygosity mapping strategy (138). The recent parental relatedness points to genuine regions of

autozygosity. Population history, e.g. historical bottlenecks or geographical isolation, also influences individual runs of homozygosity (139),(140).

A few studies have tried to analyze homozygosity in AD. Farrer and colleagues (141) studied 183 AD families in the isolated Wadi Ara region (an area in Israel populated mainly by Arab citizens of the country). Wadi Ara people has increased parental relatedness, and one of the highest prevalence rates of AD in the world. The authors presented evidence supporting the existence of potential recessively acting loci on chromosome 9 (141). Follow-up of disease-associated homozygous regions showed a strong effect and nominal significance for markers in the *AGER/NOTCH4/AGPAT1* genomic regions in chromosome 6 (142).

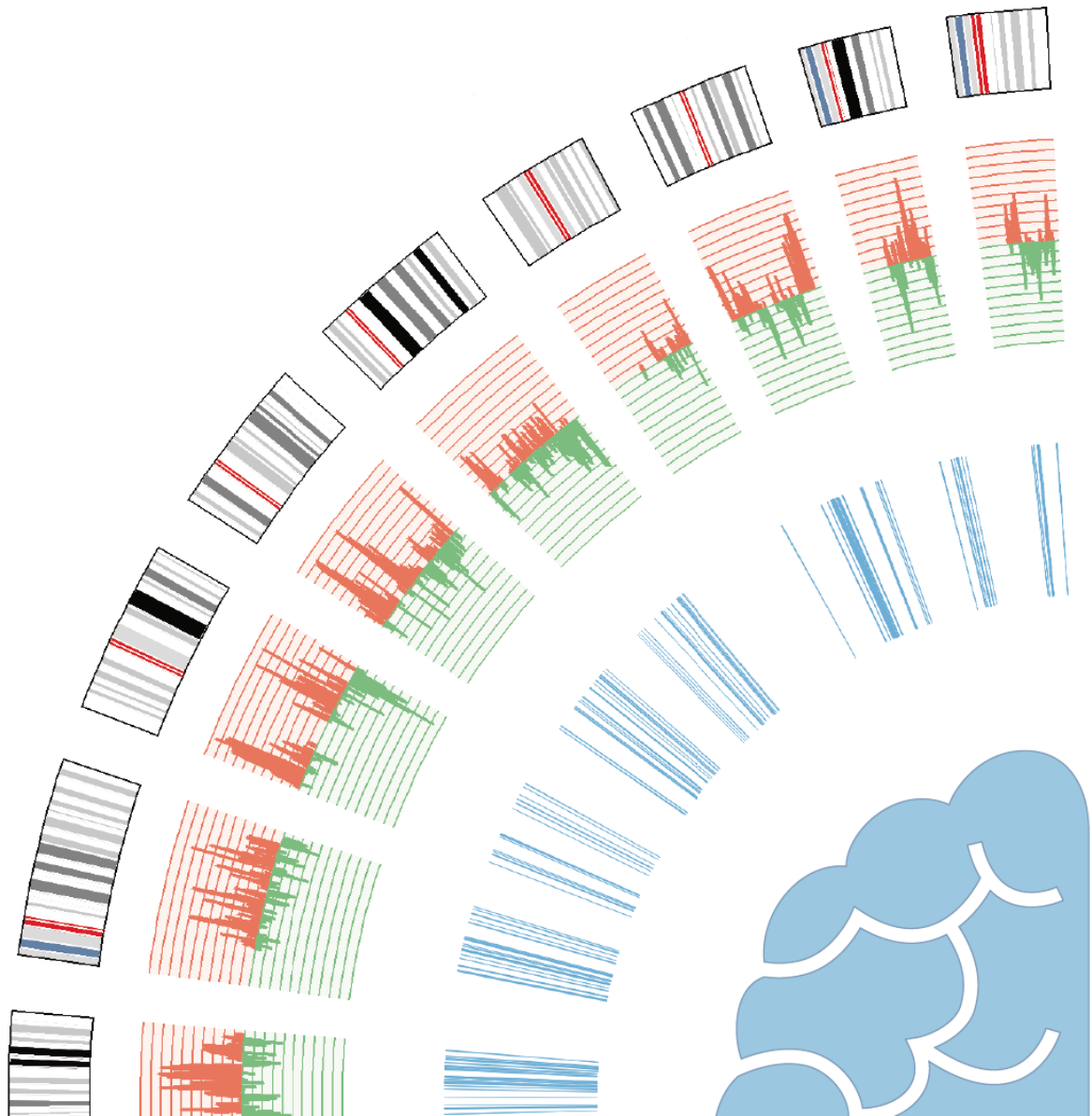
Additionally, the homozygosity mapping in a consanguineous EOAD family of European ancestry, and subsequent sequencing of candidate regions in affected individuals, pointed to the *CTSF* gene as a potential recessive locus (143),(144). Mutations in the *CTSF* locus cause type B Kufs disease, a neurodegenerative disease characterized mainly by the abnormal accumulation of lipopigments in lysosomes. Thus, further neuropathological explorations in affected individuals will be required to discard Kufs pathology before AD pathology can be accepted.

Recently, it has been demonstrated that ROHs are also found ubiquitously even in outbred populations (145),(146). In fact, an increased number of ROHs has been reported in AD cases compared to controls in Caribbean Hispanic and African-American individuals from outbred populations (147),(148). Despite that, this association has not been extended to European populations (149),(150). There might be several reasons for this. First, ROH patterns are population specific. Second, it has been estimated that large sample sizes (12,000–65,000) are required to detect an excess of homozygosity in outbred populations (151). Thus, previous studies have been severely underpowered. Additonal reasons have made the identification of recessive loci for AD challenging. Further follow-up of candidate ROHs in sequencing data is a necessary step for definitively mapping a genuine recessive locus, but it has been poorly assessed. There are a limited number of deeply characterized consanguineous families, due to difficulties in finding familial information for sporadic AD individuals.

Assessing the impact of inbreeding on the genetic architecture of AD remains a challenge. Consequently, delineating the scale of inbreeding to identify regions with potential recessive effects in AD seems a pertinent step. In this work, we present new findings emerging after a comprehensive exploration of ROHs in AD. The largest study reported to date for this specific hypothesis (**see Section 3.4**).

CHAPTER 2

HYPOTHESIS AND AIMS



The evolution of genetic studies from the first linkage studies to current GWAS or large sequencing efforts has determined a portion AD heritability. Despite that, a large percentage of the genetic variance for Alzheimer's disease remains unexplained. Thus, the general aim of this thesis is to contribute to the identification of the genetic basis underlying AD.

The specific hypothesis and aims of each section are defined as follows:

Section 3.1 analyzes the role of *APOE Ɛ4*, the strongest genetic risk factor for AD, in the continuum of AD and in relation to other AD genetic findings. It includes two subsections:

Section 3.1.1 studies the effect of *APOE Ɛ4* in a new cohort of Spanish SCD individuals, the FACEHBI population, and its relationship with cortical Aβ accumulation.

In recent years, the study of the biological characteristics of individuals in preclinical stages of AD has drawn much attention. It supposed that such study will bring new opportunities to detect individuals converting to AD before the clinical onset of dementia syndrome. Accordingly, we generated a new dataset of SCD individuals to explore the biomarkers, cognition and lifestyle of these subjects, the FACEHBI population. Our first work with this population hypothesized that if this population really comprised a pool of individuals who will come to have AD, the *APOE Ɛ4* allele might contribute to the risk of SCD, and might be involved in brain changes prior to disease onset.

Appendix 1 is a review in which the genetics for MCI and SCD populations is addressed. In this dissertation, we also performed the first meta-analysis evaluating the role of *APOE* Ɛ4 in the risk of conversion from healthy status to SCD and MCI. Thus, **Appendix 1** forms a supplementary part of **Section 3.1.1**.

Section 3.1.2 evaluates the existence of long-range linkage disequilibrium (LRLD) patterns in the *APOE* genomic region, and their impact on other AD genetic findings mapped in chromosome 19, where the *APOE* co-localizes with *ABCA7* and *CD33* loci.

The *APOE* gene is located in a complex genomic region. Since the discovery of the *APOE* locus as an AD genetic risk factor, additional loci have been linked with AD across this genomic region. Despite that, most of them showed inconsistent replication results. We hypothesize that differential LRLD patterns across different populations could be affecting the *APOE* region and genetic signals in its neighboring loci, such as the *CD33* locus.

Section 3.2 investigates the effect of genes related to neuronal apoptosis in AD by applying a candidate gene approach.

AD is characterized by neuronal death and progressive neuroinflammation (2). Although the mechanisms of neuronal death in AD are not fully understood, it has been suggested that the tumor necrosis factor (TNF) is involved in mediating programmed cell death, i.e. apoptosis through the death receptor pathway (152). Thus, we hypothesize that the genetic variation in genes controlling apoptotic processes via death receptors, including their cognate ligands and anti-apoptotic molecules, could be mediating a part of the genetic risk for AD.

Section 3.3 applies a hypothesis-free strategy to uncover new genetic risk variants and pathways associated with AD phenotype. Although the emergence of new genomic technologies and the availability of larger datasets for AD have enabled the discovery of new genetic findings, to date, the most of AD heritability remains unknown. We hypothesize that to identify new genetic risk factors and pathways for AD, the following are necessary: 1) generating new genomic resources for the disease; and 2) exploring the impact of the pathological heterogeneity of AD in genetic findings.

As part of the current work, we have generated the largest Spanish GWAS for AD, the GR@ACE Stage I project. **Appendix 2** explores the technical aspects of the new GWAS dataset. It evaluates resolution in imputation procedures of the GR@ACE genotyping chip with respect to other commercial alternatives. **Section 3.3.1** has three major goals: presenting GWAS results from GR@ACE Stage I; addressing the impact of clinical variability in genetic discoveries; and testing whether similar biological pathways operate in different subgroups of AD patients.

Section 3.4 studies the role of recessive inheritance in AD.

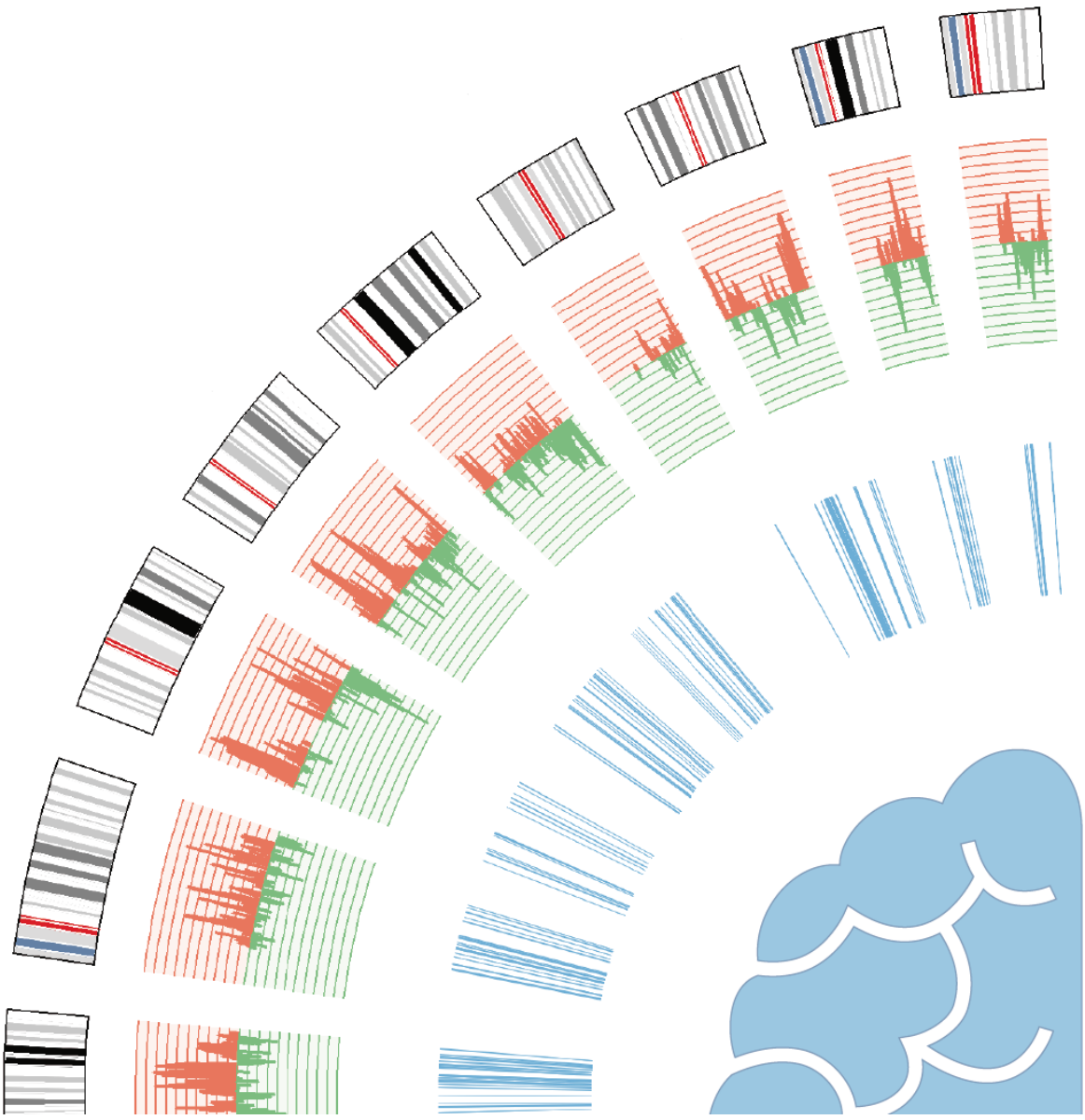
We hypothesize that new recessive genetic variants could emerge with the exploration of runs of homozygosity in AD. With the objective of identifying a novel recessive locus for AD, we propose prioritizing ROHs emerging from inbred individuals captured from an outbred population.

Chapter 4 summarizes the main findings of this thesis and put them in the broader of Alzheimer disease context.

Chapter 5 provides the final conclusions of this thesis.

CHAPTER 3

PUBLICATIONS



Section 3.1. *APOE*, The strongest genetic risk factor for AD

3.1.1. Publication I

Exploring *APOE* genotype effects on AD risk and beta-amyloid burden in individuals with subjective cognitive decline: the FACEHBI study baseline results.

Sonia Moreno –Grau[a], O. Rodríguez-Gomez[a], A. Sanabria[a], A. Perez-Cordon[a], D. Sanchez-Ruiz[a], C. Abdelnour[a], S. Valero[a,b], I. Hernandez[a], M. Rosende-Roca[a], A. Mauleon[a], L. Vargas[a], A. Lafuente[a], S. Gil[a], M. A. Santos-Santos[a,c], M. Alegret[a], A. Espinosa[a], G. Ortega[a], M. Guitart[a], A. Gailhajanet[a], I. de Rojas[a], O. Sotolongo-Grau[a], S. Ruiz[a], N. Aguilera[a], J. Papasey[a], E. Martin[a], E. Peleja[a], Alzheimer's Disease Neuroimaging Initiative*, F. Lomeña[d], F. Campos[d], A. Vivas[e], M.Gomez-Chiari[e], M.A. Tejero[e], J. Giménez[e], M. Serrano-Ríos[f,g], A. Orellana[a], L. Tarraga[a], A. Ruiz[a*], M.Boada[a], on behalf of the FACEHBI study group.

[a]Research Center and Memory Clinic of Fundació ACE, Institut Català de Neurociències Aplicades, 08028 Barcelona, Spain [b]Department of Psychiatry, Hospital Universitari Vall d'Hebron, CIBERSAM, Universitat Autònoma de Barcelona, Barcelona, Spain [c]Cognition and brain plasticity group [Bellvitge biomedical research institute- IDIBELL], L'Hospitalet de Llobregat, Barcelona, Spain [d]Servei de Medicina Nuclear, Hospital Clínic i Provincial Barcelona, Spain [e]Departament de Diagnòstic per la Imatge Clínica Corachan, Barcelona, Spain [f]CIBERDEM-Spanish Biomedical Research Centre in Diabetes and Associated Metabolic Disorders, Madrid, Spain [g]Instituto de Investigación Sanitaria del Hospital Clínico San Carlos (IdISSC), Madrid, Spain

---

ABSTRACT

Introduction: Subjective cognitive decline (SCD) has been proposed as a potential preclinical stage of Alzheimer's disease (AD). Nevertheless, the genetic and biomarker profile of SCD individuals remains mostly unexplored.

Methods: We evaluated *APOE* Ɛ4's effect in the risk of presenting SCD, using the FACEHBI SCD cohort and Spanish controls, and performed a meta-analysis addressing the same question. We assessed the relationship between *APOE* dosage and brain amyloid burden in the FACEHBI SCD and ADNI cohorts.

Results: Analysis of the FACEHBI cohort and the meta-analysis demonstrated SCD individuals presented higher allelic frequencies of *APOE* Ɛ4 with respect to controls. *APOE* dosage explained 9% (FACEHBI cohort) and 11% (FACEHBI and ADNI cohorts) of the variance of cerebral amyloid levels.

Discussion: The FACEHBI sample presents *APOE* Ɛ4 enrichment, suggesting that a pool of AD patients is nested in our sample. Cerebral amyloid levels are partially explained by the *APOE* allele dosage, suggesting that other genetic or epigenetic factors are involved in this AD endophenotype.

Keywords: Subjective cognitive decline, Alzheimer's disease, *APOE* alleles, amyloid burden, PET.

Alzheimer's and Dementia, May, 2018; 14(5), 634-343

doi: 10.1016/j.jalz.2017.10.005

INTRODUCTION

Neuropathological changes of Alzheimer's disease (AD) evolve several years before the onset of clinical symptoms (1),(2). Therefore, it is believed that characterization of earlier stages of AD, i.e. mild cognitive impairment (MCI) and subjective cognitive decline (SCD), could beget new strategies to diagnose and treat the disease earlier (3).

A SCD population represents a subset of cognitively normal individuals with self-reported cognitive impairment (4). It is suggested that a prodromal AD subgroup could be nested in a SCD population. Nevertheless, there is a scarcity of research studying both genetic and biomarker profiles of SCD individuals. New studies might help to improve the identification of those SCD individuals at risk of AD.

Thus far, presence of the *APOE* Ɛ4 allele, the mayor genetic risk factor for AD (5), has been the only genetic marker associated with risk of suffering SCD (6). In a previous meta-analysis with 6,824 individuals, we were able to estimate that *APOE* Ɛ4 was significantly associated with risk of suffering SCD [OR = 1.15 (1.02-1.30); p = 0.03] (6). Although the contribution of other genes has been explored, i.e. *IL1B* or *TNF* (7), the risk of presenting a SCD diagnosis or of converting from SCD to AD has not been associated with other genetic signals (7),(8).

Traditional AD-biomarker research has been focused on assessing cerebral amyloid aggregation and neuronal pathology, both of which are considered classical neuropathological hallmarks of AD (9). It has been demonstrated that *APOE* alleles contribute to the biological modulation of Aβ clearance (10). Furthermore, genome wide association studies (GWAS) evaluating the cerebral amyloid burden endophenotype have reinforced *APOE's* role in amyloid accumulation (11), (12). Despite the inextricable link between *APOE* and Aβ burden in AD patients, none of the reported endophenotype GWAS included SCD cohorts (11), (12). Thus *APOE's* role in amyloid burden in the SCD population remains poorly examined.

Several studies have detected higher cerebral amyloid burden in AD and MCI patients compared to cognitively normal individuals (13), (14), (15) as well as in AD compared to MCI patients (13), but studies evaluating amyloid pathology in SCD have provided inconsistent results (16), (15). However, when *APOE* genotypes are considered, more predictable results are observed, i.e. SCD individuals who are *APOE* Ɛ4 carriers have shown higher amyloid PET uptake than SCD or healthy control Ɛ4 non-carriers (17). Understanding genetic and biomarker profiles, as well as their interaction, in SCD individuals, will allow integration of multiple variables influencing AD risk i.e.

APOE status, age and amyloid burden, and enhance the discrimination of subjects at risk of conversion to AD.

Here we analyse *APOE* genotypes, baseline cerebral amyloid profile and their relationship in the Fundacio ACE Healthy Brain Initiative (FACEHBI) cohort, which included 200 SCD individuals enrolled in a long-term longitudinal study of cognition, biomarkers and lifestyle (18). We evaluated *APOE* Ɛ4's effect in the risk of presenting SCD using a case-control design including the FACEHBI SCD cohort and a cohort of Spanish population-based controls and afterwards by performing a meta-analysis including studies addressing the same question. To explore whether SCD individuals who carry the *APOE* Ɛ4 allele present increased risk of suffering MCI or AD compared to non-carrier SCD individuals, we calculated odds ratios using Fundacio ACE SCD, MCI, and AD cohorts. To explore the relationship between *APOE* and brain amyloid burden, first we analysed the FACEHBI data and then extended the analysis by including an independent cohort from the Alzheimer disease neuroimaging initiative (ADNI) study. Finally, we explored whether the effect of *APOE* dosage on brain amyloid burden was homogenous across clinical diagnoses (controls, SCD, MCI, AD) using the FACEHBI and ADNI cohorts.

SUBJECTS AND METHODS

Subjects

With the objective to explore the effect of *APOE* Ɛ4 in risk of presenting SCD we used 200 SCD individuals recruited from the FACEHBI study and 3,032 population-based controls recruited by a cross-sectional epidemiological survey, described below. Additionally, we sought to estimate the risk conferred by this genotype in SCD population to suffer MCI or AD, using the FACEHBI sample, the ACE MCI cohort (1,170 MCI patients) and the ACE AD cohort (2,517 AD patients), all of them recruited by Fundacio ACE (Supplementary Figure 1). To avoid population stratification, all individuals were selected to be of white Mediterranean ancestry with registered Spanish ancestors (for two generations).

The FACEHBI cohort

The FACEHBI cohort comprises 200 individuals diagnosed with SCD (mean age, 65.8 ± 7.1 years; 62.5% women), which are embedded in a long-term observational study (18). The sample has been obtained from two different sources: individuals referred by their physicians to our memory clinic for study of cognitive impairment and individuals who came to our institution through an Open House Initiative (OHI). SCD was defined as the coexistence of cognitive complaints and a score of ≥8 on MFE-30, the Spanish version

of the Memory Failures in Everyday Life Questionnaire (20); MMSE ≥ 27; CDR=0; and performance on Fundació ACE Neuropsychological Battery (NBACE) (19) within the normal range for age and educational level. Further description of inclusion and exclusion criteria is provided by Rodriguez-Gomez, et.al.(18). All participants gave written consent and the protocol was approved by the ethics committee of the Hospital Clinic i Provincial (Barcelona, Spain) (EudraCT: 2014-000798-38).

All subjects were screened at baseline for brain amyloidosis with Florbetaben [18F] radio tracer using PET (FBB-PET). PET images were acquired after administration of single slow intravenous bolus (6 sec/mL) of 300Mbq of FBB (NeuraCeq©), in a total volume of up to 10mL, during 20 minutes.

Genomic DNA was obtained from 200µl of human whole blood using commercial methods. High resolution melting procedures were performed to determine *APOE* genotypes. PCR reaction were carried out in a final volume of 5µl, using 11ng of genomic DNA, 0.3µM of each primer and 2.65µl of 2X SYBR Fast Master Mix (Kapa Biosystems). PCR conditions were a denaturation step at 95 °C for 2 min, 33 cycles at 95 °C for 10 s, and at 69 °C for 30 s. Melting curves were 95 °C for 15 s (ramping rate 5.5 °C s-1), 45 °C for 15 s (ramping rate of 5.5°C s−1) and 95 °C for 15 s (ramping rate of 5.5°C s−1). Fluorometric register was performed at one acquisition register per each degree Celsius. Melting peaks and genotype calls were obtained using the Eco Real-Time PCR system (Illumina).

The ACE MCI cohort

We included a sample of 1,170 MCI patients (mean age, 75.9 ± 7 years; 64.5% women) recruited and assessed at the Fundació ACE Diagnostic Unit (Barcelona, Spain) between January 2006 and July 2013. A diagnosis of MCI was assigned according to the Petersen criteria (20), (21) and the classification of Lopez et al. (22), (23) All subjects had a CDR of 0.5, and were assessed using the MMSE; the Hachinski Ischemia Scale; the BDRS; and the Neuropsychiatric Inventory Questionnaire (NPI-Q). DNA was extracted using standard procedures and conventional real-time PCR procedures (Applied Biosystems) were used to obtain *APOE* genotypes. See supplementary material from Lacour, et al. (24) for detailed information.

The ACE AD cohort

We included 2,517 AD cases (mean age, 81.6 ± 16.3 years, 70.9% women), which were referred for evaluation of cognitive impairment by their primary care physicians or primary care neurologist. Diagnosis of dementia and type of dementia are established by consensus according to

DSM-IV criteria for dementia and NINCDS-ADRDA criteria for possible or probable AD. Further information of inclusion criteria and genotyping procedures are provided by Seshadri, et.al. (25) and Boada, et. al. (26).

Population-based controls

We used 3,032 Spanish controls (mean age 54, ± 11.7 years, 61.8% women) with unknown cognitive status recruited from a cross-sectional population-based epidemiological survey to describe the prevalence of cardiovascular risk factors in the general population, previously described (27), (28).Survey procedures were adapted from the World Health Organization MONICA Project (WHO MONICA) protocol. DNA extraction and *APOE* genotyping procedures were previously described by Seshadri, et.al.(25).

The ADNI series

To validate our correlation analysis between *APOE* genotypes and brain amyloid in the FACEHBI cohort in an independent dataset, we used 182 healthy controls (HC), 103 SCD, 460 MCI and 144 AD participants with available amyloid PET and *APOE* data from the ADNI study (http://adni.loni.usc.edu/) (29). Informed consent was obtained according to the Declaration of Helsinki. The ADNI PET core processes Florbetapir [18F] (AV45) PET images according to previously described methods(30).

Statistical analysis and Meta-analysis

1. Risk analysis of the APOE locus in SCD

To investigate the effect of carrying an *APOE* Ɛ4 allele in the risk of presenting SCD we performed allelic frequency comparisons with a Chi-square test between 200 SCD individuals and 3,032 Spanish population-based controls. Similarly, to explore whether SCD individuals carriers of *APOE* Ɛ4 present increased risk of suffering MCI or AD we calculated odds ratios using Fundacio ACE SCD, MCI, and AD cohorts. Logistic regression analysis (additive model) was used to perform adjustments per: 1) gender; and 2) gender and age. All statistical analyses were performed using PLINK 1.9 software (http://www.cog-genomics.org/plink2/)(31).

In addition, meta-analysis techniques were used to estimate *APOE* Ɛ4's effect in the risk of presenting SCD. We have updated a previous meta-analysis conducted by our group, which included studies published before 2015 and a total of 6,824 individuals(6), with studies published before July 2017. Literature search in Pubmed was performed using *APOE* and SCD terms. We selected the studies meeting the following criteria: (1) case/control studies or longitudinal studies where it is possible to distinguish a sub-population of SCD individuals and a sub-population of healthy controls; (2) studies that provide a complete definition of the participants; (3) studies that provided an

odds ratio (OR) with 95% confidence interval (CI) as well as the p-value or provide sufficient data to calculate them. Studies with overlapping samples were excluded. After the incorporation of data from the FACEHBI cohort, a total of 12,183 individuals were included (Supplementary Figure 2). SCD definition and recruitment strategy for each included study has been detailed in Supplementary Table 1. Effects were determined using the inverse variant method (fixed-effects model). In the case of heterogeneity, the DerSimonian and Liard method (random-effects model) was used. Heterogeneity was considered significant when $I^2 > 50\%$ and $p < 0.05$. Pooled effects, forest and funnel plots were obtained using Metafor package from R.

2. Effect of the APOE locus on brain amyloid burden

We used baseline individual standardized uptake values ratios (SUVR) from FACEHBI and ADNI. First, we tested for normality of FBB and AV45 global SUVR measures and found they were not normally distributed. Thus, we decided to log-transform the data to conduct the consecutive analyses. *APOE* genotypes were codified according to the Ɛ4 allele dosage, i.e. Ɛ4Ɛ4 = 2; Ɛ3Ɛ4=1; Ɛ3Ɛ3 = 0; Ɛ2Ɛ4 = 0; Ɛ2Ɛ3 = -1; Ɛ2Ɛ2 = -2. The present analyses were carried out using SPSS 20.0 (SPSS Statistics 20, IBM Corporation, Somers, NY).

2.1. Global SUVR according to APOE dosage in the FACEHBI cohort

We performed an ANOVA to look for differences in global SUVR across *APOE* dosage (5 groups) in the FACEHBI cohort. Post-hoc analysis, including Bonferroni correction for multiple comparisons were performed to check differences between groups. *APOE* Ɛ2Ɛ2 genotype was not considered in post-hoc testing because of its small sample size (n=1). Ggplot2 package from R was used to depict differences between groups.

2.2. The relationship between APOE dosage and brain amyloid burden

We sought to estimate the relationship between *APOE* Ɛ4 allele dosage and brain amyloid burden in the FACEHBI SCD cohort and in an independent cohort from ADNI. The ADNI cohort included individuals with different clinical diagnoses (HC n=182, SCD n=103, early MCI n=303, late MCI n=157, and AD n=144) to explore whether the relationship changed across different clinical categories. Since FACEHBI and ADNI use different radioactive tracers, we decided to standardise log-transformed amyloid SUVR to be able to compare results. We used linear regression model with SUVR as the dependent variable and *APOE* dosage as the independent variable, to check this correlation. Ggplot2 package from R was used to depict the individual slopes per

status. Finally, two additional methods were used to explore whether there was a homogeneous correlation across differential clinical categories; 1) Meta-analysis of the effect of *APOE* Ɛ4 on brain amyloid burden across different clinical diagnoses; criterions for selection of the meta-analyses model was described previously. Correlation coefficients were pooled using Meta packages from R; and 2) the evaluation of an interaction term using the general linear model across the entire sample (n = 1,089). Clinical categories were coded as follows, HC = 0; SCD = 1; EMCI = 2; LMCI = 3; AD = 4.

We examined the relationship between age and brain amyloid burden using the same analyses described for *APOE* dosage. Age at baseline was used for the present analysis.

RESULTS

Risk analysis of *APOE* locus in SCD

APOE markers (rs7412 and rs429358) followed the Hardy Weinberg equilibrium. Enrichment in the allelic frequencies of *APOE* Ɛ2 and Ɛ4 was detected in the FACEHBI sample with respect to the control population. Genotypic and allelic frequencies for the studied cohorts are reported in Supplementary table 2.

Higher risk of suffering SCD was identified for HC *APOE* Ɛ4 carriers with respect to non-carriers [OR = 1.61 (1.21 – 2.16); p = 0.001]. In the case of *APOE* Ɛ2, a non-significant risk effect was detected [OR = 1.20 (0.82 – 1.77); p = 0.34] (Table 1). Similarly, when the comparison was conducted between SCD individuals and MCI patients, the risk of suffering MCI was not significantly increased in SCD subjects carriers of *APOE* Ɛ4 compared to non-carriers although a risk trend was observed [OR = 1.28 (0.95 – 1.72); p = 0.098] and a protective association was detected for *APOE* Ɛ2 [OR =0.53 (0.34 – 0.82); p = 0.003] . In the case of SCD individuals and AD patients, SCD individuals carriers of *APOE* Ɛ4 present an increased risk of suffering AD [OR = 1.95 (1.47 – 2.60); p = 2.69×10^{-6}], in contrast to *APOE* Ɛ2's effect [OR = 0.42 (0.28 – 0.63); p = 1.90×10^{-5}] (Table 1). No major difference in the effect size of *APOE* Ɛ4 and *APOE* Ɛ2 were detected after adjusting for gender. After adjusting for gender and age, the effect for *APOE* Ɛ2 in MCI and AD groups decreased and the significance level increased (Table 1). This effect was expected taking into account that age is an independent risk factor for AD and the presence of major differences in age among the studied cohorts.

The meta-analysis (n = 12,183) evaluating *APOE*'s effect in the risk of presenting SCD showed increased risk of suffering SCD in HC carriers of the *APOE* Ɛ4 allele [OR = 1.18 (1.06 – 1.31); p =

TABLE 1. Adjusted and unadjusted effects of *APOE* Ɛ4 in the risk of SCD diagnosis and risk to suffer MCI or AD from SCD status.

	%Genotype Carriers (n)	Unadjusted		Adjusted per Gender		Adjusted per Gender and Age	
	APOE - Ɛ4 *APOE* - Ɛ2	*APOE*-Ɛ4 OR (95%CI) *P* value	*APOE*-Ɛ2 OR (95%CI) *P* value	*APOE*-Ɛ4 OR (95%CI) *P* value	*APOE*-Ɛ2 OR (95%CI) *P* value	*APOE*-Ɛ4 OR (95%CI) *P* value	*APOE*-Ɛ2 OR (95%CI) *P* value
HC– SCD	18.4(557) – 26(52) 12.1(396) – 14.5(29)	1.61 (1.21–2.16) 0.001	1.20 (0.82 – 1.77) 0.34	1.61 (1.20 – 2.14) 0.0012	1 19 (0.82 – 1.75) 0 35	1.62 (1.20 – 2 19) 0.0016	1.31 (0.76 -1.67) 0.54
SCD–MCI	26(52) – 32.4(379) 14.5(29) – 8.1(94)	1.28 (0.95 – 1.72) 0.098	0.53 (0.34 – 0.82) 0.003	1.26 (0.95 – 1.69) 0.11	0 53 (0.35 – 0.82) 0.004	1.44 (1.05 – 1 97) 0.02	0.82 (0.49– 1.40) 0.47
SCD–AD	26(52) – 44.7(1,125) 14.5(29) – 6.4(63)	1.95 (1.47 – 2.60) 2.69e-06	0.42 (0.28 – 0.63) 1.90e-05	1.98 (1.49 – 2.64) 2.69e-06	0.44 (0.30 – 0.66) 7.38e-05	2.2 (1.59 – 3.05) 2.28e-06	0.53 (0.31– 0.91) 0.02

OR: Odds Ratio; HC: Healthy Controls; SCD: Subjective cognitive decline; MCI: Mild cognitive impairment; AD: Alzheimer's Disease.

0.002] (Figure 1). *APOE* Ɛ4's effect showed non-significant heterogeneity between studies (I^2=39%, p = 0.06). Funnel plot is shown in Supplementary Figure 3. Egger test for funnel plot asymmetry presented a p value = 0.99, supporting that publication bias is not present in our meta-analysis. Sub-population analysis only for Caucasians [1.17 (1.05 – 1.29); p = 0.004] (I^2=42%, p = 0.06) showed that the pooled effect was not modified when additional population were introduced.

Effect of *APOE* locus on brain amyloid burden

Global SUVR according to APOE dosage in the FACEHBI cohort

Global SUVR was significantly different across *APOE* dosage groups in the FACEHBI cohort. Post-hoc analysis revealed that those groups comprising *APOE* Ɛ4 carriers were driving this difference (Figure 2).

The relationship between APOE dosage and brain amyloid burden

APOE Ɛ4 dosage explained 9% of the variance in brain amyloid burden (R^2 = 0.09; p = 1.70 x 10^{-5}) in the FACEHBI cohort (Table 2). When we performed the model taking into account *APOE* allele dosage and age, 15% of the variance was explained (R^2 = 0.15; p = 5.25 x 10^{-8}).

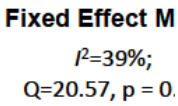


FIGURE 1. Forest plot for the effect of APOE ε4 genotype in risk to be diagnosed with SCD

ADNI, Alzheimer's Disease Neuroimaging Initiative; APOE, apolipoprotein E; CI, confidence interval; OR, odds ratio; SCD, subjective cognitive decline. For further information about Czech brain aging study, LASA, Vallecas Study, and PATH Study, see references (45), (46), (47) and (48), respectively

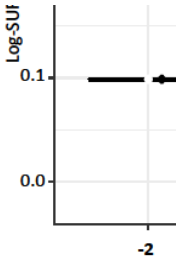


FIGURE 2. Effect of APOE allele dosage in cerebral amyloid burden for the FACEHBI cohort

Mean is represented in white.*APOE*, apolipoprotein E; SUVR, standardized uptake value ratio

TABLE 2. Determination coefficient and level of significance between *APOE* allele dosage and cerebral amyloid burden

Study	ADNI	FACEHBI	ADNI	ADNI	ADNI	ADNI
Status	HC	SCD	SMC	EMCI	LMCI	AD
N	182	200	103	303	157	144
Age (mean ±SD)	73.4 ±6.3	65.8 ± 7.1	72.2 ±5.6	71.3 ± 7.4	72.2 ± 7.5	74.4 ± 8.1
APOE ε4 carriers % (n)	28.6 (52)	26.0 (52)	31.1 (32)	43.2 (131)	57.9 (91)	67.4 (97)
R^2; *P* value	0.05; .002	0.09; 1.70e25	0.12; 0.004	0.14; 1.95e211	0.18; 2.86e28	0.09; 2.00e24

AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; *APOE*, apolipoprotein; EMCI, early mild cognitive impairment; FACEHBI, Fundacio ACE Healthy Brain Initia- tive; HC, healthy control; LMCI, late mild cognitive impairment; PET, positron emission tomography; SCD, subjective cognitive decline; SMC, subjective memory impairment

APOE Ɛ4 dosage explained 12% of the variance (R^2 = 0.12; p = 0.004) in the SCD ADNI cohort, which is in accordance with our findings (Table 2). Next, we investigated the relationship between amyloid accumulation and *APOE* allele dosage across the entire ADNI series, comprising HC, SCD, MCI, and AD cohorts. The highest correlation was detected in late onset MCI (LMCI) individuals (R^2 = 0.18; p = 2.86 x 10^{-8}). Moreover we observed an upward trend in the coefficient of determination from HC to LMCI. However, the correlation experimented a decrease when the analysis was performed in AD individuals (R^2 = 0.09; p = 2.00 x 10^{-4}) (Table 2). Individual slopes per clinical categories are shown in Supplementary Figure 4

The meta-analysis showed that the correlation between *APOE* dosage and amyloid burden was not significantly heterogeneous across clinical groups (I^2 = 7%; p = 0.372) (Figure 3), and *APOE* Ɛ4 allele dosage explained 11% of the variance in amyloid SUVR [R (CI95%) = 0.33 (0.28 – 0.40), p < 0.001] (Figure 3). Although, a significant interaction between the effect of *APOE* dosage and clinical status across the pooled data was detected (R^2 = 0.27; p = 0.02), the proportion of the variance explained only experimented a subtle change when the model was run without considering the interaction factor (R^2 = 0.26).

Finally, we decided to explore the effect of age on amyloid burden. The meta-analysis of the correlation coefficient between age and amyloid burden detected high heterogeneity (I^2 = 82.7%; p < 0.0001) (Figure 3). In addition, advanced age was not correlated with higher cerebral amyloid accumulation [R (CI95%) = 0.11 (-0.03 – 0.26); p = 0.13] (Figure 3). A non-significant interaction term was observed between age and clinical status for

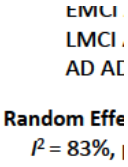


FIGURE 3. Forest plot for (A) the effect of APOE ε4 and (B) the age, in amyloid burden across clinical categories.
AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; *APOE*, apolipoprotein; EMCI, early mild cognitive impairment; FACEHBI, Fundacio ACE Healthy Brain Initia- tive; HC, healthy control; LMCI, late mild cognitive impairment; PET, positron emission tomography; SCD, subjective cognitive decline; SMC, subjective memory impairment.

predicting variance in brain amyloid levels (R^2 = 0.19; p = 0.18). The proportion of variance explained when the model did not consider the interaction was 17% (R^2 = 0.17).

DISCUSSION

SCD has been identified as a risk factor to suffer AD (32) and has been suggested as a potential preclinical stage of the disease (33). Despite this, the genetic and biomarker profile of SCD individuals remains mostly unexplored.

In this study, we report increased risk of suffering SCD from healthy status in carriers of the *APOE* Ɛ4 allele using the FACEHBI SCD sample, which supports that a pool of AD patients are nested in this cohort. We performed a large meta-analysis that demonstrated *APOE* Ɛ4 is a genetic risk factor of presenting SCD. Furthermore, *APOE* Ɛ4 carriers with SCD presented a significantly increased risk of AD diagnosis. The finding of *APOE* Ɛ2 allele enrichment in the FACEHBI cohort(not observed in other datasets, i.e. ADNI) [OR (CI95%) = 0.53 (0.37 – 0.74); p = 2.30 x 10^{-4})] was unexpected taking into account the protective role of *APOE* Ɛ2 in AD (34). There could be two reasons for this. Higher

frequency of *APOE* Ɛ2 allele has been detected in individuals with white matter hyperintensities (WMH) (35),(36). Taking into account that there is an increment of WMH in subjects with vascular dementia (37), the enrichment of *APOE* Ɛ2 in the FACEHBI population could be underlying a pool of subjects with brain vascular alterations. Random chance could also be underlying this variation in Ɛ2 allelic frequency.

The SCD cohorts used in this meta-analysis present a certain level of heterogeneity (I^2 = 39%; p = 0.06) which could be due to differences in the selection of SCD and healthy controls individuals among studies. SCD comprises individuals who will convert to AD, individuals who will convert to other dementias and others who will remain healthy. Moreover, other factors could be leading to this heterogeneity, i.e. differences in recruitment strategies, assessment (38) and SCD definitions across studies (39), which researchers must carefully control for and standardize. In this sense, significant efforts are underway to design neuropsychological tools sensitive to subtle cognitive changes and to identify genetic and biomarker profiles in SCD subjects. The integration of genetic, neuropsychological and biomarker profiles seems necessary to characterize SCD individuals and improve reproducibility of studies.

In an effort to improve characterisation of SCD individuals, we combined *APOE* and amyloid burden information. We detected statistically significant differences in cerebral amyloid burden between *APOE* allele dosage stratums in the FACEHBI sample, which is in accordance with previous findings (40),(41). The variation in brain amyloid levels is partially explained by *APOE* genotype in our series. This finding was replicated in the ADNI dataset by us and others (11).

The contribution of other genetic factors to cortical Aβ levels remains mostly unexplored and a large proportion of the variance remains unexplained. Some studies describe models that consider the contribution of other genetic markers, i.e. *APOE* Ɛ4 and *BCHE*-rs509208 explained 15% of the variance (11) and *IL1RAP*-rs12053868 7.1% (12). However, when the model considered the polygenic risk score (PGRS) of genes identified in the Stage I of the International Genomics of Alzheimer's Project (IGAP) study only 1% of the variance was explained (42). The involvement of non-amyloid processes in AD could be explaining this inconsistency. Either way, it seems necessary to include the genetic markers associated with the amyloid cascade in the model. The finding of substantial overlap between genes associated to AD pathology with those driving cerebral amyloid burden, would support the use of this trait as an AD endophenotype.

After exploring the correlation between *APOE* and amyloid burden across clinical categories, a decrease in the correlation coefficient was observed in the AD group. This result could be explained by the effect of atrophy in amyloid PET measurements. Some works have proposed partial-volume effect (PVE) can be distorting PET signal and diminishing the accuracy of the measure (43). On the other hand, a subtle interaction effect was detected between *APOE* and clinical categories predicting amyloid burden. This interaction was not detected using the meta-analysis strategy, maybe because it might be underpowered. In that scenario, adjusting for clinical category is mandatory when individuals with different clinical status are merged for studying the brain amyloid burden endophenotype with GWAS methodology.

In the same way, we also evaluated the role of age in determining the Aβ brain burden endophenotype across the AD continuum. Age has been suggested as a non-genetic risk factor for AD (44) and several studies have pointed to it as a risk factor for cerebral amyloid burden (15). Conversely, we did not detect a correlation between age and Aβ load, and an inverse correlation was detected in the AD group, which can be caused by PVE. Older individuals might present an advanced stage of the disease, showing greater atrophy which could be disturbing PET measurements. In addition, an interaction effect was not detected between age and clinical category predicting the change of amyloid burden, which it is not unexpected considering that age did not modulate this endophenotype. AD is related with both advanced age (44) and the presence of neuritic plaques (9). This dual association can generate a confounding relation when age and Aβ are analysed independently from AD. Hence, we propose that age acts as a confounding factor, which is not directly associated with the Aβ endophenotype.

There are potential limitations in the present study. We were not able to genotype *APOE* alleles at the same moment for different cohorts. Despite that, we check that both *APOE* markers (rs7412 and rs429358) were following Hardy Weinberg expectations for each comparison. In addition, to evaluate the correlation between *APOE* dosage and amyloid burden we used data from the FACEHBI and ADNI samples, which use different radioactive tracers for amyloid PET scans. In an effort to control for discrepancies between studies, we standardized the log-transformed SUVR measure. Finally, the present correlation analyses were not adjusted by time of disease duration, which could be an interesting point for future studies. Despite that, taking into account that analyses were conducted separately per clinical stratum, mayor differences are not expected.

In summary, the present data support the role of *APOE* Ɛ4 as a risk factor of SCD and its involvement in brain amyloid burden in SCD subjects. Amyloid-PET is an instrumental measure of the brain amyloid burden endophenotype in SCD subjects but the modelling of this important trait will require further integration of other genetic and epigenetic factors

REFERENCES

1. Jack CR, Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol. 2010 Jan;9(1):119–28.

2. Jack CR, Knopman DS, Jagust WJ, Petersen RC, Weiner MW, Aisen PS, et al. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. Lancet Neurol. 2013 Feb;12(2):207–16.

3. Sperling RA, Jack CR, Aisen PS. Testing the right target and right drug at the right stage. Sci Transl Med . 2011 Nov 30;3(111):111cm33.

4. Jessen F, Amariglio RE, van Boxtel M, Breteler M, Ceccaldi M, Chételat G, et al. A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. Alzheimers Dement . 2014 Nov;10(6):844–52.

5. Corder E, Saunders A. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science (80-). 1993;8(14):41–3.

6. Moreno-Grau S, Ruiz A. Genome research in pre-dementia stages of Alzheimer's disease. Expert Rev Mol Med . 2016;18:e11.

7. Lau S, Bates KA, Sohrabi HR, Rodrigues M, Martins G, Dhaliwal SS, et al. Functional effects of genetic polymorphism in inflammatory genes in subjective memory complainers. Neurobiol Aging . 2012 Jun;33(6):1054–6.

8. Laws SM, Clarnette RM, Taddei K, Martins G, Paton A, Almeida OP, et al. Association between the presenilin-1 mutation Glu318Gly and complaints of memory impairment. Neurobiol Aging. Jan;23(1):55–8.

9. Blennow K, de Leon MJ, Zetterberg H. Alzheimer's disease. Lancet. 2006 Jul 29;368(9533):387–403.

10. Castellano JM, Kim J, Stewart FR, Jiang H, DeMattos RB, Patterson BW, et al. Human apoE Isoforms Differentially Regulate Brain Amyloid-Peptide Clearance. Sci Transl Med. 2011 Jun 29;3(89):89ra57-89ra57.

11. Ramanan VK, Risacher SL, Nho K, Kim S, Swaminathan S, Shen L, et al. APOE and BCHE as modulators of cerebral amyloid deposition: a florbetapir PET genome-wide association study. Mol Psychiatry. 2014 Mar;19(3):351–7.

12. Ramanan VK, Risacher SL, Nho K, Kim S, Shen L, McDonald BC, et al. GWAS of longitudinal amyloid accumulation on 18F-florbetapir PET in Alzheimer's disease implicates microglial activation gene IL1RAP. Brain. 2015 Aug 11;138(Pt 10):3076–88.

13. Rowe CC, Ellis KA, Rimajova M, Bourgeat P, Pike KE, Jones G, et al. Amyloid imaging results from the Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging. Neurobiol Aging. 2010 Aug;31(8):1275–83.

14. Villemagne VL, Burnham S, Bourgeat P, Brown B, Ellis KA, Salvado O, et al. Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. Lancet Neurol. 2013 Apr;12(4):357–67.

15. Jansen WJ, Ossenkoppele R, Knol DL, Tijms BM, Scheltens P, Verhey FRJ, et al. Prevalence of Cerebral Amyloid Pathology in Persons Without Dementia. JAMA. 2015 May 19;313(19):1924.

16. Amariglio RE, Becker JA, Carmasin J, Wadsworth LP, Lorius N, Sullivan C, et al. Subjective cognitive complaints and amyloid burden in cognitively normal older individuals. Neuropsychologia. 2012 Oct;50(12):2880–6.

17. Risacher SL, Kim S, Nho K, Foroud T, Shen L, Petersen RC, et al. APOE effect on Alzheimer's disease biomarkers in older adults with significant memory concern. Alzheimer's Dement. 2015 Dec;11(12):1417–29.

18. Rodriguez-Gomez O, Sanabria A, Perez-Cordon A, Sanchez-Ruiz D, C. Abdelnour, S. Valero, et al. A Prospective Study of Risk Factors, Biomarkers and Cognition in a Cohort of Individuals with Subjective Cognitive Decline. Study Rationale and Research Protocols. Jpad. 2016;

19. Alegret M, Espinosa A, Valero S, Vinyes-Junqué G, Ruiz A, Hernández I, et al. Cut-off Scores of a Brief Neuropsychological Battery (NBACE) for Spanish Individual Adults Older than 44 Years Old. Brucki S, editor. PLoS One. 2013 Oct 16;8(10):e76436.

20. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. Arch Neurol . 1999 Mar;56(3):303–8.

21. Petersen RC. Mild cognitive impairment as a diagnostic entity. J Intern Med. 2004 Sep;256(3):183–94.

22. Lopez OL, Jagust WJ, Dulberg C, Becker JT, DeKosky ST, Fitzpatrick A, et al. Risk factors for mild cognitive impairment in the Cardiovascular Health Study Cognition Study: part 2. Arch Neurol. 2003 Oct;60(10):1394–9.

23. Lopez OL, Kuller LH, Becker JT, Dulberg C, Sweet RA, Gach HM, et al. Incidence of dementia in mild cognitive impairment in the cardiovascular health study cognition study. Arch Neurol . 2007 Mar 1;64(3):416–20.

24. Lacour A, Espinosa A, Louwersheimer E, Heilmann S, Hernández I, Wolfsgruber S, et al. Genome-wide significant risk factors for Alzheimer's disease: role in progression to dementia due to Alzheimer's disease among subjects with mild cognitive impairment. Mol Psychiatry. 2017 Jan 15;22(1):153–60.

25. Seshadri S, Fitzpatrick A, L. Ikram A, DeStefano AL, Gudnason V, Boada M. Genome-wide Analysis of Genetic Loci Associeted with Alzheimer's Disease. JAMA. 2011;303(18):1832–40.

26. Boada M, Antúnez C, Ramírez-Lorca R, DeStefano A, González-Pérez A, Gayán J, et al. ATP5H/KCTD2 locus is associated with Alzheimer's disease risk. Mol Psychiatry. 2014 Jun;19(6):682–7.

27. Martínez-Larrad MT, Fernández-Pérez C, González-Sánchez JL, López A, Fernández-Alvarez J, Riviriego J, et al. Prevalence of the metabolic syndrome (ATP-III criteria). Population-based study of rural and urban areas in the Spanish province of Segovia]. Med Clin. 2005 Oct 15;125(13):481–6.

28. Lorenzo C, Serrano-Rios M, Martinez-Larrad MT, Gabriel R, Williams K, Gonzalez-Villalpando C, et al. Prevalence of hypertension in Hispanic and non-Hispanic white populations. Hypertens (Dallas, Tex 1979) . 2002 Feb;39(2):203–8.

29. Weiner M, Aisen P, Jack Jr C, Jaugust W, Trojanowski J, Shaw L, et al. The Alzheimer's disease neuroinmaging iniciative: Progress report and future plans. Alzheimer's Dement. 2010;6(3):202–11.

30. Jagust WJ, Bandy D, Chen K, Foster NL, Landau SM, Mathis CA, et al. The Alzheimer's Disease Neuroimaging Initiative positron emission tomography core. Alzheimer's Dement. 2010 May;6(3):221–9.

31. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007 Sep 9;81(3):559–75.

32. Jessen F, Wiese B, Bachmann C, Eifflaender-Gorfer S, Haller F, Kölsch H, et al. Prediction of Dementia by Subjective Memory Impairment. Arch Gen Psychiatry. 2010;67(4):414–22.

33. Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011 May;7(3):280–92.

34. Corder EH, Saunders AM, Risch NJ, Strittmatter WJ, Schmechel DE, Gaskell PC, et al. Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. Nat Genet. 1994 Jun;7(2):180–4. 35. Gesierich B, Opherk C, Rosand J, Gonik M, Malik R, Jouvent E, et al. *APOE ε* 2 is associated with white matter hyperintensity volume in CADASIL. J Cereb Blood Flow Metab. 2016 Jan 29;36(1):199–203.

36. Lemmens R, Gorner A, Schrooten M, Thijs V. Association of Apolipoprotein E 2 With White Matter Disease but Not With Microbleeds. Stroke . 2007 Apr 1;38(4):1185–8.

37. Prins ND, Scheltens P. White matter hyperintensities, cognitive impairment and dementia: an update. Nat Rev Neurol. 2015 Feb 17;11(3):157–65.

38. Rodríguez-Gómez O, Abdelnour C, Jessen F, Valero S, Boada M. Influence of Sampling and Recruitment Methods in Studies of Subjective Cognitive Decline. Tales A, Jessen F, Butler C, Wilcock G, Phillips J, Bayer T, editors. J Alzheimer's Dis. 2015 Sep 24;48(s1):S99–107.

39. Abdulrab K, Heun R. Subjective Memory Impairment. A review of its definitions indicates the need for a comprehensive set of standardised and validated criteria. Eur Psychiatry. 2008 Aug;23(5):321–30.

40. Zwan MD, Villemagne VL, Doré V, Buckley R, Bourgeat P, Veljanoski R, et al. Subjective Memory Complaints in APOE ε4 Carriers are Associated with High Amyloid-β Burden. Caracciolo B, editor. J Alzheimer's Dis. 2015 Nov 25;49(4):1115–22.

41. Bangen KJ, Clark AL, Werhane M, Edmonds EC, Nation DA, Evangelista N, et al. Cortical Amyloid Burden Differences Across Empirically-Derived Mild Cognitive Impairment Subtypes and Interaction with APOE ε4 Genotype. J Alzheimers Dis. 2016;52(3):849–61.

42. Mormino EC, Sperling RA, Holmes A, Buckner R, de Jager P, Smoller J, et al. Polygenic Risk of Alzheimer'S Disease Is Associated With Early and Late Life Processes. Alzheimer's Dement. 2016;12(7):P50–1.

43. Rullmann M, Dukart J, Hoffmann K-T, Luthardt J, Tiepolt S, Patt M, et al. Partial-Volume Effect Correction Improves Quantitative Analysis of 18F-Florbetaben -Amyloid PET Scans. J Nucl Med. 2016 Feb 1;57(2):198–203.

44. Guerreiro R, Bras J. The age factor in Alzheimer's disease. Genome Med . 2015 Oct 20;7(1):106.

45. Pařízková M, Andel R, Lerch O, Marková H, Gažová I, Vyhnálek M, et al. Homocysteine and Real-Space Navigation Performance among Non-Demented Older Adults. J Alzheimer's Dis 2016;55:951–64.

46. Fernández-Blázquez MA, Ávila-Villanueva M, Maestú F, Medina M. Specific Features of Subjective Cognitive Decline Predict Faster Conversion to Mild Cognitive Impairment. J Alzheimer's Dis 2016;52:271–81.

47. Dik MG, Jonker C, Comijs HC, Bouter LM, Twisk JW, van Kamp GJ, et al. Memory complaints and APOE-epsilon4 accelerate

cognitive decline in cognitively normal elderly. Neurology 2001;57:2217–22.

48. Cherbuin N, Sargent-Cox K, Easteal S, Sachdev P, Anstey KJ. Hippocampal Atrophy Is Associated with Subjective Memory Decline: The PATH Through Life Study. Am J Geriatr Psychiatry 2015;23:446–55.

ACKNOWLEDGEMENT

We would like to thank patients and controls who participated in this project. We also want to thank the sponsors supporting the FACEHBI project (Grifols SA, Piramal AG, Araclon Biotech and Fundació ACE). We are indebted to Trinitat Port-Carbó and her family for their support of Fundació ACE research programs. Fundació ACE collaborates with the Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED, Spain) and is one of the participating centers of the Dementia Genetics Spanish Consortium (DEGESCO). Agustín Ruiz has received support from the EU/EFPIA Innovative Medicines Initiative Joint Undertaking ADAPTED Grant No. 115975 and by grants PI13/02434 and PI16/01861. Acción Estratégica en Salud, integrated in the Spanish National R + D + I Plan and financed by ISCIII (Instituto de Salud Carlos III)-Subdirección General de Evaluación and the Fondo Europeo de Desarrollo Regional (FEDER- "Una manera de Hacer Europa"), by Fundación bancaria "La Caixa" and Grifols SA (GR@ACE project). Part of the data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbV ie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd. and its affiliated company, Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research provides funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. **The present work was performed as part of the Biochemistry, Molecular Biology and Biomedicine doctoral program of S. Moreno-Grau at Universitat Autònoma de Barcelona (Barcelona, Spain)**.

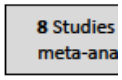
The FACEHBI study group: Abdelnour C[1], Aguilera N[1], Alegret M[1], Berthier M[2], Boada M[1], Buendia M[1], Bullich S[3], Campos F[4], Cañabate P[1], Cuevas C[1], de Rojas I[1], Espinosa A[1], Gailhajenet A[1], Diego S[1], Gil S[1], Giménez J[5], Gismondi R[3], Gómez-Chiari M[5], Guitart M[1], Hernández-Olasagarre B[1], Hernández I[1], Ibarria, M[1], Lafuente A[1], Lomeña F[4], Martín E[1], Martínez J[1], Mauleón A[1], Monté G[1], Moreno M[1], Moreno-Grau S[1], Núñez L[6], Orellana A[1], Ortega G[1], Páez A[6], Pancho A[1], Pavía J[4], Pelejà E[1], Pérez-Cordon A[1], Pérez-Grijalba V[7], Pesini P[7], Preckler S[1], Rodríguez-Gómez O[1], Romero J[7], Rosende-Roca M[1], Ruiz A[1], Ruiz S[1], Sanabria A[1], Sánchez-Ruiz D[1], Santos-Santos MA[1], Sarasa M[7], Sotolongo-Grau O[1], Tárraga L[1], Tejero MA[5], Torres M[6], Valero S[8], Vargas L[1], Vivas A[5].

(1.Fundacio ACE. Alzheimer Treatment and research Center. Barcelona, Spain; 2.Cognitive Neurology and Aphasia Unit (UNCA). University of Malaga. 3.PiramalImaging GmbH, Berlin, Germany; 4.Servei de Medicina Nuclear, Hospital Clínic I Provincial, Barcelona, Spain; 5.Departament de Diagnòstic per la Imatge. Clínica Corachan, Barcelona, Spain; 6. Grifols®; 7.Araclon Biothech®. Zaragoza, Spain; 8.Psychiatry department, Hospital Universitari Vall d Hebron, CIBERSAM, Universitat Autonoma de Barcelona, Barcelona, Spain).

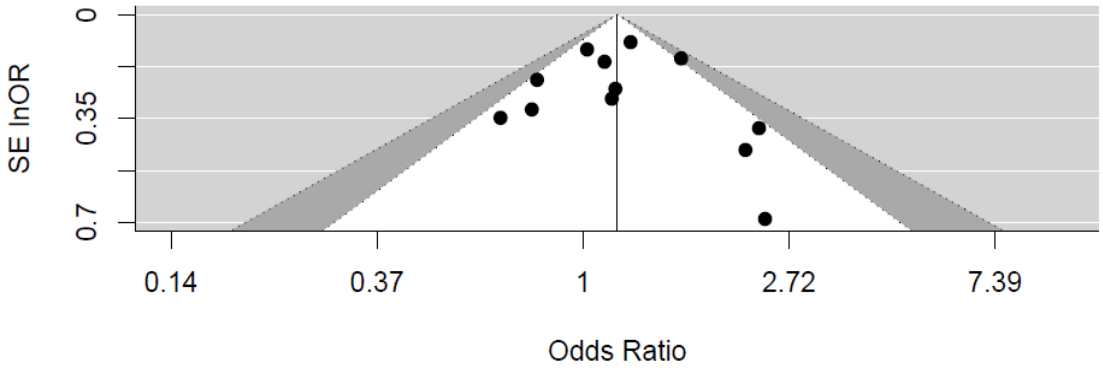
SUPPLEMENTARY FIGURES

SUPPLEMENTARY FIGURE 1. Flow chart diagram for cohorts used in evaluation of APOE ε4 effect in risk associated with SCD condition.

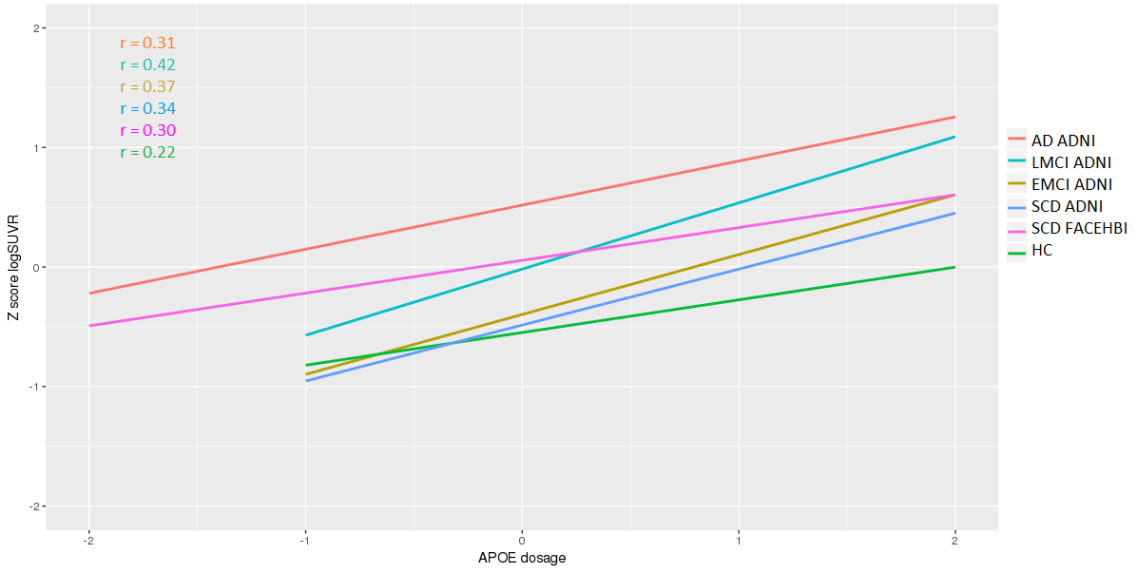
AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; APOE, apolipoprotein; FACEHBI, Fundacio ACE Healthy Brain Initiative; HC, healthy control; SCD, subjective cognitive decline



SUPPLEMENTARY FIGURE 2. Flow chart diagram of literature search and study selection for meta-analysis the effect of *APOE* ε4 inrisk of be diagnosed with SCD.
APOE, apolipoprotein; FACEHBI, Fundacio ACE Healthy Brain Initiative;CSF, cerebrospinal fluid; SCD, subjective cognitive decline.



SUPPLEMENTARY FIGURE 3. Funnel plot for meta-analysis of the effect of *APOE* ε4 in risk of be diagnosed with SCD



SUPPLEMENTARY FIGURE 4. Correlation between *APOE* allele dosage and brain amyloid burden per clinical category.

AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; APOE, apolipoprotein; EMCI, early mild cognitive impairment; FACEHBI, Fundacio ACE Healthy Brain Initiative; HC, healthy control; LMCI, late mild cognitive impairment; SCD, subjective cognitive decline.

SUPPLEMENTARY TABLE 2. Genotype and allelic frequencies of *APOE* markers across the Spanish datasets.

SUPPLEMENTARY TABLES

SUPPLEMENTARY TABLE 1. Recruitment strategies and SCD definition for the cohorts used in the meta-analysis of *APOE* and SCD risk

Study	Recruitment strategy	SCD definition
ADNI	ADNI database.	Individuals with Cognitive Change Index > 16, no informant-based complaint of memory impairment or decline, and normal cognitive performance on the LM-Delayed Recall and MMSE.
African-Caribbeans	A community study. UK population sampled by case register.	Subjective memory complaints were assessed using questions from the Geriatric Mental State schedule. Participants scoring ≥4 on an adapted scale (maximum score 10) were defined as having subjective memory impairment.
AgeCoDe	Case register. Patients were contacted by mail.	Individuals with memory complaints with normal age-, gender- and education-adjusted cognitive performance, assessed with the Subjective Memory decline scale.
AIBL	Telephone screening to appeal for volunteer or volunteered after treating physician.	Healthy individuals (CDR ≤ 0.5) with memory complaints who did not fulfil criteria for MCI or dementia.
APOE Study	Volunteer appealing through local media ads into the Arizona APOE cohort. Individuals with history of demntia.	Individuals with any score greater than zero was considered "positive" for endorsed decline on both the MANS-self and MANS-informant questionnaire.
Czech Brain Aging Study	i) Referred to the Memory Clinic by general practitioners or neurologists due to memory complaints from the patient or the informant ii) Recruited from staff and patients' relatives iii) Recruited from the University of the Third Age	Individuals with normal activities of daily living (FAQ≤5 and CDR≤ 0.5) and who did not meet diagnostic criteria for aMCI or dementia, but report subjective memory complaints.
IMH. Pekin	Case register of the Dementia Care & Research Center of Pekin University.	Individuals who did not fulfil criteria for dementia or MCI but did complain of memory decline. The scores on the MMSE and the clinical memory scale in participants with subjective cognitive impairment were within the age, gender and education adjusted norms and their social and daily functioning was intact.
LASA	Community based prospective study. Population register of 3 regions of Netherlands to select people between 55 and 85y. Interview subjects at home.	Individuals with self-reported cognitive deterioration, who were assessed using one question and additional MMSE and AVLT tests.
The Vallecas Project	Community based longitudinal study. Volunteers appealing by radio and TV campaigns.	Self-rated presence of cognitive deterioration using two criteria: i) a positive response to any yes/no-type complaint question, and ii) scores above 1 on the scale of cognitive complaints (SCD).
NHS	Prospective study. Female registered nurses aged 30–55 years, residing in 11 US states, completed a mailed questionnaire about their health and lifestyle	Individuals with self-reported memory impairment, evaluated by a questionnaire on 7 specific "yes/no" questions, and the current level of cognitive function (including memory) measured by neuropsychological testing. Excluded individuals with depression.
OPH	i) Referred to the Memory Clinic ii) Volunteer appealing through advertisements in local newspapers seeking people with memory complaints	Individuals with self-reported memory complained who were not fulfilled criteria for dementia and not present psychiatric disorders, assessed with MMSE and CAMCOG neuropsychological battery.
PATH W1	Random recruitment through the electoral roll of Canberra and Queanbeyan to participate in a longitudinal study of ageing	Individuals with self-reported memory complaint using one question, and who did not fulfil criteria for dementia.
FACE	i) Referred to the Memory Clinic ii) Open House Initiative	SCD was defined as the coexistence of cognitive complaints and a score of ≥8 on MFE-30, the Spanish version of the Memory Failures in Everyday Life Questionnaire (20); MMSE ≥ 27; CDR=0; and performance in NBACE within the normal range for age and educational level

	Population-based controls	FACEHBI cohort	ACE MCI cohort	ACE AD cohort
ε2ε2 % (n)	0.6 (17)	0.5 (1)	0.2 (2)	0.2 (5)
ε2ε3 % (n)	10.4 (316)	12.5 (25)	6.1 (71)	4.6 (117)
ε2ε4 % (n)	1.1 (33)	1.5 (3)	1.8 (21)	1.6 (41)
ε3ε3 % (n)	70.6 (2142)	61.0 (122)	61.4 (718)	50.5 (1270)
ε3ε4 % (n)	16.3 (495)	21.0 (42)	26.7 (312)	37.3 (939)
ε4ε4 % (n)	1.0 (29)	3.5 (7)	3.9 (46)	5.8 (145)
Total	3032	200	1170	2517
***APOE* ε2 allelicfrequency**	6.3%	7.5%	4.1%	3.3%
***APOE* ε3 allelic frequency**	84%	77.8%	77.6%	71.4%
***APOE* ε4 allelic frequency**	9.7%	14.7%	18.2%	25.3%

Section 3.1. *APOE*, The strongest genetic risk factor for AD

3.1.2. Publication II

Genome-wide significant risk factors in chromosome 19 and the APOE locus

Sonia Moreno-Grau[1], Isabel Hernández[1], Stefanie Heilmann-Heimbach[2,3], Susana Ruiz[1], Maitée Rosende-Roca[1], Ana Mauleón[1], Liliana Vargas[1], Octavio Rodríguez- Gómez[1], Montserrat Alegret[1], Ana Espinosa[1], Gemma Ortega[1], Nuria Aguilera[1], Carla Abdelnour[1], Alzheimer's Disease Neuroimaging Initiative[*], Silvia Gil[1], Wolfgang Maier[4,5], Oscar Sotolongo-Grau[1], Lluís Tárraga[1], Alfredo Ramirez[2,4,6], Jesús López-Arrrieta[7], Carmen Antúnez[8], Manuel Serrano-Ríos[9], Mercè Boada[1] and Agustín Ruiz[1]

1 Research Center and Memory Clinic of Fundació ACE, Institut Català de Neurociències Aplicades, Univesitat Internacional de Catalunya, Barcelona, Spain 2 Institute of Human Genetics, University of Bonn, Bonn, Germany 3 Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany 4 Department of Psychiatry and Psychotherapy, University of Bonn, Bonn, Germany 5 German Center for Neurodegenerative Diseases, DZNE, Bonn, Germany 6 Department of Psychiatry and Psychotherapy, University of Cologne, Cologne, Germany 7 Memory Unit, University Hospital La Paz-Cantoblanco, Madrid, Spain 8 Dementia Unit, University Hospital Virgen de la Arrixaca, Murcia, Spain 9 Centro de Investigación Biomédica en Investigación Biomédica en Red de Diabetes y Enfermedades Metabólicas Asociadas, CIBERDEM, Spain, Hospital Clínico San Carlos, Madrid, Spain

ABSTRACT

The apolipoprotein E (*APOE*) gene on chromosome 19q13.32, was the first, and remains the strongest, genetic risk factor for Alzheimer's disease (AD). Additional signals associated with AD have been located in chromosome 19, including *ABCA7* (19p13.3) and *CD33 (*19q13.41). The *ABCA7* gene has been replicated in most populations. However, the contribution to AD of other signals close to *APOE* gene remains controversial. Possible explanations for inconsistency between reports include long range linkage disequilibrium (LRLD). We analysed the contribution of *ABCA7* and *CD33* loci to AD risk and explore LRLD patterns across *APOE* region. To evaluate AD risk conferred by *ABCA7* rs4147929:G>A and *CD33* rs3865444:C>A, we used a large Spanish population (1796 AD cases, 2642 controls). The *ABCA7* rs4147929:G>A SNP effect was nominally replicated in the Spanish cohort and reached genome-wide significance after meta-analysis [odds ratio (OR)=1.15, 95% confidence interval (95% CI)=1.12–1.19; $P = 1.60 \times 10^{-19}$]. *CD33* rs3865444:C>A was not associated with AD in the dataset. The meta-analysis was also negative (OR=0.98, 95% CI=0.93–1.04; *P*=0.48). After exploring LRLD patterns between *APOE* and *CD33* in several datasets, we found significant LD (D' >0.20; *P* <0.030) between *APOE*-Ɛ2 and *CD33* rs3865444C>A in two of five datasets, suggesting the presence of a non-universal long range interaction between these loci affecting to some populations. In conclusion, we provide here evidence of genetic association of the *ABCA7* locus in the Spanish population and also propose a plausible explanation for the controversy on the contribution of *CD33* to AD susceptibility.

Keywords: Late Onset Alzheimer's disease, *ABCA7*, *APOE*, *CD33*, Linkage Disequilibrium.

Oncotarget May 15; 2018; 9(37): 24590–24600.

doi:10.18632/oncotarget.25083

INTRODUCTION

The development of novel approaches in genetics has led to the identification of 30 genetic determinants of Late Onset Alzheimer's disease (LOAD) (1). The most important risk locus for LOAD remains on the long arm of chromosome 19 containing the *APOE* locus, which was the first, and is the strongest, risk factor for Alzheimer's disease (AD) (2). Three *APOE* diplotypic alleles (Ɛ2, Ɛ3, and Ɛ4) are defined, which result from the combination of two single nucleotide polymorphisms (SNPs), rs7412:C>T and rs429358:C>T. While the Ɛ2 allele of *APOE* is determined by the minor allele of rs7412:C>T and is protective for LOAD(3), the minor allele of rs429358:C>T defines the Ɛ4 allele, which increases the risk of LOAD (2) by up to four-fold.

Interestingly, additional loci on chromosome 19 have been proposed, namely *TOMM40* (4), *EXOC3L2* (5), *CD33* (6), *PLD3* (7), and *ABCA7* (6),(8). While compelling genetic data have confirmed *ABCA7*, on the short arm of chromosome 19, as an *APOE*-independent AD risk factor (9),(10), inconsistent results have been reported for the other four genes, after adjustments for the effects of *APOE* (11),(12). Particular interest deserves the *CD33* locus which was reported in 2011 using meta-analysis of genome-wide association studies (GWAS) (6) and, its effect was nominally replicated in an additional sample (13). However, the association signal of *CD33* did not reach genome-wide significance in the International Genomics Alzheimer's Project (IGAP) meta-analysis (9), the largest case-control study for LOAD. Thus, further investigations are required to confirm or refute the potential association of *CD33* with LOAD.

Several possibilities may explain the inconsistency in the results of genetic association studies of regions around the *APOE* locus such as *CD33*, including: the large contribution of the *APOE* allele to AD risk, the existence of long range linkage disequilibrium (LRLD) regions in different populations, the existence of hidden familial cases and population inbreeding. However, a definitive explanation for the real driving-force behind these findings is lacking.

Thus, we explored the genetic contribution of *ABCA7* and *CD33*, including their interactions with *APOE,* to AD risk in a large independent AD case-control cohort composed by unrelated individuals who have at least two Spanish ancestors reported. Furthermore, we conducted LRDL analyses across several dataset to evaluate its influence on *CD33* association to AD.

SUBSJECTS AND METHODS

Replication study in the Spanish population: subjects and genotyping

The Spanish sample comprised 1796 unrelated sporadic AD patients (mean age, 82.1 ± 7.9 years, 70.2% women) and 2642 healthy controls (mean age, 54.1 ± 11.6 years, 64.3% women) recruited at Fundació ACE, Institut Català de Neurociències Aplicades (Barcelona, Spain); Unidad de Memoria, Hospital Universitario La Paz-Cantoblanco (Madrid, Spain); Hospital Clínico Sán Carlos Unidad de Demencias, Hospital Universitario Virgen de la Arrixaca (Murcia, Spain) and Neocodex S.L. (Supplementary Table 4). Sample characteristics were previously described by Antunez et al. (14). Briefly, all AD patients fulfilled Diagnostic and Statistical Manual of Mental Disorders IV criteria for dementia and were diagnosed according to National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's disease and Related Disorders Association criteria for possible and probable AD. Ethics committees from each referral centre approved the research protocol. All participants provided written informed consent.

Standard methods were used to isolate DNA. The SNP rs4147929:G>A, this is hg19 chr19:g.1063444A>G located in the *ABCA7* gene and the SNP rs3865444:C>A, this is hg19 chr19:g.51727962C>A located in *CD33* gene ,were genotyped using Sequenom technology (Sequenom, California, USA), as previously described. (15). Primer sequences and assay conditions for the genotyped SNPs are available upon request.

APOE rs7412C>T and rs429358:C>T markers were genotyped using real-time PCR. Primers design was previously described by Calero et al (16). Briefly, PCR reactions were performed in a final volume of 5µl, using 11ng of genomic DNA, 0.3 µM of each amplification primer and 2.65µl of 2X SYBR Fast Master Mix (Kapa Biosystems). We used an initial denaturation step of 95 °C for 2 min, followed by 33 cycles of 95 °C for 10 s, and 69 °C for 30 s. Melting curves were 95 °C for 15 s (ramping rate 5.5 °C s), 45 °C for 15 s (ramping rate of 5.5°C s^{-1}) and 95 °C for 15 s (ramping rate of 5.5°C s^{-1}). In the last step of each melting curve, a continuous fluorimetric register was performed by the system at one acquisition register per each degree Celsius. Melting peaks and genotype calls were obtained by using the Eco Real-Time PCR system (Illumina).

Statistical analysis and meta-analysis

Comparisons of allele frequencies between cases and controls were performed using Chi-square tests. Logistic regression analysis (additive model) was used to adjust for: 1) sex and *APOE* Ɛ4, and 2) age, sex, and *APOE* Ɛ4. Stratification was also

conducted according to the presence or absence of the *APOE* Ɛ4 allele. All statistical analyses were performed using PLINK 1.9 software (http://www.cog-genomics.org/plink2/) (17).

Meta-analysis techniques were used to estimate: *ABCA7* rs4147929:G>A and *CD33* rs3865444:C>A effects across studies. Meta-analysis datasets comprise new data generated, samples overlapping with IGAP were excluded (n = 3,994), data from IGAP (9) and available studies published (Supplementary Table 5). Briefly, we carried out a literature search in Pubmed for studies published before March 2018. The search terms were: *ABCA7* and Alzheimer's disease; and *CD33* and Alzheimer's disease, respectively. Only studies meeting the following criteria were included: (1) case/control studies evaluating the effect of rs4147929:G>A or rs3865444:C>A markers in AD's risk; (2) studies that provided an odds ratio with 95% confidence interval as well as the p-value or provide sufficient data to calculate them. Reviews were excluded. We included 1 article for *ABCA7* (18) (32 articles were excluded due to rs4147929:G>A was not genotyped; 1 presented sample overlapping; 66 were not case/control studies and 3 presented limited access). Thus, the meta-analysis sample size for *ABCA7* rs4147929:G>A comprises 182,208 individuals. In case of *CD33*, we included 12 articles (13), (19), (20), (21), (22), (23), (24), (25), (26), (27), (28), (29) (6 studies did not genotype rs3865444:C>A; 5 presented sample overlapping; and 43 were not case/control studies). Thus, the meta-analysis sample size for *CD33* rs3865444:C>A comprises 90,913 individuals. Meta-analysis was conducted using the inverse variant method (fixed-effects model) but in the case of heterogeneity, the DerSimonian and Liard method (random-effects model) was used. Heterogeneity was considered significant when $I^2 > 50\%$ and $p < 0.05$. Meta-analysis results and forest plots were generated using OpenMeta (http://www.cebm.brown.edu/openmeta/).

Linkage disequilibrium analysis using GWAS datasets

Patterns of LD were studied in data from the present study and four non-imputed GWAS datasets: the Alzheimer's Disease Neuroimaging Initiative (ADNI) longitudinal study (30); the Genotype-Phenotype Alzheimer Disease Association (GenADA) study (31); the National Institute of Aging (NIA) Genetic Consortium for Late Onset Alzheimer's disease study (32); and the Murcia study (14) (Supplementary Table 6). The genome assembly for the four non-imputed GWAS datasets was NCBI36/hg18.

LD measures (D' and r^2) and tests of the significance of LD were calculated between the *APOE* rs429358:C>T or rs7412:C>T markers and *CD33* rs3865444:C>A using Plink 1.9 software

(http://www.cog-genomics.org/plink2/) (17) and the R statistics package. LRLD was accepted where D' ≥ 0.2 and $P < 0.05$. With the objective of discarding LD results generated by chance, we performed bootstrapping analyses; we calculated LD between *APOE* rs7412:C>T or rs429358:C>T and 10,000 random markers, which presented minor allele frequencies (MAF) ≥ 0.20 and ≤ 0.40 (since *CD33* rs3865444:C>A, MAF = 0.30) and that did not localize to chromosome 19.

To further investigate these results, two additional methods were employed. First, the pattern of disequilibrium across the whole of chromosome 19 was calculated according to the methods of Dawson and colleagues (33); including only markers with MAFs ≥ 0.2. Briefly, we considered 1.7Mb window (1.6Mb overlap) and calculated average values of D' and r^2 for all marker pairs, which were separated by at least 50kb and at most 500kb. D' and r2 were calculated using Plink software 1.9 (17). Second, association analysis was conducted in Plink 1.9 software (17) according to the presence or absence of the Ɛ2 allele. In this second approach, markers in strong LD with the Ɛ2 allele will exhibit stronger associations. P values were adjusted using the Bonferroni correction method.

Unadjusted, stratified, and meta-analysis models were explored (using the methods described above). In the stratification analysis, subjects were classified into three groups: carriers of Ɛ2 or Ɛ4 alleles, and carriers of the Ɛ3Ɛ3 genotype. Ɛ2Ɛ4 genotype carriers were excluded from this analysis.

Estimation of inbreeding and population structure

Wright's population inbreeding coefficient (*F*) was calculated according to heterozygote reduction, with regard to Hardy-Weinberg expectations, according to the formulae: $F = 1 - Ho/He$, where Ho is the frequency of heterozygotes observed in the sample population and He is the frequency of heterozygotes expected under Hardy Weinberg. The final inbreeding calculation responds to the mean of *F* per number of markers included in the analysis. Markers were included in this analysis if they were common markers for all the dataset, had MAF > 0.2 (34), and HWE > 0.001. Individuals with less than 99% of available genotypes were excluded. Inbreeding calculations for F.ACE dataset are very imprecise compared to other datasets available, due to it comprises a small number of SNPs. Thus, inbreeding calculation with a higher number of SNPs was also performed excluding F.ACE dataset.

PCs analysis was conducted to discard LRLD was caused by differential population structure. Plink 1.9 software was used to perform the analysis. PCs analysis was conducted in 19,979 markers, which were common between studies, and which

presents low linkage disequilibrium (LD) (r2<0.3). In addition, long range LD regions were excluded to the analysis.

RESULTS

Replication results and meta-analysis

SNPs included in this study were in Hardy-Weinberg equilibrium ($P > 0.05$). A significant risk effect was observed for the minor allele of *ABCA7* rs4147929:G>A (OR=1.15; 95% CI=1.04–1.27) in the Spanish population. *ABCA7* effect survived co-variation and *APOE*-Ɛ4 stratification analyses (Table 1). Meta-analysis including new data generated, the IGAP results and studies fulfilling inclusion criteria (n = 182 208), re-affirmed a genome-wide significant association for *ABCA7* (OR=1.15; 95% CI=1.12–1.19, Figure 1A). For *CD33*, we could not replicate the originally reported protective effect of rs3865444:C>A SNP. In the *APOE*-Ɛ4 stratified analysis, *CD33* did not modulate susceptibility to AD in any stratum (Table 1). Our meta-analysis also revealed heterogeneity and a non-significant association (Figure 1B).

Linkage disequilibrium analysis

Next, we sought to explore potential reasons for the lack of replication of the association of *CD33* rs3865444:C>A with AD in some populations. Specifically, patterns of LD were studied in five datasets, comprising data from the present study and four non-imputed GWAS datasets.

We detected significant LRLD between *APOE* rs7412:C>T and *CD33* rs3865444:C>A in two of the five datasets analysed (D' ≥ 0.2) (Table 2). The probability of these results being spurious was 0.011 for the Murcia dataset and 0.002 for the NIA data (Supplementary Table 4). However, the other datasets did not display LRLD: ADNI, D' = 0.17 (P = 0.11); GenADA, D' = 0.14 (P = 0.004); and F.ACE, D' = 0.01 (P = 0.86) (Table 2). In contrast, only one dataset exhibited significant LD between *APOE* rs429358:C>T and *CD33* rs3865444:C>A; however it did not reach the lower threshold for LD (D' ≥ 0.2). We next tested the LRLD hypothesis by assessing the pattern of LD along chromosome 19. This alternative strategy highlighted common areas of high LD in all datasets, particularly at positions 10.7–13.5 Mb, 42–45.1 Mb, and 47.8–50 Mb (Figure 2A). In addition, the Murcia and ADNI dataset exhibited the largest segments of LD at positions 9.7–10.7 Mb and 9.8–13.8 Mb, respectively. ADNI dataset also presents the largest LD segment at positions 41.8–51.1Mb, the latter of which contained the *APOE* gene at position 50.1 Mb (Figure 2B). However, this strategy did not reveal a general LD between *APOE* and *CD33*. This observation is supporting the notion that long range LD between *CD33* and *APOE*, if exist, would be affecting to a discrete fraction of chromosomes in some human populations.

TABLE 1. Association between *ABCA7* rs4147929-A or *CD33* rs3865444-T and LOAD in unadjusted, adjusted and *APOE*-Ɛ4 stratified models

GENE marker	Unadjusted	Adjusted (sex and APOE-ε4)	Adjusted (age, sex and APOE-ε4)	Stratification per APOE-ε4	
				Carriers	Non-Carrier
	OR CI95% P value	OR CI95% P value	OR CI95% P value	OR CI95% P value	OR CI95% P value
ABCA7 rs4147929	1.148 1.036 – 1.27 0.008	1.190 1.070 – 1.324 0.0013	1.175 0.94 – 1.47 0.1546	1.225 1.006 – 1.492 0.043	1.167 1.028– 1.324 0.017
CD33 rs3865444	0.986 0.897 – 1.084 0.771	0.976 0.885-1.076 0.6287	0.920 0.750 –1.127 0.4182	1.157 0.970 – 1.380 0.104	0.899 0.797 – 1.014 0.083

LOAD, Late Onset Alzheimer's disease; OR, Odds Ratio; CI, Confidence Interval.

Finally, an association analysis approach according to the presence or absence of the *APOE* Ɛ2 genotype also supported that long range association between *APOE* and *CD33* is not very common in studied populations. After Bonferroni correction, the majority of significant signals mapped to a region of chromosome 19, 231 kb upstream of *APOE* rs7412C>T (Supplementary Table 4).

APOE stratification analysis and meta-analysis of GWAS data

Stratification analysis revealed a significant protective association for *CD33* rs3865444:C>A only in Ɛ4 allele carriers in the GenADA dataset (OR = 0.69; *P* = 0.009) (Table 3). This protective trend for *CD33* rs3865444:C>A was also observed in Ɛ2 carriers of Spanish origin (Murcia study: OR = 0.49, 95% CI, 0.19–1.25; F.ACE: OR = 0.68, 95% CI 0.45–1.03).

Interestingly, the minor allele of *CD33* rs3865444:C>A was associated with increased susceptibility to AD in Ɛ4 carriers in the same Spanish datasets (Murcia study: OR = 1.13, 95% CI, 0.79–1.63; F.ACE: OR = 1.15, 95% CI, 0.97–1.37). Opposite effect directions were observed in the NIA and ADNI datasets. After meta-analysis, homozygous Ɛ3Ɛ3 carriers were identified as the most homogeneous stratum for *CD33* effects, while Ɛ4 allele carriers were the most heterogeneous (Table 3). Finally, we excluded hidden relatedness and population stratification as a potential cause for LRLD by analysing inbreeding (Supplementary Table 6) and conducting principal components (PCs) analysis across the datasets (Supplementary Figure 1).

DISCUSSION

Confirmatory data on AD associations with loci neighbouring *APOE* are still lacking, despite several replication efforts. In our

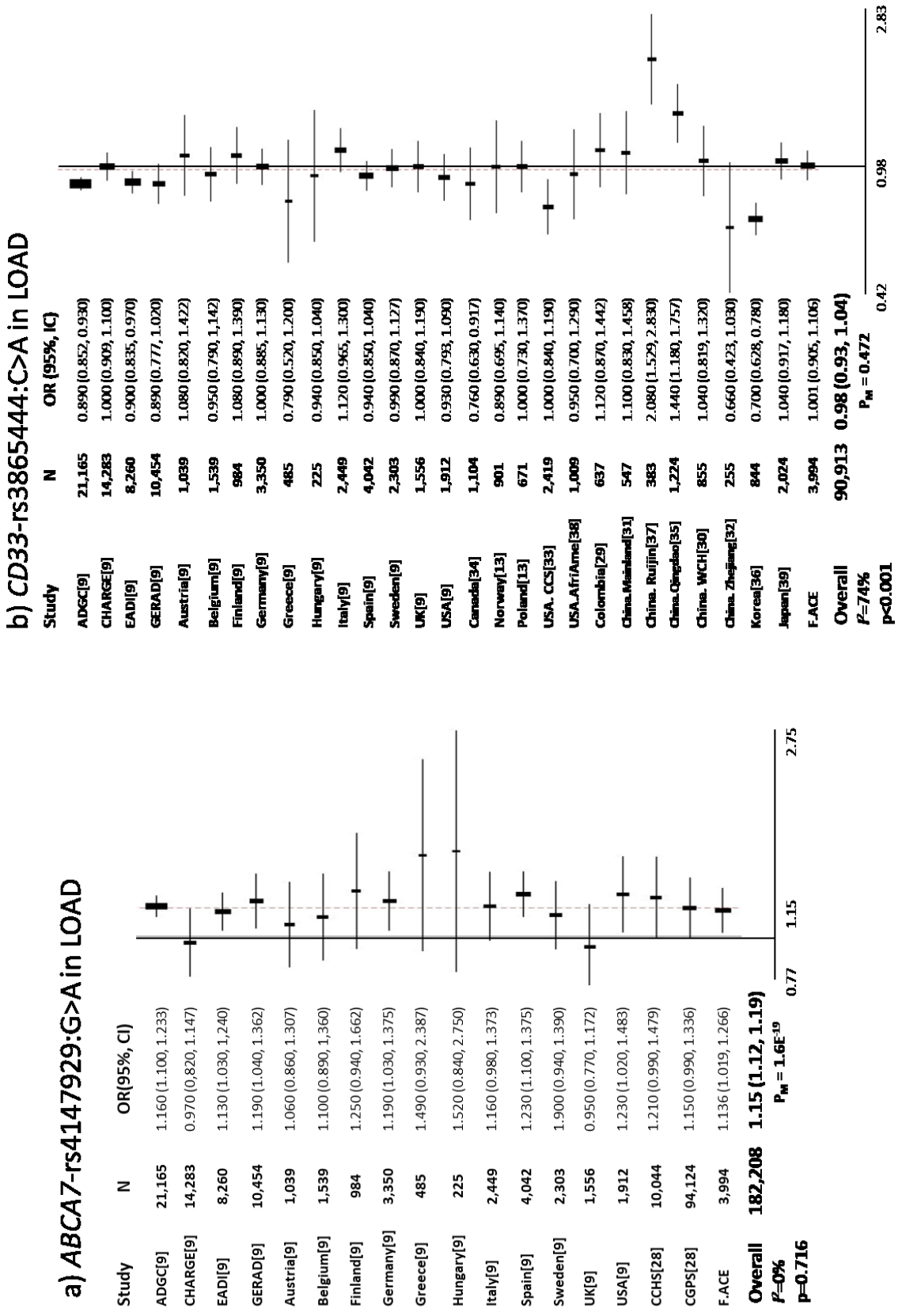


FIGURE 1. Forest plots for (A) ABCA7 rs4147929:G>A and (B) CD33 rs3865444:C>A. ABCA7 rs4147929:G>A and CD33 rs3865444:C>A

TABLE 2. Linkage disequilibrium results and test of significance for LD between APOE rs7412:C>T or rs429358:C>T

Dataset	LD				
	N	D'	r2	Chi sq	*P* value
***APOE* rs429358 – *CD33* rs3865444**					
ADNI	358	- 0.087	0.001	0.86	0.350
GenADA	1555	- 0.105	0.001	5.32	**0.020**
Murcia	1088	+ 0.036	5.00e-04	1.16	0.280
NIA	1789	+ 0.015	2.00e-04	0.75	0.380
F_ACE	4402	+ 0.013	8.00e-05	0.70	0.400
***APOE* rs7412 – *CD33* rs3865444**					
ADNI	358	+ 0.174	0.003	2.62	0.105
GenADA	1555	+ 0.138	0.002	8.00	**0.004**
Murcia	1088	- 0.312	0.002	5.75	**0.016**
NIA	1789	- 0.228	0.001	4.70	**0.030**
F_ACE	4402	+ 0.005	3.25e-06	0.03	0.860
***APOE* rs429358 – *APOE* rs7412**					
ADNI	358	- 1	0.020	13.79	**2.00×10^{-4}**
GenADA	1555	- 1	0.020	63.78	**1.44×10^{-15}**
Murcia	1088	- 1	0.010	23.42	**1.30×10^{-6}**
NIA	1789	- 1	0.030	93.22	**4.68×10^{-22}**
F_ACE	4402	- 1	0.010	91.06	**1.39×10^{-21}**

we clearly replicated the *ABCA7* signal ($P = 1.60 \times 10^{-19}$), which is located far from the *APOE* locus. However, we failed to replicate the *CD33* association in the Spanish population. Both observations are independent and consistent with previous IGAP observations (9).

ABCA7 marker was, firstly, identified such an AD locus by Hollingworth et. al. (6) and Naj et al. (8), in a large cohort of individuals (n=29,544). Following, this association was supported by IGAP consortium (9) and confirmed in non-European populations (10). Our finding is in accordance with previously studies. It reinforces the role of *ABCA7* as a candidate gene for AD and warrants its functional characterisation. *ABCA7* is involved in lipid metabolism (35) and apoptotic cell clearance (36). Its loss has been related to deficient macrophage clearance of amyloid plaques (37) and to accelerated amyloid-beta production (35).

CD33 locus appeared such as a promising signal (6) especially as a candidate target for immune-related therapies (38). Despite that, the signal disappeared after the IGAP meta-analysis (9), casting doubts on its real contribution to AD. In that scenario, we

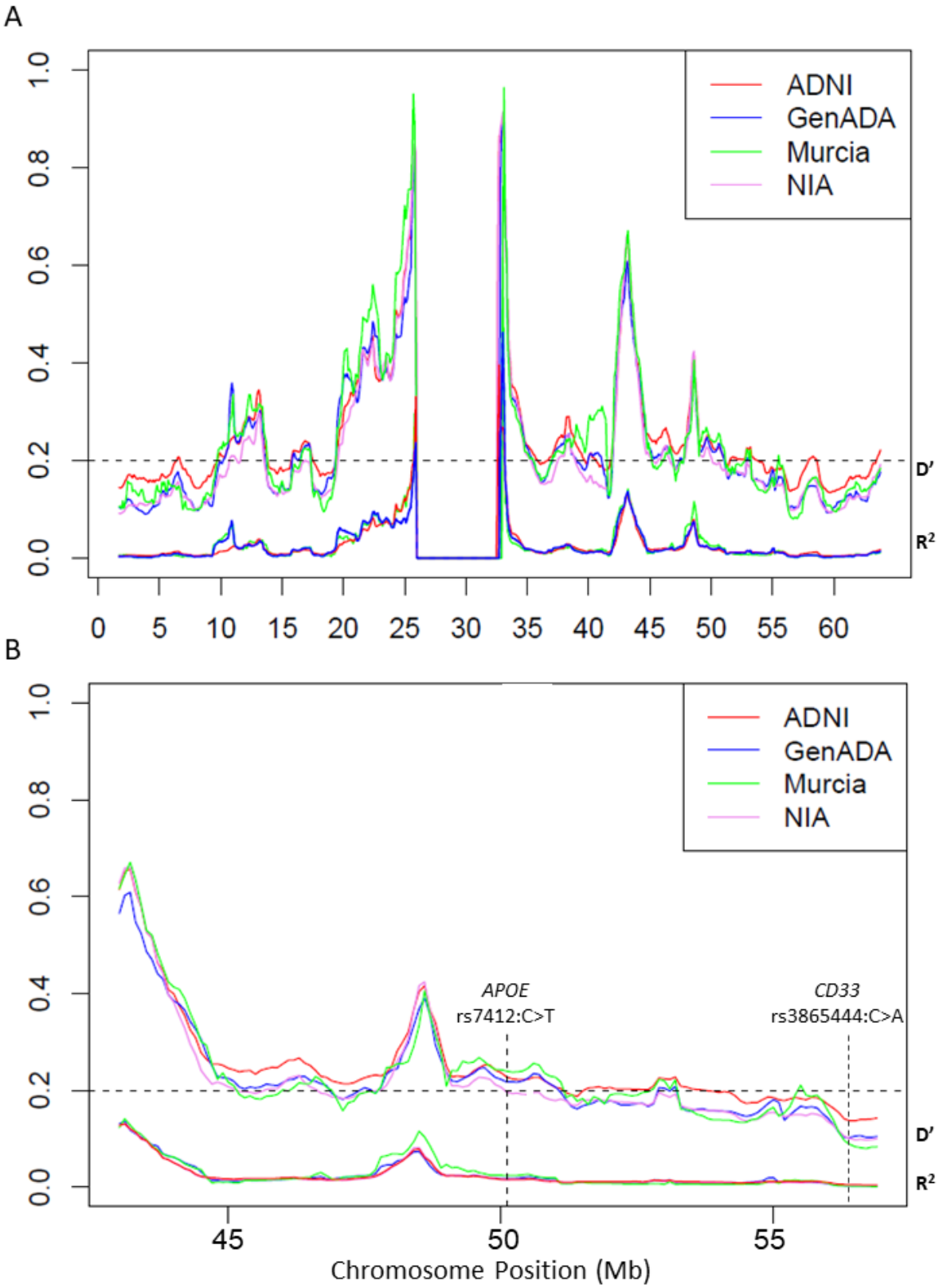


FIGURE 2. Linkage disequilibrium patterns across (A) chromosome 19 and (B) APOE region

Average D' (top groups) and r2 coefficients (bottom groups) plotted in sliding windows containing all common polymorphisms separated by 50 and 500 kb in successive 1.7-Mb segments (1.6-Mb overlap). Genome assembly NCBI36/hg18

TABLE 3.. Stratification per *Ɛ2, Ɛ4 APOE* alleles and *Ɛ3Ɛ3 APOE* genotype for five studied dataset and meta- analysis

Dataset	Unadjusted model OR (CI95%) *P* value	Ɛ2 allele carriers OR (CI95%) *P* value	Ɛ4 allele carriers OR (CI95%) *P* value	Ɛ3Ɛ3 genotype carriers OR (CI95%) *P* value
ADNI	0.865 (0.627 –1.193) 0.377	2.200 (0.553 – 8.76) 0.256	0.979 (0.580 – 1.654) 0.938	0.706 (0.419 – 1.187) 0.188
GenADA	0.864 (0.742 – 1.006) 0.06	0.863 (0.483 – 1.544) 0.620	0.697 (0.540 – 0.899) **0.005**	1.085 (0.862 – 1.366) 0.487
Murcia	0.938 (0.763 – 1.155) 0.550	0.492 (0.193 – 1.252) 0.130	1.133 (0.788 – 1.629) 0.498	0.849 (0.635 – 1.135) 0.269
NIA	0.956 (0.829 – 1.103) 0.540	1.356 (0.660 – 2.784) 0.401	0.873 (0.691 – 1.104) 0.257	0.908 (0.714 – 1.154) 0.430
F.ACE	0.986 (0.897 – 1.084) 0.772	0.682 (0.453 – 1.027) 0.066	1.153 (0.967 – 1.376) 0.112	0.928 (0.818 – 1.054) 0.25
Meta-Analysis	0.947 (0.887 – 1.010) 0.098	0.850 (0.650 – 1.111) 0.234	0.946 (0.765 – 1.170) 0.609	0.932 (0.849 – 1.023) 0.141
Heterogeneity	I^2 = 0% *P* = 0.658 Fixed	I^2 = 32% *P* = 0.206 Fixed	I^2 = **66%** *P* = **0.020** Random	I^2 = 0% *P* = 0.523 Fixed

believed that the *CD33* locus appears to be affected by the "*APOE* curse"; i.e., the impossibility of determining whether additional AD loci truly exist around *APOE*. Similar to *CD33*, conflicting data exist for *TOMM40*'523, *EXOC3L2*-597668, and *PLD3* (V232M) (11), (12). Several strategies have been attempted to tackle this issue; for example, using phylogenetic analysis, the *TOMM40*'523 poly-T marker was shown to be associated with age-at-onset in the *APOE* Ɛ3 subgroup (4). However, this finding has not been extensively replicated (39).

The lack of association between AD and *CD33* prompted us to search for an underlying explanation. Our data suggested significant, weak and non-universal, LRLD between the *APOE*-Ɛ2 allele and *CD33* rs3865444:C>A. Of course, there are limitations on these observations. First of all, very low levels of *r2* were detected, however the intrinsic properties of *r2* prompts to consider *D'* such as more informative measure for assessing historical recombination in a given population (40). Furthermore, we are dealing with relatively weak *D'* values, roughly 0.3, which means that 30% of the chromosomes will be carrying the long LD tract in specific populations. This could be compromising the capacity to detect LRLD when a direct measure of LD between studied markers is not determined.

Second, LRLD patterns differ across populations and are dependent on many factors, including admixture or migration, genetic drift, chromosome inversions, epistatic selection and hitchhiking effects (41). Furthermore, differential natural selection pressures across genomic regions, depending on specific geographical or environmental conditions, can lead to differential patterns of allele micro-heterogeneity. Although differential population structure was not identified in our GWAS datasets analysis, there is compelling evidence that micro-stratification cannot be detected by standard methods (42) and, therefore, this remains a potential limitation. Thus, the existence of undetected population sub-structure, with different LRLD patterns, could act as confounding factor, explaining the divergent observations and lack of replication between *CD33* and AD across studies (9). Of note, the genuine or spurious character of the association would remain masked under LRLD patterns, being its appearance highly dependent of the population structure.

An additional explanation for the lack of association between *CD33* and AD could be that the original *CD33* signals were simply chance findings. We feel that this possibility is less likely because under the assumption of a random association between *CD33* and AD, the chance of observing effects in opposite directions in independent studies would have the same probability. This latter observation is clearly not the situation reported in the literature to date.

Importantly, we detected LRLD upstream of the *APOE* locus using a previously described method (33). Confirmation of LRLD around *APOE* is relevant for several reasons. First, the data will assist in clarifying whether or not reported AD signals are genuine. Second, differential LRLD across populations may reveal the existence of structural variations, such as large inversions, insertions, or deletions, unequally affecting human populations. Structural variants have been implicated in the aetiology of the majority of multifactorial diseases (43).

In summary, we confirm *ABCA7* is associated with LOAD. However, we could not confirm the association between *CD33* and AD. Our data suggest that LRLD between *CD33* markers and the *APOE* alleles might explain the observed lack of consistency of *CD33* signal. Further studies using independent populations are required to clarify whether LRLD interferes with real associations between loci around *APOE* and AD.

REFERENCES

1. Moreno-Grau S, Ruiz A. Genome research in pre-dementia stages of Alzheimer's disease. Expert Rev Mol Med. 2016;18:e11.

2. Corder E, Saunders A. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science (80-). 1993;8(14):41–3.

3. Corder EH, Saunders AM, Risch NJ, Strittmatter WJ, Schmechel DE, Gaskell PC, et al. Protective effect of apolipoprotein E type 2 allele for late

onset Alzheimer disease. Nat Genet. 1994 Jun;7(2):180–4.

4. Roses AD, Lutz MW, Amrine-Madsen H, Saunders AM, Crenshaw DG, Sundseth SS, et al. A TOMM40 variable-length polymorphism predicts the age of late-onset Alzheimer's disease. Pharmacogenomics J. 2010 Oct;10(5):375–84.

5. Seshadri S, Fitzpatrick AL, Ikram MA, DeStefano AL, Gudnason V, Boada M, et al. Genome-wide analysis of genetic loci associated with Alzheimer disease. JAMA. 2010 May;303(18):1832–40.

6. Hollingworth P, Harold D, Sims R, Gerrish A, Lambert J-C, Carrasquillo MM, et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nat Genet. 2011 May;43(5):429–35.

7. Cruchaga C, Karch CM, Jin SC, Benitez BA, Cai Y, Guerreiro R, et al. Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer ' s disease. Nature. 2014;505(7484):550–4.

8. Naj AC, Jun G, Beecham GW, Wang L-S, Vardarajan BN, Buros J, et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nat Genet. 2011 May;43(5):436–41.

9. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet. 2013 Dec;45(12):1452–8.

10. Reitz C, Jun G, Naj A, Rajbhandary R, Vardarajan BN, Wang L-S, et al. Variants in the ATP-binding cassette transporter (ABCA7), apolipoprotein E ϵ4,and the risk of late-onset Alzheimer disease in African Americans. JAMA. 2013 Apr 10;309(14):1483–92.

11. Chouraki V, Seshadri S. Genetics of Alzheimer's disease. Adv Genet. 2014 Jan;87:245–94.

12. Heilmann S, Drichel D, Clarimon J, Fernández V, Lacour A, Wagner H, et al. PLD3 in non-familial Alzheimer's disease. Nature. 2015 Apr 1;520(7545):E3–5.

13. Carrasquillo MM, Belbin O, Hunter TA, Ma L, Bisceglio GD, Zou F, et al. Replication of EPHA1 and CD33 associations with late-onset Alzheimer's disease: a multi-centre case-control study. Mol Neurodegener . 2011 Jan;6(1):54.

14. Antúnez C, Boada M, González-Pérez A, Gayán J, Ramírez-Lorca R, Marín J, et al. The membrane-spanning 4-domains, subfamily A (MS4A) gene cluster contains a common variant associated with Alzheimer's disease. Genome Med. 2011 Jan;3(5):33.

15. Ruiz A, Heilmann S, Becker T, Hernández I, Wagner H, Thelen M, et al. Follow-up of loci from the International Genomics of Alzheimer's Disease Project identifies TRIP4 as a novel susceptibility gene. Transl Psychiatry . 2014 Jan;4:e358.

16. Calero O, Hortigüela R, Bullido M, Calero M. Apolipoprotein E genotyping method by Real Time PCR, a fast and cost-effective alternative to the TaqMan® and FRET assays. J Neurosci Methods. 2009;183(2):238–40.

17. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet . 2007 Sep 9;81(3):559–75.

18. Kjeldsen EW, Tybjærg-Hansen A, Nordestgaard BG, Frikke-Schmidt R. ABCA7and risk of dementia and vascular disease in the Danish population. Ann Clin Transl Neurol. 2018 Jan;5(1):41–51.

19. Moreno DJ, Ruiz S, Ríos Á, Lopera F, Ostos H, Via M, et al. Association of GWAS Top Genes With Late-Onset Alzheimer's Disease in Colombian Population. Am J Alzheimers Dis Other Demen. 2017 Feb 13;32(1):27–35.

20. Zhang D-F, Li J, Wu H, Cui Y, Bi R, Zhou H-J, et al. CFH Variants Affect Structural and Functional Brain Changes and Genetic Risk of Alzheimer's Disease. Neuropsychopharmacology. 2015 Aug 5.

21. Jiao B, Liu X, Zhou L, Wang MH, Zhou Y, Xiao T, et al. Polygenic Analysis of Late-Onset Alzheimer's Disease from Mainland China. PLoS One. 2015;10(12):e0144898.

22. Mao Y-F, Guo Z-Y, Pu J-L, Zhang Y-XC-R. Association of CD33 and MS4A cluster variants with Alzheimer's disease in East Asian Populations. Neurosci Lett. 2015 Oct 8.

23. Ebbert MTW, Ridge PG, Wilson AR, Sharp AR, Bailey M, Norton MC, et al. Population-based Analysis of Alzheimer's Disease Risk Alleles Implicates Genetic Interactions. Biol Psychiatry. 2014 May 1;75(9):732–7.

24. Omoumi A, Fok A, Greenwood T, Sadovnick AD, Feldman HH, Hsiung G-YR. Evaluation of late-onset Alzheimer disease genetic susceptibility risks in a Canadian population. Neurobiol Aging. 2014 Apr;35(4):936.e5-12.

25. Tan L, Yu J-T, Zhang W, Wu Z-C, Zhang Q, Liu Q-Y, et al. Association of GWAS-linked loci with late-onset Alzheimer's disease in a northern Han Chinese population. Alzheimers Dement . 2013 Sep;9(5):546–53.

26. Chung SJ, Lee J-H, Kim SY, You S, Kim MJ, Lee J-Y, et al. Association of GWAS top hits with late-onset Alzheimer disease in Korean population. Alzheimer Dis Assoc Disord . Jan;27(3):250–7.

27. Deng Y-L, Liu L-H, Wang Y, Tang H-D, Ren R-J, Xu W, et al. The prevalence of CD33 and MS4A6A variant in Chinese Han population with Alzheimer's disease. Hum Genet. 2012 Jul;131(7):1245–9.

28. Logue MW. A Comprehensive Genetic Association Study of Alzheimer Disease in African Americans. Arch Neurol. 2011 Dec 1;68(12):1569.

29. Miyashita A, Koike A, Jun G, Wang L-S,

Takahashi S, Matsubara E, et al. SORL1 is genetically associated with late-onset Alzheimer's disease in Japanese, Koreans and Caucasians. PLoS One. 2013 Jan;8(4):e58618.

30. Weiner M, Aisen P, Jack Jr C, Jaugust W, Trojanowski J, Shaw L, et al. The Alzheimer's disease neuroinmaging iniciative: Progress report and future plans. Alzheimer dement. 2010;6(3):202–11.

31. Li H, Wetten S, Li L, St Jean PL, Upmanyu R, Surh L, et al. Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. Arch Neurol. 2008 Jan;65(1):45–53.

32. Wijsman EM, Pankratz ND, Choi Y, Rothstein JH, Faber KM, Cheng R, et al. Genome-wide association of familial late-onset Alzheimer's disease replicates BIN1 and CLU and nominates CUGBP2 in interaction with APOE. PLoS Genet. 2011 Feb;7(2):e1001308.

33. Dawson E, Abecasis GR, Bumpstead S, Chen Y, Hunt S, Beare DM, et al. A first-generation linkage disequilibrium map of human chromosome 22. Nature. 2002 Aug 1;418(6897):544–8.

34. Zhang Q, Calus MPL, Guldbrandtsen B, Lund MS, Sahana G. Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. BMC Genet. 2015;16:88.

35. Sakae N, Liu C-C, Shinohara M, Frisch-Daiello J, Ma L, Yamazaki Y, et al. ABCA7 Deficiency Accelerates Amyloid-β Generation and Alzheimer's Neuronal Pathology. J Neurosci. 2016 Mar 30;36(13):3848–59.

36. Jehle AW, Gardai SJ, Li S, Linsel-Nitschke P, Morimoto K, Janssen WJ, et al. ATP-binding cassette transporter A7 enhances phagocytosis of apoptotic cells and associated ERK signaling in macrophages. J Cell Biol. 2006 Aug 14;174(4):547–56.

37. Kim WS, Li H, Ruberu K, Chan S, Elliott DA, Low JK, et al. Deletion of ABCA7 increases cerebral amyloid-β accumulation in the J20 mouse model of Alzheimer's disease. J Neurosci. 2013 Mar 6;33(10):4387–94.

38. Bradshaw EM, Chibnik LB, Keenan BT, Ottoboni L, Raj T, Tang A, et al. CD33 Alzheimer's disease locus: altered monocyte function and amyloid biology. Nat Neurosci. 2013 May 23;16(7):848–50.

39. Cruchaga C, Nowotny P, Kauwe JSK, Ridge PG, Mayo K, Bertelsen S, et al. Association and expression analyses with single-nucleotide polymorphisms in TOMM40 in Alzheimer disease. Arch Neurol. 2011 Aug;68(8):1013–9.

40. Mueller JC. Linkage disequilibrium for different scales and applications. Brief Bioinform . 2004 Dec;5(4):355–64.

41. Ardlie KG, Kruglyak L, Seielstad M. Patterns of Linkage Disequilibrium in the Human Genome. Nat Rev Genet. 2002;3:299–309.

42. Campbell CD, Ogburn EL, Lunetta KL, Lyon HN, Freedman ML, Groop LC, et al. Demonstrating stratification in a European American population. Nat Genet. 2005 Aug 24;37(8):868–72.

43. Beckmann JS, Estivill X, Antonarakis SE. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. Nat Rev Genet. 2007 Aug;8(8):639–46.

AVAILABILITY OF THE DATA AND MATERIAL

Fundacio ACE datasets used during the current study are available from the corresponding author on reasonable request. Murcia study data analysed are available upon request to Carmen Antunez, Manuel Serrano-Rios and Agustin Ruiz. The GENADA and NIA datasets that support the findings of this study are available from dbGaP but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. The ADNI datasets analysed during the current study are available in the ADNI repository, http://adni.loni.usc.edu/ .

ACKNOWLEDGEMENTS

We would like to thank patients and controls who participated in this project. We are indebted to Trinitat Port-Carbó and her family for their support of Fundació ACE research programs. Fundació ACE collaborates with the Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED, Spain) and is one of the participating centers of the Dementia Genetics Spanish Consortium (DEGESCO). The genotypic and associated phenotypic data used in the study, 'Multi-Site Collaborative Study for Genotype-Phenotype Associations in Alzheimer's Disease (GenADA)', were provided by GlaxoSmithKline, R&D Limited. The datasets used for analyses described in this manuscript were obtained from dbGaP through dbGaP accession numbers phs000219.v1.p1 and phs000168.v1.p1. Funding support for the 'Genetic Consortium for Late Onset Alzheimer's Disease' was provided through the Division of Neuroscience, NIA. The Genetic Consortium for Late Onset Alzheimer's Disease included a GWAS, funded as part of the Division of Neuroscience, NIA. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the Genetic Consortium for Late Onset Alzheimer's Disease. Furthermore, part of the data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbV ie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd. and its affiliated company, Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies;

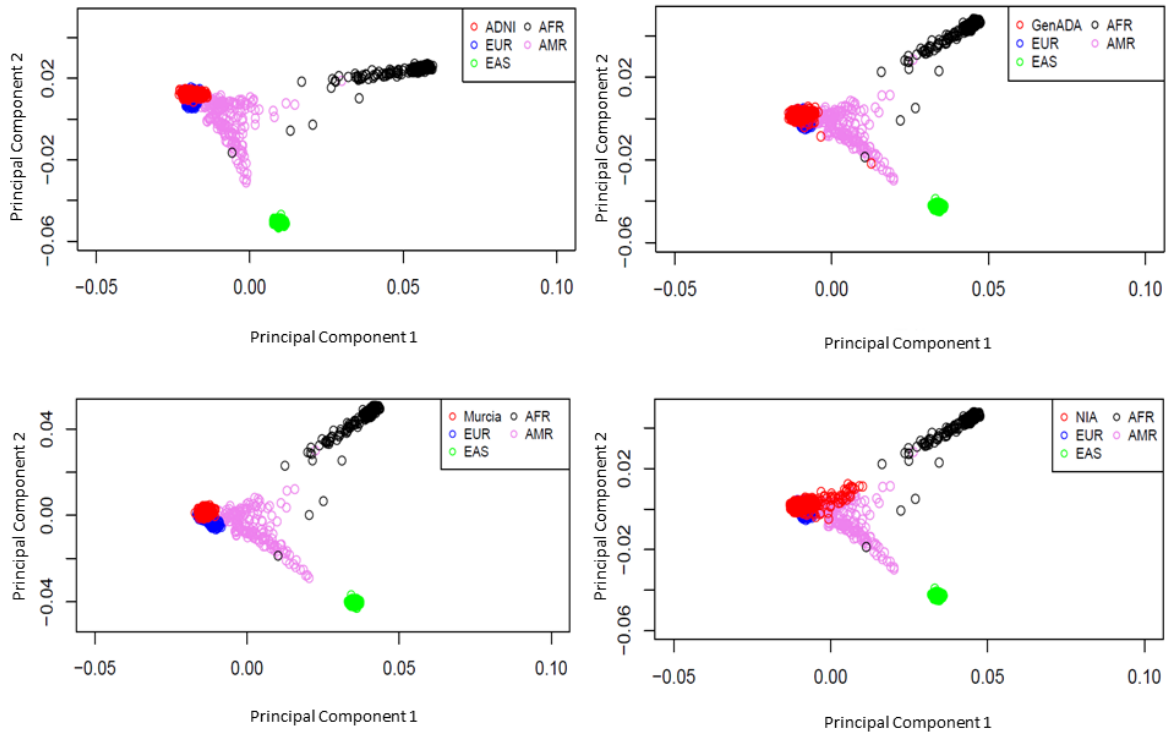
Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research provides funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. **The present work was performed as part of the Biochemistry, Molecular Biology and Biomedicine doctoral program of S. Moreno-Grau at Universitat Autònoma de Barcelona (Barcelona, Spain)**.

AUTHOR'S CONTRIBUTION

SMG and ARu contributed to the study design and analysed data. SMG prepared the first draft of the manuscript. IH, SR, MR, AM, LV, ORG, MA, AE, GO, MT, CA, DS and JLA contributed to phenotyping, database curation and patients selection. CA, MSR and ARu carried out Murcia study. SMG and OSG contributed to data processing methods. SHH and ARa carried out genotyping procedures. WM, LT, MB and ARu participated in funds procurement. ARa and ARu were responsible for the critic revision of the manuscript. All authors read and approved the final manuscript.

Conflict of Interest: None. The authors declare that they have no competing interest.

SUPPLEMENTARY FIGURES



SUPPLEMENTARY FIGURE 1. Characteristics of the population structure in four analysed GWAS datasets
EUR: Europeans; AFR: African; AMR: Americans; EAS: East Asian.

SUPPLEMENTARY TABLES

SUPPLEMENTARY TABLE 1. Full description of the Spanish replication sample.

Dataset	Cases				Controls			
		Mean Age (SD)	% Women	%*APOE* Ɛ4	N	Mean Age (SD)	% Women	%*APOE* Ɛ4
FACE	617	82.3 (7.7)	70.4%	45.1%	125	62.3 (7.7)	75.2%	25.6%
CBC	59	78.4 (8.5)	69.4%	39.6%	0	-	-	-
CSC		-	-	-	2002	47.7 (10.6)	50.3%	18.4%
NXC		-	-	-	515	67.1 (4.7)	96.3%	17.8%
Murcia	0	78.4 (8.3)	70%	50%	0	-	-	-
TOTAL	796	82.1 (7.9)	70.2%	44.7%	2642	54.1 (11.6)	64.3%	18.6%

FACE: Fundacio ACE. Institut Català de neurociències (Barcelona, Spain); CBC: Unidad de Memoria, Hospital Universitario La Paz-Cantoblanco (Madrid, Spain); CSC: Proyecto Viva Segovia. Hospital Clínico Sán Carlos; NXC: Neocodex; Murcia: Unidad de Demencias, Hospital Universitario Virgen de la Arrixaca (Murcia, Spain).

SUPPLEMENTARY TABLE 2. Full description of the datasets used in the meta-analysis assessing *ABCA7* and *CD33* susceptibility in LOAD

***ABCA7*-rs4147929:G>A Meta-analysis**

Country/Consortium	Study	Ancestry	N cases	N Controls	N total
ADGC	Lambert 2013	European	10,273	10,892	21,165
CHARGE	Lambert 2013	European	1,315	12,968	14,283
EADI	Lambert 2013	European	2,243	6,017	8,260
GERAD	Lambert 2013	European	3,177	7,277	10,454
Austria	Lambert 2013	European	210	829	1,039
Belgium	Lambert 2013	European	878	661	1,539
Findland	Lambert 2013	European	422	562	984
Germany	Lambert 2013	European	972	2,378	3,350
Greece	Lambert 2013	European	256	229	485
Hungary	Lambert 2013	European	125	100	225
Italy	Lambert 2013	European	1,729	720	2,449
Spain	Lambert 2013	European	2,121	1,921	4,042
Sweeden	Lambert 2013	European	797	1,506	2,303
UK	Lambert 2013	European	490	1,066	1,556
USA	Lambert 2013	European	572	1,340	1,912
CCHS	Kjeldsen 2018	European	349	9,695	10,044
CGPS	Kjeldsen 2018	European	613	93,511	94,124
Fundacio ACE*	--	European	1,500	2,494	3,994
Total	--	European	28,042	154,166	182,208

***CD33*-rs3865444:C>A Meta-analysis**

Dataset	Study	Ancestry	N cases	N Controls	N Total
ADGC	Lambert 2013	European	10,273	10,892	21,165
CHARGE	Lambert 2013	European	1,315	12,968	14,283
EADI	Lambert 2013	European	2,243	6,017	8,260
GERAD	Lambert 2013	European	3,177	7,277	10,454
Austria	Lambert 2013	European	210	829	1,039
Belgium	Lambert 2013	European	878	661	1,539
Findland	Lambert 2013	European	422	562	984
Germany	Lambert 2013	European	972	2,378	3,350
Greece	Lambert 2013	European	256	229	485
Hungary	Lambert 2013	European	125	100	225
Italy	Lambert 2013	European	1,729	720	2,449
Spain	Lambert 2013	European	2,121	1,921	4,042
Sweeden	Lambert 2013	European	797	1,506	2,303
UK	Lambert 2013	European	490	1,066	1,556
USA	Lambert 2013	European	572	1,340	1,912
Canada	Omouni 2014	European	524	580	1,104
Norway	Carrasquillo 2011	European	346	555	901
Poland	Carrasquillo 2011	European	483	188	671
USA. CCS.	Ebbert 2014	European	326	2,093	2,419
USA.Afri-Amer	Logue 2014	African American	513	496	1,009
Colombia	J.D Moreno 2017	Hispanic American	280	357	637
China. Mainland	Jiao 2015	East Asia	229	318	547
China. Ruijin	Tan 2012	East Asia	190	193	383
China.Qingdao	Deng 2012	East Asia	612	612	1,224
China. WCH	Zhang 2015	East Asia	380	475	855
China. Zhejian	Mao 2015	East Asia	126	129	255
Korea	Chung	East Asia	290	554	844

Dataset	Study	Ancestry	N cases	N Controls	N Total
Japan	Miyashita 2013	East Asia	1,008	1,016	2,024
Fundacio ACE*	--	European	1,500	2,494	3,994
Total	--	Heterogeneous	32,387	58,526	90,913

*Fundacio ACE sample available for meta-analysis excludes samples overlapping with IGAP.

SUPPLEMENTARY TABLE 3. Number of variants and individuals in non-imputated GWAS datasets.

Dataset	N Variants	N cases	N controls	N Total
ADNI	553,938	164	194	358
GenADA	438,752	782	773	1,555
Murcia	198,431	319	769	1,088
NIA	524,084	987	802	1,789

SUPPLEMENTARY TABLE 4. Probability to detect false positive results after LD analysis under specified criterions (D' ≥ 0.2; D' ≥ 0.2 and P ≤ 0.05)

Dataset	P value (D' ≥ 0.2)	P value (D' ≥ 0.2; p < 0.05)
***APOE* rs429358:C>T**		
ADNI	0.649	0.058
GenADA	0.043	0.002
Murcia	0.559	0.011
NIA	< 0.001	< 0.001
***APOE* rs7412:C>T**		
ADNI	0.383	0.164
GenADA	0.132	0.003
Murcia	0.150	0.011
NIA	0.096	0.002

SUPPLEMENTARY TABLE 5. Results for the association analysis taking as phenotype the presence of Ɛ2 genotype. *CD33* rs3865444:C>A effect for *APOE*-Ɛ2 phenotype. Total number of genome-wide significant SNPs and range of physical distances covered.

Dataset	CD33 rs3865444:C>A Effect	Bonferroni Correction (p)	Significant SNPs (Bp range)
ADNI	OR =1.44 CI =0.79 – 2.62 p = 0.23	$9.02x10^{-8}$	2 (chr19:50,100,676-50,103,909)
GenADA	OR =1.26 CI =0.96 – 1.64 p = 0.09	$1.14x10^{-7}$	6 (chr19:49,868,180-50,103,909)
Murcia	OR =0.69 CI =0.49 – 0.98 p = 0.03	$2.50x10^{-7}$	2 (chr19:49,868,180-50,103,909)
NIA	OR =0.82 CI =0.63 – 1.08 p = 0.16	$9.54x10^{-8}$	10 (chr19:49,929,652-50,103,909)
F_ACE	OR =1.01 CI =0.85 – 1.20 p = 0.91	NA	NA

Bonferroni Correction (P) = 0.05/SNPs; NA = Not available.

SUPPLEMENTARY TABLE 6. Inbreeding coefficients calculated between common SNPs between datasets and 1000 random SNPs

Dataset	2 SNPS		1000SNPs/1000times		
	N	***F***	**N**	***F***	**SD**
ADNI	358	-0.019	335	0.002	0.001
Genada	1542	0.015	1555	0.004	0.0008
Murcia	1052	0.034	972	0.004	0.001
NIA	1778	0.017	1778	0.004	0.0007
F_ACE	4283	0.009	NA	NA	NA

Section 3.2. Candidate gene approach for uncovering new AD genes

3.2.1. Publication III

Evaluation of Candidate Genes Related to Neuronal Apoptosis in Late-Onset Alzheimer's Disease

Sonia Moreno-Grau [a,b], Bruna Barneda[b,c,d], Paulina Carriba[b,c,d], Juan Marín[e], Oscar Sotolongo-Grau[a], Isabel Hernández[a], Maitée Rosende-Roca[a], Ana Mauleón[a], Liliana Vargas[a], Ana Espinosa[a], Montserrat Alegret[a], Octavio Rodriguez[a], Gemma Ortega[a], Maria Victoria Fern´andez[a], Jesús López-Arrieta[f] , Lluís Tárraga[a], Mercè Boada[a], Carmen Antúnez[e], Joaquin López[b,c,d], Agustín Ruiz[a,*] and Joan Xavier Comella[b,c,d]

[a]Memory Clinic of Fundació ACE, Institut Català de Neurociències Aplicades, Barcelona, Spain;[b]Institut de Recerca de l'Hospital Universitari de la Vall d'Hebron (VHIR), Passeig Vall d'Hebron, Barcelona, Spain; [c]Institut de Neurociències, Departament de Bioqu´ımica i Biologia Molecular, Facultat de Medicina, Universitat Autònoma de Barcelona, Campus de Bellaterra (Edifici M), Bellaterra, Spain; [d]Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED), Spain; [e]Dementia Unit, University Hospital Virgen de la Arrixaca, Murcia, Spain; [f]Memory Unit, University Hospital La Paz-Cantoblanco, Madrid, Spain

---

ABSTRACT

The objective of this study was to identify genetic variation in genes encoding death receptors and signals that modulate their activity. After conducting a meta-analysis with five previous genome-wide association studies (GWAS) and aggregated data, the most significant signals, (*TNF* locus: rs2395488, rs2534672, and rs9267445; and *FASLG* locus: rs730278), were replicated in 1,046 cases and 372 controls. The rs2395488 and rs2534672 markers showed a modest protective effect (OR = 0.849, p = 0.49780; OR = 0.687, p = 0.11335), in contrast to rs730278 marker (OR = 1.146, p = 0.17212), which did not follow the previous effect direction, in any case it reached the significance level. Final meta-analysis, adding the replication sample, confirmed these observations. We concluded that *FASLG* marker is not etiologically linked to Alzheimer's disease (AD). However, single nucleotide polymorphisms (SNPs) around *TNF* locus require further analyses in order to explain the association between AD and human leukocyte antigen (*HLA*).

Keywords: Alzheimer's disease, genetics, genome-wide association study, meta-analysis, apoptosis, death receptors, *FASLG, TNF*

Journal of Alzheimer's Disease, 2015; 45(2), 621-629

doi: 10.3233/JAD-142721

INTRODUCTION

Late-onset Alzheimer's disease (LOAD) is the most common Alzheimer's disease (AD) phenotype. However, despite the advances provided by comprehensive genome scans, mechanisms involved in the age at onset (AAO) or disease progression remain unknown.

For over a decade, neuroinflammation has been suggested to play a crucial role in the development of AD (1). In that context, the tumor necrosis factor's (TNF) capacity to induce apoptosis seems to be mainly responsible for the neurotoxic effect, leading to apoptotic cell death (2).

Accumulated evidence suggests that programmed cell death (apoptosis) is involved in AD neuronal death (3). Its initiation is dependent on two major pathways: intrinsic through mitochondria and extrinsic through death receptors (DRs). DRs are a part of the TNF receptor superfamily (4), and play a crucial role in transmitting the death signals. DRs are activated through binding to their cognate ligands, mainly Fas ligand (FasL) and TNF. Evidence shows that DR ligands, FasL, and TNF expressions are increased in AD brains, such as TNF receptor 1 (3), (5), (6). Moreover, apoptosis is strictly controlled by a group of intracellular proteins (antiapoptotic proteins) that inhibit apoptotic process and promote survival. Studies have shown a reduction in antiapoptotic protein levels, such as cFLIP, BIRC3, and BIRC4, in AD cases (3), (6), (7) suggesting an upregulation of DR-related signaling and a down regulation of its natural inhibitors in AD brains, i.e., promoting the context necessary for the cells to die.

Based on the relation between neuroinflammation and cell death, we have hypothesized that genetic variation in genes leading and controlling apoptotic processes could be relevant in sporadic AD etiology. To assess our hypothesis, we have developed a candidate gene meta-analysis strategy using five previously published GWAS and a wet replication of selected candidate signals.

MATERIAL AND METHODS

Data standardization phase: datasets selection

All the GWAS included in this study were previously meta-analyzed and ranked using the ranking results of Antúnez et al. (5) and Boada et al. (6). The GWAS included: the ADNI study (7), the GenADA study (8), the TGEN study (9), the NIA study (10), and the Murcia study (5). Furthermore, we aggregated data from Harold et al. 2009 (11), Seshadri et al. 2010 (12), Hu et al. 2011 (13) and Naj et al. 2011 (14).

In silico phase: identification of SNP markers within selected candidate genes

SNP selection

We intended to select SNPs located at the genomic regions of the genes codifying for: (i) death receptors: TNFR1 (*TNFRSF1A* at 12p13.2), DR3 (*TNFSRF25* at 1p36.2), DR4 (*TNFSRF10A* at 8p21), DR6 (*TNFSRF21* at 6p21.1), and TWEAKR (*TNFSRF12A* at 16p13.3); (ii) its cognate ligands: FasL (*FASLG* at 1q23), TNF-α (*TNF* at 6q21.3); and (iii) antiapoptotic proteins: cFLIP (*CFLAR* at 2q33-q34), FAIM (*FAIM1* at 3q22.3), A20 (*TNFAIP3* at 6q23), BAG4 (*BAG4* at 8p11.23), BIRC proteins (*BIRC* cluster), and FAIM2 (*FAIM2* at 12q13), following a previously described approach (15).

Association studies

The most promising signals were selected after association studies, as previously described (5). A final meta-analysis, which included previous GWAS and the final replica, was performed on the selected markers using Plink and the command metan in Stata 12 (College Station, TX, USA). Alzheimer's disease genetics consortium (ADGC) p values were aggregated (when available) to observe statistical consistency of the results with the meta-GWAS results reported herein.

Validation phase: replication study

Subjects' selection

The replication study was conducted using 1,418 subjects: 1,046 sporadic AD patients [312 males, mean age = 80.7 years (standard deviation, SD = 8.62); and 724 females, mean age = 82.4 years (SD = 7.85)], and 372 unrelated controls with unknown cognitive status [183 males, mean age = 53.5 years (SD = 11.24); and 185 females, mean age = 54.4 years (SD = 10.09)]. The AD cases were recruited from the Fundació ACE-Institut Català de Neurociències Aplicades, Barcelona (Spain). Control subjects were retrieved from the Neocodex Biobank (16). Details of the inclusion criteria, recruitment, and testing have been previously described (15).

DNA extraction and replication genotyping

DNA extraction was performed as previously described (15). Genotyping of the selected markers was performed using real-time PCR in a LightCycler 480 System (Roche Diagnostics). Details of the primers and probes used for genotyping protocol are available upon request. The reaction mixture contained 1 μL of genomic DNA, 10 μL TaqMan GTXpress Master Mix buffer 2× concentration, 0.5 μL TaqMan genotyping assay 40× concentration, and 8.5 μL nuclease-free water to obtain a final volume of 20 μL. PCR conditions were as follows: initial step of 95 °C for 10 min to activate the Taq enzyme, followed by 40 cycles of 95 °C for 15 s and 60 °C for 1 min. Finally, a fluorometric register was performed.

Statistical Analysis

To explore the genotype distribution and deviation of Hardy–Weinberg (HW) equilibrium, we employed the online resource at the Institute for Human Genetics, Munich, Germany (http://ihg.gsf.de). The following unconditional logistic regression models were used to adjust the Spanish sample by age, sex, and APOE ε4 status: (i) a full model adjusted by age, gender, and *APOE* ε4; (ii) an intermediate model adjusted by gender and *APOE* ε4, and (iii) an unadjusted model. In addition, to check the effects of each marker on AAO for AD, a linear regression analysis was conducted using: adjusted model by gender and *APOE* status and an unadjusted model. Both analyses were run using SPSS 18.0 software. Power calculations were performed using the PS Power and Sample Size Calculations software.

RESULTS

In silico exploration of candidate genes

Association analysis of the initially selected 1,501 SNPs, localized in the genomic region of candidate genes (Supplementary Table 1), showed that 78 SNPs reached a nominal statistical significance (Supplementary Table 2). The most significant results were obtained in *TNF* region for three markers: rs2905726 (p = 0.00008), rs2534672 (p = 0.00186), and rs9267445 (p = 0.0025). A fourth best result for the marker rs730278 (p = 0.00331) was localized near *FASLG* gene outside the *HLA* (Supplementary Table 2).

Replication study of selected SNPs

A validation case-control study was designed to test whether the SNPs showing the most significant results could be replicated in our series. A total of 1,046 AD cases and 372 unrelated controls were genotyped for the four selected markers: rs2395488 (rs2905726 proxy), rs2534672, rs9267445, and rs730278.

The observed genotypic distribution was in accordance with the HW equilibrium for all SNPs, except for rs9267445 marker ($p = 2.834e^{-33}$), which was excluded from further analysis (data not shown). The analysis of the genotype data (Table 1) for rs2395488 and rs2534672 markers showed that homozygosis for the minor allele induced a protective effect even when the association did not reach the significance level. In contrast, the rs730278 marker showed a modest risk effect that did not follow the effect direction of previous meta-analysis (OR = 0.892 vs. OR = 1.146 obtained during replication). Covariate adjustments to test whether age, gender, and *APOE* ε4 genotype could influence the magnitude of the effect size indicated that the adjustments inverted the tendency of the effect for rs2395488 and rs2534672 markers (OR = 1.132, p = 0.556; OR = 1.016, p = 0.939). The authors concluded that observed effects did not resist covariation analysis by age (Table 1).

TABLE 1. rs2395488, rs2535672, and rs730278 effects using different statistical models in the validation study.

Marker	Subjects		Statistical Methods		
	Case n = 1046	Controls n = 372	Allele Frequency OR (CI95%) *P* value	Homozygous OR (CI95%) *P* value	Allele positivity OR (CI95%) *P* value
rs2395488 GG/GA/AA	78/424/534	27/123/157	0.960 0.786–1.171 0.68624	0.849 0.530–1.362 0.49780	0.984 0.763–1.269 0.90092
rs2534672 GG/GC/CC	60/389/568	30/141/195	0.882 0.729–1.067 0.19599	0.687 0.430-1.096 0.11335	0.901 0.709–1.146 0.39622
rs730278 TT/TC/CC	81/411/528	23/133/196	1.146 0.942–1.393 0.17212	1.307 0.800–2.137 0.28389	1.171 0.918–1.494 0.20434

OR: odds ratio; CI: confidence interval.

Results of the linear regression between AAO and AD indicated that the per-allele effect of target markers to modify AAO was almost null. The differences between the two models used were null for rs2395488 and very low for rs2535672 and rs730278. The marker rs730278 showed some influence on AAO, even when it did not reach significant levels ($\beta = -0.682$; $p = 0.101$).

Results of the meta-analysis showed a consistent effect of markers across studies. Despite this, none of these markers reached the study-wide significance level. The rs730278 meta-analysis results confirmed, as we had previously observed, that our data had an inverse effect on risk although it did not considerably modify the averaged OR obtained during the meta-analysis (Figure 1).

DISCUSSION

A large evidence supports that chronic inflammation impairs neuronal function and leads to neurodegeneration (17). In that context, there exists a proapoptotic tendency and a dysregulation of antiapoptotic mechanisms. However, a limited number of genetic studies analyzing genes involved in neuronal apoptosis exist (21), (22), (23).

Our exploration of the candidate genes revealed a weak association of *TNF* region with AD. Importantly, our results are in accordance with the results obtained by ADGC case-control datasets (14)

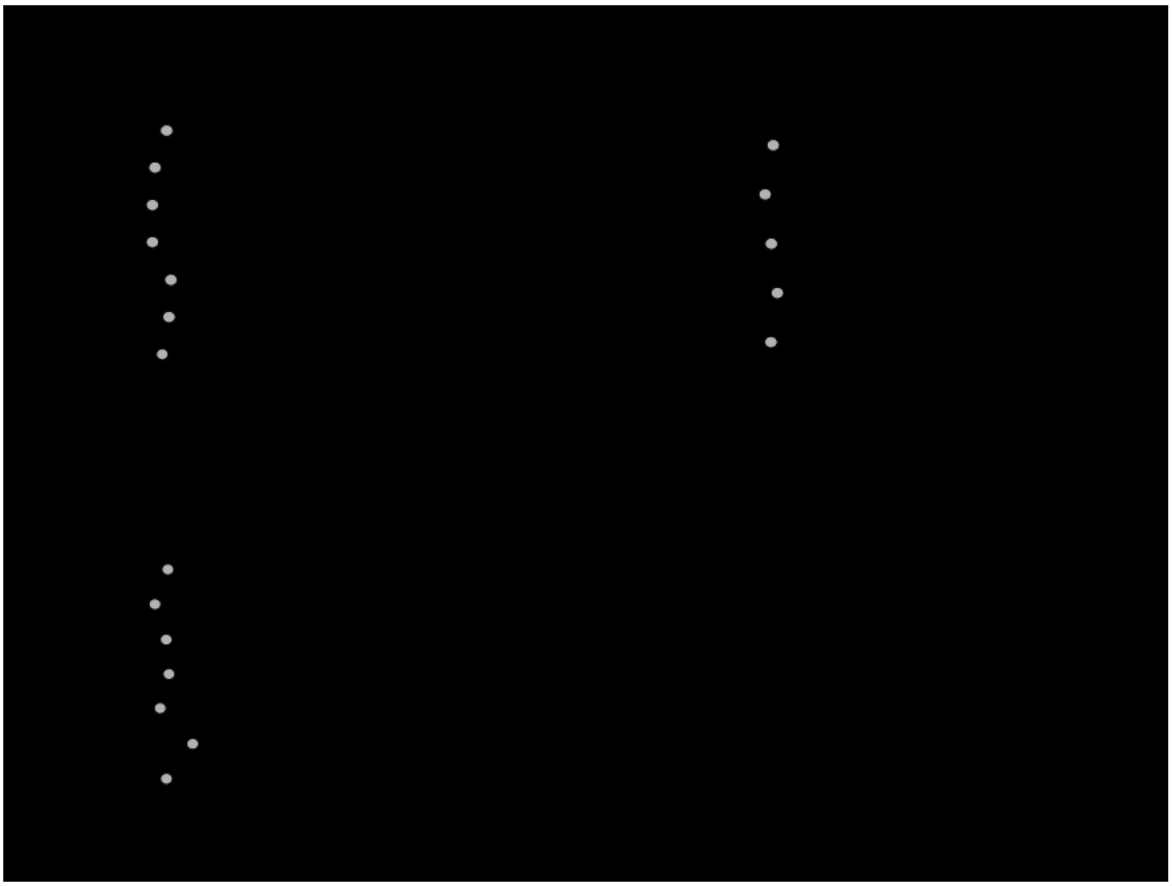


FIGURE 1. Fixed-effect model meta-analysis and Forest plot of rs2395488 (a), rs2534672 (b), and rs730278 (c)

Unfortunately, none of the SNPs selected for replication showed a nominal statistically significant association with the disease. No *FASLG* polymorphism has been previously associated with AD progression, suggesting that preselected SNPs could be false positives. With respect to SNPs on *TNF* gene, a risk haplotype and a polymorphism in -850 nucleotide promoter have been associated with AD (22), (23), (24); however, a major tendency for non-association survived (18). Taking into account that the most of the studies, including ours, were developed in small sample, which suggests that the sample size could be affecting the power of the study.

Despite our best results are close to the *TNF* locus, it is difficult to know which genes are actually responsible for the signal detected as *TNF* maps within the complex genetic region of *HLA*'s class III region. Early works established a putative association of AD with *HLA* region (19). Recently, a multicentric study (IGAP) related *HLA-DRB5-DRB1* region with AD (20), reinforcing the inflammatory pathway for the disease. This region has also been associated with other neurodegenerative diseases, such as multiple sclerosis (MS) (21), Parkinson's disease (PD) (22), and frontotemporal dementia (23). At this stage, we thought that our signal could be related

with the signal previously detected by IGAP in *HLA-DRB5-DRB1* region. However, LD between *TNF* region (rs2395488) and *HLA-DRB5-DRB1* (rs9271192) was almost null ($r^2 = 0.001$; D = 0.103). Thus, signals detected around *TNF* could track an independent locus unrelated to *HLA-DRB5-DRB1* region. In fact, there are several proximal markers (<500 kb), which have been associated with other neurodegenerative pathologies, such as age-related macular degeneration (31), (32), (33), (34), (35). However, none of these markers is in high LD with rs2395488 (data not shown). Thus, our study does not have enough positive results in support of neuroinflammatory hypothesis, despite that neuroinflammation in AD is still supported by large scientific evidence (2), (36), (37).

Whether inflammatory processes are a part of the cause or the response to the disease is still not known. However, more studies highlight the implication of genetic variations of the immune system as a cause of LOAD (38), (39). In that sense, the relevance of *TNF* locus in AD merits for deeper analysis of *HLA* region in order to identify the loci behind observed association. Hence, scientific and technological efforts are still needed to elucidate the way in which AD genetic variants may contribute to the disease's neurodegenerative mechanism.

REFERENCES

1 Akiyama H, Barger S, Barnum S, Bradt B Inflammation and Alzheimer's disease. Neurobiol Aging. 2000;21, 383–421.

2 Combs CK, Karlo JC, Kao SC, Landreth GE beta-Amyloid stimulation of microglia and monocytes results in TNFalpha-dependent expression of inducible nitric oxide synthase and neuronal apoptosis. 2001;J Neurosci. 21(4), 1179–88.

3 Zhao M, Cribbs DH, Anderson AJ, Cummings BJ, Su JH, Wasserman AJ, et al. The Induction of the TNF Death Domain Signaling Pathway in Alzheimer's Disease Brain. Neurochem Res. 2003;28(2), 307–318.

4 Lavrik I, Golks A, Krammer PH Death receptor signaling. J Cell Sci. 2005;118(2), 265–7.

5 Monte SM De, Sohn YK, Wands JR. Correlates of p53- and Fas(CD95)- mediated apoptosis in Alzheimer ' s disease. J Biol Sci. 1997;152(1), 73–83.

6 Engidawork E, Gulesserian T, Yoo BC, Cairns N, Lubec G. Alteration of caspases and apoptosis-related proteins in brains of patients with Alzheimer's disease. Biochem Biophys Res Commun. 2001; 281(1), 84–93.

7 Pugazhenthi S, Wang M, Pham S, Sze C-I, Eckman CB. Downregulation of CREB expression in Alzheimer's brain and in Aβ-treated rat hippocampal neurons. Mol Neurodegener. 2011;6, (60).

8 Antúnez C, Boada M, González-Pérez A, Gayán J, Ramírez-Lorca R, Marín J, et al. The membrane-spanning 4-domains, subfamily A (MS4A) gene cluster contains a common variant associated with Alzheimer's disease. Genome Med. 2011;3(5), 33.

9 Boada M, Antúnez C, Ramírez-Lorca R, DeStefano A, González-Pérez A, Gayán J, et al. ATP5H/KCTD2 locus is associated with Alzheimer's disease risk. Mol Psychiatry. 2014;19(6), 682–7.

10 Weiner M, Aisen P, Jack Jr C, Jaugust W, Trojanowski J, Shaw L, et al. The Alzheimer's disease neuroinmaging iniciative: Progress report and future plans. Alzheimer dement. 2010;6(3), 202–11.

11 Li H, Wetten S, Li L, St Jean PL, Upmanyu R, Surh L, et al. Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. Arch Neurol.2008;65(1), 45–53.

12 Reiman E, Webster J, Myers A, Hardy J, Dunckley T, Zismann V, et al. GAB2 alleles modify Alzheimer's risk in APOE ε4 carriers. Neuron. 2008;54(5), 713–720.

13 Wijsman EM, Pankratz ND, Choi Y, Rothstein JH, Faber KM, Cheng R, et al. Genome-wide association of familial late-onset Alzheimer's disease replicates BIN1 and CLU and nominates CUGBP2 in interaction with APOE. PLoS Genet. 2011;7(2), e1001308.

14 Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet. 2009;41(10), 1088–93.

15 Seshadri, S. Fitzpatrick, AL. Ikram, A. DeStefano, AL. Gudnason, V. Boada M, et al. Genome-wide Analysis of Genetic Loci Associeted with Alzheimer's Disease. JAMA. 2011;303(18), 1832–1840.

16 Hu X, Pickering E, Liu YC, Hall S, Fournier H, Katz E, et al. (2011) Meta-analysis for genome-wide association study identifies multiple variants at the BIN1 locus associated with late-onset Alzheimer's disease. PLoS One. 2011;6(2), e16616.

17 Naj AC, Jun G, Beecham GW, Wang L-S, Vardarajan BN, Buros J, et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nat Genet. 2011;43(5), 436–41.

18 Martinez-Mir A, González-Pérez A, Gayán J, Antúnez C, Marín J, Boada M, et al. Genetic study of neurexin and neuroligin genes in Alzheimer's disease. J Alzheimers Dis. 2013;35(2), 403–12.

19 Gayán J, Galan JJ, González-Pérez A, Sáez ME, Martínez-Larrad MT, Zabena C, et al. Genetic structure of the Spanish population. BMC Genomics. 2010;11, 326.

20 Maccioni RB, Rojo LE, Fernández J, Kuljis RO. The role of neuroimmunomodulation in Alzheimer's disease. Ann N Y Acad Sci. 2009;1153, 240–6.

21 Erten-Lyons D. The FAS gene, brain volume and disease progression in Alzheimer's disease. Alzheimer's & dementia. 2011;6, 118–124.

22 Gnjec A, D'Costa KJ, Laws SM, Hedley R, Balakrishnan K, Taddei K, et al. Association of alleles carried at TNFA -850 and BAT1 -22 with Alzheimer's disease. J Neuroinflammation. 2008; 5, 36.

23 Edgünlü TG, Ozge A, Yalın OÖ, Kul S, Erdal ME, et al. A Study of the Impact of Death Receptor 4 (DR4) Gene Polymorphisms in Alzheimer's Disease. Balkan Med J. 2008;30(3), 268–72.

24 Collins JS, Perry RT, Watson B, Harrell LE, Acton RT, Blacker D, et al. Association of a haplotype for tumor necrosis factor in siblings with late-onset Alzheimer disease: the NIMH Alzheimer Disease Genetics Initiative. Am J Med Genet. 2000;96(6), 823–30.

25 Di Bona D, Candore G, Franceschi C, Licastro F, Colonna-Romano G, Cammà C, et al. Systematic review by meta-analyses on the possible role of TNF-alpha polymorphisms in association with Alzheimer's disease. Brain Res Rev. 2009;61(2), 60–8.

26 Harris JM, Cumming AM, Craddock N, St Clair D, Lendon CL. Human leucocyte antigen-A2 increases risk of Alzheimer's disease but does not affect age of onset in a Scottish population. Neurosci Lett. 2000;294(1), 37–40.

27 Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet. 2013;45(12), 1452–8.

28 Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos N, Moutsianas L, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature 2011;476(7359), 214–9.

29 Hamza TH, Zabetian CP, Tenesa A, Laederach A, Montimurro J, Yearout D, et al. Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. Nat Genet. 2010;42(9), 781–5.

30 Ferrari R, Hernandez DG, Nalls M, Rohrer JD, Ramasamy A, Kwok JBJ, et al. Frontotemporal dementia and its subtypes: a genome-wide association study. Lancet Neurol. 2014;13(7), 686–99.

31 Sobrin L, Ripke S, Yu Y, Fagerness J, Bhangale TR, Tan PL, et al. Heritability and genome-wide association study to assess genetic differences between advanced age-related macular degeneration subtypes. Ophthalmology. 2012;119(9), 1874–85.

32 Chen W, Stambolian D, Edwards AO, Branham KE, Othman M, et al. Genetic variants near TIMP3 and high-density lipoprotein-associated loci influence susceptibility to age-related macular degeneration. Proc Natl Acad Sci. U. S. A. 2010;107(16), 7401–6.

33 Cipriani V, Leung H-T, Plagnol V, Bunce C, Khan JC, Shahid H, et al. Genome-wide association study of age-related macular degeneration identifies associated variants in the TNXB-FKBPL-NOTCH4 region of chromosome 6p21.3. Hum Mol Genet. 2012;21(18), 4138–50.

34 Fritsche LG, Chen W, Schu M, Yaspan BL, Yu Y, Thorleifsson G, et al. Seven new loci associated with age-related macular degeneration. Nat Genet. 2013;45(4), 433–9, 439e1–2.

35 Naj AC, Scott WK, Courtenay MD, Cade WH, Schwartz SG, Kovach JL, et al. Genetic factors in nonsmokers with age-related macular degeneration revealed through genome-wide gene-environment interaction analysis. Ann Hum Genet. 2013;77(3), 215–31.

36 Morimoto K, Horio J, Satoh H, Sue L, Beach T, Arita S, et al. Expression profiles of cytokines in the brains of Alzheimer's disease (AD) patients compared to the brains of non-demented patients with and without increasing AD pathology. J Alzheimers. Dis. 2011;25(1), 59–76.

37 Patel NS, Paris D, Mathura V, Quadros AN, Crawford FC, Mullan MJ. Inflammatory cytokine levels correlate with amyloid load in transgenic mouse models of Alzheimer's disease. J Neuroinflammation. 2005;2(1), 9.

38 Jones L, Holmans PA, Hamshere ML, Harold D, Moskvina V, Ivanov D, et al. Genetic evidence implicates the immune system and cholesterol metabolism in the aetiology of Alzheimer's disease. PLoS One. 2010;5(11), e13950.

39 Lambert J-C, Grenier-Boley B, Chouraki V, Heath S, Zelenika D, Fievet N, et al. Implication of the immune system in Alzheimer's disease: evidence from genome-wide pathway analysis. J Alzheimers Dis. 2010;20(4), 1107–18.

ACKNOWLEDGMENTS

We would like to thank the patients and control subjects who participated in this project. This work was funded by the Spanish Government's "Ministerio de Sanidad y Consumo" (CIBERNED, CB06/05/1104), "Ministerio de Economía y Competitividad" (SAF2010-19953), and "Generalitat de Catalunya" (*Suport als Grups de Recerca Consolidats* 2009SGR346) to JXC. PC was awarded a "Beatriu de Pinos" postdoctoral grant from the "Generalitat de Catalunya" cofinanced by the FP7-People-COFUND Program. We are indebted to Trinitat Port-Carbó and her family for their support of the Fundació ACE research programs. Fundació ACE collaborates with the Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED, Spain) and is one of the participating centers of the Dementia Genetics Spanish Consortium (DEGESCO). CIBERNED is an Instituto de Salud Carlos III ISCIII Project. Agustín Ruiz is supported by grant PI13/02434 (Acción Estratégica en Salud, Instituto de Salud Carlos III (ISCIII), Ministerio de Economía y Competitividad, Spain), and Obra Social "La Caixa" (Barcelona, Spain). **This study has been performed as part of the Biochemistry, Molecular Biology, and Biomedicine doctoral program of S. Moreno-Grau at Universitat Autònoma deBarcelona (Barcelona, Spain).**

SUPPLEMENTARY TABLES

The Supplementary Table 1 can be accessed scanning the following code.

Supplementary Table 1. Results for the 1,501 candidate markers tested in the meta-analysis of five GWAS and aggregated data.

SUPPLEMENTARY TABLE 2. Number of SNPs described within candidate gene and number of significant SNPs after meta-analysis.

Gene	Chromosome	Number of SNPs *(BP position range)	Number of significant SNPs *(Top SNP; p value)
FASLG	1	83 *(170794135–170991017)	5 *(rs730278; p = 0.00331)
TNFRSF25	1	57 *(6300329–6494762)	0
CFLAR	2	66 *(201612546–201801640)	0
*BIRC*_Cluster	3	103 *(101600386–101795189)	2 *(rs10936200; p = 0.03337)
FAIM1	3	84 *(139701399–139894326)	0
TNF	6	304 *(31554067–31613866)	44 *(rs2905726; p = 0.00008)
TNFIAP3	6	102 *(138108380–138299661)	0
TNFRSF21	6	221 *(47200215–47446050)	9 *(rs9463301; p = 0.00495)
BAG4	8	29 *(38096889–38252950)	1 *(rs2517388; p = 0.00883)
TNFRSF10A	8	177 *(23097027–23293928)	8 *(rs11780345; p = 0.00365)
TNFRSF1A	12	81 *(6202571–6398513)	1 *(rs3764875; p = 0.02395)
FAIM2	12	111 *(48446397–48640704)	6 rs11169172; p = 0.02362)
TNRFSF12A	16	83 *(2903262–3095007)	2 *(rs4786370; p = 0.01725)

BP: Base pair; SNPs: Single nucleotide polymorphisms.

3.3. Hypothesis free approach for uncovering new AD genes

3.3.1. Publication IV

Genome-wide association analysis of dementia and its clinical endophenotypes reveal novel loci associated with Alzheimer's disease and three causality networks: the GR@ACE project

Sonia Moreno-Grau[a,b], Itziar de Rojas[a], Isabel Hernández[a,b], Inés Quintela[c], Laura Montrreal[a], Montserrat Alegret[a,b], Begoña Hernández-Olasagarre[a], Laura Madrid[d], Antonio González-Perez[d], Olalla Maroña[c], Maitée Rosende-Roca[a], Ana Mauleón[a], Liliana Vargas[a], Asunción Lafuente[a], Carla Abdelnour[a,b], Octavio Rodríguez-Gómez[a,b], Silvia Gil[a], Miguel Ángel Santos-Santos[a], Ana Espinosa[a,b], Gemma Ortega[a,b], Ángela Sanabria[a,b], Alba Pérez-Cordón[a], Pilar Cañabate[a,b], Mariola Moreno[a], Silvia Preckler[a], Susana Ruiz[a,b], Nuria Aguilera[a], Juan Antonio Pineda[e], Juan Macías[e], Emilio Alarcón-Martín[a,f], Oscar Sotolongo-Grau[a], GR@ACE/DEGESCO consortium*, Alzheimer's Disease Neuroimaging Initiative*, Marta Marquié[a], Gemma Monté-Rubio[a], Sergi Valero[a,b], Alba Benaque[a], Jordi Clarimón[b,g], Maria Jesus Bullido[b,h,i], Guillermo García-Ribas[j], Pau Pástor[k], Pascual Sánchez-Juan[b,l], Victoria Álvarez[m,n], Gerard Piñol-Ripoll[b,o], Jose Maria García-Alberca[p], José Luis Royo[f], Emilio Franco[q], Pablo Mir[b,r], Miguel Calero[b,s,t], Miguel Medina[b,s], Alberto Rábano[b,s,u], Jesús Ávila[b,w], Carmen Antúnez[x], Luis Miguel Real[e,f], Adelina Orellana[a], Ángel Carracedo[c,y], María Eugenia Sáez[d], Lluís Tárraga[a,b], Mercé Boada[a,b] and Agustín Ruiz[a,b]**.

a. Research Center and Memory clinic Fundació ACE. Institut Català de Neurociències Aplicades. Universitat Internacional de Catalunya. Barcelona Spain. b. CIBERNED Network Center for Biomedical Research in Neurodegenerative Diseases National Institute of Health Carlos III Spain. c. Grupo de Medicina Xenómica Centro Nacional de Genotipado (CEGEN-PRB3-ISCIII). Universidade de Santiago de Compostela Santiago de Compostela Spain. d. CAEBI. Centro Andaluz de Estudios Bioinformáticos Sevilla Spain. e. Unidad Clínica de Enfermedades Infecciosas y Microbiología. Hospital Universitario de Valme Sevilla Spain. f. Dep. of Surgery Biochemistry and Molecular Biology School of Medicine. University of Málaga. Málaga Spain g. Memory Unit Neurology Department and Sant Pau Biomedical Research Institute Hospital de la Santa Creu i Sant Pau Universitat Autònoma de Barcelona Barcelona Spai. h. Centro de Biologia Molecular Severo Ochoa (C.S.I.C.-U.A.M.) Universidad Autonoma de Madrid Madrid Spain. i. Instituto de Investigacion Sanitaria Hospital la Paz (IdIPaz) Madrid Spain. j. Hospital Universitario Ramón y Cajal Madrid Spain. k. Fundació per la Recerca Biomèdica i Social Mútua Terrassa and Memory Disorders Unit Department of Neurology Hospital Universitari Mutua de Terrassa University of Barcelona School of Medicine Terrassa Barcelona Spain. l. Neurology Service 'Marqués de Valdecilla' University Hospital (University of Cantabria and IDIVAL) Santander Spain. m. Laboratorio de Genética Hospital Universitario Central de Asturias Oviedo. n. Instituto de Investigación Biosanitaria del Principado de Asturias (ISPA). o. Unitat Trastorns Cognitius Hospital Universitari Santa Maria de Lleida Institut de Recerca Biomèdica de Lleida (IRBLLeida) Lleida España. p. Alzheimer Research Center & Memory Clinic. Andalusian Institute for Neuroscience. Málaga Spain. q. Unidad de Demencias Servicio de Neurología y Neurofisiología. Instituto de Biomedicina de Sevilla (IBiS) Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla Seville Spain. r. Unidad de Trastornos del Movimiento Servicio de Neurología y Neurofisiología. Instituto de Biomedicina de Sevilla (IBiS) Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla Seville Spain. s. CIEN Foundation Queen Sofia Foundation Alzheimer Center Madrid Spain. t. Instituto de Salud Carlos III (ISCIII) Madrid Spain . u. BT-CIEN. w. Department of Molecular Neuropathology Centro de Biología Molecular Severo Ochoa (CBMSO) Consejo Superior de Investigaciones Científicas (CSIC)/Universidad Autónoma de Madrid (UAM). x. Unidad de Demencias. Hospital Clínico Universitario Virgen de la Arrixaca. y. Fundación Pública Galega de Medicina Xenómica- CIBERER-IDIS Santiago de Compostela Spain.

****Corresponding autor.**

A**BSTRACT**

Introduction: Large variability between Alzheimer's disease (AD) cases might impact genetic discoveries and complicate dissection of underlying biological pathways.

Methods: GR@ACE is a genome-wide study of dementia and its clinical endophenotypes, defined according to AD's clinical certainty and vascular burden. We assessed the impact of known AD loci across endophenotypes to generate loci categories. We incorporated gene co-expression data and conducted pathway analysis per category. Finally, to evaluate the effect of heterogeneity in genetic studies, GR@ACE series were meta-analyzed with additional GWAS datasets.

Results: We classified known AD loci into three categories, which might reflect the disease clinical heterogeneity. Vascular processes were only detected as a causal mechanism in probable AD. Meta-analysis revealed the *ANKRD31*-rs4704171 and *NDUFAF6*-rs10098778, and confirmed *SCIMP*-rs7225151 and *CD33*-rs3865444.

Discussion: The regulation of vasculature is a prominent causal component of probable AD. GR@ACE meta-analysis revealed novel AD genetic signals, strongly driven by the presence of clinical heterogeneity in the AD series.

Keywords: Alzheimer's disease, vascular pathology, cerebral amyloid angiopathy, GWAS, biological pathway

Alzheimer's and Dementia. 2019 Aug 13. pii: S1552-5260(19)35117-9

doi: 10.1016/j.jalz.2019.06.4950

INTRODUCTION

Dementia is an age-related clinical syndrome that devastates cognitive abilities and interferes in elderly people's daily activities. Although its incidence is decreasing due to improvements to public health systems and control of cardiovascular risk factors (1), its prevalence is steadily increasing due to rising life expectancy of human populations (2).

Dementia is linked to many underlying pathologies, with Alzheimer's disease (AD) as the most common condition. Clinical AD is a heterogeneous syndrome. AD brain autopsies have shown that roughly 80% of clinical AD patients present brain vascular pathology (3) in addition to the common neuropathological AD hallmarks: amyloidosis, neurofibrillary tangles (NFTs), and cerebral amyloid angiopathy (CAA) (4). In fact, brain vascular pathology has been shown to be an important risk factor for AD that accelerates cognitive decline (5) and lowers the threshold for clinical diagnosis of AD (6). In that context, it has been suggested that dementia is represented by a gradient of neurodegenerative and vascular components (7); from pure AD forms, with a strong neurodegenerative component, to pure vascular dementia (VaD) cases and in-between mixed pathologies, representing the co-existence of both, neurodegenerative and vascular components (7). Despite that, whether there are differential biological routes operating under different levels of vascular burden in clinical AD patients remains mostly unknown.

In the search for the etiology of AD, genetic factors play a pivotal role. Two forms of the disease can be differentiated according to individual genetic background. The Mendelian form is an uncommon disorder that mainly affects families with early-onset AD (EOAD) (<65 years old), whereas the polygenic form is a complex disorder mainly appearing in sporadic cases with late-onset AD (LOAD) (>65 years old). Highly penetrant mutations detected in EOAD families have been pinpointed to three genes: *APP* (8), *PSEN1* (9), and *PSEN2* (10); leading to the establishment of the amyloid hypothesis as a potential causal mechanism for the disease (11).

LOAD heritability falls in the range of 13%–80% (12), (13). Although *APOE* ε4 was the first to be discovered and still remains the strongest genetic risk factor for AD (14), almost 40 additional genetic variants have been identified (15), (16), (17) using genome-wide association studies (GWAS) and large sequencing projects. Among the biological pathways underlying genetic hits, the roles of the immune system, cholesterol, APP, and Tau metabolism have been highlighted (15). Despite these, current genetic findings account for 31% of LOAD heritability (18), and the biological picture of AD is still poorly understood. Several reasons can explain this. Among them, the presence of

clinical and neuropathological heterogeneity between AD cases in genetic studies, as recently demonstrated (12), might compromise the power to detect genuine genetic associations and decrease the estimates of risk attributed to genetic variation.

The neuropathological variability in clinical AD cases comprises a wide spectrum, from those with concomitant vascular brain disease to those with a pure AD phenotype, as previously proposed by Viswanathan et al. (7). This large heterogeneity might hamper the identification of functional categories of genes underlying differential biological routes to dementia and might impact AD genetic studies. To gain insight on the causality networks behind AD clinical subgroups and to explore their impact in large GWAS, we conducted the Genome Research at Fundacio ACE (GR@ACE) study. This is a GWAS of dementia and its clinical endophenotypes defined according to AD's clinical certainty and the burden of vascular co-morbidity. GR@ACE is a unique genomic resource comprising the largest number of dementia cases diagnosed in a single memory clinic to date. First, we determined whether we could identify categories of known LOAD genes linked to clinical subgroups of AD cases. Next, we explored whether these categories suggested different biological routes. Finally, to assess the impact of these clinical subgroups of AD cases in GWAS findings, we meta-analyzed the GR@ACE data with independent GWAS series.

METHODS

Subjects

Gr@ACE cohort and phenotype definitions

The GR@ACE study comprises 4,120 AD cases and 3,289 control individuals (Table 1). Cases were recruited from Fundació ACE, Institut Català de Neurociències Aplicades (Catalonia, Spain). Diagnoses were established by a multidisciplinary working-group, including neurologists, neuropsychologists, and social workers, according to the DSM-IV criteria for dementia and to the National Institute on Aging and Alzheimer's Association's (NIA-AA) 2011 guidelines for defining AD (19). See Supplementary Material. In the present study, we considered AD cases as dementia individuals diagnosed with probable or possible AD at any moment of their clinical course.

We took advantage of this wide clinical definition to refine AD cases. Considering the dementia spectrum proposed by Viswanathan et al. (7), we classified Gr@ACE AD patients according to the degree of clinical certainty for AD phenotype and the presence of vascular co-morbidity, from "pure" clinical AD cases to mixed and vascular enriched cases. This approach was feasible due to Fundació ACE's endorsement of a primary and a secondary etiologic diagnosis as well as

TABLE 1. GR@ACE demographic characteristic and endophenotype definitions

Phenotype	Primary Diagnostic	Secondary Diagnostic	N	Mean Age ± (SD)	Women %	*APOE* ε4 %
Controls	--	--	3289	54.3 ± 14.4	48.9	21.4
VaD^{++}	VaD	Pss AD	373	80.1 ± 5.5	54.9	25.0
VaD^{+}	VaD/ Pss AD	VaD/ Pss AD	1168	80.4 ± 6.3	65.0	32.8
AD	Pr/Pss AD at any time in medical history		4120	79.0 ± 7.5	69.6	40.1
AD^{+}	Pr/Pss AD	Pr/Pss AD	3797	79.2 ± 7.5	70.6	41.2
AD^{++}	Pr AD	Pr/Pss AD	2611	78.8 ± 7.9	72.8	44.6
AD^{+++}	Pr AD	Pr AD	1854	79.0 ± 8.0	74.6	47.0

VaD = Vascular Dementia; Pss AD = Possible AD; Pr AD = Probable AD.

routine follow-up evaluations (20) (Supplementary Methods). Using the entire clinical chart of each subject, we differentiated five clinical subgroups of patients representing the GR@ACE endophenotypes: 1) the AD^{+++} endophenotype comprises individuals with a last clinical diagnosis of probable AD in both primary and secondary diagnoses (n = 1,854); 2) the AD^{++} includes individuals diagnosed with probable AD either in the primary or the secondary diagnosis (n = 2,611); 3) the AD^{+} encompasses patients diagnosed with probable or possible AD either in the primary diagnosis or in the secondary diagnosis (n = 3,797); 4) the VaD^{+} includes patients diagnosed with vascular dementia (VaD) or possible AD in the primary diagnosis (n = 1,168); and 5) the VaD^{++} comprises patients diagnosed with probable or possible vascular dementia in the primary diagnosis (n = 373) (Table 1) (Supplementary Figure 1). VaD patients were defined according to NINDS-AIREN criteria (21).

Control individuals were recruited from three centers: Fundació ACE (Barcelona, Spain), Valme University Hospital (Seville, Spain), and the Spanish National DNA Bank Carlos III (University of Salamanca, Spain) (www.bancoadn.org). Written informed consent was obtained from all participants. The Ethics and Scientific Committees have approved this research protocol (Acta 25/2016. Ethics Committee. H. Clinic i Provincial, Barcelona, Spain).

Replication sample

With the objective to successfully replicate novel GWAS findings, we used an independent Spanish sample of 1,943 AD cases [(Mean age = 79.2; SD = 7.6); 66.3% women] and 3,016 controls [(Mean age = 52.8; SD = 15.2); 46% women] presenting similar characteristics to GR@ACE cohort

and with available genetic data. All AD cases were examined at a single-site, Fundació ACE, Institut Català de Neurociències Aplicades (Catalonia, Spain), and were assessed by applying the same criteria previously explained. The sample composition of dementia cases was comprised of 30.1% of AD^{+++} cases (n = 584), 53.9% of AD^{++} (n = 1,048), 91.9% of AD^{+} (n = 1,783), 23.0% of VaD^{+} (n = 447), and 4.6% of VaD^{++} (n = 89). Control individuals were selected from the Spanish population available at three centers, previously described.

GWAS genotyping, quality control, imputation, and statistical analysis

Participants were genotyped using the Axiom 815K Spanish Biobank Array (Thermo Fisher). Genotyping was performed in the Spanish National Center for Genotyping (CeGEN, Santiago de Compostela, Spain) (Supplementary Methods).

We removed samples with genotype call rates <97%, excess heterozygosity, duplicates, samples genetically related to other individuals in the cohort, or sample mix-up (PIHAT >0.1875). If a sex discrepancy was detected, the sample was removed unless the discrepancy was safely resolved. To detect population outliers of non-European ancestry (>6SD from European population mean), principal component analysis was conducted using SMARTPCA from EIGENSOFT 6.1.4 (Figure 1) (Supplementary Methods).

We excluded variants with a call rate <95% or that grossly deviated from Hardy-Weinberg equilibrium in controls (P-value $\leq 1x10^{-6}$); we also excluded markers with a different missing rate between case and control (P-value $< 5\times10^{-4}$ for the difference) or a minor allele frequency (MAF) <0.01. Imputation was carried out using the Haplotype Reference Consortium panel in the Michigan Imputation Server (https://imputationserver.sph.umich.edu). Only common markers (MAF >0.01) with a high imputation quality ($R^2 > 0.30$) were selected for downstream analyses.

The GWAS was performed for GR@ACE as a whole and for each endophenotype, [(N_{AD+++} = 5,143); (N_{AD++} = 5,900); (N_{AD+} = 7,086); ($N_{AD\ dementia}$ = 7,409); (N_{VaD+} = 4,487); and (N_{VaD++} = 3,662)]. The GWAS was conducted for genotype dosages using an additive genetic model with PLINK 1.9. A model including the top four PCs as covariates was used for the discovery stage because it exhibited the lowest inflation and optimal power compared to alternative tested models (Supplementary Figure 2). Further description is provided in Supplementary Methods. Power analysis was performed using QUANTO software v1.2.4 (22) to model the impact on statistical power at different MAFs and effect sizes in available case-control cohorts. Results were depicted using the ggplot2 package in R. Analyses

were performed for a GWAS experiment that would meet criteria for genome-wide significance ($p < 5 \times 10^{-8}$) and for a replication experiment that would meet $p < 0.05$ (Supplementary Figure 3). Calculations were performed considering a disease prevalence of 1.73%, according to registers in the Spanish population.

GR@ACE GWAS data has been deposited into the European Genome-phenome Archive (EGA) (https://ega-archive.org), which is hosted by the EBI and the CRG, under accession number EGAS00001003424.

Genetic exploration of GR@ACE clinical endophenotypes and enrichment analysis

With the objective to explore whether different biological routes operate under different levels of vascular burden in clinical AD patients, first, we classified GR@ACE dementia cases to cover the dementia spectrum previously proposed by Viswanathan et al. (7). Second, we extracted the effect (odds ratio, OR) for known LOAD genetic variants (MAF >1%). See included variants in Supplementary Table 1. Then, we explored whether previously identified LOAD variants were more strongly associated with a specific subgroup of AD patients. We quantified the strength of the association for each variant across endophenotypes, named henceforth global effect change. It was calculated as the absolute difference between variant OR for extreme endophenotypes (VaD^{++} vs AD^{+++}). According to the global effect change and direction of the enrichment, i.e., from VaD^{++} to AD^{+++} or from AD^{+++} to VaD^{++}, we classified LOAD genetic variants into three categories. Thus, category A includes variants with an increase in the association effect from VaD^{++} to AD^{+++} endophenotypes and a global effect change >0.05; and category B variants with an increase in the association effect from AD^{+++} to VaD^{++} and a global effect change >0.05. Category C comprises variants not fulfilling criteria for categories A or B (Supplementary Table 1). Finally, we assessed the biological pathways underlying each category. We incorporated data from gene co-expression for each specific gene category using the GeneFriends tool (http://genefriends.org/) and performed pathway analysis of top co-expressed genes using the overrepresentation enrichment method in WebGestalt (http://www.webgestalt.org/option.php). See Supplementary Methods. With the objective to further explore potential additional trends in category C, we performed a specific sub-analysis in this category, widely described in Supplementary Methods. To validate previous gene classification, which strongly determines the pathway analysis results, we conducted a stringent sub-analysis. See Supplementary Methods and Supplementary Figure 4.

Meta-analysis: datasets and association analysis

To explore the impact of the different clinical endophenotypes in GWAS findings, we combined the whole GR@ACE GWAS dataset, which represents a dementia set of samples and its endophenotypes with (1) genotype-level data from nine additional GWAS series (N = 13,826), available through dbGaP (https://www.ncbi.nlm.nih.gov/gap) that we processed by applying identical quality control and imputation procedures to those described for the GR@ACE cohort (Supplementary Table 2); and (2) aggregated summary statistics available from the International Genomics Alzheimer's Project (IGAP) (http://web.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php) (23), including IGAP stages I (N final = 61,571) and IGAP I & II (N final = 81,455) (Supplementary Methods). Meta-analyses were conducted using the inverse variant method in METAL software (https://genome.sph.umich.edu/wiki/METAL). The LD Score calculations, clumping, and conditional analysis are described in Supplementary Methods.

Replication of genome-wide significant findings

We then explored genome-wide significant (GWS) signals in an independent cohort of Spanish ancestry (N = 4,959). We extracted variants of interest from GWAS data, which was genotyped and processed applying similar methods to those explained for the GR@ACE study. Finally, meta-analysis including the discovery stage, named Stage I, and the replica dataset, Stage II, were performed as previously described. Results were interpreted according to American Statistical Association (ASA) guidelines (24), (25).

Biological interpretation of meta-GWAS signals

Gene expression quantitative trait locus (eQTL) analysis was conducted to link meta-GWAS top signals to genes. Markers with moderate-to-high LD ($r^2 \geq 0.6$) with the novel lead markers were identified using LDlink (26) for European population and were included in this analysis. We used brain (n = 11) and whole blood (n = 1) tissues from the GTEx v7 repository (https://www.gtexportal.org/home) for mapping cis-eQTLs (Supplementary Table 3). As an extension of GTEx tissue eQTL mapping, we explored brain eQTLs for GWS genomic regions using additional databases available via FUMA (27). We also performed functional annotation for GWS markers, chromatin interaction and gene-based analysis using similar criteria to those previously described by Jansen et al. (17) (See Supplementary Methods).

RESULTS

GR@ACE genome-wide association study

After quality control and imputation, the GR@ACE study encompassed 7,409 unrelated individuals from the Spanish population and 7.7 million variants (λ_{GC} = 1.03). The *APOE*-rs429358 marker was the only one to have a GWS association [OR = 2.27 (2.06–2.50); p = $1.25x10^{-62}$] (Figure 1). Four additional LOAD variants displayed statistically significant evidence of replication (*BIN1*-rs6733839, *MAPT*-rs2732703, *MS4A2*-rs983392, and *PICALM*-rs10792832) and nine additional markers presented a consistent direction for the effect (Supplementary Table 1). *MAPT* marker association remains significant in *APOE* ε4 non-carriers, but the effect size was stable in both strata (Supplementary Table 4). GWAS of clinical endophenotypes showed that *CNTNAP2*-rs117834366 was associated with the VaD^{++} endophenotype (Table 2). This marker is in complete linkage equilibrium with *CNTNAP2*-rs114360492 (r^2 = 0), previously reported in GWAS of AD by proxy (17). See results in Supplementary Results (Supplementary Figures 5 and 6; Supplementary Table 5).

Genetic exploration of GR@ACE clinical endophenotypes and enrichment analysis

To explore whether clinical AD subgroups, representing GR@ACE endophenotypes, reflected variations in the underlying biological pathways driving dementia, we classified LOAD genetic variants into three categories. Category A comprised variants strongly related to the purest form of clinical AD (i.e., subjects with probable AD in primary and secondary diagnoses). The most prominent locus of this category was *APOE*-rs429358 [AD^{+++} OR = 2.92 (2.60–3.27), p-value = 9.26×10^{-75}; VaD^{++} OR (95%) = 1.27 (1.02–1.59), p-value = 0.04]. Other loci included in category A were *CR1*, *BIN1*, *MEF2C*, *MS4A2*, *PICALM*, *MAPT,* and *CD33*. In contrast, category B comprised variants with the strongest effect observed in subjects with AD mixed with vascular disease (*SORL1*, *ADAM10*, *CASS4*, *ATP5H*, and *ACE*) (Supplementary Table 1). Further description is provided in Supplementary Results. Category C comprised a group of variants with effects in all clinical endophenotypes (Figure 2). Sub-analysis for category C is shown in Supplementary Results and Supplementary Table 6.

Next, we explored biological pathways for each gene category. Note that the regulation of vasculature development and blood vessel morphogenesis were only detected for genes in category A, which is more closely related to pure AD ($p = 2.03 \times 10^{-7}$, $p = 1.90 \times 10^{-6}$, respectively) (Table 3). Additional categories indicated immune system pathways (category B, $p = 2.07 \times 10^{-7}$; category C, $p = 5.77 \times 10^{-15}$) (Table 3).

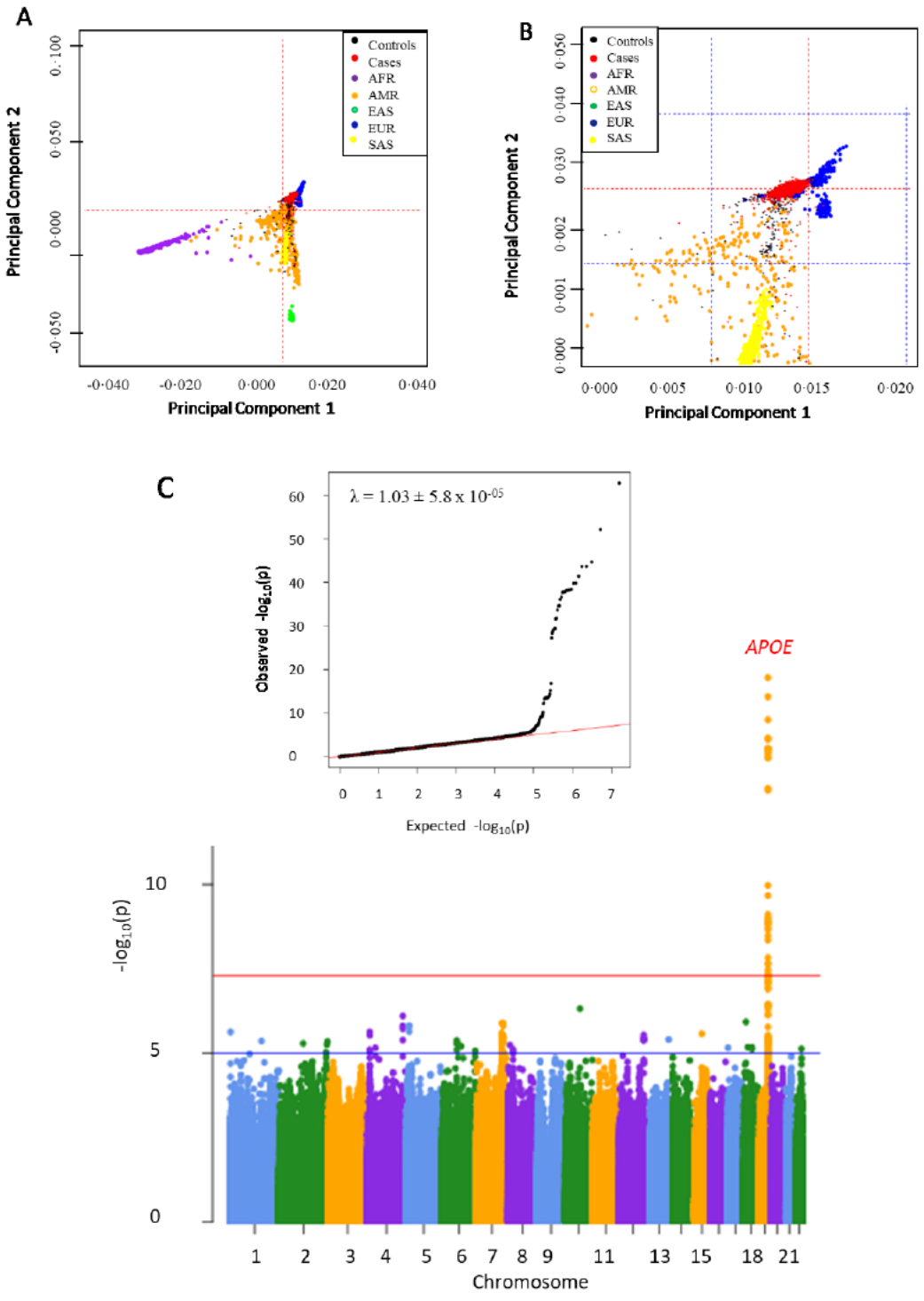


FIGURE 1. Results of genome-wide association analysis for the GR@ACE data set (N = 7409). (A) Principal component analysis; (B) principal component analysis centered in European population; (C) QQplot for the discovery model, adjusted for first four PCs; (D) Manhattan plot for genome-wide results.

AFR, African; AMR, Admixed American; EAS, East Asian; EUR, European; SAS, South Asian

able 2. Association results for lead single-nucleotide polymorphisms reaching genome-wide significance.

Marker	Near Locus	Position CHR:BP	A1/ A2	MAF	Discovery Stage	Stage I OR (CI95%)	*P*	Stage II (Replica) OR (CI95%)	*P*	Stage I+ OR (CI95%)
117834366 [a]	*CNTNAP2*	7:147634891	A/G	0·011	GR@ACE VaD^{++}	6.03 (3.22 – 11.2)	1.91 x 10^{-8}	1.09 (0.66 – 1.78)	0.73	2.08 (1.42 – 3.07)
s4704171 [b]	*ANKRD31*	5:74368254	C/T	0·123	GR@ACE + dbGaP	1.19 (1.12 – 1.27)	2.78 x 10^{-8}	1.10 (0.98 – 1.11)	0.09	1.18 (1.11 – 1.24)
10098778 [b]	*TP53INP1/ NDUFAF6*	8:95992020	C/T	0·470	GR@ACE + IGAP I&II	0.94 (0.91 – 0.96)	2.54 x 10^{-8}	0.96 0.89 – 1.05)	0.40	0.94 (0.92 – 0.96)
s7225151 [b]	*SCIMP*	17:5137047	A/G	0·126	GR@ACE + IGAP I&II	1.10 (1.06 – 1.14)	8.98 x 10^{-8}	1.14 (1.00 – 1.30)	0.04	1.10 (1.06 – 1.14)
s7225151 [b]	*SCIMP*	17:5137047	A/G	0·126	GR@ACE AD^{+++} + IGAP I&II	1.11 (1.07 – 1.15)	1.12 x 10^{-8}	1.07 (0.88 – 1.30)	0.49	1.11 (1.15 – 1.07)

e threshold for genome-wide significance was 5 x 10^{-8}. [a]Imputed Variant; rs117834366 *r2*=0.68; [b]Genotyped variant. Position = GRCh37/hg19 coordinate omosome:base pair; A1 = Minor allele; A2 = Mayor Allele; MAF = Minor allele frequency obtained from the GR@ACE study (N = 7,409). Stage I corresponds to th ge; Stage II corresponds to the replica; Stage I+II corresponds to the meta-analysis of the discovery and the replica. Stage I+II for rs117834366 corresponds to the m GR@ACE VaD^{++} and the entire replica (N=4,959).

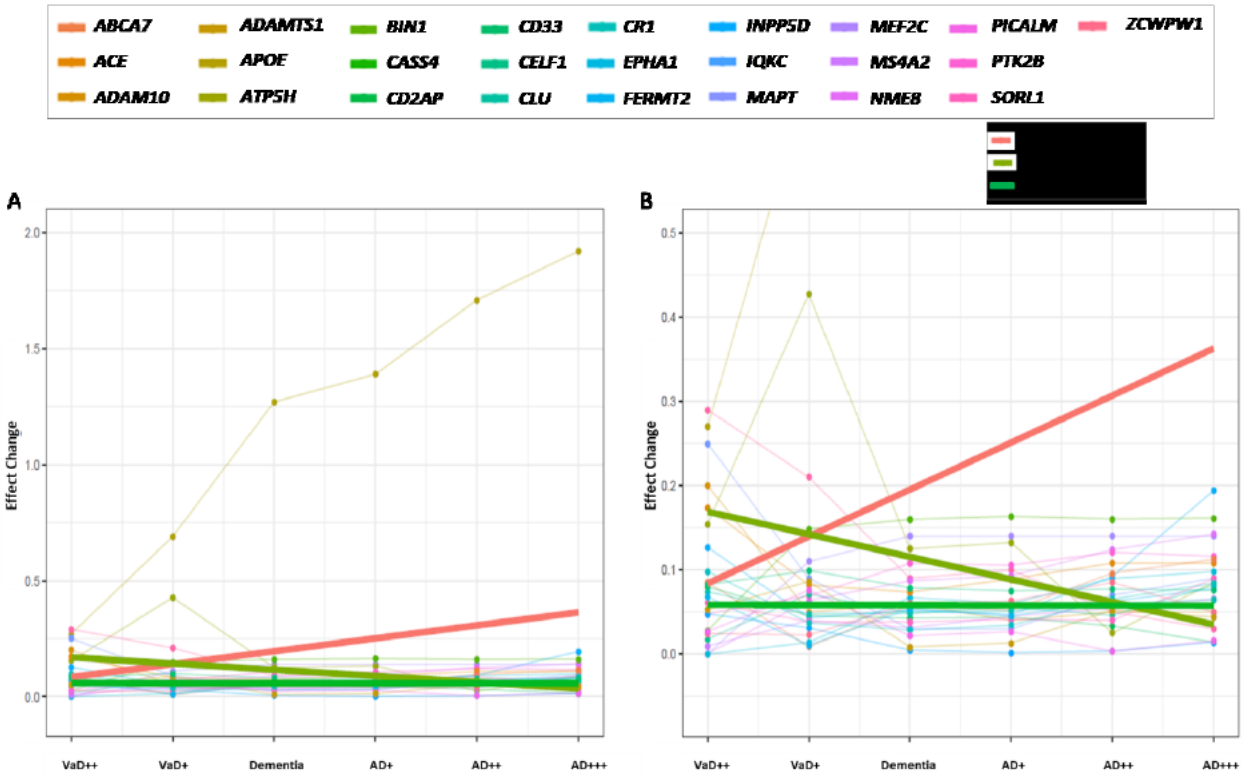


FIGURE 2. (A) Enrichment trend per genetic marker and gene category across GR@ACE endophenotypes. (B) Graph centered in effect change range 0–0.5.
Effect change per endophenotype = Variant Odds ratio–variant null effect; Enrichment trend per category was obtained applying a linear regression using ggplot2 in R.AD, Alzheimer's disease; VaD, vascular dementia.

Finally, with the aim of validating previous results, we conducted a sub-analysis by classifying LOAD genetic variants with more stringent classification criteria (Supplementary Methods). Again, *APOE*, *CR1*, *MEF2C*, *MS4A2,* and *PICALM* loci were found in category A; *SORL1* and *CASS4* were in category B; and additional AD loci were in category C (Supplementary Figure 4). Regulation of vasculature development was exclusively identified as the top pathway in category A ($p = 2.14 \times 10^{-7}$) when we restricted the analysis to include those loci co-expressing with at least 4 LOAD genes (Supplementary Table 7). Further replication in cohorts with available neuropathological data would be recommended.

Meta-analysis of GR@ACE study with other datasets

To assess the impact of sample composition in AD GWAS, and look for new AD loci, we first combined the GR@ACE dataset with nine additional genomic databases that had genotypic level data available. Subtle genomic inflation was detected, mainly explained by polygenicity ($\lambda_{GC} = 1.10$; LD score intercept = 1.04). Five regions were associated with LOAD (Figure 3); of these four,

TABLE 3. Top ten biological pathways per gene category

Gene Ontology Pathway	Top 10 co-regulated pathways for Category A	P-value
GO:1901342	regulation of vasculature development	2.03×10^{-7}
GO:0060326	cell chemotaxis	2.59×10^{-7}
GO:0048771	tissue remodeling	6.77×10^{-7}
GO:0050865	regulation of cell activation	1.14×10^{-6}
GO:0007159	leukocyte cell-cell adhesion	1.21×10^{-6}
GO:0048514	blood vessel morphogenesis	1.90×10^{-6}
GO:0003012	muscle system process	2.54×10^{-6}
GO:0002764	immune response-regulating signaling pathway	3.48×10^{-6}
GO:0032103	positive regulation of response to external stimulus	3.91×10^{-6}
GO:0010959	regulation of metal ion transport	4.36×10^{-6}
Gene Ontology Pathway	**Top 10 co-regulated pathways for Category B**	**P-value**
GO:0009620	response to fungus	2.02×10^{-7}
GO:0050886	endocrine process	3.58×10^{-7}
GO:0002443	leukocyte mediated immunity	$5.47x\ 10^{-7}$
GO:0050865	regulation of cell activation	1.52×10^{-5}
GO:0031349	positive regulation of defense response	8.42×10^{-5}
GO:0032103	positive regulation of response to external stimulus	1.00×10^{-4}
GO:0002250	adaptive immune response	1.30×10^{-4}
GO:0098542	defense response to other organism	2.00×10^{-4}
GO:1901568	fatty acid derivative metabolic process	2.24×10^{-4}
GO:0050900	leukocyte migration	2.57×10^{-4}
Gene Ontology Pathway	**Top 10 co-regulated pathways for Category C**	**P-value**
GO:0007159	leukocyte cell-cell adhesion	5.77×10^{-15}
GO:0050865	regulation of cell activation	4.37×10^{-14}
GO:0002764	immune response-regulating signaling pathway	1.33×10^{-12}
GO:0002253	activation of immune response	3.96×10^{-12}
GO:0002443	leukocyte mediated immunity	4.34×10^{-12}
GO:0002274	myeloid leukocyte activation	7.78×10^{-12}
GO:0002250	adaptive immune response	1.24×10^{-11}
GO:0002263	cell activation involved in immune response	7.07×10^{-11}
GO:0022407	regulation of cell-cell adhesion	5.40×10^{-9}
GO:0070661	leukocyte proliferation	1.22×10^{-8}

(*APOE*-rs429358, *PICALM*-rs10792832, *MS4A2*-rs983392 and *BIN1*-rs6733839) have been previously linked to AD (Supplementary Table 8), and one is a new GWAS finding [*ANKDR31*-rs4704171; OR = 1.19 (1.12–1.27); $p = 2.78 \times 10^{-8}$] (Table 2).

Then, we conducted a genome-wide meta-analysis combining the GR@ACE study with IGAP stage I. We identified 12 LOAD genomic regions reaching GWS. *CD33*-rs3865444, which did not reach GWS in the IGAP meta-analysis, was significantly associated with LOAD [OR =

0.92 (0.89–0.95); p = 3.61 x 10^{-8}] (Supplementary Figure 7). Among the top suggestive signals, we detected *HBEGF*-rs4150233 [OR = 0.92 (0.90–0.95); p = 5.10 x 10^{-8}] previously identified by a transethnic GWAS (28).

Next, meta-analysis of the whole GR@ACE dataset with IGAP I and II enabled the identification, for the first time, of *NDUFAF6*-rs10098778 as a GWS signal [OR = 0.94 (0.91–0.96); p = 2.54 x 10^{-8}). When we combined GR@ACE AD^{+++} endophenotype with IGAP I and II, we also detected *SCIMP*-rs7225151 [OR = 1.11 (1.07–1.15); p = 1.12 x 10^{-8}] (Table 2) (Supplementary Figure 8). It was previously reported as a genome-wide suggestive signal by IGAP (23). Recently, *SCIMP*-rs113260531, which is in complete LD with our lead marker (r^2 = 1), was associated with AD (17).

Replication of genome-wide significant findings

Finally, we tested for replication of the new signals in an independent sample of 4,959 Spanish individuals. The *CNTNAP2*-rs117834366, detected in the GWAS of GR@ACE VaD^{++} endophenotype, had a p value of 0.79 with similar effect direction to that reported previously, but strongly deflated in the entire replica sample [OR = 1.09 (0.66–1.78); p = 0.79] (Table 3). Analysis of the sub-specific VaD phenotype in the replica (N = 89) would be highly inaccurate.

In the exploration of meta-GWAS findings, we observed that the *ANKDR31*-rs4704171-C marker increased the risk of AD [OR = 1.10 (0.98–1.25); p = 0.09; Power = 33%]. Although the expected effect is in line with previous data, the precision of the estimate in this replica differs, ranging from a 2% decrease (a small negative association), to a 25% increase. Of note, the result emerging from the meta-analysis of the replica with the discovery sample (n = 26,194) is compatible with its potential role in dementia [OR = 1.18 (1.11–1.24); p = 1.15 x 10^{-8}] (Table 3). See forest plot in Supplementary Figure 9.

We observed a similar effect direction in the *NDUFAF6*-rs10098778 marker to that reported in the discovery stage, with interval estimates ranging from a risk decrease of 11% to a risk increase of 5% [OR = 0.96 (0.89–1.05); p = 0.40; Power = 16%]. This signal had a p = 2.32 x 10^{-8} in the final meta-analysis, including the whole GR@ACE dataset, the replica and IGAP I and II (n = 91,373) (Supplementary Figure 9).

SCIMP-rs7225151 showed a risk effect in the whole replica. Limits of the interval were consistent with a positive association [OR = 1.14 (1.01–1.29); p = 0.047; Power = 39%; n cases = 1,943]. In the AD^{+++} endophenotype, the marker presented the same positive risk effect direction [OR = 1.07 (0.88–1.30)], although

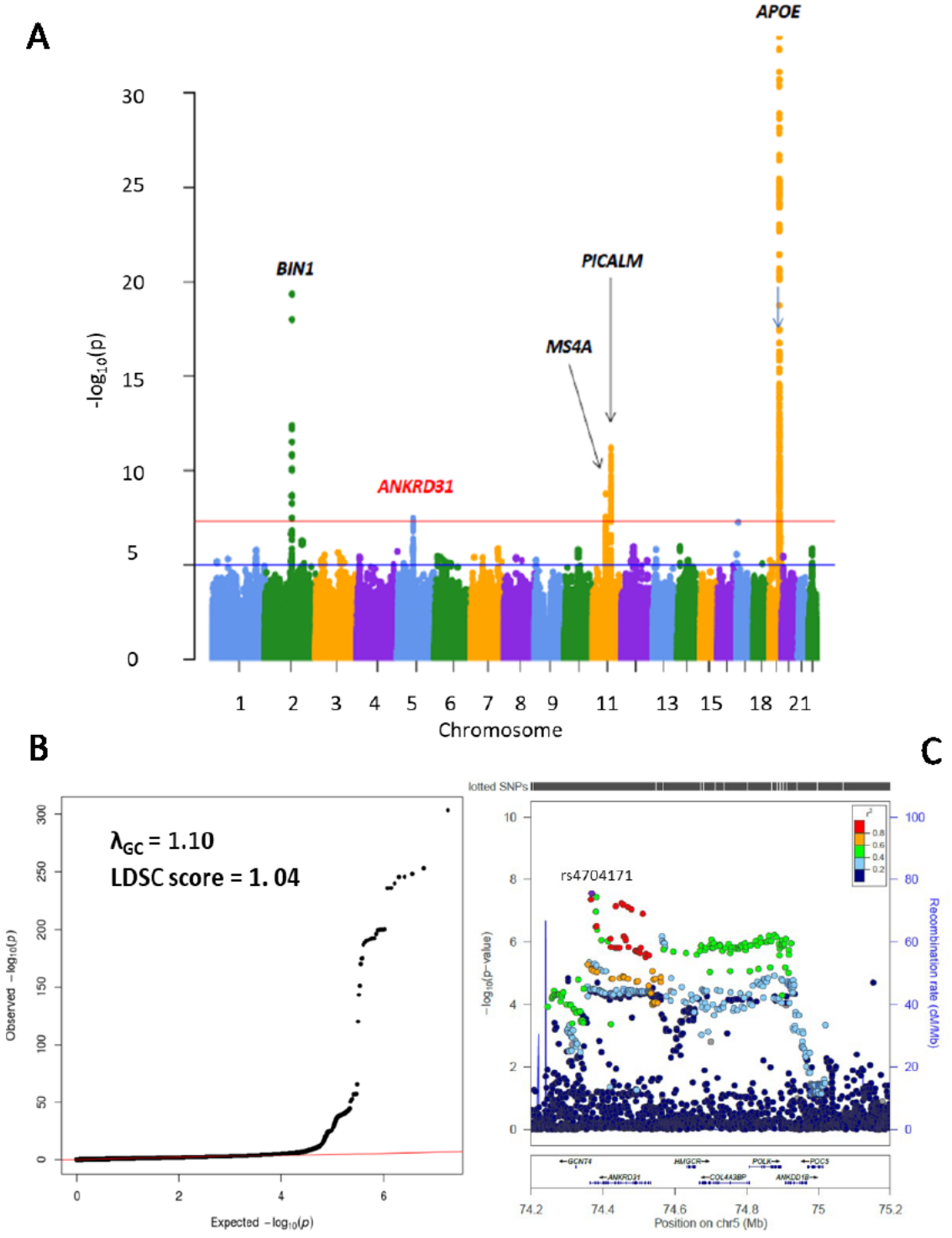


FIGURE 3. A) Results of genome-wide association analysis for GR@ACE meta-analysis with nine additional databases (N = 21,235). (B) QQplot. (C) Associations of the region centered on rs4704171 located in the *ANKRD31* locus and containing the *HMGCR* locus.

had p = 0.49, which could be mainly explained by by a reduction of the sample size [Power = 19%; n cases = 584]. Our results for both meta-analyses, the final meta-analysis of the whole GR@ACE dataset (n = 91,373) and the final meta-analysis of GR@ACE AD^{+++} endophenotype (n = 85,055) were compatible with a potential effect of this marker in AD (Table 3). See forest plots in Supplementary Figure 9.

Biological interpretation of meta-GWAS signals

To identify candidate genes and potential causal variants within novel meta-GWAS regions, we conducted cis-eQTL mapping. The rs2335107 marker located in the *ANKRD31* locus (chr5:74,451,693) was associated with the cortical expression of the long non-coding RNA (lncRNA) CTD-2235C13.3 ($p = 1.26 \times 10^{-5}$). This variant is located 83.4kb from the meta-GWAS lead SNP (rs4704171, chr5:74,368,254) and both are in complete LD ($r^2 = 1$). The *CTD-2235C13.3* gene is located 1.6kb from the *HMGCR* locus and its function is unknown. The *NDUFAF6* region mapped for NDUFAF6 RNA cortical expression ($p = 5.56 \times 10^{-6}$) and for TP53INP1 RNA blood expression ($p = 1.17 \times 10^{-10}$). Finally, rs73976325 (chr17:5,123,227), located in the *SCIMP* locus, to 13.8kb from the meta-GWAS top signal (rs7225151, chr17:5,137,047), mapped to brain cis-acting eQTL for AC012146.1 lincRNA ($p = 2.15 \times 10^{-7}$). Two additional markers were pinpointed to blood eQTLs, *SCIMP*-rs6502851 ($p = 3.89 \times 10^{-08}$) and *RABEP*-rs59277121 ($p = 3.89 \times 10^{-8}$) (Supplementary Table 3). In an additional prioritization strategy, combining information from positional mapping, eQTL, chromatin interaction and genome-wide association gene base analysis via FUMA, pointed to *ANKDR31* and *POLK* for the *ANKDR31* genomic region, as well as *NDUFAF6* and *TP53INP*1 for the*NDUFAF6* region (Supplementary Table 9). Further description is provided in supplementary information and Supplementary Table 9.

DISCUSSION

We present a comprehensive genome-wide association study of AD dementia cases. This represents the first pilot study exploring the genetics and underlying biological pathways of subgroups of AD patients, defined according to vascular burden. We showed differential biological routes underlying clinical endophenotypes and demonstrated how these differential subgroups of AD patients impact GWAS discoveries. The GR@ACE study represents a unique genomic resource because all affected cases were diagnosed in a single memory clinic using the same screening and diagnostic techniques. This might limit potential sources of clinical variation between study participants, recently demonstrated in a large meta-GWAS (12).

Based on the increase in evidence suggesting that vascular brain pathology can act concomitantly with AD to produce a more rapid cognitive decline (5), we explored the effect of known LOAD loci across different levels of vascular burden in dementia patients using only clinical definitions. Our basic idea was to dissect, from a molecular point of view, the model previously proposed by Viswanathan et al. (7). We observed three categories of loci, which might reflect the disease's clinical heterogeneity, from vascular and mixed forms to a "purer" AD phenotype. Intriguingly, we detected vascular processes to be the main causal mechanism in clinically pure AD and found the immune system pervasively detected across the three categories. Although both pathways have been previously associated with LOAD by network analysis (29), this is the first study to show that the association with the vascular system is conducted by AD-specific clinical subgroup.

It should be noted that the present study used clinical criteria to define the AD cases (19), but recently the classification of AD has evolved. In 2018 the National Institute of Aging-Alzheimer's Association (NIA-AA) proposed a novel research framework for the biological classification of AD based on the presence *in vivo* of biomarker-evidences for amyloid (A), tau (T) and neurodegeneration (N), as surrogate of the neuropathological state of an individual (30). The AT(N) biomarker system allows the classification of individuals into three categories: those with a normal biomarker profile, those with biomarkers compatible with AD-change, and those with biomarkers compatible with non-AD pathological changes (30). Using the NIA-AA approach, the generation of subgroups of AD patients considering vascular pathology would be unfeasible, as nowadays the ATN classification does not integrate measures of vascular dysfunction. Taking into account that most of dementia cases are caused by mixed pathologies, the current system seems deeply incomplete to study the probable interaction between neurodegeneration and vascular dysfunction. This idea has also been claimed by others (31), (32). Thus, we encourage other groups to contrast the proposed loci classification, which was based on the GR@ACE clinical endophenotypes, but using well-powered GWAS cohorts with available neuropathological data.

Silent changes occur in brain microvasculature during AD progression. In fact, CAA is a well-recognized AD pathological feature characterized by the accumulation of amyloid proteins in the walls of small cerebral vessels. CAA has been proposed to compromise the perivascular drainage of Aβ from the brain to the peripheral system (33). Almost all AD brains harbor CAA pathology to some extent, although *in vivo* most CAA cases remain undiagnosed, even using the

validated Boston criteria (34). Mendelian mutations of the *APP* gene have been found in CAA and AD (8), (35). *APOE* Ɛ4 and *CR1* have been associated with an increased risk of CAA (36), (37). In particular, distinct AD loci have been associated with capillary and non-capillary CAA (38). Between them, *APOE* Ɛ4 was strongly related to capillary CAA (38). These links make it conceivable that a potential genetic overlap exists between CAA and AD, and suggest that CAA pathology could represents an underlying process for AD. In that context, we think that intrinsic alterations to the vasculature could contribute to disease pathogenesis in more pure forms of AD, explaining our results. Conversely, in AD individuals with evident cerebrovascular lesions comprising mixed forms, the additional role of cardiovascular risk factors such as hypertension, atherosclerosis, or arteriosclerosis should be considered, as these could point to a systemic pathological state leading to vascular damage and dementia. This would agree with the limited genetic correlation between neurodegenerative and other neurologic disorders (12), as well as with results coming from heterochronic parabionts in aging models (39).

Understanding the role of vasculature pathology in AD seems a pertinent step. In that scenario, CAA would be a key AD hallmark. CAA represents the unique identified link between the vascular and amyloid hypotheses, but it has been completely neglected in the original hypothesis formulation.

From a clinical point of view, placing each patient somewhere along the disease spectrum proposed by Viswanathan is complex (7). A deep understanding of heterogeneity in AD seems necessary to design better genetic studies, which must drive the discovery of novel loci and, ultimately, innovative targets for AD therapies. In this study, we explored how clinical heterogeneity might impact GWAS findings by integrating distinct GWAS datasets with either the GR@ACE cohort as a whole or its endophenotypes. We found several new GWS signals that seem strongly dependent on the sample composition. For example, after combining IGAP Stages I and II with the entire GR@ACE dataset, we identified genetic signals in the *NDUFAF6* genomic region, but not in *SCIMP* region. When this exercise was conducted using GR@ACE endophenotypes, the *SCIMP* signal was detected using the clinically "pure" AD GR@ACE endophenotype. It should be noted that the power to detect *SCIMP* signal in the meta-analysis with GRA@ACE dementia and GR@ACE AD^{+++} was 75 and 70%, respectively. We tried to replicate this finding in a purer AD dataset without clinical mixed dementia cases, but the available number of clinical AD^{+++} cases might be compromising the statistical power to replicate (N cases = 584; Power = 19%). Despite that, we think that using

specific clinical subgroups of the AD population might empower genetic studies to detect genes associated with specific disease axes.

An alternative strategy is taking advantage of clinical heterogeneity. Specifically, heterogeneity might play a dual role in genetic studies. Although it might decrease the power to detect genes associated with more specific clinical subgroups, incorporating detailed clinical AD definitions can also promote identifying genes shared with other conditions or co-pathologies such as small vessel disease (SVD). In fact, this was the case for the *ATP5H* locus, which was previously found to be associated with AD (40) and more recently found in relation to SVD (41). We think that the same could apply to the *ANKRD31* finding. *ANKRD31* encodes a protein containing ankyrin-repeats, which is linked with neurodevelopmental disorders (42). Of note, *ANKRD31* GWAS signal mapped to the brain eQTL of a lncRNA, located 1.6kb from the *HMGCR* locus and residing in the *COL4A3BP* gene. The *HMGCR* locus is one of the most important co-regulators of cholesterol biosynthesis, and it is the therapeutic target of statins. The *COL4A3BP* gene is involved in lipid transport (43). Several studies have linked *HMGCR* polymorphisms and AD risk or age at onset for AD (44), and the cholesterol pathway has been identified to be a biological route shared between AD and SVD. Interestingly, markers in the *POLK* locus, associated by genome-wide association gene based analysis on this study, and located in the same disequilibrium block of *ANKDR31* (Figure 3), jointly conferred risk for AD and plasma levels of LDL (45). Considering prior findings, our results are consistent with this genomic region having a role in mixed dementia. The reported genetic signal should be considered a highly probable finding, although independent replications are still required.

In the present work, *NDUFAF6* signals reached GWS for the first time. This finding presented the same effect direction in the independent sample (Power = 16%), and remains as GWS after the final meta-analysis. Our lead GWAS marker is in high LD with *NDUFAF6*-rs4735340, the top suggestive signal reported by Kunkle et al. (15) at this region (r^2 = 0.95, for CEU population). Despite that, there are subtle differences in LD estimates for the Iberian population (r^2 = 0.87), suggesting that the genetic architecture of the Spanish population could be helping to pinpoint the region of interest. This region is in the close vicinity of *TP53INP1*, previously associated with AD by a gene-based approach (46). Considering these findings and results emerging from eQTL analysis, it would be feasible that a regulatory element for *NUDFAF6* resides upstream of the *TP53INP1* locus. We also detected that the *SCIMP* signal was mainly conducted by a specific group of AD cases. This signal was

reported to be a suggestive signal by IGAP (23), and a proxy of it (*SCIMP*-rs113260531) reached GWS recently (17). *SCIMP* is involved in mitochondrial function. *SCIMP* regions has been involved in several eQTLs, from uncharacterized cortical lncRNA to blood eQTLs in *SCIMP* and *RABEP1* loci, both associated with immune system function (47), (48). The *CD33* locus remains a controversial LOAD locus because large meta-GWAS were unable to replicate this signal (23), but here it reached GWS. We previously proposed that cryptic population sub-structure could explain the divergent observations for this locus (49).

Note that the lack of definitive neuropathological data for AD cases is a severe limitation of the present study. Clinical definitions have important uncertainties, and diagnosis misclassifications sometimes occur. Hence, some AD individuals included in enriched AD endophenotypes may present concomitant vascular brain disease. The generation of large histopathological GWAS cohorts with associated quantitative data on each pathological hallmark is the ultimate solution to tackling the intrinsic heterogeneity in AD. Unfortunately, there are few examples of neuropathological cohorts: only one GWAS has investigated the genetics of CAA, with *APOE* being the unique GWS signal (36). In this study, a small number of AD cases evolved to vascular dementia during follow-up. Large cross-sectional clinical GWAS cannot control diagnostic changes occurring in clinical practice. Clinical diagnosis is a dynamic variable, so understanding the genetic profiles of subgroups of patients evolving to other pathologies would provide powerful information.

It should be considered that there is a limitation in reducing the sample size by splitting the cohort in different endodophenotypes, instead of combining them. Despite that, spanning the spectrum of dementia individuals to generate clinical endophenotypes provided us a versatile design, which let us explore the effect of heterogeneity in GWAS and replicate the main findings of pathway analysis using an alternative strategy. The limited number of VaD cases in subgroup analysis limits the accuracy of gene categorization and pathway analysis. Finally, the exact effector genes for LOAD genetic findings remain unclear. This is a severe limitation to pathway analysis that can only be circumvented by isolating the causative mutations. Independent replication will be needed to corroborate our new GWS signals. In that sense, the selection of specific patient groups might lead to successful replication studies.

The assessment of heterogeneity has important implications for gene discovery, the development of treatments, and their appropriate use in individual patients. The GR@ACE cohort provides useful genomic information, as it accounts

for potential sources of variability and contains different subgroups of cases. This enabled us to analyze the LOAD genetic landscape in terms of clinical endophenotypes. Our efforts to disentangle the mechanistic pathways operating under clinical subgroups of patients revealed that vasculature regulation may be an essential part of the causative mechanism of LOAD. Finally, our exploration of AD genetics highlights the relevance of sample composition in genetic discoveries. Considering sample composition in the design of genetic studies might lead to the identification of genetic profiles, which can help clinicians distinguish subsets of patients within the disease spectrum and promote novel therapy targets for AD.

REFERENCES

1. Baumgart M, Snyder HM, Carrillo MC, Fazio S, Kim H, Johns H. Summary of the evidence on modifiable risk factors for cognitive decline and dementia: A population-based perspective. Alzheimer's Dement . 2015 Jun;11(6):718–26.
2. Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi HM. Forecasting the global burden of Alzheimer's disease. Alzheimer's Dement . 2007 Jul;3(3):186–91.
3. Toledo JB, Arnold SE, Raible K, Brettschneider J, Xie SX, Grossman M, et al. Contribution of cerebrovascular disease in autopsy confirmed neurodegenerative disease cases in the National Alzheimer's Coordinating Centre. Brain . 2013 Sep 1;136(9):2697–706.
4. Serrano-Pozo A, Frosch MP, Masliah E, Hyman BT. Neuropathological Alterations in Alzheimer Disease. Cold Spring Harb Perspect Med . 2011 Sep 1;1(1):a006189–a006189.
5. Regan C, Katona C, Walker Z, Hooper J, Donovan J, Livingston G. Relationship of vascular risk to the progression of Alzheimer disease. Neurology . 2006 Oct 24;67(8):1357–62.
6. Arvanitakis Z, Capuano AW, Leurgans SE, Bennett DA, Schneider JA. Relation of cerebral vessel disease to Alzheimer's disease dementia and cognitive function in elderly people: a cross-sectional study. Lancet Neurol . 2016 Aug ;15(9):934–43.
7. Viswanathan A, Rocca WA, Tzourio C. Vascular risk factors and dementia: How to move forward? Neurology . 2009 Jan 27;72(4):368–74.
8. Goate A, Chartier-Harlin MC, Mullan M, Brown J, Crawford F, Fidani L, et al. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. Nature . 1991 Feb 21;349(6311):704–6.
9. Sherrington R, Rogaev EI, Liang Y, Rogaeva EA, Levesque G, Ikeda M, et al. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. Nature . 1995 Jun 29;375(6534):754–60.
10. Sherrington R, Froelich S, Sorbi S, Campion D, Chi H, Rogaeva EA, et al. Alzheimer's disease associated with mutations in presenilin 2 is rare and variably penetrant. Hum Mol Genet . 1996 Jul;5(7):985–8.
11. Hardy JA, Higgins GA. Alzheimer's disease: the amyloid cascade hypothesis. Science . 1992 Apr 10;256(5054):184–5.
12. Brainstorm Consortium V, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, et al. Analysis of shared heritability in common disorders of the brain. Science . 2018 Jun 22;360(6395):eaap8757.
13. Avramopoulos D. Genetics of Alzheimer's disease: recent advances. Genome Med . 2009 Jan;1(3):34.
14. Corder E, Saunders A. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science (80-) . 1993;8(14):41–3.
15. Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. Nat Genet . 2019 Mar 28;51(3):414–30.
16. Sims R, van der Lee SJ, Naj AC, Bellenguez C, Badarinarayan N, Jakobsdottir J, et al. Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. Nat Genet . 2017 Jul 17;49(9):1373–84.
17. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. Nat Genet . 2019 Jan 7;
18. Ridge PG, Hoyt KB, Boehme K, Mukherjee S, Crane PK, Haines JL, et al. Assessment of the genetic variance of late-onset Alzheimer's disease. Neurobiol Aging . 2016 May;41:200.e13-200.e20.
19. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement . 2011 May;7(3):263–9.

20. Boada M, Tárraga L, Hernández I, Valero S, Alegret M, Ruiz A, et al. Design of a comprehensive Alzheimer's disease clinic and research center in Spain to meet critical patient and family needs. Alzheimers Dement . 2014 May;10(3):409–15.

21. Román GC, Tatemichi TK, Erkinjuntti T, Cummings JL, Masdeu JC, Garcia JH, et al. Vascular dementia: diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop. Neurology . 1993 Feb;43(2):250–60.

22. Gauderman WJ. Sample size requirements for matched case-control studies of gene-environment interaction. Stat Med . 2002 Jan 15;21(1):35–50.

23. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet . 2013 Dec;45(12):1452–8.

24. Wasserstein RL, Lazar NA. The ASA's Statement on *p* -Values: Context, Process, and Purpose. Am Stat . 2016 Apr 2;70(2):129–33.

25. Wasserstein RL, Schirm AL, Lazar NA. Moving to a World Beyond " $p < 0.05$." Am Stat . 2019 Mar 29;73(sup1):1–19.

26. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants: Fig. 1. Bioinformatics . 2015 Nov 1;31(21):3555–7.

27. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun . 2017 Dec 28;8(1):1826.

28. Jun GR, Chung J, Mez J, Barber R, Beecham GW, Bennett DA, et al. Transethnic genome-wide scan identifies novel Alzheimer's disease loci. Alzheimers Dement . 2017 Jul;13(7):727–38.

29. Zhang B, Gaiteri C, Bodea L-G, Wang Z, McElwee J, Podtelezhnikov AA, et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. Cell . 2013 Apr 25;153(3):707–20.

30. Jack CR, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimer's Dement . 2018 Apr;14(4):535–62.

31. Sweeney MD, Montagne A, Sagare AP, Nation DA, Schneider LS, Chui HC, et al. Vascular dysfunction-The disregarded partner of Alzheimer's disease. Alzheimers Dement . 2019 Jan 1;15(1):158–67.

32. Hachinski V. Dementia: Paradigm shifting into high gear. Alzheimer's Dement . 2019 Apr 9;

33. Weller RO, Boche D, Nicoll JAR. Microvasculature changes and cerebral amyloid angiopathy in Alzheimer's disease and their potential impact on therapy. Acta Neuropathol . 2009 Jul 22;118(1):87–102.

34. Charidimou A, Shoamanesh A, Al-Shahi Salman R, Cordonnier C, Perry LA, Sheth KN, et al. Cerebral amyloid angiopathy, cerebral microbleeds and implications for anticoagulation decisions: The need for a balanced approach. Int J Stroke . 2018 Feb 9;13(2):117–20.

35. Sakai K, Yamada M. Cerebral amyloid angiopathy]. Brain Nerve . 2014 Jul;66(7):827–35.

36. Beecham GW, Hamilton K, Naj AC, Martin ER, Huentelman M, Myers AJ, et al. Genome-wide association meta-analysis of neuropathologic features of Alzheimer's disease and related dementias. PLoS Genet . 2014 Sep;10(9):e1004606.

37. Biffi A, Shulman JM, Jagiella JM, Cortellini L, Ayres AM, Schwab K, et al. Genetic variation at CR1 increases risk of cerebral amyloid angiopathy. Neurology . 2012 Jan 31;78(5):334–41.

38. Mäkelä M, Kaivola K, Valori M, Paetau A, Polvikoski T, Singleton AB, et al. Alzheimer risk loci and associated neuropathology in a population-based study (Vantaa 85+). Neurol Genet . 2018 Feb;4(1):e211.

39. Villeda SA, Plambeck KE, Middeldorp J, Castellano JM, Mosher KI, Luo J, et al. Young blood reverses age-related impairments in cognitive function and synaptic plasticity in mice. Nat Med . 2014 Jun 4;20(6):659–63.

40. Boada M, Antúnez C, Ramírez-Lorca R, DeStefano A, González-Pérez A, Gayán J, et al. ATP5H/KCTD2 locus is associated with Alzheimer's disease risk. Mol Psychiatry . 2014 Jun;19(6):682–7.

41. Traylor M, Adib-Samii P, Harold D, Alzheimer's Disease Neuroimaging Initiative the ADN, International Stroke Genetics Consortium (ISGC), UK Young Lacunar Stroke DNA resource TISGC, Dichgans M, et al. Shared genetic contribution to Ischaemic Stroke and Alzheimer's Disease. Ann Neurol. 2016 Feb 23 ;79(5):739.

42. Lucariello M, Vidal E, Vidal S, Saez M, Roa L, Huertas D, et al. Whole exome sequencing of Rett syndrome-like patients reveals the mutational diversity of the clinical phenotype. Hum Genet . 2016;135(12):1343–54.

43. Hanada K, Kumagai K, Yasuda S, Miura Y, Kawano M, Fukasawa M, et al. Molecular machinery for non-vesicular trafficking of ceramide. Nature . 2003 Dec 18;426(6968):803–9.

44. Leduc V, De Beaumont L, Théroux L, Dea D, Aisen P, Petersen RC, et al. HMGCR is a genetic modifier for risk, age of onset and MCI conversion to Alzheimer's disease in a three cohorts study. Mol Psychiatry . 2015 Jul 15;20(7):867–73.

45. Broce IJ, Tan CH, Fan CC, Jansen I, Savage JE, Witoelar A, et al. Dissecting the genetic relationship between cardiovascular risk factors and Alzheimer's disease. Acta Neuropathol . 2018 Nov 9;

46. Escott-Price V, Bellenguez C, Wang L-S, Choi S-H, Harold D, Jones L, et al. Gene-wide analysis

detects two new susceptibility genes for Alzheimer's disease. PLoS One . 2014 Jan;9(6):e94661.

47. Luo L, Bokil NJ, Wall AA, Kapetanovic R, Lansdaal NM, Marceline F, et al. SCIMP is a transmembrane non-TIR TLR adaptor that promotes proinflammatory cytokine production from macrophages. Nat Commun . 2017;8:14133.

48. Delunardo F, Margutti P, Pontecorvo S, Colasanti T, Conti F, Riganò R, et al. Screening of a microvascular endothelial cDNA library identifies rabaptin 5 as a novel autoantigen in Alzheimer's disease. J Neuroimmunol . 2007 Dec;192(1–2):105–12.

49. Moreno-Grau S, Hernández I, Heilmann-Heimbach S, Ruiz S, Rosende-Roca M, Mauleón A, et al. Genome-wide significant risk factors on chromosome 19 and the APOE; locus. Oncotarget . 2018 May 14;9(37):24590–600.

The GR@ACE study group

Abdelnour C[1,2], Aguilera N[1], Alarcon E[1,3], Alegret M[1,2], Benaque A[1], Boada M[1,2], Buendia M[1], Cañabate P[1,2], Carracedo A[4,5], Corbatón A[6], de Rojas I[1], Diego S[1], Espinosa A[1,2], Gailhajenet A[1], García González P[1], Gil S[1], Guitart M[1], González Pérez A[7], Hernández I[1,2], Ibarria, M[1], Lafuente A[1], Macias J[8], Maroñas O[4], Martín E[1], Martínez MT[6], Marquié M[1], Mauleón A[1], Monté-Rubio G[1], Montrreal L[1], Moreno-Grau S[1,2], Moreno M[1], Orellana A[1], Ortega G[1,2], Pancho A[1], Pelejà E[1], Pérez-Cordon A[1], Pineda JA[8], Preckler S[1], Quintela I[3], Real LM[3,8], Rodríguez-Gómez O[1,2], Rosende-Roca M[1], Ruiz A[1,2], Ruiz S[1,2], Sáez ME[7], Sanabria A[1,2], Santos-Santos MA[1], Serrano-Rios M[6], Sotolongo-Grau O[1], Tárraga L[1,2], Valero S[1,2], Vargas L[1]

(1. Research Center and Memory clinic Fundació ACE. Institut Català de Neurociències Aplicades. Universitat Internacional de Catalunya. Barcelona, Spain; 2. CIBERNED, Center for Networked Biomedical Research on Neurodegenerative Diseases, National Institute of Health Carlos III, Ministry of Economy and Competitiveness, Spain; 3. Dep. of Surgery, Biochemistry and Molecular Biology, School of Medicine. University of Málaga. Málaga, Spain; 4. Grupo de Medicina Xenómica, Centro Nacional de Genotipado (CEGEN-PRB3-ISCIII). Universidade de Santiago de Compostela, Santiago de Compostela, Spain; 5. Fundación Pública Galega de Medicina Xenómica-CIBERER-IDIS, Santiago de Compostela, Spain; 6. Centro de Investigación Biomédica en Red de Diabetes y Enfermedades Metabólicas Asociadas, CIBERDEM, Spain, Hospital Clínico San Carlos, Madrid, Spain; 7. CAEBI. Centro Andaluz de Estudios Bioinformáticos., Sevilla, Spain; 8. Unidad Clínica de Enfermedades Infecciosas y Microbiología. Hospital Universitario de Valme, Sevilla, Spain)

DEGESCO consortium

Adarmes-Gómez AD[1,2], Alarcón-Martín E[3,4], Álvarez I[5], Álvarez V[6,7], Amer-Ferrer G[8], Antequera M[9], Antúnez C[9], Baquero M[10], Bernal M[11], Blesa R[2,12], Boada M[2,3], Buiza-Rueda D[1,2], Bullido MJ[2,13,14], Burguera JA[10], Calero M[2,15,16], Carrillo F[1,2], Carrión-Claro M[1,2], Casajeros MJ[17], Clarimón J[2,12], Cruz-Gamero JM[4], de Pancorbo MM[18], de Rojas I[3], del Ser T[14], Diez-Fairen M[5], Fortea J[2,12], Franco E[11], Frank-García A[2,14,19], García-Alberca JM[20], Garcia Madrona S[16], Garcia-Ribas G[16], Gómez-Garre P[1,2], Hernández I[2,3], Hevilla S[20], Jesús S[1,2], Labrador Espinosa MA[1,2], Lage C[2,21], Legaz A[9], Lleó A[2,12], López de Munáin A[22], López-García S[2,21], Macias D[1,2], Manzanares S[9,23], Marín M[11], Marín-Muñoz J[9], Marín T[20], Marquié M[3], Martín Montes A[2,13,19], Martínez B[9], Martínez C[7,24], Martínez V[9], Martínez-Lage Álvarez P[25], Medina M[2,14], Mendioroz Iriarte M[26], Menéndez-González M[7,27], Mir P[1,2], Molinuevo JL[28], Monté-Rubio G[3], Montrreal L[3], Moreno-Grau S[2,3], Orellana A[3], Pastor AB[15], Pastor P[5], Pérez Tur J[2,29,30], Periñán-Tocino T[1,2], Piñol Ripoll G[2,31], Rábano A[2,15,32], Real de Asúa D[33], Rodrigo S[11], Rodríguez-Rodríguez E[2,21], Royo JL[4], Ruiz A[2,3], Sanchez del Valle Díaz R[34], Sánchez-Juan P[2,21], Sastre I[2,13], Sotolongo-Grau O3, Tárraga L[2,3], Valero S[2,3], Vicente MP[9], Vivancos L[9]

(1. Unidad de Trastornos del Movimiento, Servicio de Neurología y Neurofisiología. Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville, Spain; 2. CIBERNED, Network Center for Biomedical Research in Neurodegenerative Diseases, National Institute of Health Carlos III, Spain; 3. Research Center and Memory clinic Fundació ACE. Institut Català de Neurociències Aplicades. Universitat Internacional de Catalunya. Barcelona, Spain; 4. Dep. of Surgery, Biochemistry and Molecular Biology, School of Medicine. University of Málaga. Málaga, Spain; 5. Fundació per la Recerca Biomèdica i Social Mútua Terrassa, and Memory Disorders Unit, Department of Neurology, Hospital Universitari Mutua de Terrassa, University of Barcelona School of Medicine, Terrassa, Barcelona, Spain; 6. Laboratorio de Genética Hospital Universitario Central de Asturias, Oviedo; 7. Instituto de Investigación Biosanitaria del Principado de Asturias (ISPA); 8. Department of Neurology, Hospital Universitario Son Espases, Palma, Spain; 9. Unidad de Demencias. Hospital Clínico Universitario Virgen de la Arrixaca; 10. Servei de Neurologia, Hospital Universitari i Politècnic La Fe; 11. Unidad de Demencias, Servicio de Neurología y Neurofisiología. Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville, Spain; 12. Memory Unit, Neurology Department and Sant Pau Biomedical Research Institute, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain; 13. Centro de Biologia Molecular Severo Ochoa (C.S.I.C.-U.A.M.), Universidad Autonoma de Madrid, Madrid, Spain; 14. Instituto de Investigacion Sanitaria Hospital la Paz' (IdIPaz), Madrid, Spain; 15. CIEN Foundation, Queen Sofia Foundation Alzheimer Center, Madrid, Spain; 16. Instituto de Salud Carlos III (ISCIII), Madrid, Spain; 17. Hospital Universitario Ramón y Cajal; Madrid, Spain; 18. BIOMICs, País Vasco; Centro de Investigación Lascaray. Universidad del País Vasco UPV/EHU; 19. Neurology Service, Hospital Universitario La Paz (UAM), Madrid, Spain; 20. Alzheimer Research Center & Memory Clinic. Andalusian Institute for Neuroscience. Málaga, Spain; 21. Neurology Service, Marqués de Valdecilla University Hospital (University of Cantabria and IDIVAL), Santander, Spain; 22. Hospital Donostia de San Sebastían; 23. Fundación para la Formación e Investigación Sanitarias de la Región de Murcia; 24. Servicio de Neurología -Hospital de Cabueñes-Gijón; 25. Centro de Investigación y Terapias Avanzadas. Fundación CITA-alzheimer; 26. Navarrabiomed; 27. Servicio de Neurología -Hospital Universitario Central de Asturias, Oviedo; 28. Barcelona βeta Brain Research Center – Fundació Pasqual Maragall; 29. Unitat de Genètica Molecular. Institut de Biomedicina de València-CSIC; 30. Unidad Mixta de Neurologia Genètica. Instituto de Investigación Sanitaria La Fe; 31. Unitat Trastorns Cognitius, Hospital Universitari Santa Maria de Lleida, Institut de Recerca Biomédica de Lleida (IRBLLeida), Lleida, España. 32. BT-CIEN; 33. Hospital Universitario La Princesa, Madrid, Spain; 34. Hospital Clínic Barcelona)

ACKNOWLEDGEMENTS

We would like to thank patients and controls who participated in this project. The Genome Research @

Fundació ACE project (GR@ACE) is supported by Fundación bancaria "La Caixa," Grifols SA, Fundació ACE, and ISCIII (Ministry of Health, Spain). We also want to thank the private sponsors who support the basic and clinical projects of our institution (Piramal AG, Laboratorios Echevarne, Araclon Biotech S.A., and Fundació ACE). We are indebted to the Trinitat Port-Carbó legacy and her family for their support of Fundació ACE research programs. Fundació ACE is a participating center in the Dementia Genetics Spanish Consortium (DEGESCO). A.R. and M.B. receive support from the European Union/EFPIA Innovative Medicines Initiative Joint Undertaking ADAPTED and MOPEAD projects (Grants No. 115975 and 115985, respectively). M.B. and A.R. are also supported by national grants PI13/02434, PI16/01861, and PI17/01474. Acción Estratégica en Salud is integrated into the Spanish National R + D + I Plan and funded by ISCIII (Instituto de Salud Carlos III)-Subdirección General de Evaluación and the Fondo Europeo de Desarrollo Regional (FEDER- "Una manera de Hacer Europa"). LMR is supported by Consejería de Salud de la Junta de Andalucía (Grant PI-0001/2017). Control samples and data from patients included in this study were provided in part by the National DNA Bank Carlos III (www.bancoadn.org, University of Salamanca, Spain) and Hospital Universitario Virgen de Valme (Sevilla, Spain); they were processed following standard operating procedures with the appropriate approval of the Ethical and Scientific Committee. **The present work was performed as part of the Biochemistry, Molecular Biology and Biomedicine doctoral program of S. Moreno-Grau at Universitat Autònoma de Barcelona (Barcelona, Spain).**

Data collection and sharing for this project was partially funded by the **Alzheimer's Disease Neuroimaging Initiative (ADNI)** (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging and the National Institute of Biomedical Imaging and Bioengineering, as well as through generous contributions from the following: AbbVie; the Alzheimer's Association; the Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research provides funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study was coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

The **AddNeuroMed** data are from a public-private partnership supported by EFPIA companies and SMEs as part of InnoMed (Innovative Medicines in Europe), an integrated project funded by the European Union of the Sixth Framework program priority FP6-2004-LIFESCIHEALTH-5. Clinical leads responsible for data collection are Iwona Kłoszewska (Lodz), Simon Lovestone (London), Patrizia Mecocci (Perugia), Hilkka Soininen (Kuopio), Magda Tsolaki (Thessaloniki) and Bruno Vellas (Toulouse). Imaging leads are Andy Simmons (London), Lars-Olad Wahlund (Stockholm) and Christian Spenger (Zurich). Bioinformatics leads are Richard Dobson (London) and Stephen Newhouse (London).

Funding support for the **Alzheimer's Disease Genetics Consortium (ADGC)** was provided through the NIA Division of Neuroscience (U01-AG032984).

The genotypic and associated phenotypic data used in the study "Multi-Site Collaborative Study for Genotype-Phenotype Associations in Alzheimer's Disease (**GenADA**)" were provided by GlaxoSmithKline, R&D Limited. The datasets used for the analyses described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/gap through dbGaP accession number phs000219.

The **Mayo Clinic Alzheimer's Disease Genetic Studies**, led by Dr. Nilüfer Ertekin-Taner and Dr. Steven G. Younkin at the Mayo Clinic in Jacksonville, FL, used samples from the Mayo Clinic Study of Aging, the Mayo Clinic Alzheimer's Disease Research Center and the Mayo Clinic Brain Bank. Data collection was supported through funding by NIA grants P50 AG016574, R01 AG032990, U01 AG046139, R01 AG018023, U01 AG006576, U01 AG006786, R01 AG025711, R01 AG017216 and R01 AG003949, NINDS grant R01 NS080820, the CurePSP Foundation and support from the Mayo Foundation.

The **Neocodex-Murcia** study was funded by the Fundación Alzheimur (Murcia), the Ministerio de Educación y Ciencia (Gobierno de España), Corporación Tecnológica de Andalucía, Agencia IDEA (Consejería de Innovación, Junta de Andalucía), the Diabetes Research Laboratory and the Biomedical Research Foundation. University Hospital Clínico San Carlos has been supported by CIBER de *Diabetes y Enfermedades Metabólicas Asociadas* (CIBERDEM); CIBERDEM is an ISCIII Project.

The **ROS/MAP** study data were provided by the Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago. Data collection was supported through funding by NIA grants P30AG10161, R01AG15819, R01AG17917, R01AG30146, R01AG36836, U01AG32984 and U01AG46152, the Illinois Department of Public Health and the Translational Genomics Research Institute

The **TGEN** study was supported by Kronos Life Sciences Laboratories, the National Institute on Aging (Arizona Alzheimer's Disease Center grants P30 AG19610 and RO1 AG023193, the Mayo Clinic Alzheimer's Disease Center grant P50 AG16574 and the Intramural Research Program), the National Alzheimer's Coordinating Center (U01 AG016976) and the state of Arizona.

The results published here are in part based on data obtained from the AMP-AD Knowledge Portal accessed at http://dx.doi.org/doi:10.7303/syn2580853

We thank the International Genomics of Alzheimer's Project (IGAP) for providing summary results data for these analyses. The investigators within IGAP contributed to the design and implementation of IGAP and/or provided data but did not participate in analysis or writing of this report. IGAP was made possible by the generous participation of the control subjects, the patients, and their families. The i–Select chip was funded by the French National Foundation on Alzheimer's disease and related disorders. EADI was supported by the LABEX (laboratory of excellence program investment for the future) DISTALZ grant, Inserm, Institut Pasteur de Lille, Université de Lille 2 and the Lille University Hospital. GERAD was supported by the Medical Research Council (Grant n 503480), Alzheimer's Research UK (Grant n 503176), the Wellcome Trust (Grant n 082604/2/07/Z) and German Federal Ministry of Education

and Research (BMBF): Competence Network Dementia (CND) grant n 01GI0102, 01GI0711, 01GI0420. CHARGE was partly supported by the NIH/NIA grant R01 AG033193 and the NIA AG081220 and AGES contract N01–AG–12100, the NHLBI grant R01 HL105756, the Icelandic Heart Association, and the Erasmus Medical Center and Erasmus University. ADGC was supported by the NIH/NIA grants: U01 AG032984, U24 AG021886, U01 AG016976, and the Alzheimer's Association grant ADGC–10–196728.

SUPPLEMENTARY METHODS

GR@ACE cohort and phenotype definitions

All AD patients included in the GR@ACE study received a thorough structured neurological evaluation that included: history, examination, Mini-Mental State Examination (MMSE) (1), Blessed Dementia Rating Scale (BDRS) (2), (3), Neuropsychiatric Inventory-questionnaire (NPI-Q) (4), Tinnetti scale for gait and balance(5), Clinical Dementia Rating (CDR) (6), Global Deterioration Scale (GDS) (7) scoring and Hachinski Ischemia Scale (8). The Fundacio ACE neuropsychological battery (NBACE) (9) was administered to all patients. The NBACE includes measures of cognitive information processing speed, orientation, attention, verbal learning and memory, language, visuoperception, praxis and executive functions. Family members or caregivers are interviewed by a social worker.

Endophenotype approach was feasible due to Fundació ACE's endorsement both, a primary and a secondary diagnosis, as well as routine follow-up evaluations. The secondary diagnosis might be the same or an additional clinical condition that explains the clinical symptoms. Follow-up visits enabled controlling for the progression of the disease and identifying comorbidities and sometimes diagnosis changes, thereby providing a longitudinal landscape for each individual's clinical evolution (Supplementary figure 1). VaD patients were defined according to NINDS-AIREN criteria (10).

GWAS genotyping

Peripheral blood was taken from all individuals to isolate germline DNA from leukocytes. DNA extraction was performed automatically according to standard procedures using the Chemagic system (Perkin Elmer). Extensive DNA quality control was conducted. Only samples reaching DNA concentrations >10ng/µl and presenting high integrity were included for genotyping. Samples and controls were randomized across sample plates to avoid batch effects.

The genotyping array, Axiom 815K Spanish Biobank array, is an adaptation of the Axiom Biobank Genotyping Array, but also contains population-specific rare variations observed in the Spanish population. DNA samples were genotyped according to the manufacturer's instructions (Axiom™ 2.0 Assay Manual Workflow). The Axiom 2.0 assay interrogates biallelic SNPs and simple indels in a single assay workflow. Starting with 200 ng of genomic DNA, the samples were processed through a manual target preparation protocol followed by automated processing of the array plates in the GeneTitan Multi-Channel instrument. Target preparation

involved DNA amplification, fragmentation, purification and resuspension of the target in a hybridization cocktail. The hyb-ready targets were then transferred to the GeneTitan Multi Channel instrument for automated, hands-free processing including hybridization, staining, washing and imaging. CEL files were generated using the GeneTitan Multi Channel instrument. To achieve higher genotyping performance, quality control for samples and plates was performed using the Affymetrix power tool (APT) 1.15.0 software following Axiom Data Analysis Workflow. Briefly, sample quality was determined based on the resolution of AT and GC channels in a group of non-polymorphic SNPs (resolution > 0.82). Passing samples were genotyped for sample QC. Samples with a call rate greater than 97% and plates with an average call rate above 98.5% were included for final SNP calling. Quality samples were jointly called to minimize batch effects. Best quality markers were selected for downstream analysis (N_{SNPs} =777,649; 95.4%) using the SNPolisher R package (Thermo Fisher). To assess the sample genotyping concordance, we intentionally re-sampled 200 samples and determined a concordance rate of 99.5%.

GWAS quality control

Principal component analysis was conducted excluding markers with moderate-to-high linkage disequilibrium (LD) (r2>0.3) and long-range LD regions from PCA analysis using the "indep-pairwise" option of PLINK 1.9(11) (https://www.cog-genomics.org/plink2).

GWAS Statistical analysis

We tested the association between markers and the targeted phenotypes with and without covariates. To establish the discovery model, we conducted an exploratory step using the GR@ACE dataset. We evaluated the performance of five different models: a) Model A, unadjusted; b) Model B, unadjusted and excluding all individuals matching with Hispanic American ancestry; c) Model C, adjusted per four main PCs; d) Model D, adjusted per top four PCs, gender and age; and e) Model E, adjusted per top four PCs, gender, age and *APOE* status (Supplementary figure 2). We explored the QQ plot and the genomic inflation factor (λ), using the Qqman and GenABEL packages (http://www.genabel.org/packages/GenABEL) from R, to test for covariates in the regression model.

Genetic exploration of GR@ACE clinical endophenotypes and enrichment analysis

We explored the biological pathways underlying each gene category. First, we extracted the significant co-expressed genes ($p < 0.05$) for each category using GeneFriends in human microarray datasets containing 26,113 experimental conditions and 19,080 genes (12). Second, to avoid unspecific pathway

detection, we selected co-expressed genes with ≥2 LOAD loci per category, and ranked the results from the highest to the lowest number of co-occurrences. Third, we set the maximum number of co-expressed genes for pathway analysis to 150. Thus, *category A* and *category C* includes 150 co-expressed genes (Bonferroni p-value = 3.33 x 10^{-4}); and *category B*, 99 (Bonferroni p-value = 5.05 x 10^{-4}). Next, we explored the biological pathways underlying each co-expression on the ranked list by applying an overrepresentation enrichment method in WebGestalt (13), while using Gene Ontology as a reference database for functional annotations of non-redundant biological pathways.

Next, we performed a sub-classification and pathway analysis for variants from *category C*: Subset C1 includes variants presenting a stable effect across endo-phenotypes (*ABCA7, ADAMTS1, CD2AP, CELF1, EPHA1, INPP5D, NME8, PTK2B, ZCWPW1*); and subset C2 includes variants with an inverse effect between extreme endo-phenotypes (*CLU, FERMT2, IQCK*). Both clusters presented 150 co-expressed genes for pathway analysis.

Finally, we carried out a stringent analysis to validate previous results. We applied a linear regression model using R to evaluate the effect change trend per each genetic variant across clinical endophenotypes. Dementia endophenotypes were the independent variables, coded as follows: VaD^{++} = 1; VaD^{+} = 2; Dementia = 3; AD^{+++} = 6; AD^{+} = 4; AD^{++} = 5. Effect change was the dependent variable. Effect change represent the strength of the association per variant in each endophenotype, and it is calculated as the absolute difference between variant odds ratio (OR) and variant null effect (OR = 1) (Supplementary Figure 4). Variants presenting a linear trend and a positive effect ($R2$>0.80 and β≥0.01) comprised Category A (*APOE, CR1, MEF2C, MS4A2, PICALM*). Those presenting a linear trend and a negative effect ($R2$>0.80 and β≤-0.01) comprised Category *B* (*SORL1, CASS4*). Those not fulfilling the above criteria comprised Category C (*EPHA1, ABCA7, ACE, ADAMTS1, ADAM10, ATP5H, BIN1, MAPT, CELF1, CD2AP, CD33, CLU, FERMT2, INPP5D, IQCK, NME8, PTK2B, ZCWPW1*). Pathway analysis was conducted applying identical procedure to those described previously. Category A includes 150 co-expressed genes (Bonferroni p-value = 3.33 x 10^{-4}); Category B, 116 (Bonferroni p-value = 4.31 x 10^{-4}) and Category C, 150. At this point, we were not able to detect vascular processes in Category A. To discard that this was caused by unspecific pathway detection, we restricted the analysis to include those loci co-expressing with more than 4 LOAD genes at this cluster (co-expressing loci = 43, Bonferroni p-value = 5.55 x 10^{-4}). Regulation of vascular development was identified as top pathway ($p < 2.14 \times 10^{-7}$).

Meta-analysis: datasets

To identify novel loci associated with AD, we combined the GR@ACE dataset and its endophenotypes with: 1) raw genotype data from nine additional GWAS series (N = 13,826) (Supplementary Table 2); and 2) public summary statistics from IGAP stages I (N = 54,162) and II (N = 74,046).

Genotype level data

In the first meta-analysis, we had access to raw genotyped data for 7,879 AD patients and 5,947 controls. All datasets were processed by applying the same quality control and imputation procedures as those described for the GR@ACE cohort. To exclude duplicate samples coming from different studies, we also performed a joint analysis and carried out an identity-by-descendent analysis (IBD) (Pi-hat > 0.80) of the nine cohorts and GR@CE (n = 21,235) using PLINK software 1.9.(11) The study cohorts included:

The Alzheimer's Disease Neuroimaging Initiative (ADNI).

We obtained the data used in preparing this article from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI is to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers and clinical and neuropsychological assessments can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). The ADNI study has three phases: ADNI1, ADNI GO and ADNI2. For up-to-date information, see www.adni-info.org. In the present study, we included 478 cases and 243 controls from ADNI1 and ADNI2.

The AddNeuroMed Study.

AddNeuroMed was a public-private partnership for biomarker discovery and replication in Alzheimer's disease (14,15). It was a multi-center study in Europe with the first patient enrolled in January 2006 and the last in February 2008. The study protocol was planned for a baseline assessment visit with follow-ups every 3 months for the first year, then annual visits that continued through 2013. The study enrolled a total of 258 AD, 257 MCI and 266 controls, but not all had complete data at each assessment. In the present study, we included 450 cases and 187 controls.

This dataset was downloaded from Synapse (doi:10.7303/syn2790911).

The Alzheimer's Disease Genetics Consortium (ADGC).

The National Institute on Aging (NIA) Alzheimer's Disease Centers' (ADC) cohort includes subjects ascertained and

evaluated by the clinical and neuropathology cores of the 29 NIA-funded ADCs (16). Data collection was coordinated by the National Alzheimer's Coordinating Center (NACC). The ADC cohort consists of autopsy-confirmed and clinically-confirmed AD cases, cognitively normal elders (CNEs) with complete neuropathology data who were older than 60 years at age of death, as well as living CNEs evaluated using the Uniform dataset (UDS) protocol who were documented to not have mild cognitive impairment (MCI) and were between 60 and 100 years of age at assessment. In the present study, we included 3287 cases and 1322 controls.

This study was downloaded from dbGaP (phs000372).

Multi-Site Collaborative Study for Genotype-Phenotype Associations in Alzheimer's disease and Longitudinal follow-up of Genotype-Phenotype Associations in Alzheimer's disease and Neuroimaging component of Genotype-Phenotype Associations in Alzheimer's disease (GenADA).

GenADA was a multi-site collaborative study involving GlaxoSmithKline Inc and nine medical centers in Canada, including 1000 AD patients and 1000 ethnically-matched controls in order to associate DNA sequence (allelic) variations in candidate genes with AD phenotypes (17,18). The study consisted of both retrospective and prospective data. Where possible, biological relatives with Alzheimer's (up to third-degree relationships such as cousins) and unaffected siblings of AD cases were also recruited. In the present study, we included 785 cases and 764 controls.

This study was downloaded from dbGaP (phs000219).

The Mayo Clinic LOAD genome-wide association study.

Subjects from the Mayo LOAD GWAS were selected from two clinical AD Case-Control series: Mayo Clinic Jacksonville (MCJ) and Mayo Clinic Rochester (MCR), as well as a neuropathological series of autopsy-confirmed subjects from the Mayo Clinic Brain Bank (19). All subjects from the clinical series (MCJ and MCR) were diagnosed by a Mayo Clinic neurologist; all control subjects had a Clinical Dementia Rating score of zero at the most recent time of testing; all LOAD patients had a diagnosis of probable or possible AD according to the NINCDS-ADRDA criteria (20). All ADs had definite diagnoses according to the NINCDS-ADRDA criteria and had Braak scores of ≥4.0. All non-AD Controls had Braak scores of ≤2.5; many had brain pathology unrelated to AD. . In the present study, we included 703 cases and 1066 controls.

This dataset was downloaded from Synapse (doi:10.7303/syn5550404).

The Neocodex-Murcia study.

The study included 327 sporadic AD patients and 801 controls with unknown cognitive status from the Spanish general population collected by Neocodex (21,22). AD patients were diagnosed as possible or probable AD in accordance with the criteria of the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA)(20). In the present study, we included 324 cases and 754 controls.

The Religious Orders Study and Memory and Aging Project (ROS/MAP) Study.

The Religious Orders Study (ROS) was a longitudinal clinical-pathologic cohort study of aging and Alzheimer's disease (AD) from Rush University that enrolled individuals from religious communities for longitudinal clinical analysis and brain donation (23). Participants were enrolled from more than 40 groups of religious orders (nuns, priests, brothers) across the United States. Enrolment required no known indications of dementia. Medical conditions were documented starting in 1994 by clinical evaluation or self-report. Alzheimer's disease status was determined by a computer algorithm based on cognitive test performance with a series of discrete clinical judgments made in series by a neuropsychologist and a clinician.

The Memory and Aging Project (MAP) was a longitudinal epidemiologic clinical-pathologic cohort study of common chronic conditions of aging with an emphasis on declines in cognitive and motor function and the risk of Alzheimer's disease. This study began in 1997 and was run by Rush University (23). This study was designed to complement the ROS study by enrolling individuals with a wider range of life experiences and socioeconomic status into a study of similar structure and design as ROS. The study enrolled older individuals without any signs of dementia, primarily recruiting from continuous care retirement communities throughout northeastern Illinois, USA. Diagnoses of dementia and AD were performed in an identical manner to the ROS study. . In the present study, we included 628 cases and 229 controls. This dataset was downloaded from Synapse (doi:10.7303/syn3219045).

The Translational Genomics Research Institute (TGEN) study.

The TGEN GWAS study included 643 late-onset AD cases and 404 controls from a neuropathological cohort, and 197 late-onset AD cases and 114 controls from a clinical cohort, all of which were genotyped with the Affimetrix 500 K GeneChip Array (24). In the present study, we included 741 cases and 449 controls.

TGEN investigators have provided free access to genotype data to other

researchers via Coriell Biorepositories (http://www.coriell.org/).

IGAP summary statistics

In the second meta-analysis, we combined public summary statistics from IGAP stages I and II (http://web.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php) with the GR@ACE dataset and its endophenotypes. IGAP stage I consisted of 17,008 AD cases and 37,154 controls collected from four published GWAS datasets (The European Alzheimer's disease Initiative – EADI; the Alzheimer Disease Genetics Consortium – ADGC; the Cohorts for Heart and Aging Research in Genomic Epidemiology consortium – CHARGE; the Genetic and Environmental Risk in AD consortium – GERAD) containing 7,055,881 single nucleotide polymorphisms (SNPs). IGAP stage II was a replication effort in which significantly associated variants of stage I were genotyped (N_{SNP} = 11,632) in 8,572 AD cases and 11,312 controls.

Meta-analysis: association analysis

Summary statistics for each individual dataset were combined with METAL software (25), release 2011-03-25, using inverse variance weighted meta-analysis without genomic control as default. Meta-analysis with genomic control was also explored. To define independent significant signals in the meta-analyses, we assigned variants to clusters using the clump function of PLINK software 1.9(11) based on the association p-values and the short-range LD (r^2 > 0.5; 250Kb). We used GCTA-COJO(26), version 1.91.3 beta3, to perform standard conditional analysis adjusting for the lead SNP in a region of ±500kb. We estimated the LD score intercept using LD score regression (LDSC v.1.0.0) to distinguish polygenicity from other confounding factors (27). Regional plots were generated using LocusZoom software (28).

Biological interpretation of meta-GWAS signals

Functional consequences of GWS markers obtained in the meta-analyses were assessed using FUMA (29). Briefly, FUMA uses ANNOVAR, RegulomeDB, and chromatin states to predict genomic functional elements, and CADD score to predict variant effect. Criteria to define lead SNPs was set to *r2*>0.5 in a window of 250kb, which is the same criteria used for clumping analysis. eQTL mapping performed in FUMA included the exploration of brain and blood eQTLs in BRAINEAC, CommonMind Consortium Portal, and xQTL Serve databases, in addition to GTEx v7. Chromatin interaction was performed applying similar methodology to that described by Jansen, et al. (30). Finally, gene based analysis was conducted using MAGMA, via FUMA. A total of 18,795 coding genes were included for GRACE and dbGaP meta-analysis, 861 for GRACE and IGAP12 meta-analysis, and 864 from GRACE AD^{+++} and IGAP12

meta-analysis, Bonferroni correction was set to 2.66 x 10^{-6}, 5.79 x 10^{-5} and 5.81 x 10^{-5}, respectively.

SUPPLEMENTARY RESULTS

GR@ACE genome-wide association study

Genome-wide analysis for GR@ACE clinical endophenotypes revealed a suggestive signal in AD^{+++} endophenotype [*PCBD1/UNC5B*-rs7100488; OR = 0.76 (0.69 – 0.84); p = 5.16 x 10^{-8}]. After exploring endophenotype GWAS signals, *PCBD1/UNC5B*-rs7100488 and *CNTNAP2*-rs117834366, in additional datasets, we detected a nominal significance and a consistent effect for *PCBD1/UNC5B*-rs7100488 marker only in the ADGC2 dataset [OR = 0.79 (0.61 – 1.01); p = 0.058] (Supplementary table 5).

Genetic exploration of GR@ACE clinical endophenotypes and enrichment analysis

For category B, with variants with stronger effects in AD mixed with vascular disease, *SORL1*-rs11218343 showed the strongest vascular enrichment [VaD^{++} OR (95%) = 0.71 (0.43 - 1.16), p-value = 0.168; AD^{+++} OR (95%) = 0.97 (0.77 – 1.23), p-value = 0.805], and *ADAM10*-rs593742 was the unique marker significantly associated with the VaD^{++} endophenotype [VaD^{++} OR (95%) = 0.80 (0.66 – 0.97), p-value = 0.02; AD^{+++} OR (95%) = 0.95 (0.87 – 1.05), p-value = 0.34].

The sub-analysis of Category C allowed distinguishing two sub-categories of variants residing in this cluster. The first subset comprised variants with a stable effect across all endo-phenotypes. The second subset included variants with inverse effects between extreme clinical subgroups. Vascular processes and the regulation of nervous system development were detected in top pathways in this second subset (Supplementary Table 6)

Biological interpretation of meta-GWAS signals

The combined results from four prioritization strategies, including positional mapping, eQTLs, chromatin and genome-wide association genes based analysis, showed that *ANKDR31* and *POLK* loci were the top ranked genes contained in the region of *ANKDRD31* signal. They were mapped by positional mapping, eQTLs, and gene-based analysis. Chromatin interaction did not show a significant association for regions containing markers reaching GWS neither in *ANKDRD31* nor in *POLK* (Supplementary Table 9A, 9D, 9G, 9J). Apart from FUMA analysis, our primary eQTL strategy showed a significant association for non-protein coding RNA, CTD-2235C13.3, further described in the main text. In the case of *NDUFAF6* genomic region, the top rank genes are *NDUFAF6* and *TP53INP1*, which are mapped by three out of four strategies, positional mapping, eQTLs, and gene-based

analysis (Supplementary Table 9B, 9E, 9H, 9K). Finally, in *SCIMP* region *SCIMP* locus was mapped using the four prioritization strategies, although *RABEP1* gene was also associated by these four, the chromatin interaction region did not contained GWS markers (Supplementary Table 9C, 9F, 9I, 9L).

Genome-wide association analysis for the results emerging from GRACE dementia with IGAP12 showed 133 significant genes associated with AD, 15 of them being new genomic regions, not previously associated with AD. In case of analysis of GRACE AD^{+++} with IGAP12, we detected 144 significant loci, and 3 new genomic regions, not previously associated with AD and not detected in the analysis from GRACE dementia with IGAP12. It should be noted that gene-based analysis was performed using summary statistics via FUMA. Thus, further analyses with raw genotypes are recommended to explore new regions. The association detected in the new region is driven by more than 5 markers in all cases.

REFERENCES

1. Blesa R, Pujol M, Aguilar M, Santacruz P, Bertran-Serra I, Hernández G, et al. Clinical validity of the "mini-mental state" for Spanish speaking communities. Neuropsychologia . 2001;39(11):1150–7.

2. Blessed G, Tomlinson BE, Roth M. The association between quantitative measures of dementia and of senile change in the cerebral grey matter of elderly subjects. Br J Psychiatry . 1968 Jul;114(512):797–811.

3. Peña-Casanova J, Aguilar M, Bertran-Serra I, Santacruz P, Hernández G, Insa R, et al. Normalization of cognitive and functional assessment instruments for dementia (NORMACODEM) (I): objectives, content and population. Neurologia . 1997 Feb;12(2):61–8.

4. Boada M, Cejudo JC, Tàrraga L, López OL, Kaufer D. [Neuropsychiatric inventory questionnaire (NPI-Q): Spanish validation of an abridged form of the Neuropsychiatric Inventory (NPI)]. Neurologia .;17(6):317–23.

5. Tinetti ME, Williams TF, Mayewski R. Fall risk index for elderly patients based on number of chronic disabilities. Am J Med . 1986 Mar;80(3):429–34.

6. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology . 1993 Nov;43(11):2412–4.

7. Reisberg B, Ferris SH, de Leon MJ, Crook T. The Global Deterioration Scale for assessment of primary degenerative dementia. Am J Psychiatry . 1982 Sep ;139(9):1136–9.

8. Hachinski VC, Iliff LD, Zilhka E, Du Boulay GH, McAllister VL, Marshall J, et al. Cerebral blood flow in dementia. Arch Neurol . 1975 Sep;32(9):632–7.

9. Alegret M, Espinosa A, Valero S, Vinyes-Junqué G, Ruiz A, Hernández I, et al. Cut-off Scores of a Brief Neuropsychological Battery (NBACE) for Spanish Individual Adults Older than 44 Years Old. Brucki S, editor. PLoS One . 2013 Oct 16;8(10):e76436.

10. Román GC, Tatemichi TK, Erkinjuntti T, Cummings JL, Masdeu JC, Garcia JH, et al. Vascular dementia: diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop. Neurology . 1993 Feb;43(2):250–60.

11. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet . 2007 Sep 9;81(3):559–75.

12. van Dam S, Cordeiro R, Craig T, van Dam J, Wood SH, de Magalhães J. GeneFriends: An online co-expression analysis tool to identify novel gene targets for aging and complex diseases. BMC Genomics . 2012 Oct 6;13(1):535.

13. Wang J, Duncan D, Shi Z, Zhang B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. Nucleic Acids Res . 2013 Jul 1 ;41(W1):W77–83.

14. Proitsi P, Lupton MK, Velayudhan L, Newhouse S, Fogh I, Tsolaki M, et al. Genetic Predisposition to Increased Blood Cholesterol and Triglyceride Lipid Levels and Risk of Alzheimer Disease: A Mendelian Randomization Analysis. PLoS Med. 2014;11(9).

15. Lourdusamy A, Newhouse S, Lunnon K, Proitsi P, Powell J, Hodges A, et al. Identification of cis-regulatory variation influencing protein abundance levels in human plasma. Hum Mol Genet. 2012;21(16):3719–26.

16. Naj AC, Jun G, Beecham GW, Wang L-S, Vardarajan BN, Buros J, et al. Common variants

at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nat Genet . 2011 May;43(5):436–41.

17. Filippini N, Rao A, Wetten S, Gibson RA, Borrie M, Guzman D, et al. Anatomically-distinct genetic associations of APOE e{open}4 allele load with regional cortical atrophy in Alzheimer's disease. Neuroimage. 2009;44(3):724–8.

18. Li H, Wetten S, Li L, St Jean PL, Upmanyu R, Surh L, et al. Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. Arch Neurol . 2008 Jan;65(1):45–53.

19. Carrasquillo MM, Zou F, Pankratz VS, Wilcox SL, Ma L, Walker LP, et al. Genetic variation in PCDH11X is associated with susceptibility to late-onset Alzheimer's disease. Nat Genet. 2009;41(2):192–8.

20. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. 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 Jul;34(7):939–44.

21. Gayán J, Galan JJ, González-Pérez A, Sáez ME, Martínez-Larrad MT, Zabena C, et al. Genetic structure of the Spanish population. BMC Genomics . 2010 Jan;11:326.

22. Antúnez C, Boada M, González-Pérez A, Gayán J, Ramírez-Lorca R, Marín J, et al. The membrane-spanning 4-domains, subfamily A (MS4A) gene cluster contains a common variant associated with Alzheimer's disease. Genome Med . 2011 Jan;3(5):33.

23. A. Bennett D, A. Schneider J, Arvanitakis Z, S. Wilson R. Overview and Findings from the Religious Orders Study. Curr Alzheimer Res. 2012;9(6):628–45.

24. Reiman EM, Webster JA, Myers AJ, Hardy J, Dunckley T, Zismann VL, et al. GAB2 Alleles Modify Alzheimer's Risk in APOE ε4 Carriers. Neuron. 2007;54(5):713–20.

25. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics . 2010 Sep 1;26(17):2190–1.

26. Yang J, Ferreira T, Morris AP, Medland SE, Genetic Investigation of ANthropometric Traits (GIANT) Consortium GI of AnT (GIANT), DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium DiaGRAM (DIAGRAM), et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet . 2012 Mar 18;44(4):369–75, S1-3.

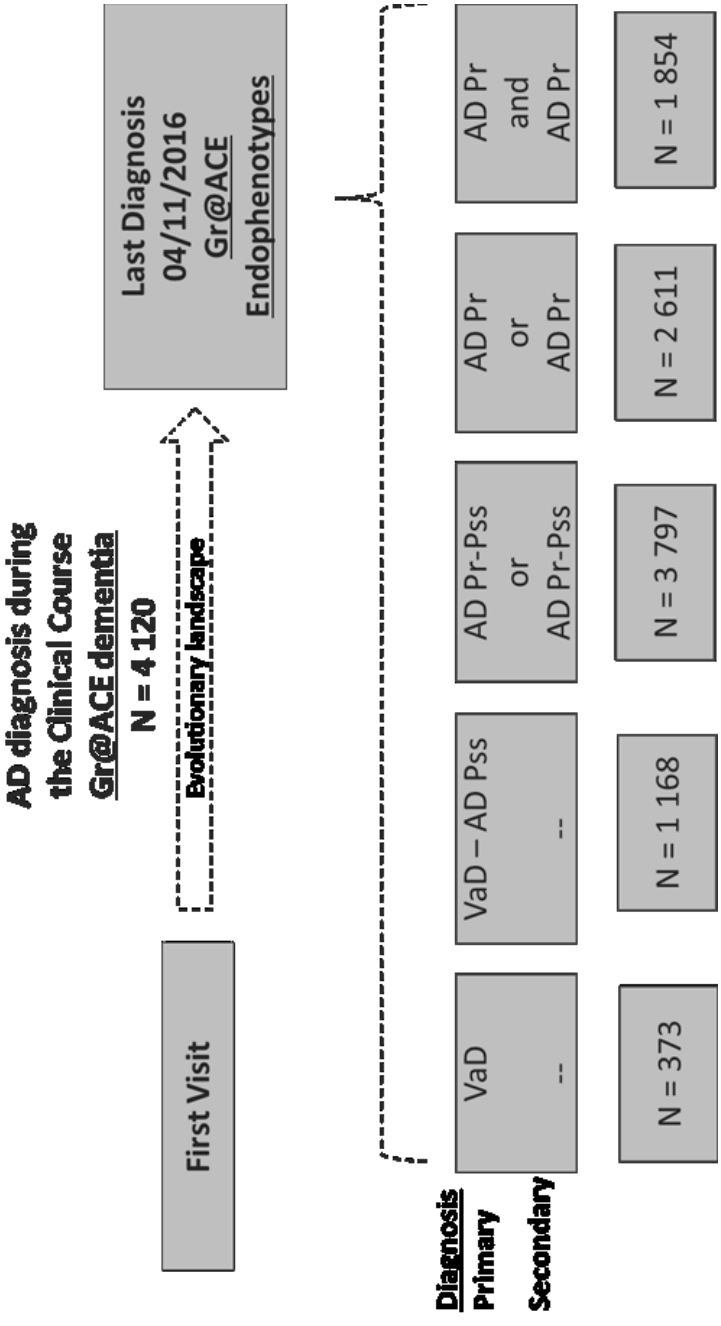
27. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet . 2015 Feb 2;47(3):291–5.

28. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics . 2010 Sep 15;26(18):2336–7.

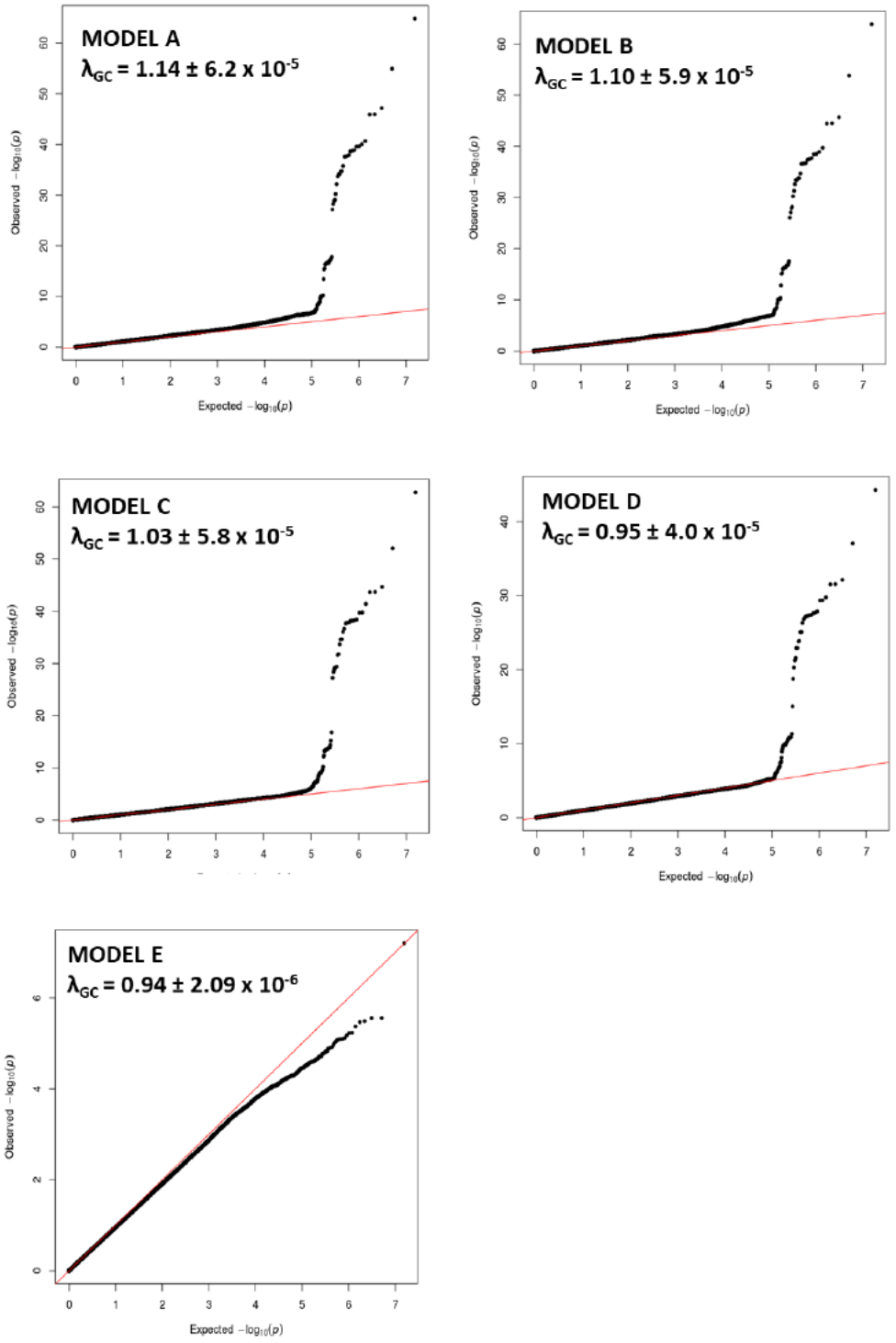
29. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun . 2017 Dec 28 ;8(1):1826.

30. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. Nat Genet . 2019 Jan 7;

SUPPLEMENTARY FIGURES

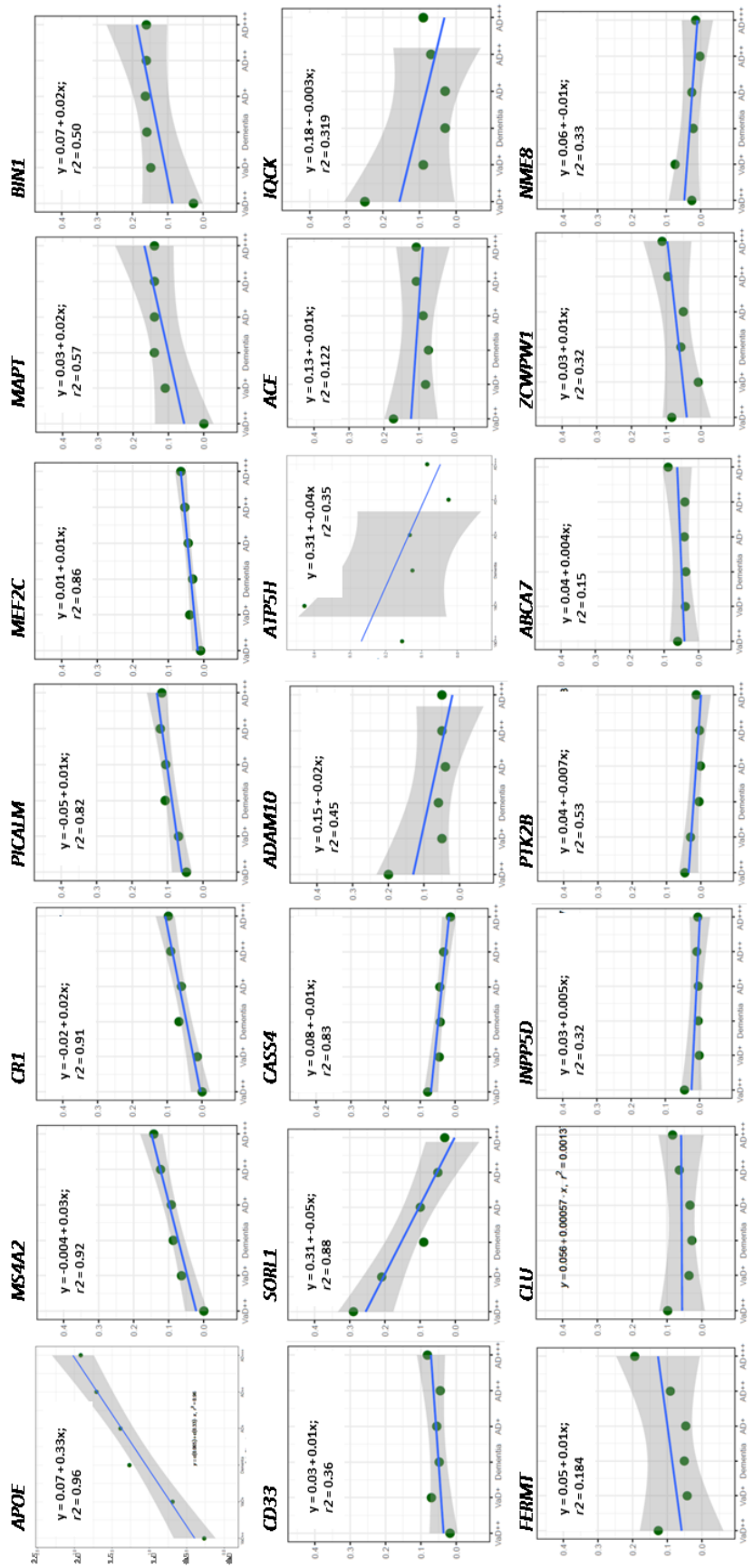


SUPPLEMENTARY FIGURE 1. Flow chart diagram for inclusion of AD patients and construction of GR@ACE clinical endophenotypes.

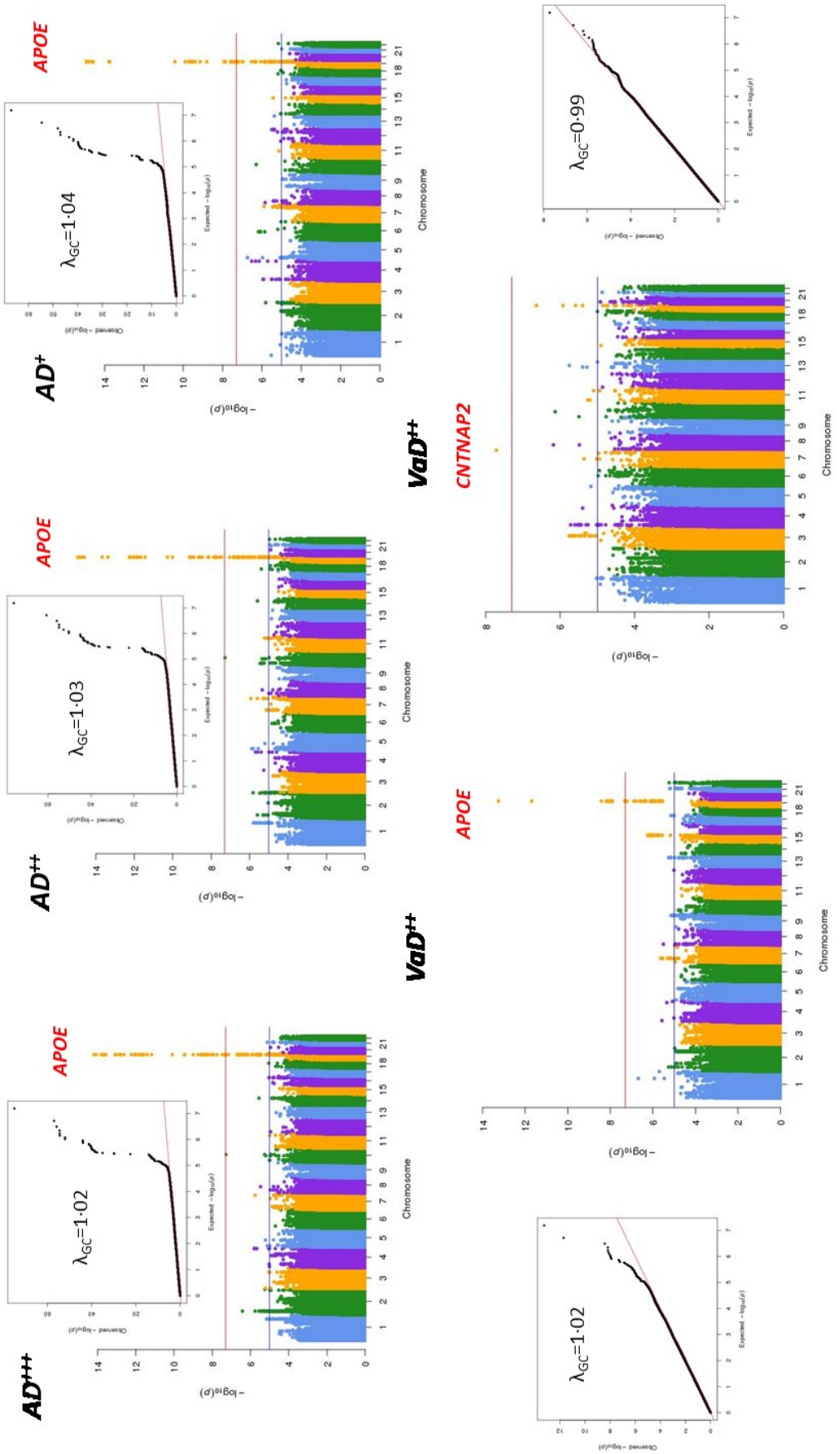


SUPPLEMENTARY FIGURE 2. Quantile-quantile plots across five regression models using the GR@ACE cohort.

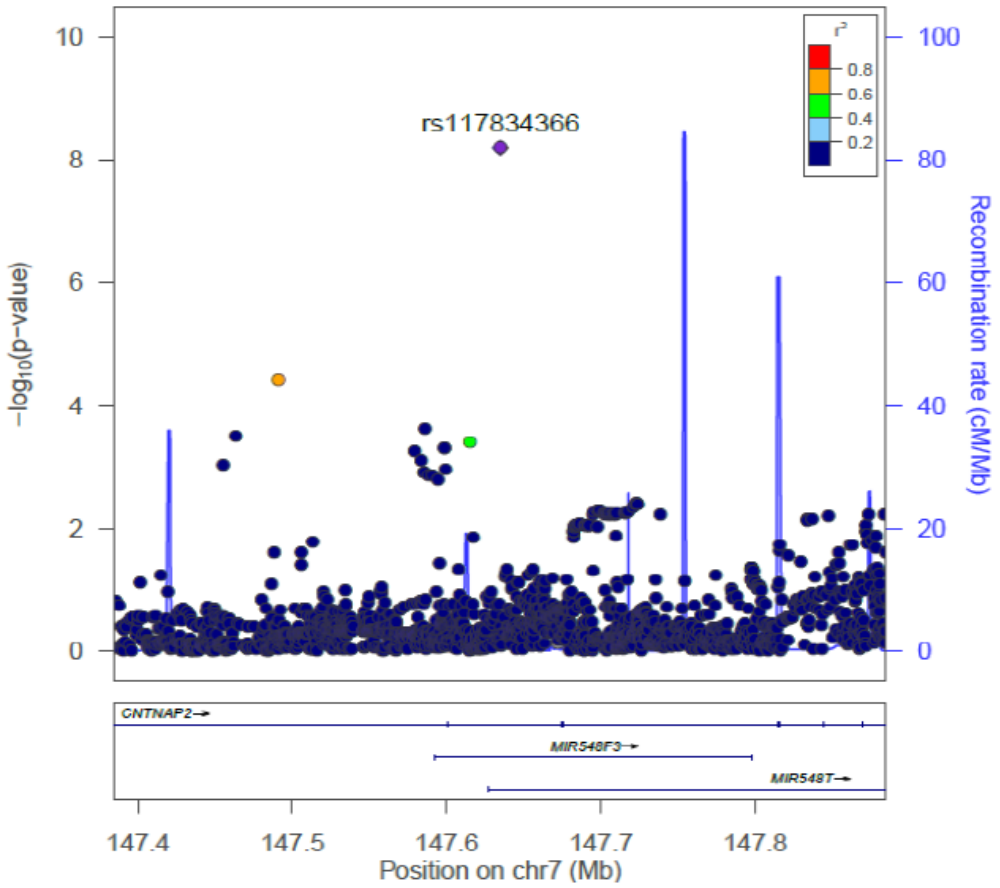
Model A, unadjusted; Model B, unadjusted and excluding all individuals matching with Hispanic American ancestry; Model C, adjusted per four main PCs; Model D, adjusted per top four PCs, gender and age; and Model E, adjusted per top four PCs, gender, age and *APOE* status



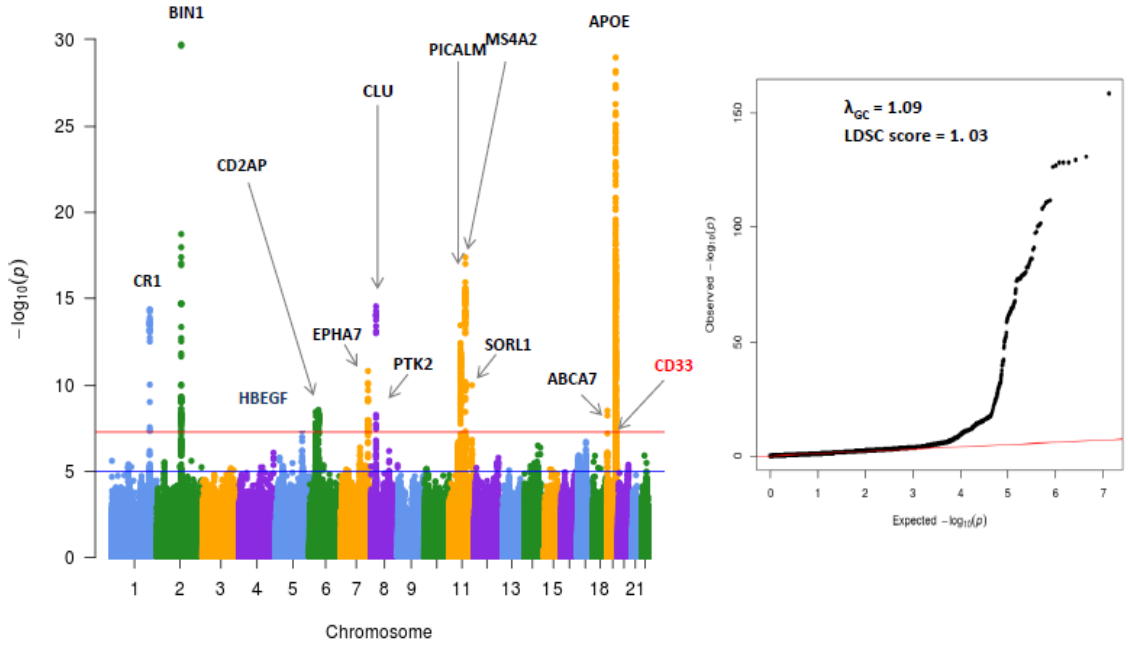
SUPPLEMENTARY FIGURE 4. Effect change for each individual LOAD genetic variant across GR@ACE clinical endophenotypes



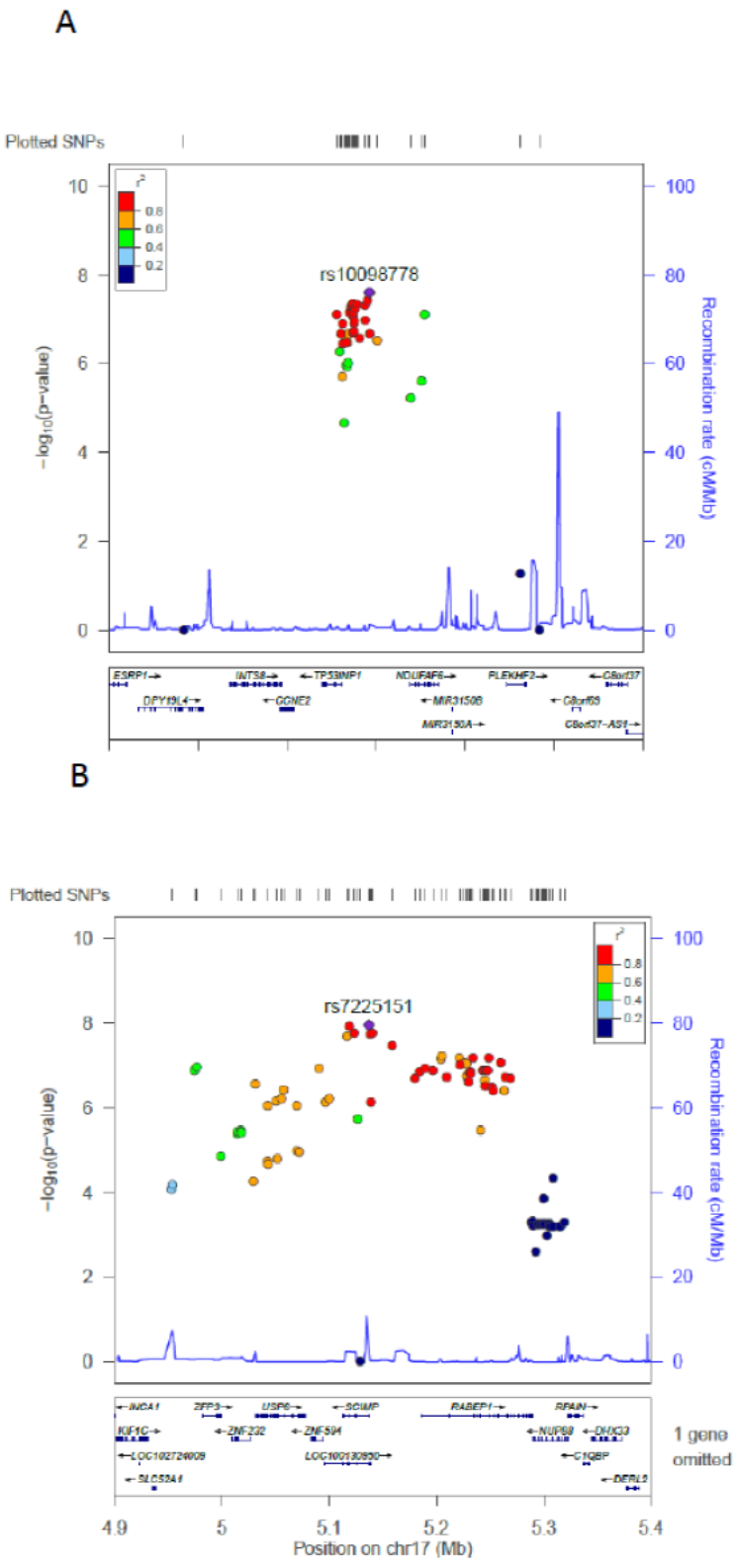
SUPPLEMENTARY FIGURE 5. Results of genome-wide association analysis for GR@ACE clinical endophenotypes amd quantile-quantile plot.s



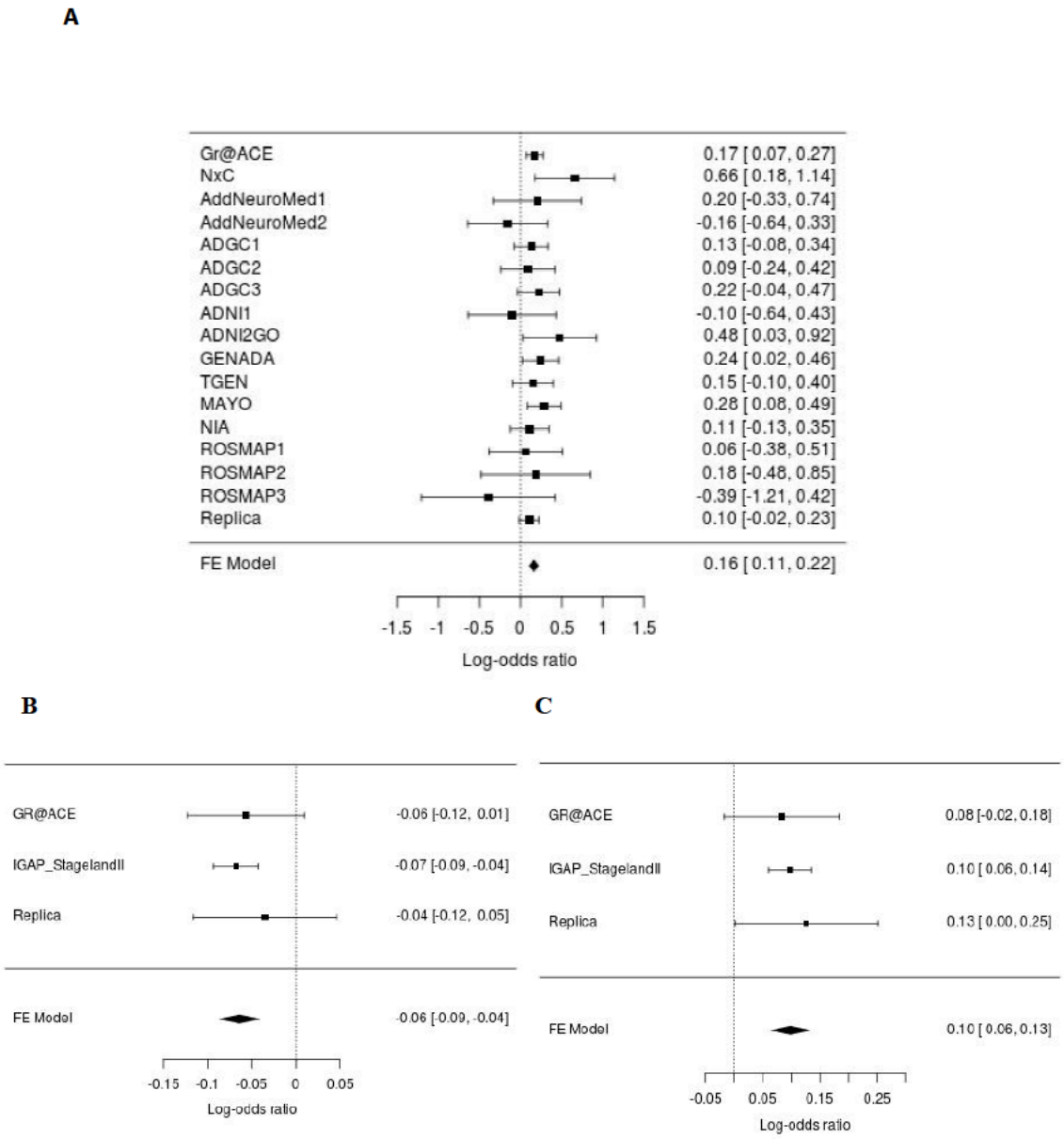
SUPPLEMENTARY FIGURE 6. Associations of the region centered on rs117834366 located in the *CNTNAP2* gene.



SUPPLEMENTARY FIGURE 7. Results of genome-wide association analysis for GR@ACE meta-analysis with IGAP Stage I. and QQplot



SUPPLEMENTARY FIGURE 8. Associations of the region centered on A) rs10098778 located in the *TP53INP1/NDUFAF6* gene; and B) on rs7225151 located in the *SCIMP* gene.



SUPPLEMENTARY FIGURE 9. Forest plot for the effect of meta genome-wide significant signals in LOAD. A) *ANKDR31*, rs4704171; B) *NDUFAF6*- rs10098778; and C) *SCIMP*-rs7225151

SUPPLEMENTARY FIGURE 3 AND SUPPLEMENTARY TABLES

Supplementary Figure 3 and Supplementary Tables can be accessed scanning the following code:

Supplementary Figure 3. Statistical power analysis. A) Analyses were performed for an experiment that would meet criteria for genome-wide significance ($p < 5 \times 10^{-8}$); and B) significance ($p < 0.05$) in GR@ACE datasets; C) Analyses were performed for an experiment that would meet criteria for genome-wide significance ($p < 5 \times 10^{-8}$) in meta-analysis.

Supplementary Table 1. Association results obtained from the GR@ACE dataset and GR@ACE endophenotypes, classified per gene categories for functional pathway analysis, and including global effect change per genetic variant.

Supplementary Table 2. Full description of dbGaP datasets used in the meta-analysis.

Supplementary Table 3. eQTL analysis for novel GWAS significant hits associated with AD

Supplementary Table 4. Association results for rs2732703-*KANSL1/MAPT* in *APOE* ε4 carriers and non-carriers across GR@ACE endophenotypes.

Supplementary Table 5. Replication results for rs7100488-*PCBD1/UNC5B* and rs117834366-*CNTNAP2* in additional datasets.

Supplementary Table 6. Top ten biological pathways per gene cluster after sub-analysis of gene Category C.

Supplementary Table 7. Top ten biological pathways per gene cluster after secondary gene categorization.

Supplementary Table 8. Association results for known LOAD loci in meta-analysis with dbGaP datasets and IGAP Stage I and Stage I+II.

Supplementary Table 9. Functional annotation of GWS variants, eQTL analysis, chromatin interaction and gene-based results performed via FUMA.

.

Section 3.4. Homozygosity mapping in Alzheimer's Disease

3.4.1. Publication V

Autosomal Recessive Alzheimer's disease (arAD): homozygosity mapping of genomic regions containing arAD loci.

Sonia Moreno-Grau.[1,2] and Agustín Ruiz [1,2]

[1]Research Center and Memory clinic Fundació ACE. Institut Català de Neurociències Aplicades. Universitat Internacional de Catalunya. Barcelona. Spain. [2]CIBERNED, Center for Networked Biomedical Research on Neurodegenerative Diseases, National Institute of Health Carlos III, Ministry of Economy and Competitiveness, Spain.

ABSTRACT

Long runs of homozygosity (ROH) are contiguous stretches of homozygous genotypes, which are a footprint of recent inbreeding and recessive inheritance. The effect of recessive-acting alleles is suggested for Alzheimer's disease (AD). However, the search for recessive variants has been poorly assessed to date. To investigate homozygosity in AD, we performed a fine-scale ROH analysis including 21,100 individuals from 10 cohorts of European ancestry (11,919 AD cases and 9,181 controls). We detected an increase of homozygosity in AD cases respect to controls [β_{FROH} (CI95%) = 0.051 (0.023 – 0.078); P = 3.25 x 10^{-4}]. None ROH was associated with AD after multiple corrections. Our top ROH was identified near to *HS3ST1* locus, previously related with AD. ROHs increasing the risk of AD were significantly overrepresented respect to ROHs increasing protection ($p < 2.20 \times 10^{-16}$). Next, to construct a homozygosity map of AD cases, we selected ROHs shared by genuine inbred AD cases extracted from an outbred population. We then used whole-exome sequencing data from 1,449 individuals from the Knight-ADRC-NIA-LOAD (KANL) cohort, to identify potential recessive variants in candidate ROHs. Candidate variants, neighbored genes previously associated with AD (rs116644203 in *ZNF282*, near to *CNTNAP2,* and rs73263258, a missense variant in *ESPR1*, near to *TP53INP1/NDUDFAF6*). We also identified a novel marker, rs117458494, mapped in *SPON1* locus which has been previously associated with amyloid metabolism. Here, we provide a research framework to look for recessive variants in AD using outbred populations. Our results showed that AD cases are enriched in homozygosity, suggesting that recessive effects may explain a proportion of the AD heritability.

Keywords: Alzheimer's disease, Runs of Homozygosity, recessive variants, inbreeding.

2019. [In preparation]

.

INTRODUCTION

Alzheimer's disease (AD) is a neurodegenerative disorder, which represents the leading cause of dementia worldwide (1). A small proportion of patients develop AD before 65 years old, known as Early Onset AD (EOAD), whereas in most persons, clinical symptoms begin after 65 years old, namely Late Onset AD (LOAD). AD presents a strong genetic component. In fact, heritability estimations for EOAD and LOAD fall in a range of 92 to 100%, and 13 to 73%, respectively (2),(3).

Specific autosomal dominant mutations have been linked to familial EOAD: mutations in *presenilin 1* (*PSEN1*) (4), *presenilin 2* (*PSEN2*) (5), and *amyloid precursor protein* (*APP)* (6). These findings were pivotal events pinpointing to the role of amyloid metabolism such a disease-causing mechanism (7). Despite that, dominant causes account for a minority of both, familial or apparently sporadic EOAD cases. It is suggested that autosomal recessive loci might cause most of EOAD cases (~90%) (2). But to date, only two homozygous mutations in *APP* gene (A673V and E693Δ) have been described (8),(9), that remain controversial.

The commonest AD phenotype, the sporadic form of LOAD, has a polygenic background. Genome-wide association studies (GWAS) and large sequencing projects have identified near to 40 genetic variants associated with LOAD risk (10), (11), (12), (13), yet all these discoveries together only explain a limited part of disease heritability (~31%) (14). Current genetic findings were discovered using an additive mode of inheritance, which overlooks the relevance of non-additive genetic components, i.e. the recessive model despite they could be explaining large part of disease heritability.

It is well-known that inbreeding increases the incidence of recessive diseases. The probability of detecting a recessive locus increases in offspring of consanguineous unions (15), because the partners share alleles inherited from a recent common ancestor. This recent parental relatedness points to genuine regions of autozygosity. Long runs of homozygosity (ROHs) – long stretches of consecutive homozygous genotypes (>1 Mb) – are a recognized signature of recessive inheritance. Thus far, they have been used for homozygosity mapping (16). Population history, e.g. historical bottleneck or geographical isolation, is also influencing homozygosity levels in individual genomes (17), (18).

To assess the role of recessive inheritance in AD, Farrer et al. (19) studied 183 families of the isolated Wadi Ara region (an area in Israel populated mainly by Arab citizens of Israel). The Wadi Ara population has increased parental relatedness and high prevalence of AD. Farrer et al. pointed to candidate regions

with potential recessively acting loci (19), (20). Using homozygosity mapping in a consanguineous EOAD family, and subsequence sequencing of candidate regions, Bras et al. suggested *CTFS* gene as a potential recessive locus (21), (22).

Recently, it has been demonstrated that ROHs are also found ubiquitously even in outbred populations (23), (24). An excess of homozygosity has been associated with AD in individuals from Caribbean-Hispanic and African-American ancestries (25), (26), (27). It suggests the presence of inbreed and potentially autosomal recessive AD (arAD) cases nested in these populations. Conversely, this association presented controversial results for individuals from European ancestry (28), (29). Several reasons might explain inconsistencies. First, ROHs patterns are highly divergent between populations. In that sense, specific recent bottlenecks, as well as the presence of cultural practices promoting endogamous marriages, in Latino groups could be increasing inbreeding at these populations (30), (31); and consequently ROHs estimations. Second, it has been estimated that large samples sizes (12,000 – 65,000) are required to detect an excess of homozygosity in outbreed populations (32). Thus, previous studies might be underpowered.

Assessing the impact of inbreeding in the genetic architecture of AD remains a challenge. The limited number of deeply characterized consanguineous families, the difficulties to find familial information for sporadic AD individuals, mainly due to the late onset of the disease, and the reduced size of intragenerational pedigrees in western countries make intricate the search for arAD loci. Furthermore, follow-up of candidate ROHs in sequencing data might be a necessary step to definitively map a recessive locus, but it has been poorly assessed to date. Considering limitations, we think that capturing the fraction of consanguineous individuals nested in AD cases from outbred population could be an efficient strategy to prioritize homozygous regions potentially harboring recessive loci.

To the best of our knowledge, this is the largest genomic data set exploring the influence of homozygosity in AD (n = 21,100). First, we investigated whether AD individuals from a European outbred population presented an excess of homozygosity respect to controls. Next, we delineated the scale of inbreeding in AD cases. To prioritize regions with potential recessive loci, we constructed a homozygosity map of genomic regions overrepresented in inbred AD cases. Finally, further exploration of several promising candidate ROHs was performed in whole exome sequencing (WES) data.

SUBJECTS AND METHODS

The overview of the proposed strategy for ROH detection and subsequent prioritization is depicted in Figure 1.

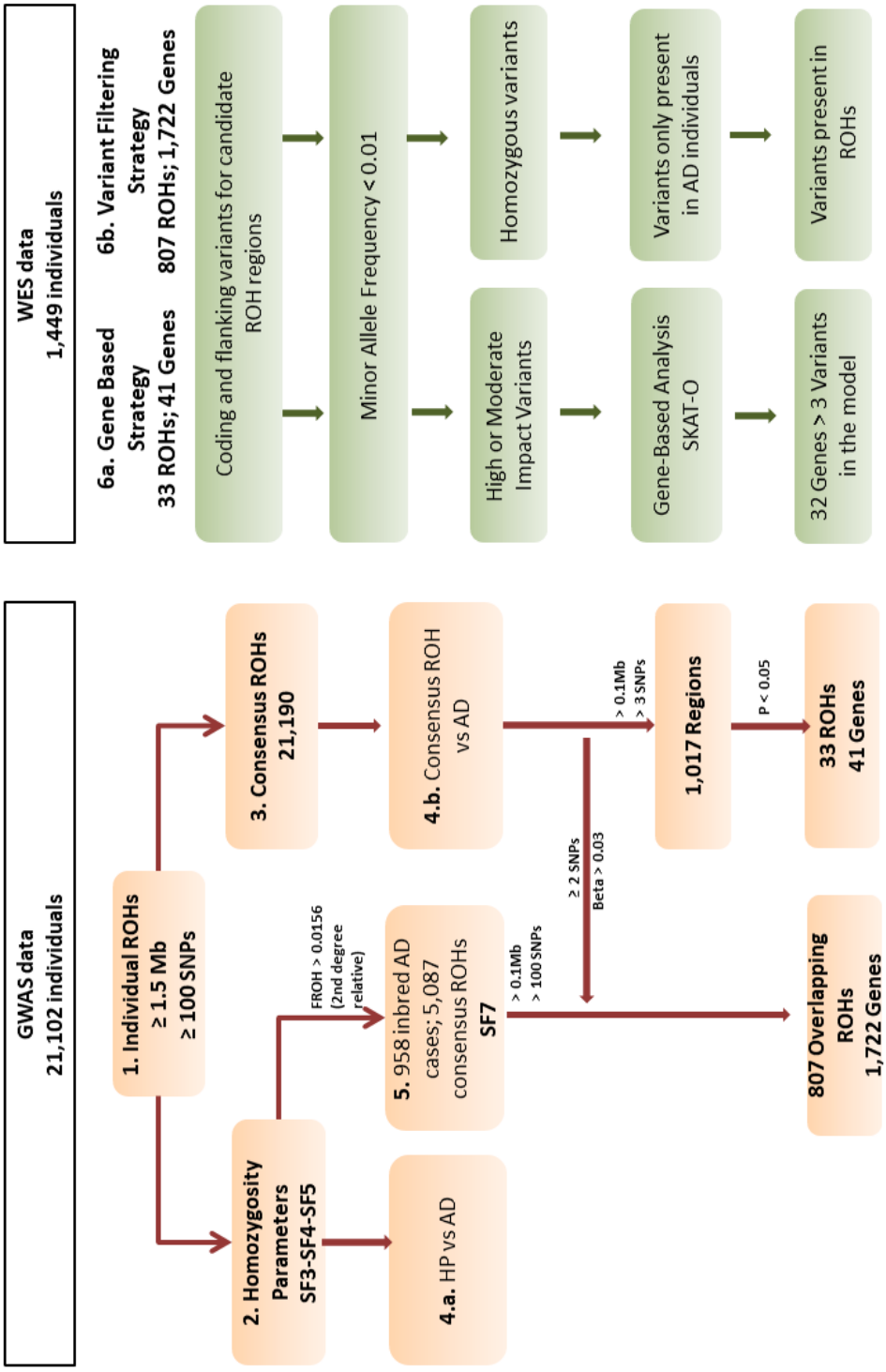


FIGURE 1. Schematic of the stepwise for ROH prioritization
SF; Supplementery Figure.; HP: Homozygosity parameters.

Genotyping data

This study includes 10 independent genome wide data sets comprising, a total sample size of 21,100 unrelated individuals (11,921 AD cases and 9,181 individual controls) from European ancestry (Supplementary Table 1). Further information on the recruitment, phenotyping and quality control for genome-wide data has been previously described (13).

Briefly, genotype-level data for each cohort was processed by applying identical quality control and imputation procedures. Individuals were excluded for low quality sample (call rate <97%), excess heterozygosity, sample duplicates, or if they were related to another sample (PIHAT > 0.1875). Individuals were excluded if sex discrepancy was detected. Population outliers from European ancestry were also removed. Variants were excluded if departures from Hardy-Weinberg equilibrium (P-value $\leq 1 \times 10^{-4}$), presented different missing rate between cases and controls (P-value $< 5 \times 10^{-4}$ for the difference), low frequency (MAF < 0.01) or low call rate < 95%. High quality variants were imputed in Michigan Server using the Haplotype reference consortium (HRC) panel (https://imputationserver.sph.umich.edu). Only common markers (MAF > 0.05) with a high imputation quality (R^2>0·90) were used for downstream analysis.

Next, we generated a merged dataset combining imputed genotypes of high quality from available data sets. We calculated identity-by-descendent (IBD) with PLINK 1.9 to generate a cohort of unrelated individuals. All possible pairs had Pi-hat < 0.1875, Z0 ≥ 0.75 and Z1 ≤ 0.25. Imputed markers with call rate > 0.95 and MAF > 0.05 in the merged dataset were selected for ROH calling (N_{SNPs} = 2,678,325).

Whole Exome Sequence (WES) data

With the objective to explore most promising ROHs candidates in sequencing data, we used the Knight-ADRC-NIA-LOAD (KANL) cohort (33). We excluded autosomal dominant familial cases and sporadic AD cases harboring well-known disease-causing mutations, as they could explain disease status. Thus, this study comprised 986 AD cases and 463 control individuals from European ancestry (See Supplementary Table 1 and Supplementary Figure 1). Of them, 488 subjects presented both, GWAS and WES data available for this study. Detailed description of cohort characteristics and quality control for WES data have been previously reported (33).

Briefly, exome libraries were prepared using Agilent's SureSelect Human All Exon kits V3 and V5 or Roche VCRome. WES samples were sequenced on a HiSeq2000 with paired ends reads, with a mean depth of coverage of 50x to 150x for WES and 30x for WGS. Fastq sequences

were aligned to the GRCh37.p13 genome reference. Variant calling was performed following GATKv.3.6 Best Practices (https://software.broadinstitute.org/gatk/) and restricted to a 100bp of padding capture region. Variants and indels within the 99.9% of the VQSR confidence interval were included in the analysis, as well as variants with allele-balance between 0.30–0.70, quality depth ≥5 for indels and ≥ 2 for SNPs and call rate > 95% (33).

1. Identification of individual ROHs

Individual ROH calling was conducted using the observational genotype-counting approach implemented in PLINK (v1.09) (https://www.cog-genomics.org/plink/1.9/), as it outperforms additional methods in ROH detection (34). ROH detection was performed for each individual study and for the merged dataset using imputed genotypes. Since included datasets were genotyped using different genotyping arrays, they shared a small fraction of directly genotyped markers. Given that it is demonstrated that lower SNP density can impact the accuracy of ROH analysis (35), we decided to use imputed genotypes of high quality to increase SNP coverage. To scan the genome we used a sliding window of 50 SNPs in 5000 Kb of length. One heterozygote and five missing calls were tolerated per window in order to manage genomic regions with small number of genotyping errors and discrete missingness. These parameters were similar to those previously described (36). The minimal number of SNPs in a ROH was set to 100 SNPs (37),(38). We empirically explored two minimal length cut-offs to consider a ROH, 1 Mb and 1.5 Mb. It is suggested that ROHs > 1Mb prevents the detection of short homozygosity stretches, which are generated by linkage disequilibrium forces in the human genome according to empirical studies (39), (40), (41). However, the ability to detect autozygous regions with ROH length set to 1Mb could be compromised. Inbreeding estimations resulting from individual ROH ≥ 1.5Mb has been most strongly correlated with inbreeding estimated from pedigree information (23), but this threshold has never been applied to AD studies. Autosomal SNPs were included in a ROH if >5% of the sliding window was homozygous, it means that at least 3 SNPs in 250 Kb from the sliding window are required to include a new marker. The maximum distance between two consecutive SNPs was set to 1000 Kb apart, and SNP density to at least 1 SNP in 50Kb.

2. Exploration of homozygosity parameters

To assess data quality and genetic architecture of detected ROH (> 1 Mb and > 1.5 Mb) in each individual study and in the whole dataset, we calculated: a) the mean of the total length of ROH, or sum of ROH (SROH), b) the average ROH length (AVROH), c) the number of ROHs (NROH), and d) ROH-based estimates of

the inbreeding coefficient, F, (FROH) per individual. AVROH is the SROH divided by NROH per subject. FROH represents the proportion of homozygous segments in the autosomal genome per individual (Equation 1), i.e. for individuals would be the SROH detected divided by a factor of 3,020,190 Kb, the total autosomal genome length according to GRCh37.p13 assembly. We further explored whether the effect of homozygosity parameters were similar when: 1) ROH length was set to 1 Mb or 1.5 Mb; and 2) the analysis was performed per dataset or in the final merged database (Supplementary Figure 2). Supplementary Table 2 and Supplementary Figure 3 demonstrated that FROH estimates derived from ROH calling at 1Mb exhibited large inflation, not allowing an accurate detection of inbreeding (Mean FROH 1Mb = 0.028; Mean FROH 1.5Mb = 0.011), which is in accordance with prior studies (23). After conducting the analysis with the 2,678,325 SNPs shared between available datasets, parameters from individual datasets and the merged dataset analyses were similar (Supplementary Figure 2 and Supplementary Table 3). After these exploratory analyses, we decided to conduct downstream analyses with ROH calling at 1.5 Mb in the merged data.

Equation 1. $$F_{ROH} = \frac{SROH\ (Kb)}{Autosomal\ Genome\ (Kb)}$$

Copy number variants (CNV), particularly hemizygous deletions, are known to cause spurious ROHs. However, prior studies have demonstrated that the impact of performing ROH calling with or without CNVs is only 0.3% in the total ROH length (23), making it highly unlikely that deletions called as ROH influence findings. To assess the impact of CNVs, specifically deletions, in our study, we also conducted ROH calling after removing common CNVs deletions , extracted from the Database of Genomic Variants (DGV) (http://dgv.tcag.ca/) (42).

3. Identification of consensus ROHs

Consensus ROHs were defined as overlapping segments between individual ROHs observed in different genomes. A consensus ROH needs a DNA segment match of at least 95% for non-missing SNP markers. Consensus ROH calling was performed using PLINK 1.9 in the merged dataset. We then extracted those consensus

ROHs with a DNA length over 100 Kb and more than 3 consecutive SNPs. These criteria were applied to prevent the detection of false positive findings.

4. Analyses

4a. Association analysis between homozygosity parameters and AD risk

To assess the quality of the data in each individual study we explored sample distribution for each of four homozygosity parameters: NROH, SROH, AVROH and FROH. Exploratory analysis was depicted with violin plots, which combine a box plot with kernel density plot, using ggplot2 package from R (Supplementary Figure 4 and 5). Inverse rank normal transformation was performed to general homozygosity parameters using "rankNorm" option of RNOmni package in R. Transformed distribution are shown in Supplementary Figure 6. To test the association of homozygosity parameters with AD status, we conducted a generalized linear model for a binominal outcome using R, for individual level data. To account with potential heterogeneity between individual studies we adjusted the model per cohort and the first four principal components (PCs) resulting from ancestry analysis. See Equation 2. Sensitivity analysis was conducted to explore the impact of age in homozygosity parameters (Supplementary Table 6).

Equation2. $Z = \beta_1$ Homozygosity Parameter $+\beta_2$ Cohort $+\beta_3$ PC1 $+\beta_4$ PC2 $+\beta_5$ PC3 $+\beta_5$ PC4 $+ e$

4b. Association analysis between consensus ROHs and AD

The association between the phenotype and all consensus ROHs was explored using a logistic model. The model was adjusted per cohort and PCs as covariates for downstream analysis.. Nonetheless, covariate models adjusted per age and gender, in addition to cohort and PCs, were also calculated. Regression-based results were corrected for multiple testing using Bonferroni correction.

Next, we sought to estimate whether there was an overrepresentation of risk ($\beta > 0$) or protective ($\beta < 0$) consensus ROHs in our association results at different levels of ROH length and SNP number. We applied a binominal test using R. We considered that under the null hypothesis of no association similar distribution would be expected for both (50/50).

5. The homozygosity map of inbred AD individuals

5a. Detection of inbred individuals

To detect the subset of inbreed individuals within the whole dataset, we used FROH estimates. This parameter has been previously shown to better correlate

with the unobserved pedigree inbreeding (32),(43). The cut-off between inbreed and non-inbreed individuals was set to F_{ROH} > 0.0156 (35). This cut-off corresponds to second-degree relative, i.e. the mean inbreeding coefficient for kinship of second cousin marriage, or closer. It is assumed that there is not different biological effects below 0.0156 respect to general population (44). We demonstrated the efficient capture of inbreed individuals in Supplementary Figure 7 which shows the inverse relationship between ROH length and ROH age. Thus, short ROHs evidence ancient origin, and long ROHs more recent origin, which might indicate ROHs emerging from consanguineous mating. Next, we explore whether the frequency of consanguinity was higher in AD cases respect to controls, we calculated odds ratio and chi square P values using epitools package from R.

5b. ROHs prioritization based on inbred AD cases

ROH detection was conducted in the subset of inbred AD cases, applying similar criteria to that previously described for the outbred population. Briefly, considering the long size of homozygous tracts for inbred individuals, there is higher probability to find a consensus ROH within consaguineous AD cases than in the general population by chance. Hence, we applied stringent criteria to define consensus ROHs within inbred AD cases. Consensus ROHs from inbred AD cases with ROH length > 100 Kb and ROH > 100 SNPs were given priority for further analysis. Shared overlapping regions between consanguineous AD cases and the whole dataset were also identified (See bash code in Supplementary Material), and selected considering their overrepresentation in cases respect to controls (β > 0.03). Prioritized regions were explored in sequencing data.

6. Candidate gene prioritization strategies using WES

6a. Gene based analysis

To prioritize genes in consensus ROHs regions, we performed a gene-based analysis (986 cases vs 463 controls) (Figure 1). To generate SNP sets, variants were filtered out according to minor allele frequency (MAF<0.01) and functional impact. The allele frequency cut-off was established according the Exome Aggregation Consortium (ExAC) non-Finnish European, Exome Sequencing project (ESP) and 1000G. Only those variants predicted to have a high or moderate effect according to SnpEff were included (45). To compute p-values per gene set, SKAT-O model were applied using R. The models were adjusted to consider the impact of the first two PCs and sex. Genes were filtered out from results if the number of SNPs included in the model was below or equal to 3.

6b. Variant filtering strategy for inbred AD cases

ROH segments emerging from consanguineous AD cases are the most promising candidates to harbor autosomal recessive variants. Therefore, we deeply explored ROHs by applying an alternative strategy, based on variant filtering. To identify candidate genes and/or mutations associated with AD, we explored 488 individuals with complementary GWAS and WES data in the present study. Because there is a low likelihood to identify any novel or causative mutation in available databases, variants with MAF > 0.01 in the Exome Aggregation Consortium (ExAC) non-Finnish European, Exome Sequencing project (ESP) and 1000G were excluded. All heterozygous variants were removed. Finally, only the variants mapped in individual ROHs were selected.

Biological significance of ROH findings

To map genes within ROHs, first we extracted all the SNPs located in ROHs region. Next, we annotated individually each SNP of them

RESULTS

RoH parameters are associated with Alzheimer disease risk.

We examined the general characteristic of the four ROH parameters (SROH, NROH, AVROH, and FROH) in 21,100 unrelated European individuals of 10 independent cohorts (Supplementary Table 1). Data distributions in each individual dataset and in the joint analysis are showed in Supplementary Table 2 and Supplementary Figure 4. All datasets exhibited similar estimations in homozygosity measures. Relationship between the mean NROH and SROH are show in Figure 2. Within the merged dataset the mean NROH was 14.6 ± 4.6, the AVROH was 2.11 ± 0.61Mb and the SROH was 31.9 ± 22.2Mb. These estimations are in accordance with that observed in European individuals (35), except for NROH parameter, which resulted higher compared to previous studies (35).

Next, we tested the association of the four parameters between AD cases and control subjects. We found that i) a larger homozygosity fraction of the genome (F_{ROH}) increased the risk to suffer AD [β_{FROH} (CI95%) = 0.051 (0.023 – 0.078); P = 3.25 x 10^{-4}] (Table 1) ii) AD individuals presented more ROH segments compared to controls [β_{FROH} (CI95%) = 0.043 (0.015 – 0.071); P = 2.48 x 10^{-3}], iii) increased average lengths of ROHs in AD cases compared with controls [β_{FROH} (CI95%) = 0.027 (0.000 – 0.055); P = 0.051] (Table 1). Results per each individual cohort are shown in Supplementary Table 4. Of note, sensitivity analysis conducted excluding known deletions, i.e. hemizygous segments (42), provided similar results (Supplementary Table 5).

TABLE 1. Effect of genome-wide homozygosity measures in Alzheimer Disease for the joint analysis

Dataset	Unadjusted		Adjusted Cohort		Adjusted Cohort, PCs	
	Beta (CI95%)	P value	Beta (CI95%)	P value	Beta (CI95%)	P value
FROH	0.059 (0.032 - 0.087)	2.02×10^{-5}	0.064 (0.036 - 0.091)	4.96×10^{-6}	0.051 (0.023 - 0.078)	3.25×10^{-4}
AVROH	0.014 (-0.001 - 0.053)	0.060	0.030 (0.003 - 0.057)	0.032	0.027 (0.000 - 0.055)	0.051
NROH	0.053 (0.027 - 0.081)	1.11×10^{-4}	0.058 (0.030 - 0.086)	3.62×10^{-5}	0.043 (0.015 - 0.071)	2.48×10^{-3}

Results for association of excess of homozygosity (FROH), average ROH lenght (AVROH), and number of ROH (NROH) with Alzheimer disease status.
OR, Odds Ratio; with 95% confindence interval (CI95%) and level of statistical significance (P value)

We also detected correlation between age and homozygosity measures in the control group ranging from 50 to 80 years old. Specifically, FROH and NROH exhibited a significant positive correlation, conversely, AVROH showed a significant inverse correlation (Supplementary Table 6). Having these findings, we decided to test the impact of acquired clonal mosaicism introduced by aging in homozygosity estimations. First, we conducted a sensitivity analysis controlling per cohort, PCs and age. The effect of FROH remains significant and stable after adjustments. The average length of ROH also remained significantly different between cases and controls [β_{AVROH} (CI95%) = 0.074 (0.040 - 0.106); $p = 2.16 \times 10^{-5}$)]. Interestingly, the number of ROH was largely age-dependent [(β_{NROH} (CI95%) = 0.010 (-0.024 - 0.044); p = 0.559]. These findings are supporting the notion that genomic somatic instability increases with age and can distort pervasively the gene-dosage of multiple loci (Supplementary Table 6).

ROH analysis of AD risk using the whole dataset

We identified 21,190 consensus ROHs in the merged dataset (N = 21,100). We then tested the association of each consensus ROH with AD status. We identified 11,974 risk associations ($\beta > 0$) and 9,216 protective associations ($\beta < 0$). Overall, we observed a highly significant over-representation of ROHs increasing the risk to suffer AD ($p < 2.20 \times 10^{-16}$) (Table 2). The same over-representation of risk associations was detected after filtering at several levels (Table 2). When the test was conducted with results adjusted per cohort,

TABLE 2. Frequency of Consensus ROHs with a Potentially Risk or Protective effect in Alzheimer's disease

	N ROH	Risk associations	Protective associations	P value	Probability of Success
Whole dataset	21190	11974	9216	$< 2.2 \times 10^{-16}$	0.56
Category A	1017	593	424	$< 2.2 \times 10^{-16}$	0.58
Category B	926	537	389	1.30×10^{-6}	0.57
Category C	858	499	359	1.98×10^{-6}	0.58
Category D	42	33	9	2.7×10^{-4}	0.79
Whole dataset / Map of Inbreed AD Cases	6636	3969	2667	$< 2.2 \times 10^{-16}$	0.60

Strategy A, ROHs > 100 kb; > 3 SNPs; Strategy B, ROHs > 100 kb; > 25 SNPs; Strategy C, ROHs > 100 kb; > 50 SNPs; Strategy D, ROHs > 100 kb; > 3 SNPs, P< 0.05

PCs, age and gender, the over-representations of risk associations still remained significant ($p < 2.20 \times 10^{-16}$).

To prevent the detection of false positive associations, we selected consensus ROHs with ≥100 Kb and ≥3 SNPs, which provided a subset of 1,017 consensus ROHs (Figure 1 and Supplementary Table 7). After correction of multiple testing (Bonferroni correction of $p = 4.92 \times 10^{-5}$), the most significantly associated ROH was detected in 57 individuals (45 AD cases vs 12 controls, β (CI95%) = 1.09 (0.48 – 1.48), $p = 9.03 \times 10^{-4}$). It expanded 115.9Kb in an intergenic region (chr4:11,189,482–11,305,456) near to *HS3ST1* locus. This region survived to age and gender adjustments (Supplementary Table 7). Importantly, this region has been previously associated with AD (46) but the recessive model has never been tested.

Using associated ROH as a reference, we explored the genes located in significant risk consensus ROHs ($p<0.05$) also in WES data (Figure 1). A total of 33 ROHs, comprising 41 genes were analyzed. Of them, 32 genes included >3 SNPs in the model (32 genes; Bonferroni correction p = 0.0015). *NECAB1* locus (chr8:91,803,921-91,971,630) presented the most significant signal in all the models tested (p = 0.01) (Supplementary Table 8) but it never reached statistical significance after multiple test correction.

Homozygosity mapping of AD using DNA segments identified in inbred cases

We detected 1,621 individuals, presenting a $F_{ROH} \geq 0.0156$, from the total sample (N =21,100) (Figure 2) (Supplementary Table 9). Interestingly, inbreeding over second degree of consanguinity was associated with higher risk to suffer AD [OR (95%, CI) = 1.12 (1.01 – 1.25); p value = 0.027), which is in line with our previous results. This supports the idea that AD consanguineous cases are

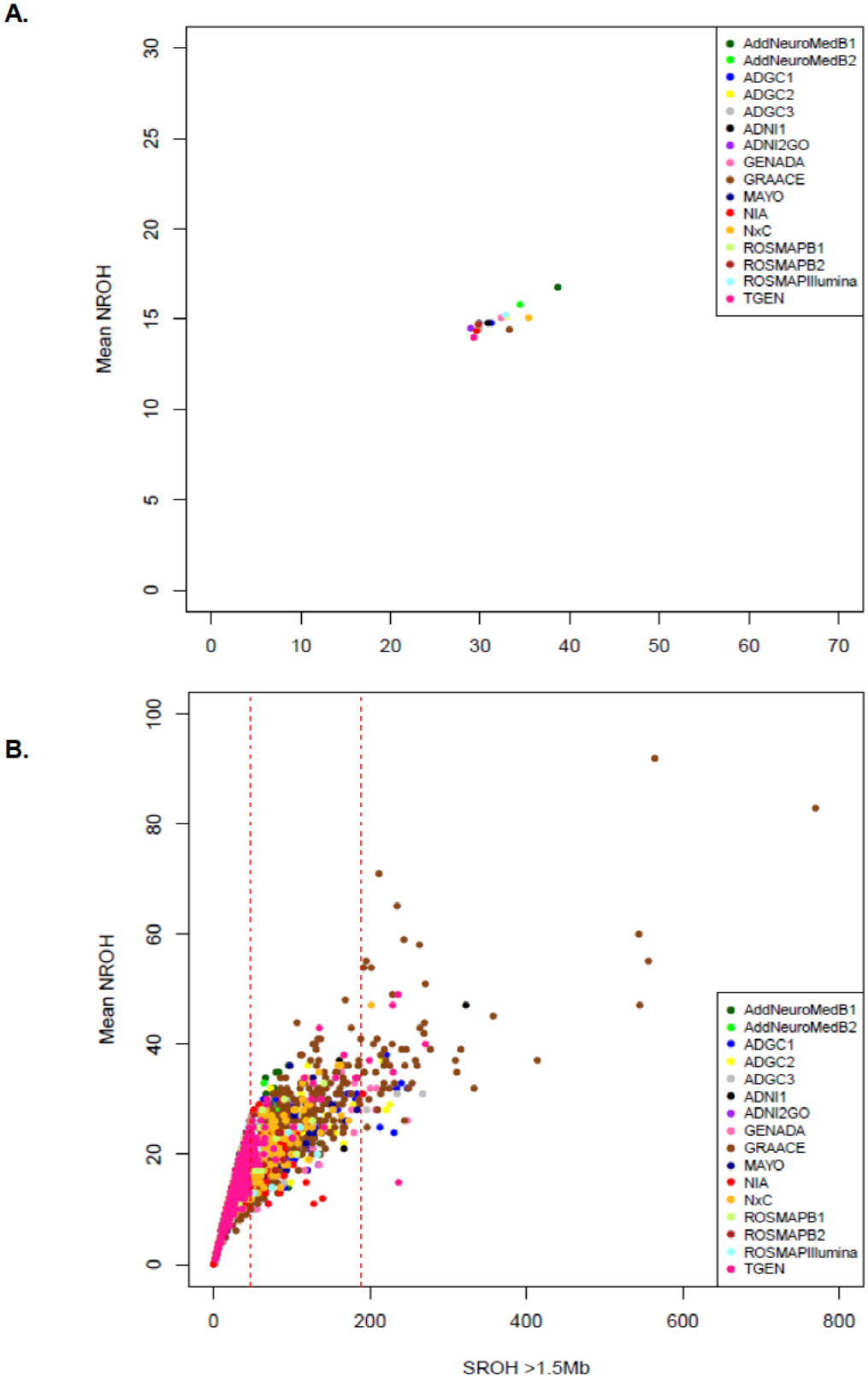


FIGURE 2. Runs of homozygosity per cohort and per individual. A) Mean number of ROHs versus mean total sum of ROHs in Mb for the 10 cohorts explored. B) Mean number of ROHs versus mean total sum of ROHs in Mb per individual explored.

Red dashed lines represent the threshold for inbreeding coefficient of 0.0156 (second cousins offspring) and 0.0625 (first cousins offspring).

overrepresented in the general AD population. Accordingly, the search for recessive loci having a role in AD can be first assessed in consanguineous cases.

After ROHs calling in inbred AD cases, we detected 5,087 pools of overlapping ROHs. Of them, we extracted consensus ROHs with ≥100 Kb and ≥100 SNPs. Then, we selected those ROHs overlapping with the whole sample, and over-represented in cases respect to controls (Figure 1). We prioritized 807 consensus homozygous segments from inbred cases (Figure 3 and Supplementary Table 10). It represented the 8.6% of the total autosomal genome and comprised 1,722 genes (Supplementary Table 11). Of them, we explored 1,136 genes with available WES data using a gene-based approach. None of them remained associated after multiple corrections ($N_{genes\ tested}$ = 1,136; p = 3.47 x 10^{-5}), our top signal was detected for *FRY* locus (p = 0.001) (Supplementary Table 11).

Considering that recessive variants are expected at low frequencies, even gene based analysis would be underpowered to detect gene candidates. Therefore, we decided to further prioritize loci by searching homozygous mutations within selected consensus ROHs from inbred AD subjects (Figure 1).We identified seven AD cases that had eight new (or extremely rare) homozygous variants in long ROH segments (Table 3). Of them, two individuals were consanguineous (F_{ROH} > 0.156). One had a potential causative missense variant (rs140790046) within the *MKX* locus, and another carried a novel variant (rs116644203) in *ZNF282* locus, which was located in an extremely large region of homozygosity (14.9 Mb) (Table 3). Furthermore, three additional homozygous variants were detected: a novel variant (rs117458494) in *SPON1* locus, previously related with amyloid metabolism (47), and two potential causative variants, carried by only the same individual, within a previously identified AD region (*TP53INP/NDUFAF6)* (13). Of them one (rs73263258) is a missense variant (c.475G>A) which encodes p.Ala159Thr change (Table 3). Further annotation and functional effect predictions for these variants are provided in Supplementary Table 12.

DISCUSSION

This study represents the largest analysis of homozygosity (ROHs) conducted in Alzheimer's disease (AD). Our estimate of excess of homozygosity in individuals with AD from European population provides firm evidence for a role of consanguinity in AD. This finding predicts that there might be a number of recessive AD loci explaining genomic variability. This statement has several implication for the design of AD genetic studies, for the better understanding of the causes of phenotypic variation in AD and

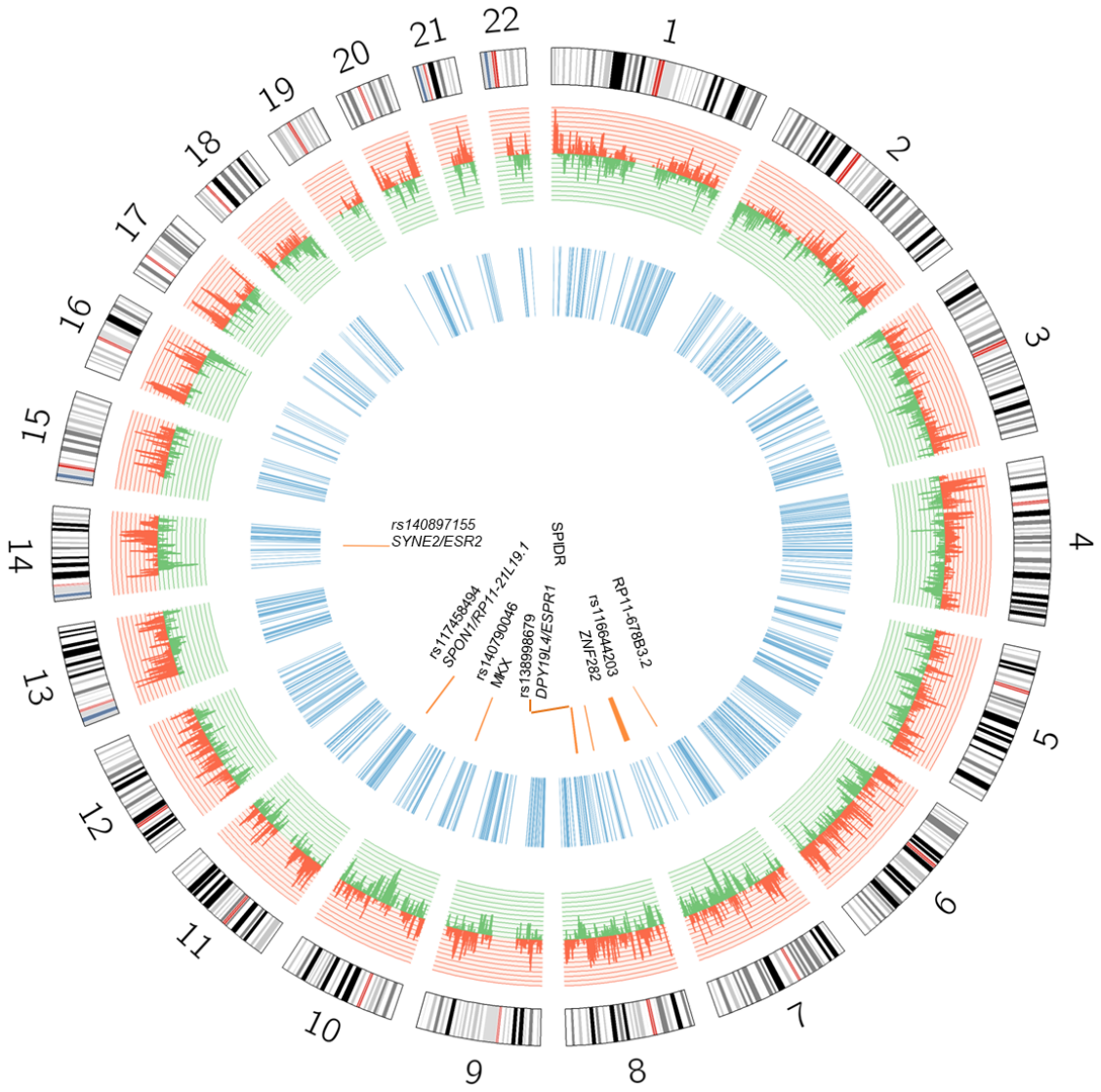


FIGURE 3. Circos plot for the prioritized regions.

Histogram for the effect of the 21,190 consensus ROHs identified in the whole sample is shown; risk ROHs associations are shown in red; protective ROHs associations are shown in green. Blue regions represent prioritized ROHs from inbred AD cases. Orange segments represent priorizated regions harboring potential recessive variants.

finally, for the search of an efficient therapeutic target.

At this study, we efficiently identified a number of potentially inbred AD cases nested an outbred population, offering a new framework for the analysis of the inbreeding component in AD. Furthermore, we demonstrated that detected consensus ROHs are enriched in risk associations. Considering our findings, we believe that recessive allelic architecture defines a portion of AD heritability, even, the accumulation of multiple, low penetrance or even pure recessive

TABLE 3. Candidate recessive variants after ROH prioritization focused in inbred AD cases

al	F_{ROH}	SROH (Kb)	CHR	ROH start	ROH end	ROH length	ROH $_{SNPs}$	Rs	Variant	Near Locus	Ref/Alt Allele	MAF	
l 1	0.0058	17422	7	53232690	55262627	2029.9	4213	NA	7:53930748	*RP11-678B3.2*	C/A	Novel	M
l 2	0.0185	55861	7	135617715	150518393	14900.7	12426	rs116644203	7:148903805	*ZNF282*	C/T	Novel	M
l 3	0.0099	29869	8	47018484	49544784	2526.301	1265	NA	8:48352954	*SPIDR*	T/A	0.004	
l 4	0.0153	46195	8	91282507	96735572	5453.1	4734	rs73263258	8:95658495	*ESPR1**	G/A	0.002	MC M
								rs138998679	8:95780656	*DPY19L4**	A/G	0.003	
l 5	0.0179	54138	10	24516308	28414933	3898.6	6248	rs140790046	10:27964291	*MKX*	T/C	0.0005	MC
l 6	0.0137	41393	11	12032272	16358722	4326.5	4473	rs117458494	11:14282085	*SPON1 RP11-21L19.1*	G/A	Novel	M
l 7	0.0115	34603	14	63906601	66333670	2427.1	2633	rs140897155	14:64688393	*SYNE2 ESR2*	G/A	0.001	MC M

o from TP53INP1/NDUFAF6

variants should play a role.

The genetic basis of human late-onset diseases has been mainly explained by selective neutrality (48) under the common disease common–variant (CD–CV) hypothesis (49). Considering that, several evolutionary theories of aging have been proposed and demonstrated to some extent, e.g. mutation accumulation (MA) and antagonistic pleiotropy (AP) (50),(51). Theoretical and experimental models for MA further support inbreeding effects for late-onset diseases (52),(53). Given that the human population is evolutionarily young, large proportion of human variation is necessarily rare (54). Thereby, late-acting alleles will also be found at low-frequencies. In that scenario, contemporary genetic models are in agreement with present results, where rare and recessive-acting variants could explain a part of the genetic basis of AD.

On the search of homozygosity patterns in AD, previous studies in populations from European and non-European ancestries have showed inconsistent results (28),(29),(27),(26). In that sense, we believe that several technical considerations must be taken into account for the analysis of ROH. First, it is suggested that the estimation of an excess of homozygosity from outbred population requires a large sample size (32), but prior studies accounted with very modest sample size (N < 6,000) (28),(29),(27),(26). Thus most probably they were underpowered to detect inbreeding effects from unrelated individuals.

Second, different scenarios should be considered for selecting shorter or longer ROHs than 1.5 Mb for the measurements and statistics, because they evidenced different aspects of demographic history. Evidence suggest that individual ROHs < 1.5 Mb might reflect LD patterns of ancient origin, rather than consanguineous cultural practices and genetic isolation, captured with ROHs > 1.5 Mb (55). Here, we detected substantial inflation in homozygosity parameters when individual ROH length was set to 1 Mb. This makes intricate the detection of genuine inbred individuals from an outbred population, and strongly confounds the interpretation of homozygosity estimations. Despite that, prior AD genetic studies, assessing the role of homozygosity, have not tested the potential effect of performing ROH calling for segments longer than 1 Mb (28),(29),(27),(26), which might be explaining, in part, initial failures.

Overall, ROH studies in AD have been several technical handicaps, which might have caused to overlook the potential inbreeding effect for this disease. Hence, we encourage other groups to conduct ROH analysis from unrelated population, but with enough sample sizes and redefining the ROHs lengths at least to 1.5Mb, to better capture the recessive component of AD.

In our ROH analysis using the whole sample (N = 21,100), the most promising consensus ROH was located in the proximity of the *HS3ST1* gene (~200 Kb), and showed a strong genetic effect. Genetic markers in the vicinity of the homozygous block (~300kb) have been previously associated with AD (56),(46). Additionally, brain expression of *HS3ST1* was differentially expressed between AD cases and controls (56). Despite these findings, the causative genetic mechanism involving this region with AD remains elusive. In that sense, we believe that high resolution mapping across the 115 Kb of the reported consensus ROH could help to positional cloning of the causative mutation.

Our study of ROH > 1.5Mb shed lights in the homozygosity component influencing AD, as it reflected recent consanguinity and/or population isolation. Inbred individuals tend to have lower survival, fertility and growth rates (57),(58),(59), as well as post-reproductive health (60). Considering that, we believe that enriching our subset in inbred cases can provide a redefined framework for investigating inbreeding effect and looking for recessive acting variants. This idea has driven the design of the present study. With the aim to increase the probability to detect regions harboring recessive-acting loci, we prioritized consensus ROH according to the homozygosity map exclusively obtained of inbred AD individuals. Candidate regions were then explored in sequencing data. Among them, variants in *MKX* and *ZNF282* genes were detected in two independent inbred AD cases. Both, *ZNF282* and *MKX* loci are encoding transcription factors (61),(62),(63). In that sense, recently, the largest WES study analyzing rare variation in AD highlighted the potential role of transcriptional regulation for this disease (64). Of note, *ZNF282* gene is mapped roughly to 800Kb from *CNTNAP2* gene, previously associated with AD (11). Autosomal recessive mutations in *CNTNAP2* loci have been also linked with epilepsy and intellectual disability (OMIM 604569).

In this study, we also found a potential recessive variant in *SPON1* locus. SPON1 has been related with the mechanistic of AD, where the APP metabolism has a central role. APP cleavage through β-secretases produces amyloid-beta (Aβ), which later accumulates in AD brains (7). It is described that SPON1 binds to APP, inhibiting its α/β cleavage (47). Further studies also reported SPON1 binding to APOE family of receptors (65), and genetic markers in this gene has been related with dementia severity (66). Taking into account prior findings and the present result, it would be biologically plausible the presence of recessive acting-variants in *APP* or in its biological partners, which directly influence amyloid cascade. Thus, we believe that *SPON1* could be considered an interesting

candidate, which deserves future resequencing efforts.

Among other candidates, we identified a missense variant (rs73263258 in *ESPR1* gene) within a long ROH of an AD patient. This gene is mapped in the close vicinity of *TP53INP1/NDUFAF6* genomic region. This region has been previously associated with AD using gene-based strategy (67). Recently our group also identified genome-wide significant markers within this region (13). It is not unexpected that genes containing common variants with small genetic effect might also be enriched in rarer variants, with higher penetrance. The existence of several genetic mechanisms acting at this region should be considered when deeply sequencing will be conducted, to pinpoint the causative variant.

Our observations are subject to limitations that need to be considered. Since datasets used in the present study were genotyped using different genotyping platforms, they shared a small proportion of directly genotyped markers. Given that lower SNP density could impact the accuracy of the study (35), we decided to perform the present analysis using imputed genotypes of high quality (imputation quality, $r2> 0.90$). To make the data optimally comparable, we generated a merged dataset including the same variants with MAF > 0.05. We also showed that ROH calling is insensitive to perform the analysis per each individual dataset or in the merged data, for a set of individuals from the same ancestral group, when we fixed the SNP set to use.

The potential impact of CNVs in ROH analysis must be taken in consideration as a potential limitation. However, when we assessed CNV impact in our analyses, no differences were found in homozygosity parameters, before and after CNVs exclusion. These results are in agreement with previous studies, suggesting that the effect of deletions in homozygosity parameters, when exist, is minimal (36).

Clonal mosaicism can also generate spurious ROHs. A direct correlation between clonal mosaicism events in peripheral blood and age >50 years, was demonstrated (68). We believe that these events might be introducing ROHs of short lengths. Consequently, an age-dependent increase in the total ROH length and the number of segments was detected. This phenomenon promotes a reduction of the total average length of ROH. In fact, it would be explaining why we identified higher than expected mean ROH number for this dataset of European population compared to prior studies (35). From our point of view, controlling the role of genome instability for late onset neurodegenerative diseases represents a challenge, due to the impossibility to detect true converters to AD before disease onset, and the difficulties in collecting biological information from the target tissue. In spite of a signature of genome instability in ROH studies might exist, our adjusted results

considering age effect are still supporting the idea that inbred individuals are overrepresented in AD population respect to controls.

In summary, we demonstrated the existence of an inbreeding effect in AD and efficiently captured a fraction of consanguineous individuals from outbred populations. The proposed method can be considered a refined strategy to investigate the role of recessive variants in AD. Considering that there is a large limitation in collecting complete information from consanguineous AD families, the identification of highly probable consanguineous AD cases in outbred populations could be important for future large-scale homozygosity mapping. Furthermore, the opportunity to explore complementary sequencing data gave an added value to this research, providing a subset of potential candidates harboring recessive variants. In any case, the proposed candidates, acting under a recessive inheritance mode, will only be confirmed when at least an additional individual harboring the same recessive mutation or a compound heterozygote is detected. We recognize our current lack of power to firmly verify arAD loci. That is why, greater efforts and larger collections of individuals with GWAS and sequencing data are needed to confirm our findings.

Understanding the dynamics of population genomics in complex diseases like AD is far from complete but ROHs analyses provide us a window to go further; and might be an alternative strategy to uncover the genetic loci underlying Alzheimer's disease

REFERENCIES

1. 2019 Alzheimer's disease facts and figures. Alzheimer's Dement . 2019 Mar 1;15(3):321–87.

2. Wingo TS, Lah JJ, Levey AI, Cutler DJ. Autosomal recessive causes likely in early-onset Alzheimer disease. Arch Neurol . 2012 Jan ;69(1):59–64.

3. Brainstorm Consortium V, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, et al. Analysis of shared heritability in common disorders of the brain. Science . 2018 Jun 22;360(6395):eaap8757.

4. Sherrington R, Rogaev EI, Liang Y, Rogaeva EA, Levesque G, Ikeda M, et al. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. Nature . 1995 Jun 29;375(6534):754–60.

5. Sherrington R, Froelich S, Sorbi S, Campion D, Chi H, Rogaeva EA, et al. Alzheimer's disease associated with mutations in presenilin 2 is rare and variably penetrant. Hum Mol Genet . 1996 Jul;5(7):985–8.

6. Goate A, Chartier-Harlin MC, Mullan M, Brown J, Crawford F, Fidani L, et al. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. Nature . 1991 Feb 21;349(6311):704–6.

7. Hardy J, Selkoe DJ. The Amyloid Hypothesis of Alzheimer's Disease: Progress and Problems on the Road to Therapeutics. Science (80-). 2002 Jul 19;297(5580):353–6.

8. Di Fede G, Catania M, Morbin M, Rossi G, Suardi S, Mazzoleni G, et al. A Recessive Mutation in the APP Gene with Dominant-Negative Effect on Amyloidogenesis. Science (80-). 2009 Mar 13;323(5920):1473–7.

9. Tomiyama T, Nagata T, Shimada H, Teraoka R, Fukushima A, Kanemitsu H, et al. A new amyloid β variant favoring oligomerization in Alzheimer's-type dementia. Ann Neurol . 2008 Mar;63(3):377–87.

10. Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. Nat Genet . 2019 Mar 28;51(3):414–30.

11. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. Nat Genet . 2019 Jan 7;

12. Sims R, van der Lee SJ, Naj AC, Bellenguez C, Badarinarayan N, Jakobsdottir J, et al. Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. Nat Genet . 2017 Jul 17;49(9):1373–84.

13. Moreno-Grau S, de Rojas I, Hernández I, Quintela I, Montrreal L, Alegret M, et al. Genome-wide association analysis of dementia and its clinical endophenotypes reveal novel loci associated with Alzheimer's disease and three causality networks: The GR@ACE project. Alzheimer's Dement . 2019 Aug 13;

14. Ridge PG, Hoyt KB, Boehme K, Mukherjee S, Crane PK, Haines JL, et al. Assessment of the genetic variance of late-onset Alzheimer's disease. Neurobiol Aging . 2016 May;41:200.e13-200.e20.

15. Garrod AE. The incidence of alkaptonuria: a study in chemical individuality. Lancet. 1902;1616:1902–11.

16. Lander E, Botstein D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. Science (80-) . 1987 Jun 19;236(4808):1567–70.

17. Jakkula E, Rehnström K, Varilo T, Pietiläinen OPH, Paunio T, Pedersen NL, et al. The Genome-wide Patterns of Variation Expose Significant Substructure in a Founder Population. Am J Hum Genet . 2008 Dec;83(6):787–94.

18. Gross A, Tönjes A, Kovacs P, Veeramah KR, Ahnert P, Roshyara NR, et al. Population-genetic comparison of the Sorbian isolate population in Germany with the German KORA population using genome-wide SNP arrays. BMC Genet . 2011 Jul 28;12(1):67.

19. Farrer LA, Bowirrat A, Friedland RP, Waraska K, Korczyn AD, Baldwin CT. Identification of multiple loci for Alzheimer disease in a consanguineous Israeli-Arab community. Hum Mol Genet . 2003 Feb 15;12(4):415–22.

20. Sherva R, Baldwin CT, Inzelberg R, Vardarajan B, Cupples LA, Lunetta K, et al. Identification of novel candidate genes for Alzheimer's disease by autozygosity mapping using genome wide SNP data. J Alzheimers Dis . 2011;23(2):349–59.

21. Clarimón J, Djaldetti R, Lleó A, Guerreiro RJ, Molinuevo JL, Paisán-Ruiz C, et al. Whole genome analysis in a consanguineous family with early onset Alzheimer's disease. Neurobiol Aging . 2009 Dec;30(12):1986–91.

22. Bras J, Djaldetti R, Alves AM, Mead S, Darwent L, Lleo A, et al. Exome sequencing in a consanguineous family clinically diagnosed with early-onset Alzheimer's disease identifies a homozygous CTSF mutation. Neurobiol Aging . 2016 Oct;46:236.e1-6.

23. McQuillan R, Leutenegger A-L, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al. Runs of Homozygosity in European Populations. Am J Hum Genet . 2008 Sep;83(3):359–72.

24. Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ. Genomic Patterns of Homozygosity in Worldwide Human Populations. Am J Hum Genet . 2012 Aug 10;91(2):275–92.

25. Vardarajan BN, Schaid DJ, Reitz C, Lantigua R, Medrano M, Jiménez-Velázquez IZ, et al. Inbreeding among Caribbean Hispanics from the Dominican Republic and its effects on risk of Alzheimer disease. Genet Med . 2015 Aug;17(8):639–43.

26. Ghani M, Sato C, Lee JH, Reitz C, Moreno D, Mayeux R, et al. Evidence of recessive Alzheimer disease loci in a Caribbean Hispanic data set: genome-wide survey of runs of homozygosity. JAMA Neurol . 2013 Oct 26;70(10):1261–7.

27. Ghani M, Reitz C, Cheng R, Vardarajan BN, Jun G, Sato C, et al. Association of Long Runs of Homozygosity With Alzheimer Disease Among African American Individuals. JAMA Neurol . 2015 Nov 1;72(11):1313–23.

28. Nalls MA, Guerreiro RJ, Simon-Sanchez J, Bras JT, Traynor BJ, Gibbs JR, et al. Extended tracts of homozygosity identify novel candidate genes associated with late-onset Alzheimer's disease. Neurogenetics . 2009 Jul 7;10(3):183–90.

29. Sims R, Dwyer S, Harold D, Gerrish A, Hollingworth P, Chapman J, et al. No evidence that extended tracts of homozygosity are associated with Alzheimer's disease. Am J Med Genet Part B Neuropsychiatr Genet . 2011 Dec;156(7):764–71.

30. Lemes RB, Nunes K, Carnavalli JEP, Kimura L, Mingroni-Netto RC, Meyer D, et al. Inbreeding estimates in human populations: Applying new approaches to an admixed Brazilian isolate. Calafell F, editor. PLoS One . 2018 Apr 24;13(4):e0196360.

31. Mooney JA, Huber CD, Service S, Sul JH, Marsden CD, Zhang Z, et al. Understanding the Hidden Complexity of Latin American Population Isolates. Am J Hum Genet . 2018 Nov 1;103(5):707–26.

32. Keller MC, Visscher PM, Goddard ME. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetics . 2011 Sep;189(1):237–49.

33. Fernández MV, Kim JH, Budde JP, Black K, Medvedeva A, Saef B, et al. Analysis of neurodegenerative Mendelian genes in clinically diagnosed Alzheimer Disease. Myers AJ, editor. PLoS Genet . 2017 Nov 1;13(11):e1007045.

34. Howrigan DP, Simonson MA, Keller MC. Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. BMC Genomics . 2011 Sep 23;12(1):460.

35. Ceballos FC, Hazelhurst S, Ramsay M. Runs of Homozygosity in sub-Saharan African populations provide insights into a complex demographic and health history. bioRxiv . 2018 Nov 14;470583.

36. Joshi PK, Esko T, Mattsson H, Eklund N, Gandin I, Nutile T, et al. Directional dominance on stature and cognition in diverse human populations. Nature . 2015 Jul 1;523(7561):459–62.

37. Lencz T, Lambert C, DeRosse P, Burdick KE, Morgan T V, Kane JM, et al. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. Proc Natl Acad Sci . 2007 Dec 11;104(50):19942–7.

38. Christofidou P, Nelson CP, Nikpay M, Qu L, Li M, Loley C, et al. Runs of Homozygosity: Association with Coronary Artery Disease and Gene Expression in Monocytes and Macrophages. Am J Hum Genet . 2015 Aug 6;97(2):228–37.

39. Shifman S, Kuypers J, Kokoris M, Yakir B, Darvasi A. Linkage disequilibrium patterns of the human genome across populations. Hum Mol Genet . 2003 Apr 1;12(7):771–6.

40. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature . 2007 Oct 18;449(7164):851–61.

41. Slatkin M. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. Nat Rev Genet . 2008 Jun 1;9(6):477–85.

42. Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, et al. An integrated map of structural variation in 2,504 human genomes. Nature . 2015 Oct 30;526(7571):75–81.

43. Gazal S, Sahbatou M, Perdry H, Letort S, Génin E, Leutenegger A-L. Inbreeding coefficient estimation with dense SNP data: comparison of strategies and application to HapMap III. Hum Hered . 2014;77(1–4):49–62.

44. Bittles AH, Black ML. Consanguinity, human evolution, and complex diseases. Proc Natl Acad Sci . 2010 Jan 26;107(suppl_1):1779–86.

45. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly. 2012 Apr 27;6(2):80–92.

46. Witoelar A, Rongve A, Almdahl IS, Ulstein ID, Engvig A, White LR, et al. Meta-analysis of Alzheimer's disease on 9,751 samples from Norway and IGAP study identifies four risk loci. Sci Rep . 2018 Dec 27;8(1):18088.

47. Ho A, Südhof TC. Binding of F-spondin to amyloid-beta precursor protein: a candidate amyloid-beta precursor protein ligand that modulates amyloid-beta precursor protein cleavage. Proc Natl Acad Sci U S A . 2004 Feb 24;101(8):2548–53.

48. KIMURA M. Evolutionary Rate at the Molecular Level. Nature . 1968 Feb;217(5129):624–6.

49. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science . 1996 Sep 13;273(5281):1516–7.

50. Wright A, Charlesworth B, Rudan I, Carothers A, Campbell H. A polygenic basis for late-onset disease. Trends Genet . 2003 Feb;19(2):97–106.

51. Rodríguez JA, Marigorta UM, Hughes DA, Spataro N, Bosch E, Navarro A. Antagonistic pleiotropy and mutation accumulation influence human senescence and disease. Nat Ecol Evol . 2017 Mar 30;1(3):0055.

52. Charlesworth B, Hughes KA. Age-specific inbreeding depression and components of genetic variance in relation to the evolution of senescence. Proc Natl Acad Sci U S A . 1996 Jun 11;93(12):6140–5.

53. Hughes KA, Alipaz JA, Drnevich JM, Reynolds RM. A test of evolutionary theories of aging. Proc Natl Acad Sci U S A . 2002 Oct 29;99(22):14286–91.

54. McClellan J, King M-C. Genetic Heterogeneity in Human Disease. Cell . 2010 Apr 16;141(2):210–7.

55. McQuillan R, Leutenegger A-L, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al. Runs of Homozygosity in European Populations. Am J Hum Genet . 2008 Sep;83(3):359–72.

56. Broce IJ, Tan CH, Fan CC, Jansen I, Savage JE, Witoelar A, et al. Dissecting the genetic relationship between cardiovascular risk factors and Alzheimer's disease. Acta Neuropathol. 2018 Nov 9;

57. Bittles AH, Neel J V. The costs of human inbreeding and their implications for variations at the DNA level. Nat Genet . 1994 Oct;8(2):117–21.

58. Stoltenberg C, Magnus P, Skrondal A, Lie RT. Consanguinity and recurrence risk of birth defects: a population-based study. Am J Med Genet. 1999 Feb 19;82(5):423–8.

59. Stoltenberg C, Magnus P, Skrondal A, Lie RT. Consanguinity and recurrence risk of stillbirth and infant death. Am J Public Health. 1999 Apr;89(4):517–23.

60. Rudan I, Rudan D, Campbell H, Carothers A, Wright A, Smolej-Narancic N, et al. Inbreeding and risk of late onset complex disease. J Med Genet . 2003 Dec;40(12):925–32.

61. Anderson DM, George R, Noyes MB, Rowton M, Liu W, Jiang R, et al. Characterization of the DNA-binding properties of the Mohawk homeobox transcription factor. J Biol Chem . 2012 Oct 12;287(42):35351–9.

62. Okumura K, Sakaguchi G, Naito K, Tamura T, Igarashi H. HUB1, a novel Kruppel type zinc finger protein, represses the human T cell leukemia virus type I long terminal repeat-mediated expression. Nucleic Acids Res . 1997 Dec 15;25(24):5025–32.

63. Kim JH, Yang CK, Stallcup MR. Downstream signaling mechanism of the C-terminal activation domain of transcriptional coactivator CoCoA. Nucleic Acids Res . 2006 May 22;34(9):2736–50.

64. Bis JC, Jian X, Kunkle BW, Chen Y, Hamilton-Nelson KL, Bush WS, et al. Whole exome sequencing study identifies novel rare and common Alzheimer's-Associated variants involved in immune response and transcriptional regulation. Mol Psychiatry. 2018 Aug 14

65. Hoe H-S, Wessner D, Beffert U, Becker AG, Matsuoka Y, Rebeck GW. F-Spondin Interaction

with the Apolipoprotein E Receptor ApoEr2 Affects Processing of Amyloid Precursor Protein. Mol Cell Biol . 2005 Nov 1;25(21):9259–68.

66. Jahanshad N, Rajagopalan P, Hua X, Hibar DP, Nir TM, Toga AW, et al. Genome-wide scan of healthy human connectome discovers SPON1 gene variant influencing dementia severity. Proc Natl Acad Sci . 2013 Mar 19;110(12):4768–73.

67. Escott-Price V, Bellenguez C, Wang L-S, Choi S-H, Harold D, Jones L, et al. Gene-wide analysis detects two new susceptibility genes for Alzheimer's disease. PLoS One . 2014 Jan;9(6):e94661.

68. Laurie CC, Laurie CA, Rice K, Doheny KF, Zelnick LR, McHugh CP, et al. Detectable clonal mosaicism from birth to old age and its relationship to cancer. Nat Genet . 2012 Jun 6;44(6):642–50.

ACKNOWLEDGEMENTS

We would like to thank patients and controls who participated in this project. The Genome Research @ Fundació ACE project (GR@ACE) is supported by Fundación bancaria "La Caixa," Grifols SA, Fundació ACE, and ISCIII (Ministry of Health, Spain). We also want to thank the private sponsors who support the basic and clinical projects of our institution (Piramal AG, Laboratorios Echevarne, Araclon Biotech S.A., and Fundació ACE). We are indebted to the Trinitat Port-Carbó legacy and her family for their support of Fundació ACE research programs. Fundació ACE is a participating center in the Dementia Genetics Spanish Consortium (DEGESCO). A.R. and M.B. receive support from the European Union/EFPIA Innovative Medicines Initiative Joint Undertaking ADAPTED and MOPEAD projects (Grants No. 115975 and 115985, respectively). M.B. and A.R. are also supported by national grants PI13/02434, PI16/01861, and PI17/01474. Acción Estratégica en Salud is integrated into the Spanish National R + D + I Plan and funded by ISCIII (Instituto de Salud Carlos III)-Subdirección General de Evaluación and the Fondo Europeo de Desarrollo Regional (FEDER- "Una manera de Hacer Europa"). LMR is supported by Consejería de Salud de la Junta de Andalucía (Grant PI-0001/2017). Control samples and data from patients included in this study were provided in part by the National DNA Bank Carlos III (www.bancoadn.org, University of Salamanca, Spain) and Hospital Universitario Virgen de Valme (Sevilla, Spain); they were processed following standard operating procedures with the appropriate approval of the Ethical and Scientific Committee. The present work was performed as part of the Biochemistry, Molecular Biology and Biomedicine doctoral program of S. Moreno-Grau at Universitat Autònoma de Barcelona (Barcelona, Spain).

Data collection and sharing for this project was partially funded by the **Alzheimer's Disease Neuroimaging Initiative (ADNI)** (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging and the National Institute of Biomedical Imaging and Bioengineering, as well as through generous contributions from the following: AbbVie; the Alzheimer's Association; the Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research provides funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study was coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

The **AddNeuroMed** data are from a public-private partnership supported by EFPIA companies and SMEs as part of InnoMed (Innovative Medicines in Europe), an integrated project funded by the European Union of the Sixth Framework program priority FP6-2004-LIFESCIHEALTH-5. Clinical leads responsible for data collection are Iwona Kłoszewska (Lodz), Simon Lovestone (London), Patrizia Mecocci (Perugia), Hilkka Soininen (Kuopio), Magda Tsolaki (Thessaloniki) and Bruno Vellas (Toulouse). Imaging leads are Andy Simmons (London), Lars-Olad Wahlund (Stockholm) and Christian Spenger (Zurich). Bioinformatics leads are Richard Dobson (London) and Stephen Newhouse (London).

Funding support for the **Alzheimer's Disease Genetics Consortium (ADGC)** was provided through the NIA Division of Neuroscience (U01-AG032984).

The genotypic and associated phenotypic data used in the study "Multi-Site Collaborative Study for Genotype-Phenotype Associations in Alzheimer's Disease (**GenADA**)" were provided by GlaxoSmithKline, R&D Limited. The datasets used for the analyses described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/gap through dbGaP accession number phs000219.

The **Mayo Clinic Alzheimer's Disease Genetic Studies**, led by Dr. Nilüfer Ertekin-Taner and Dr. Steven G. Younkin at the Mayo Clinic in Jacksonville, FL, used samples from the Mayo Clinic Study of Aging, the Mayo Clinic Alzheimer's Disease Research Center and the Mayo Clinic Brain Bank. Data collection was supported through funding by NIA grants P50 AG016574, R01 AG032990, U01 AG046139, R01 AG018023, U01 AG006576, U01 AG006786, R01 AG025711, R01 AG017216 and R01 AG003949, NINDS grant R01 NS080820, the CurePSP Foundation and support from the Mayo Foundation.

The **Neocodex-Murcia** study was funded by the Fundación Alzheimur (Murcia), the Ministerio de Educación y Ciencia (Gobierno de España), Corporación Tecnológica de

Andalucía, Agencia IDEA (Consejería de Innovación, Junta de Andalucía), the Diabetes Research Laboratory and the Biomedical Research Foundation. University Hospital Clínico San Carlos has been supported by CIBER de *Diabetes y Enfermedades Metabólicas Asociadas* (CIBERDEM); CIBERDEM is an ISCIII Project.

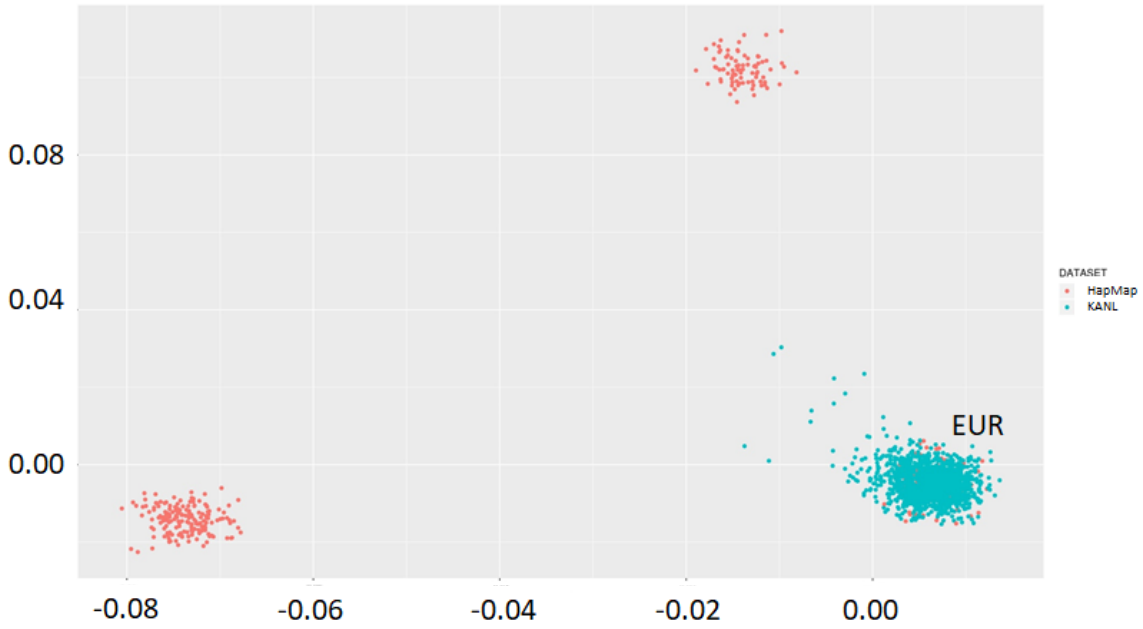
The **ROS/MAP** study data were provided by the Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago. Data collection was supported through funding by NIA grants P30AG10161, R01AG15819, R01AG17917, R01AG30146, R01AG36836, U01AG32984 and U01AG46152, the Illinois Department of Public Health and the Translational Genomics Research Institute

The **TGEN** study was supported by Kronos Life Sciences Laboratories, the National Institute on Aging (Arizona Alzheimer's Disease Center grants P30 AG19610 and RO1 AG023193, the Mayo Clinic Alzheimer's Disease Center grant P50 AG16574 and the Intramural Research Program), the National Alzheimer's Coordinating Center (U01 AG016976) and the state of Arizona.

The results published here are in part based on data obtained from the AMP-AD Knowledge Portal accessed at http://dx.doi.org/doi:10.7303/syn2580853

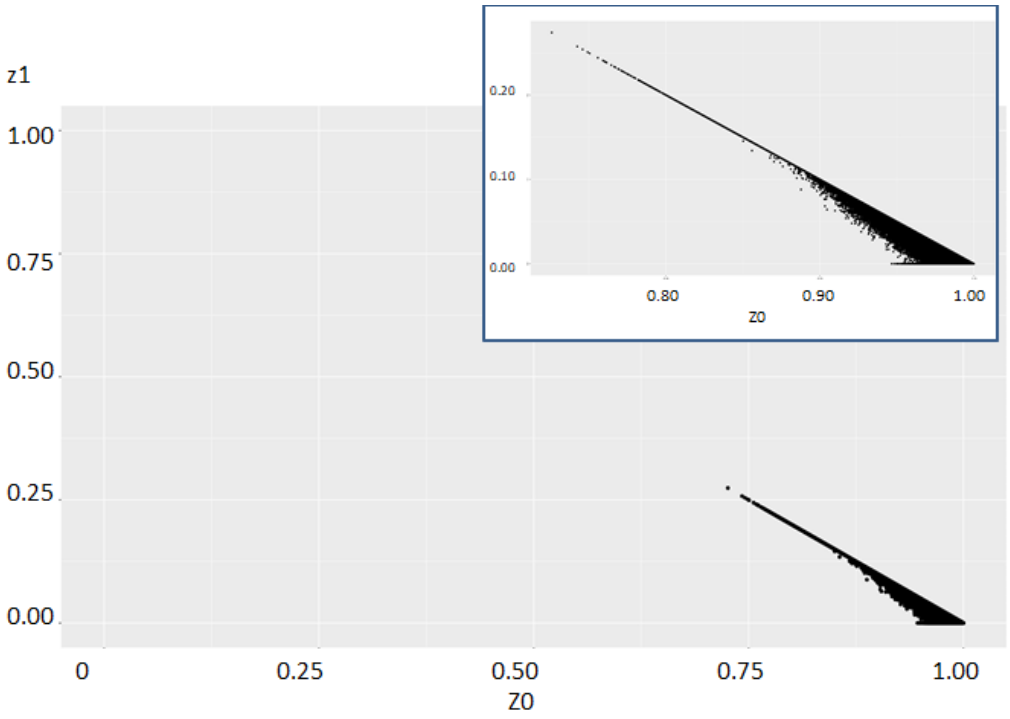
We thank the International Genomics of Alzheimer's Project (IGAP) for providing summary results data for these analyses. The investigators within IGAP contributed to the design and implementation of IGAP and/or provided data but did not participate in analysis or writing of this report. IGAP was made possible by the generous participation of the control subjects, the patients, and their families. The i–Select chip was funded by the French National Foundation on Alzheimer's disease and related disorders. EADI was supported by the LABEX (laboratory of excellence program investment for the future) DISTALZ grant, Inserm, Institut Pasteur de Lille, Université de Lille 2 and the Lille University Hospital. GERAD was supported by the Medical Research Council (Grant n 503480), Alzheimer's Research UK (Grant n 503176), the Wellcome Trust (Grant n 082604/2/07/Z) and German Federal Ministry of Education and Research (BMBF): Competence Network Dementia (CND) grant n 01GI0102, 01GI0711, 01GI0420. CHARGE was partly supported by the NIH/NIA grant R01 AG033193 and the NIA AG081220 and AGES contract N01–AG–12100, the NHLBI grant R01 HL105756, the Icelandic Heart Association, and the Erasmus Medical Center and Erasmus University. ADGC was supported by the NIH/NIA grants: U01 AG032984, U24 AG021886, U01 AG016976, and the Alzheimer's Association grant ADGC–10–196728.

SUPPLEMENTARY FIGURES

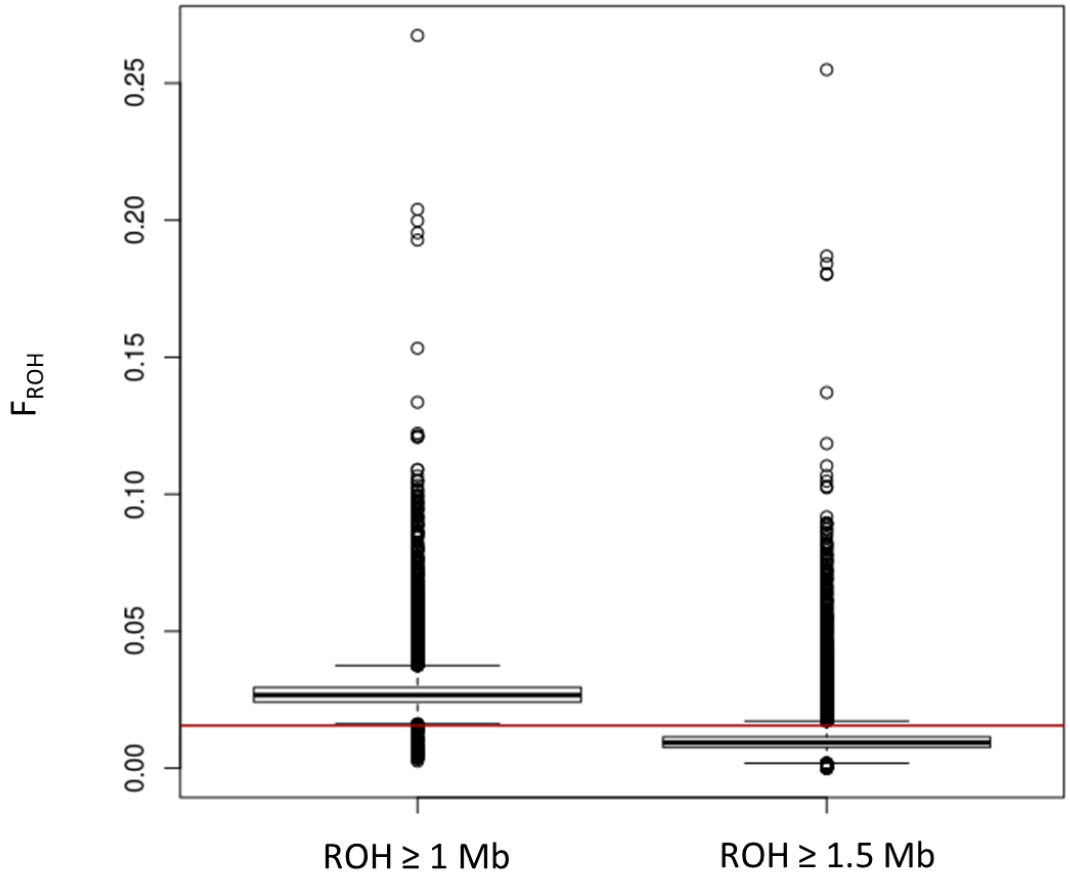


A.

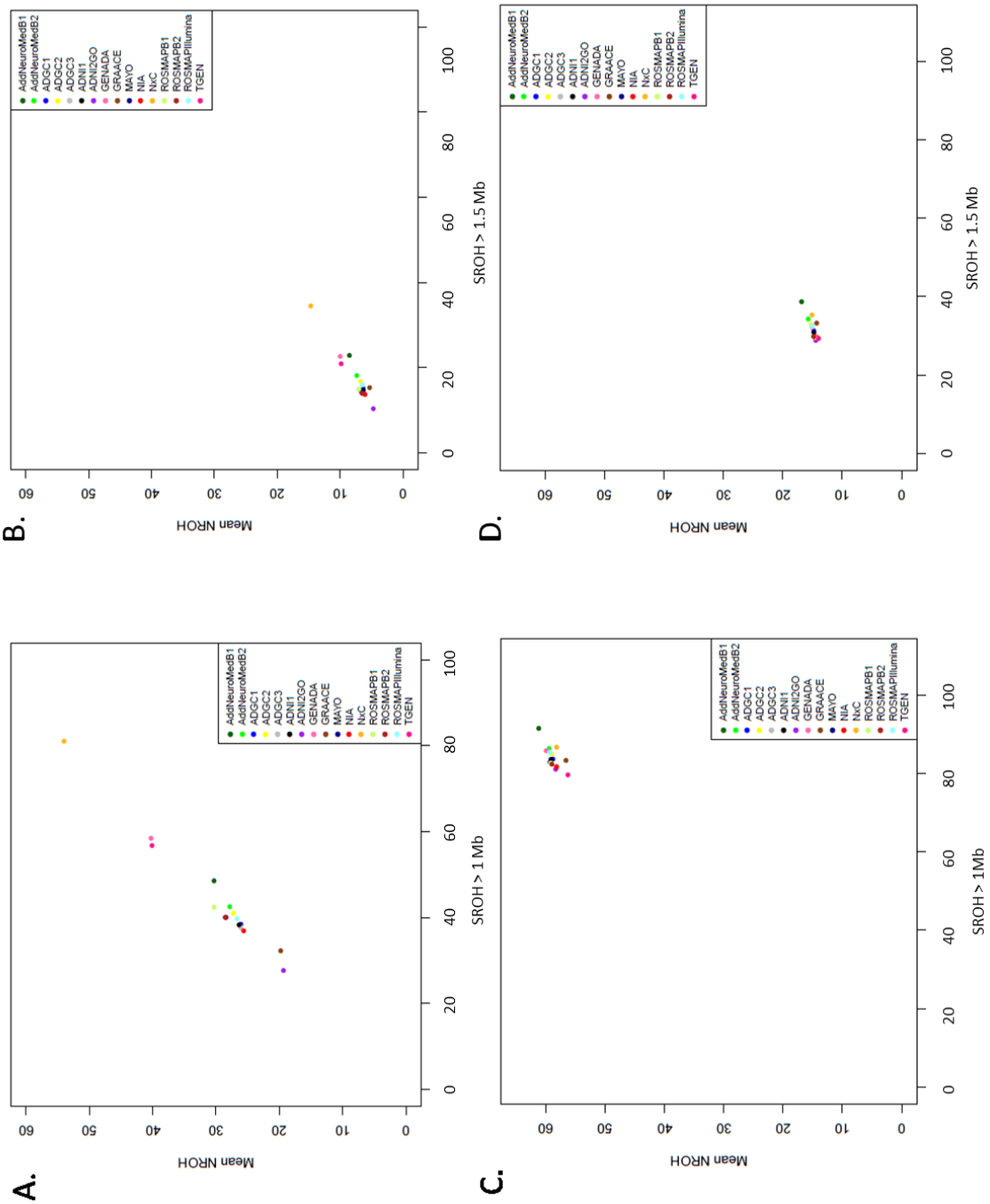
B.



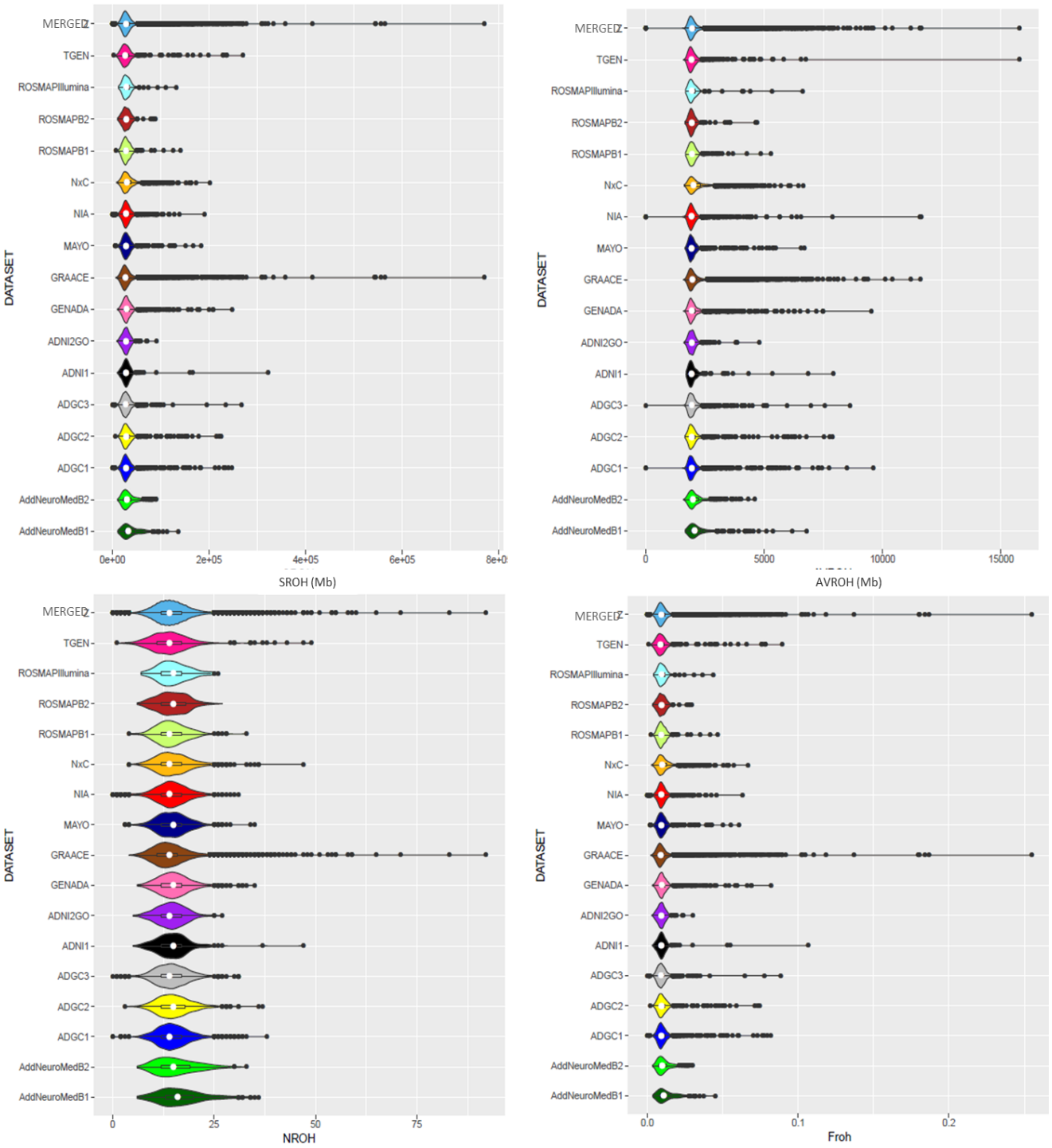
SUPPLEMENTARY FIGURE 1. Quality control for A) ancestry and B) relatedness in the exome.
All possible pairs had Pi-hat < 0.1875, a Z0 ≥ 0.75 and a Z1 ≤ 0.25. EUR; Individuals from European ancestry.



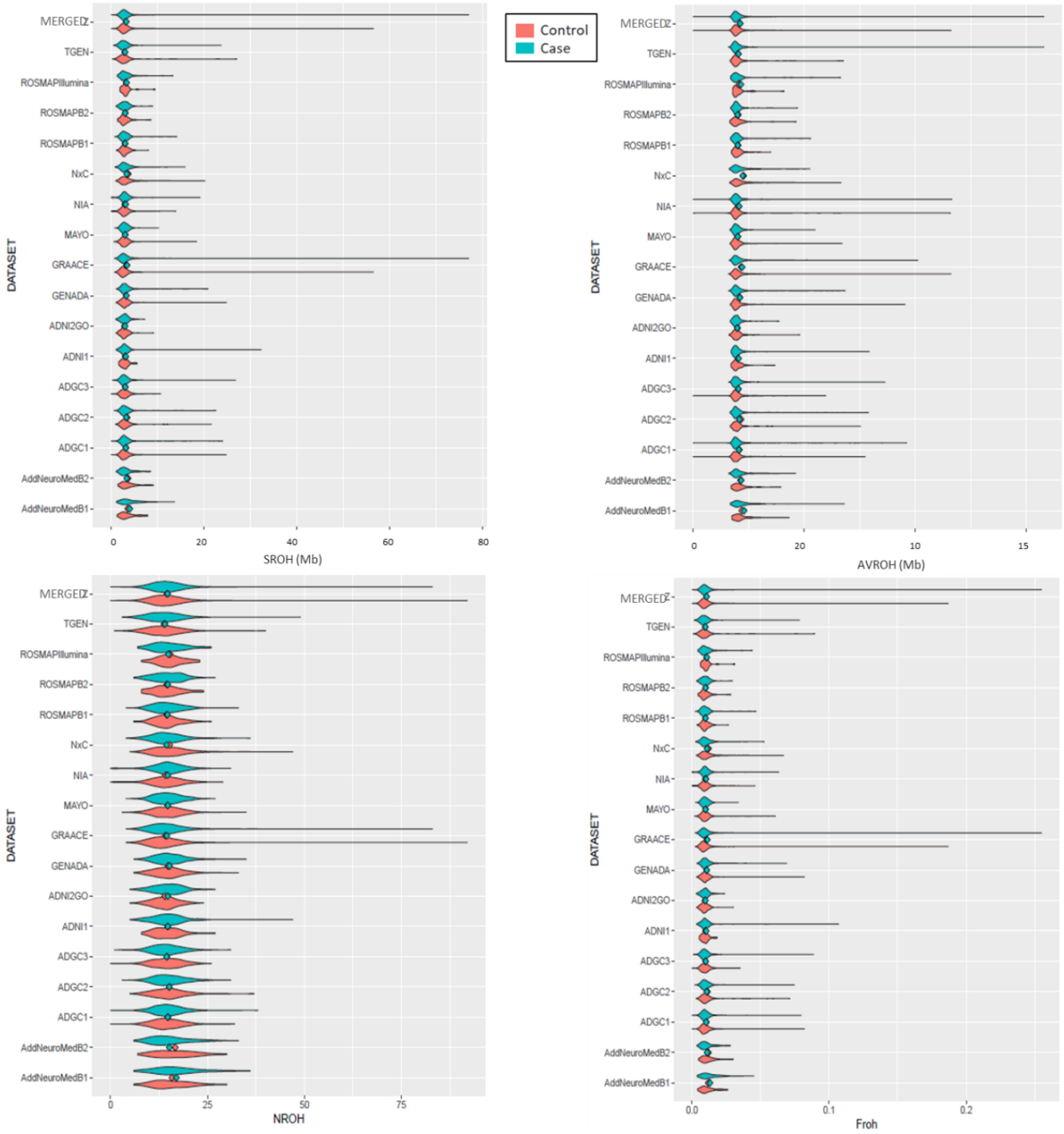
SUPPLEMENTARY FIGURE 2. Boxplot for FROH per individual at ROH calling with 1Mb and 1.5Mb.
Red line represents FROH = 0.0156 (mean inbreeding coefficient for kinship of second cousin marriage).



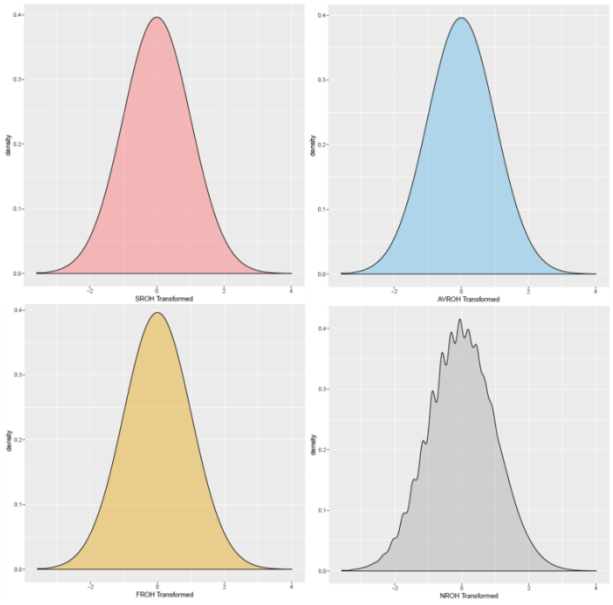
SUPPLEMENTARY FIGURE 3.Mean number of ROHs versus mean total sum of ROHs in Mb for the 10 cohorts explored, according to different ROH calling parameters.



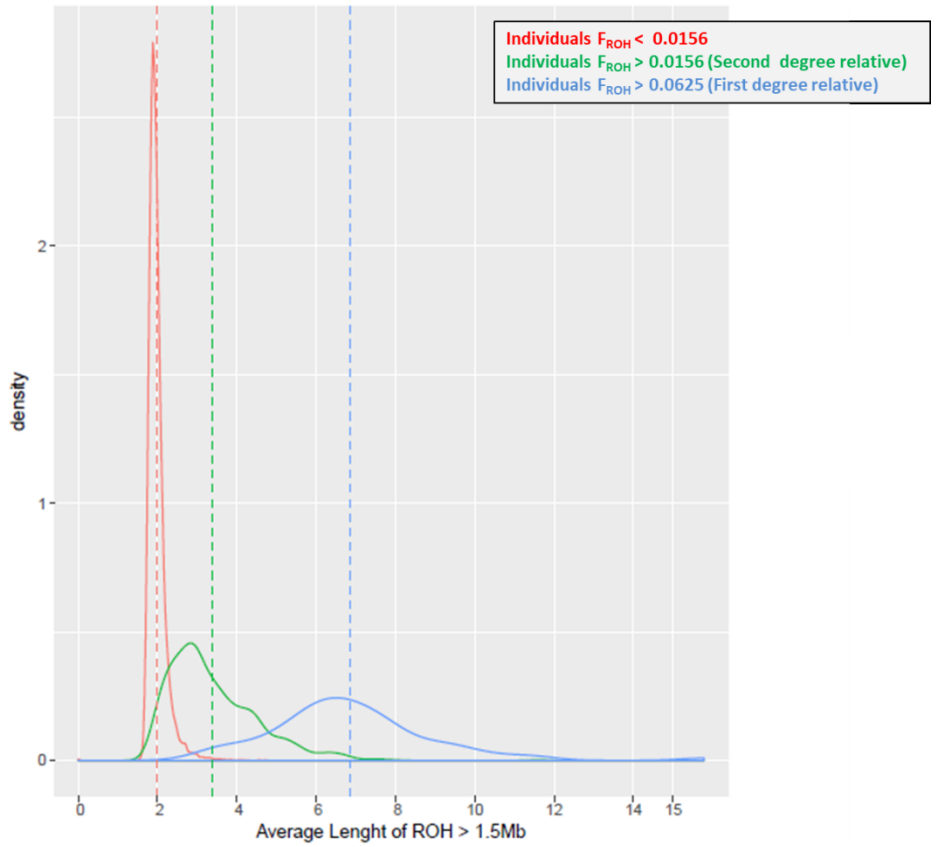
SUPPLEMENTARY FIGURE 4. Violin plots showing the distribution of ROH > 1.5 Mb within each dataset and in the merged data for the homozygosity parameters (NROH, SROH, AVROH, FROH).



SUPPLEMENTARY FIGURE 5. Violin plots showing the distribution of ROH > 1.5 Mb within each dataset and in the merged data for the homozygosity parameters (NROH, SROH, AVROH, FROH), split by case control status.



SUPPLEMENTARY FIGURE 6. Transformed distribution for the homozygosity measures.
Transformation was performed using an inverse rank normal transformation in "rankNorm" option of RNOmni package in R.



SUPPLEMENTARY FIGURE 7. Distribution for average length of individual ROH segments for: non-inbred individuals (FROH < 0.0156), second degree relative (FROH > 0.0156) and first degree relative (FROH > 0.0625)

SUPPLEMENTARY TABLES

Supplementary Tables can be accessed scanning the following code:

Supplementary Table 1. Characteristics of the cohorts used in the analysis.

Supplementary Table 2. Summary of homozygosity parameters for each individual study and the merged dataset, considering two minimal ROH lenght cut-off of 1 Mb and 1.5 Mb.

Supplementary Table 3. Summary statistics for the difference of homozygosity parameters calculated using two different methods; 1Mb and 1.5 Mb.

Supplementary Table 4. Effect of genome-wide homozygosity parameters in Alzheimer's disease per each individual dataset.

Supplementary Table 5. Effect of genome-wide homozygosity parameters in Alzheimer's disease for the joint analysis excluding deletions.

Supplementary Table 6. Effect of genome-wide homozygosity parameters in Alzheimer's disease for the joint analysis considering the effect of age.

Supplementary Table 7. Consensus ROH associated with Alzheimer's disease in the whole dataset.

Supplementary Table 8. Gene based results for genes located in consensus ROH associated with Alzheimer's disease in the whole dataset

Supplementary Table 9. Demographics for the pool of inbred individuals.

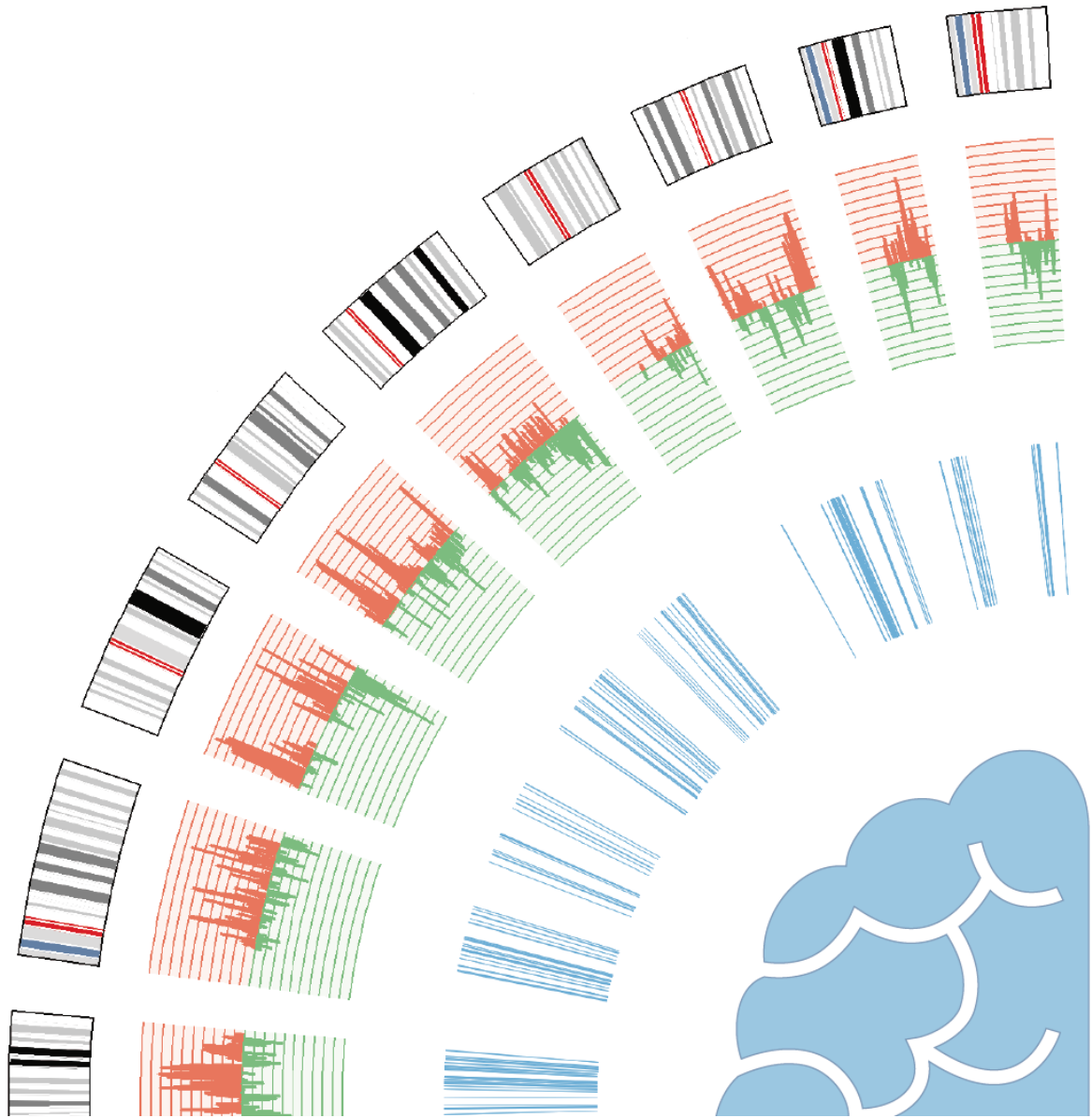
Supplementary Table 10. Consensus ROH priorizated based on the map of inbred Alzheimer's disease patients.

Supplementary Table 11. Gene based results for genes located in consensus ROH priorizated based on the map of inbred Alzheimer's disease patients.

Supplementary Table 12. Variants annotation and functional effect prediction.

CHAPTER 4

DISCUSSION



The aim of this thesis was to disentangle the genetic basis of Alzheimer disease and its related endophenotypes. Herewith, we highlight our findings (**Figure 1**), and place them in the context of genetic AD research.

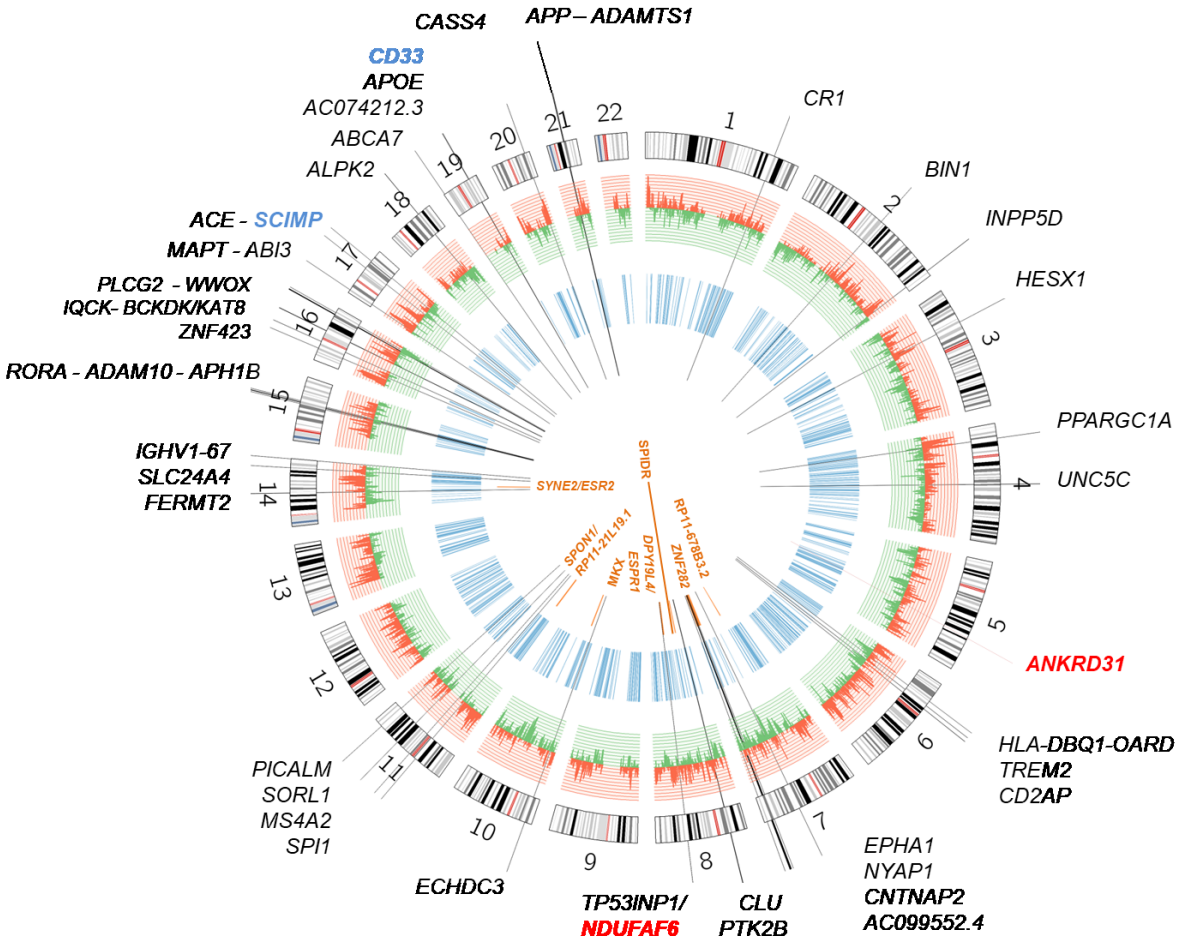


FIGURE 1. Loci associated with Late Onset Alzheimer's disease after this thesis.
The circular ideogram was created using Circos (118). New genetic markers reaching GWS at this work are highlithed in red. Previously genome-wide loci not reaching significance in the lastest IGAP or follow-up studies but reaching GWS at this work are showed in blue. Histogram for ROHs shows: risk ROHs associations in red; protective ROHs associations in green. Blue regions represent priorizated ROHs from inbred AD cases. Orange segments represent priorizated regions harboring potential recessive variants.

APOE, the strongest genetic risk factor for AD

The *APOE ε4* is the strongest genetic risk factor for AD. Here, we addressed its effect and utility in early stages of the disease (**Section 3.1.1**). Next, we investigated genetic forces disturbing *APOE* genetic region (**Section 3.1.2**).

The early effect of APOE

In **Section 3.1.1**, we evaluated whether *APOE* alleles were associated with SCD status and with cerebral Aβ accumulation in the FACEHBI cohort.

We showed that SCD individuals from the FACEHBI study are enriched in *APOE Ɛ4* and *APOE Ɛ2* alleles. This fact has important implications. The increased frequency of the *APOE Ɛ4* allele in preclinical AD supports the hypothesis that this subpopulation represents an enriched subgroup of at-risk individuals for AD. Conversely, the increased allele frequency of the *APOE Ɛ2* was an unexpected finding, considering the protective effect attributed to *Ɛ2* in AD (72). We believe that this enrichment could be caused by the presence of individuals who might convert to VaD, instead of dementia due to AD. Thus, *APOE* patterns at the preclinical stage can be used instrumentally, to enrich subsets of individuals with high probability of conversion to AD.

In an effort to further characterize our population, we also tested the association of *APOE* alleles and the cerebral Aβ endophenotype. We showed that increased burden of cerebral Aβ was mainly driven by *APOE Ɛ4* carriers, which is in accordance with other studies (153), (154). *APOE* dosage explained 11% of the variance for this trait. Despite that, there is a large part of the trait variance which remains unexplained (155), (156). We then expanded the study to analyze this correlation in the entire AD spectrum (SCD, MCI and AD). After doing so, we detected the highest correlation between cerebral Aβ and *APOE* dosage in late MCI stages, instead of AD. This effect might be explained by the influence of atrophy in AD brains (157). To counteract it, and to increase the statistical power of analyses, we recommend the use of the entire spectrum of AD subjects, i.e. SCD, MCI and AD individuals, to conduct large GWAS for the cerebral Aβ endophenotype. Given that, adjustments according to clinical classification would be pertinent.

The expansion of genetic studies beyond the AD phenotype is helping to: 1) identify AD progression genetic factors; 2) validate new proposed AD endophenotypes; and 3) select and follow up on individuals with high probability of conversion to AD. Despite that, the availability of preclinical cohorts with complete genetic or biomarker information is scarce. Further efforts are required to generate large preclinical cohorts with multiscale data.

The molecular characterization of individuals at preclinical stages has important implications for the design of clinical trials, and prevention strategies, and for the evaluation of the responses to treatments. In that sense, genetics could be used as a strategy to enrich preclinical sets of AD individuals.

The APOE curse

In **Section 3.1.2,** we introduced the concept of the *APOE* curse, or "the impossibility of determining whether additional AD loci truly exist around APOE." Since its discovery 26 years ago, controversial genetic findings have emerged in genes neighboring *APOE* (74),(158), (76), (77). Here, we investigated whether LRLD patterns are affecting genetic discoveries at this region, in chromosome 19. Thus, we designed a replication study for a well-known AD marker, *ABCA7*-rs4147929 (115), located in the opposite arm of chromosome 19, where LRLD with *APOE* is impossible; and a controversial AD marker, *CD33*-rs3865444 (115), located near to *APOE,* in the long arm of chromosome 19.

Section 3.3 reinforced the role of *ABCA7*-rs4147929 in AD, which is in line with prior studies (104), (126). Conversely, inconsistent results were observed for the *CD33*-rs3865444 marker during this study. Our replication study did not support a role for rs3865444 signal in AD. However, the meta-GWAS strategy from **Section 3.3** showed a GWS association for this marker. This genome-wide association was also observed by another group (114).

With the purpose of revealing the causes of these inconsistencies, we evaluated whether LRLD was present between *APOE* markers (rs7412 and rs429358) and *CD33* SNP (rs3865444). We detected a subtle and non-universal LRLD between *APOE ε2* and *CD33*-rs3865444 (*D'* = 0.3), which meant that a fraction, roughly 30% of individuals, in the population exhibited this LRLD.

Forces increasing long range LD patterns are diverse, i.e. mutations, genetic drift, genetic flow, chromosome inversions, and natural selection, including hitchhiking effect and epistasis (159). The effect of population-specific natural selection patterns can lead to allele micro-heterogeneity. Thus, we also explored the effect of population-specific LRLD patterns, but our data did not support an evident population substructure in the studied cohorts. Despite that, it cannot be ruled out, because it is well-known that current methods are unable to detect it (160). We also discarded an inbreeding effect affecting unequally tested populations. A deeper exploration of the forces creating LRLD across the *APOE* region will be required to definitively discard or support our findings. Sequencing of the entire region, centered on the search for local rearrangements near *APOE*, will thus be advisable.

This study represents a proof of concept, suggesting that LRLD might be present in the *APOE* region. This observation, when confirmed, could contribute to elucidating the role of *APOE* in longevity and in disease. As long as it remains unknown, the *APOE* curse will impair the confirmation of additional loci surrounding the major genetic risk factors for AD.

Strategies for gene hunting in Alzheimer's disease

Here, we have addressed several hypotheses that supported our objective of making new genetic discoveries related to AD. First, we tested whether genes related to neuronal apoptosis were potentially linked with AD (**Section 3.2**). Next, we moved to an agnostic approach, using genome-wide data (GWAS). At this level, we explored the role of 1) phenotypic and 2) genetic heterogeneity in AD (**Section 3.3** and **Section 3.4**).

Genes related with neuronal apoptosis in Alzheimer's disease

Neuronal apoptosis, a process of programmed cell death, is mainly mediated by pro-inflammatory molecules through two major pathways: mitochondria or DRs. TNF-α and Fas-L are identified as key effectors activating DRs (161). cFLIP, BIRC3 or FAIM anti-apoptotic proteins are crucial to the efficient control of apoptosis, as they can inhibit TNF- α and Fas-L mediated apoptosis. In that context, the Cell Signaling and Apoptosis Lab of VHIR observed a decrease of FAIM proteins in AD brains. This finding was the main driver motivating the present study. At this point, I, as master student in the lab, and in collaboration with Fundació ACE began a candidate gene study of apoptosis-related genes. However, our findings did not support the association of these genes with AD.

Most AD candidate gene studies have failed to discover of new genes (see **Chapter 1**). For a complex disease like AD, where the disease mechanic remains mainly unknown or poorly understood, I believed that the use of agnostic approaches can be more informative for uncovering AD etiology.

Exploring phenotypic heterogeneity in Alzheimer's disease

In **Section 3.3**, we, as part of GR@ACE and the DEGESCO consortia, performed the largest Spanish GWAS for dementia to date. This study was launched simultaneously with large AD meta-GWAS, i.e. Kunkle et al. (N = 94,437) and Jansen et al. (N = 455,258). Thus, rather than an effort to increase the sample size, and consequently the statistical power, our objective was to explore the impact of AD's clinical heterogeneity in genetic findings, and in biological routes leading to dementia.

Alzheimer's disease is a heterogeneous neuropathological disorder. Apart from the typical neuropathological hallmarks of AD (see **Chapter 1**), most AD patients have mixed pathologies, with vascular damage being highly prevalent in AD cases (11). Viswanathan et al. (162) suggested that dementia could be considered a gradient of vascular and neurodegenerative

components. We believed that, considering this scenario from a molecular point of view could be a useful strategy for disentangling disease pathways. Thus, we classified our group of dementia cases to cover the Viswanathan spectrum (162), generating five clinical endophenotypes of dementia. Next, we explored the effect of known LOAD genetic variants across dementia subgroups, and investigated whether differential biological pathways were operating across the different subgroups of detected loci.

Notably, these categories let us identify different biological pathways for dementia. Immune-related pathways were detected across all categories, but pathways related to vasculature alterations were only detected in the purest forms of clinical AD.

There is considerable evidence supporting the role of vasculature in AD. First, the accumulation of amyloid beta in the wall of cerebral vessel, CAA, is a recognized hallmark of AD (8). Capillary and non-capillary types of CAA have been defined, being capillary CAA, more frequently found in AD. Second, mutations in the *APP* gene cause both the Mendelian form of CAA (163) and early-onset AD (40). This genetic overlap confirms a role for the *APP* locus in vascular amyloidosis. Third, the *APOE Ɛ4* allele and markers in the *CR1* gene have been associated with CAA. Particularly, *APOE Ɛ4* is related with capillary CAA (164). Fourth, knock-out models of the *APOE* gene cause atherosclerotic lesions in blood vessels (165), (166), (167). Although this might be an indirect effect of the *APOE* gene in the regulation of cholesterol metabolism (168), it should be remembered that cholesterol metabolism is an AD-related pathway (115). It has recently been demonstrated that cholesterol efflux efficiently controls angiogenesis (169). Finally, vascular traits, such as arteriosclerosis and atherosclerosis, are associated with AD risk (12). Taken together, these pieces of evidence reinforce the causative role of vascular alterations in AD. We believe this role is carried out via two potential pathways: the intrinsic vascular damage mediated by the amyloid, and the accumulation of more systemic age-related vascular insults, leading to dementia.

Given that CAA represents the unique identified link between the vascular and amyloid hypotheses, investigating its role in disease mechanics seems pertinent. The concept of a “vessel on a chip” has recently been expanded to generate a tridimensional artificial model of CAA, which includes endothelial cells, smooth muscle cells, and astrocytes, enabling further study of CAA biology (170). Using this model, it would be possible to study the role of known AD genetic variants in CAA, for example by their gene editing, as well as by silencing or knockout of target genes. It is reported that Aβ40 has a direct effect on vessels, promoting vasoconstriction (171). Taking into account that CAA is characterized by the accumulation of Aβ40 in the wall of the vessels, exploring its effects in the context of AD would be useful. Finally, the monitoring of biomarkers for vascular damage in preclinical individuals could help

to test which alteration happens first *in vivo*, vascular or amyloid plaques. Despite that, the National Institute of Aging-Alzheimer's Association (NIA-AA) has not yet integrated measures of vascular dysfunction into the proposed core AD biomarkers (172).

Deep understanding of heterogeneity in AD seems necessary to design better genetic studies. Our meta-GWAS strategy revealed two new GWS signals, markers in the *ANKDR31* and *TP53INP1/NDUFAF6* genes (mentioned below), which are highly dependent on the group of AD cases, or the clinical endophenotype. In addition, two additional signals were confirmed in our data; *SCIMP*-rs7225151 and *CD33*-rs3865444, which has been mentioned above.

For example, we were able to detect the *ANKDR31* signal using all the dementia cases, instead of group of probable AD patients alone. The *ANKDR31* gene has been related with neurodevelopmental disorders (173), and very recently, with longevity (174). Our *ANKRD31* GWAS signal was mapped to a brain eQTL in a lncRNA, located 1.6kb from the *HMGCR* locus and residing in the *COL4A3BP* gene. The *HMGCR* gene is the most important co-regulator of cholesterol synthesis, and is with LD with our marker. Polymorphism in the *HMGCR* locus has been previously associated with AD and age at onset of the disease (175). Moreover, cholesterol metabolism is among top pathways linked to AD pathology (115). Our post-GWAS analysis, which also included results from eQTLs, chromatin interactions and gene-based analysis pointed to the *POLK* gene as a potential causal gene in this region. Markers in the *POLK* locus have previously been linked with AD and LDL plasma levels (176). Considering prior findings, our results are consistent with the role of this genomic region in mixed dementia, although independent replications are required. In that context, deeper post-GWAS analysis will ultimately be needed to pinpoint the causal variant and the causal gene.

The performance of large meta-GWAS in neuro-pathologically confirmed cases would be the next step in confirming specific disease pathways and new GWS findings. In that scenario, conducting genetic studies for each dementia-associated neuro-histopathological hallmark, could provide a holistic way to construct "gene categories". Next, we could test the association of genes per category with different types of dementia, or even with different types of neurodegenerative disease, using a Mendelian randomization strategy. These strategies would be useful: 1) to redefine the genuine effect of known AD loci; 2) to better understand the biological complexity operating in AD; and 3) to find new loci linked with specific dementia axes. In addition, testing the genetic overlap between specific gene categories would be recommended to explore the shared biology between neurodegenerative diseases.

The search of AD recessive alleles

In **Section 3.4**, we looked for a footprint of recessive inheritance in AD. With that aim and using an outbred population, we conducted the largest ROH study for AD to date (N = 21,100). First, we explored whether homozygosity parameters, i.e. NROH, SROH, AVROH and FROH, were associated with AD respect to controls, and investigated the role of consensus ROHs in the whole sample. Next, we prioritized homozygosity segments overrepresented in consanguineous AD cases to look for potential recessive variants.

Our results show that there is an excess of homozygosity in AD cases respect to controls, supporting the claim that consanguineous AD patients are overrepresented in this population. This reinforces the role of recessive inheritance in AD, and creates a new scenario for studying AD genetics beyond the classical GWAS studies or the search of dominant loci.

Previous ROH studies in AD for European and non-European populations have shown inconsistent results (149),(147),(150),(148). When ROH analysis is performed from an outbred population, several technical considerations should be taken into account: 1) ROH patterns across the human population; 2) sample size; and 3) individual ROH length. ROH patterns are highly dependent of local population history and cultural practices. Large sample size is required to detect an excess of homozygosity from an outbred population (151). However, previous ROH studies of AD have used modest samples (N < 6,000) (149), (147),(150),(148), compromising their statistical power. Second, in **Section 3.4**, we demonstrated that establishing the individual length of ROH to 1 Mb, the threshold used in previous AD studies, inflates homozygosity parameters, complicating their interpretation. It is suggested that individual ROHs < 1.5 Mb might reflect LD patterns of ancient origin, rather than the consanguineous cultural practices and genetic isolation captured with ROHs ≥ 1.5 Mb (177). This, could, in part, explain previous inconsistent results in the field (149), (147). Hence, we decided to conduct our analysis by redefining the length of individual ROH segments from 1 Mb to 1.5 Mb, to better capture the consanguinity component.

ROH analysis in the whole sample revealed a homozygosity region of 115.9 Kb (chr4:11,189,482–11,305,456), in the proximity of the *HS3ST1* gene. GWS AD markers have been mapped near to *HS3ST1* gene (rs6448807, chr4: 11,676,144; and rs13113697; chr4:11,711,232) (107),(178). Different brain expression of the *HS3ST1* gene has been reported between AD cases with respect to controls (178). To date, however, the causative mutation and the molecular mechanisms of this association remain elusive. We believe that a recessive acting variant located in the homozygosity segment could explain the reported associations. High-resolution mapping in this region would be required to confirm this rising hypothesis.

Accordingly, to increase our chance to detect a recessive variant, we prioritized the homozygosity segments based exclusively on the homozygosity map observed in inbred AD cases. First, we efficiently captured inbred individuals from an outbred population, offering a new framework for analyzing the inbreeding component in AD. Next, we selected candidate regions to explore them in the context of whole-exome sequencing data. Our prioritization strategy is outlined in **Section 3.4**. This strategy let us select eight rare (MAF <1%) and homozygous variants located in candidate ROHs, carried exclusively by seven AD cases (see **Section 3.4**). We highlighted the potential role of a novel variant in the *SPON1* gene. SPON1 protein inhibits APP cleavage through β-secretase (179), probably influencing the amyloid cascade. The amyloid cascade has been previously described in **Chapter 1**. Intriguingly, it has been reported that SPON1 protein also binds to APOE receptors (180), and *SPON1* genetic variants have been related with dementia severity (181). Considering previous data, the presence of recessive acting-variants in partners of APP would be biologically plausible. We believe that the *SPON1* gene could be considered an interesting candidate, which deserves future confirmatory resequencing efforts.

It must be noted that, we did not explore whether inbred AD cases present differential clinical characteristics respect to non-inbred cases. However, we believe that their deeper characterization would be of interest in future genotype-phenotype correlations.

Here, we also detected a direct correlation between the ROH number and the total ROH length, from peripheral blood, with age. Considering that clonal mosaicism can emerge as age-dependent copy number variations (CNVs) (182), and that CNVs, specifically hemizygous deletions, can be captured in ROH studies, we believe this finding might be explained by the presence of genome instability in aged populations, which has previously been reported (182). Thus, it supports the idea that, in late-onset disease, ROHs can be caused by a genuine inbreeding component, but also by the increasing genome instability intrinsic to the aging process. The integration of a refined map of CNVs in our data, or perhaps the crossing of ROHs, CNVs, and a Hardy-Weinberg disequilibrium maps, would be the next step in differentiating both types of segments. Conducting a GWAS for homozygosity parameters could be considered an additional strategy to distinguish both types of segments. In that context, we believe that results for AVROH and results for NROH could differentiate these two aspects of ROHs, the homozygosity component and the somatic instability, respectively. New questions also arise from this finding, e.g. how many of the previously identified AD regions can represent regions of somatic instability? Although we are unable to answer this yet, tissue-specific, or cell-specific whole sequencing could help us to address this emerging concept in AD research.

Clonal mosaicism could also explain why we found a higher number of ROHs in our study than that expected for European populations (183). From our point of view, although a signature of genome instability in ROH studies might exist, our adjusted association results, which consider the age effect, still support the idea that inbred individuals are overrepresented in the AD population as compared to controls. Controlling the role of genome instability for late-onset neurodegenerative disease is a challenge, due to the impossibility of detecting true converters to AD before disease onset and the difficulties of collecting biological information from the target tissue.

Here, we demonstrated the effect of inbreeding in AD and efficiently captured the subset of consanguineous cases nested in our population. The opportunity to explore complementary sequencing data provided a subset of potential candidates harboring recessive variants. However, a larger collection of individuals with GWAS and sequencing data would be needed to confirm our findings. Altogether, ROH analyses have provided an alternative strategy to uncover the genetic heterogeneity underlying Alzheimer's disease.

New Findings in TP53INP1/NDUFAF6 genomic region

In this research, *NDUFAF6* signals reached genome-wide significance for the first time (**Section 3.3**). Our GWAS top hit (rs10098778 in chr8:95,992,020) supports the potential AD involvement of *TP53INP1* and *NDUFAF6* loci. Intriguingly, we observed that our top marker resides in a consensus ROH (chr8:95,843,111–chr8:96,071,272) detected in 32 AD cases and 20 controls from the whole sample (**Section 3.4**). When ROH analysis was conducted only with inbred AD cases, the consensus ROH was expanded from chr8:95,578,201 to chr8:96,071,272. At this point, we were able to identified two rare variants, rs73263258 (chr8:95,658,495) in the *ESPR1* gene and rs138998679 (chr8:95,780,656) in the *DPY19L4* gene, carried by the same individual in a long ROH segment (5.4 Mb) across the region (**Figure 2**).

The candidate region includes six genes (*ESPR1*, *DYP19L4*, *INTS8*, *CCNE2, TP53INP1* and *NDUFAF6*). Our post-GWAS analysis further supports a role for *TP53INP1* or *NDUFAF6* genes. Several mechanisms could be explaining detected signals at this locus. For example, a synthetic association could be causing a GWAS signal. Conversely, several genetic mechanisms mediating the risk to AD could be operating at the same locus, e.g. additive and recessive models or even regional genetic instability. Detailed re-exploration of this region will be required to elucidate the causal mechanism.

Extended haplotype analysis in a set of AD cases could help to investigate whether our GWAS marker is linked with *ESPR1- DPY19L4* detected variants. Subsequently, a recessive model for

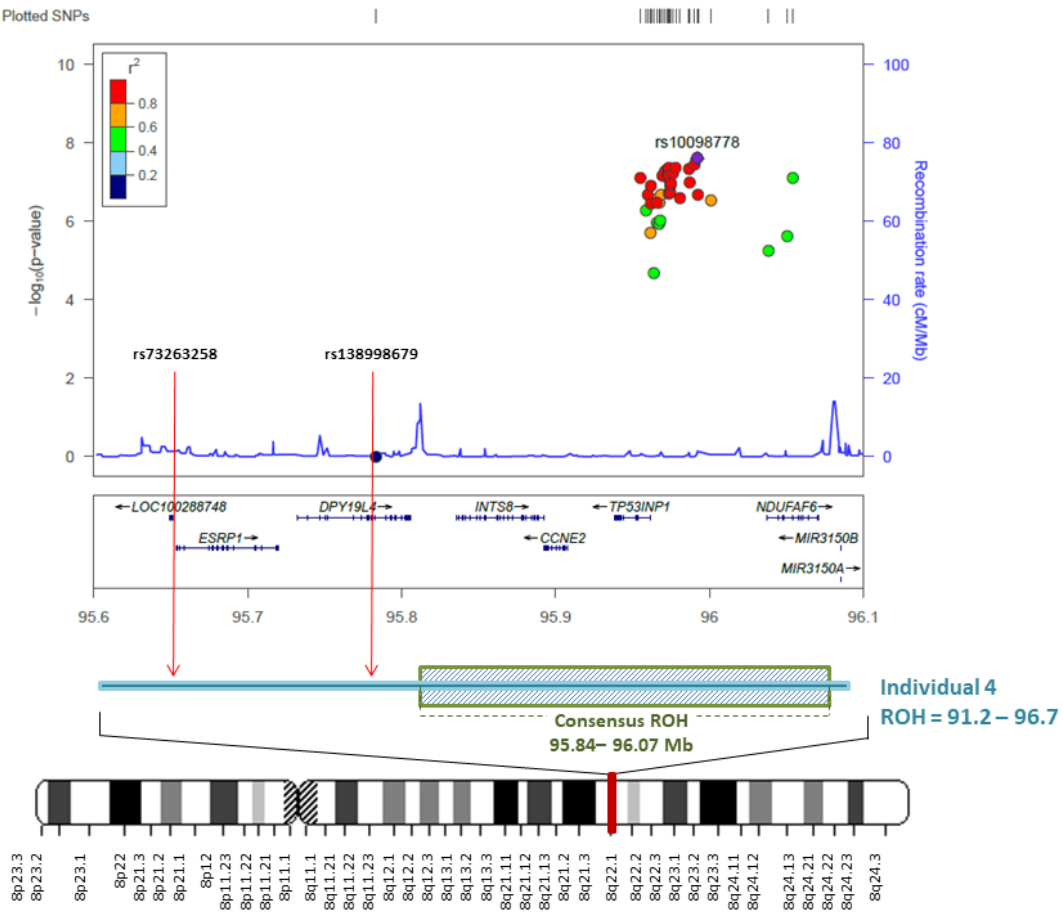


FIGURE 2. Genetic findings in the *TP53INP1/NDUFAF6* genetic region.
LocusZoom plot shows GWS markers identified at this work. Homozygosity tract is showed in blue. Red arrows points to candidate recessive variants at this region.

the variants in that haplotype could be tested in a cases vs control dataset. Conducting the analysis specifically for the Spanish population should be considered to help identify casual mechanisms. The Spanish population is genetically more diverse than western European cohorts. For example, our lead GWAS marker is in high LD with *NDUFAF6*-rs4735340, the top suggestive signal reported by IGAP (115) in this region (r^2 = 0.95, for CEU population). Despite that, there are subtle differences in LD estimates for the Iberian population (r^2 = 0.87). We believe that substructure of the Spanish population could help to fine-mapping of this AD region.

What to do now? Old questions new data

What is the causal variant? What is the target gene? Thousands of sentinel variants have been linked to complex disease or traits using GWAS. Despite that, the causative variants and/or the

target genes remain mostly unknown. Thus, GWAS is a successful approach to delineate a genomic region, but this is just the beginning. In that context, intense post-GWAS studies are required. In this new post-GWAS era, extensive population cohorts should be selected for deep multi-omics and single cell genotyping and phenotyping.

Our work shows the need for generating complementary data coming from GWAS and sequencing technologies to explore recessiveness. Without the generation of complementary data, narrowing down from a genomic locus to a single variant seems really intricate. In addition, I believe that AD genetics needs to recover basic genetic concepts, and expand them beyond the simple strategy to increase the sample size. Considering evolutionary forces affecting human late disease seems necessary to better understanding AD genetics.

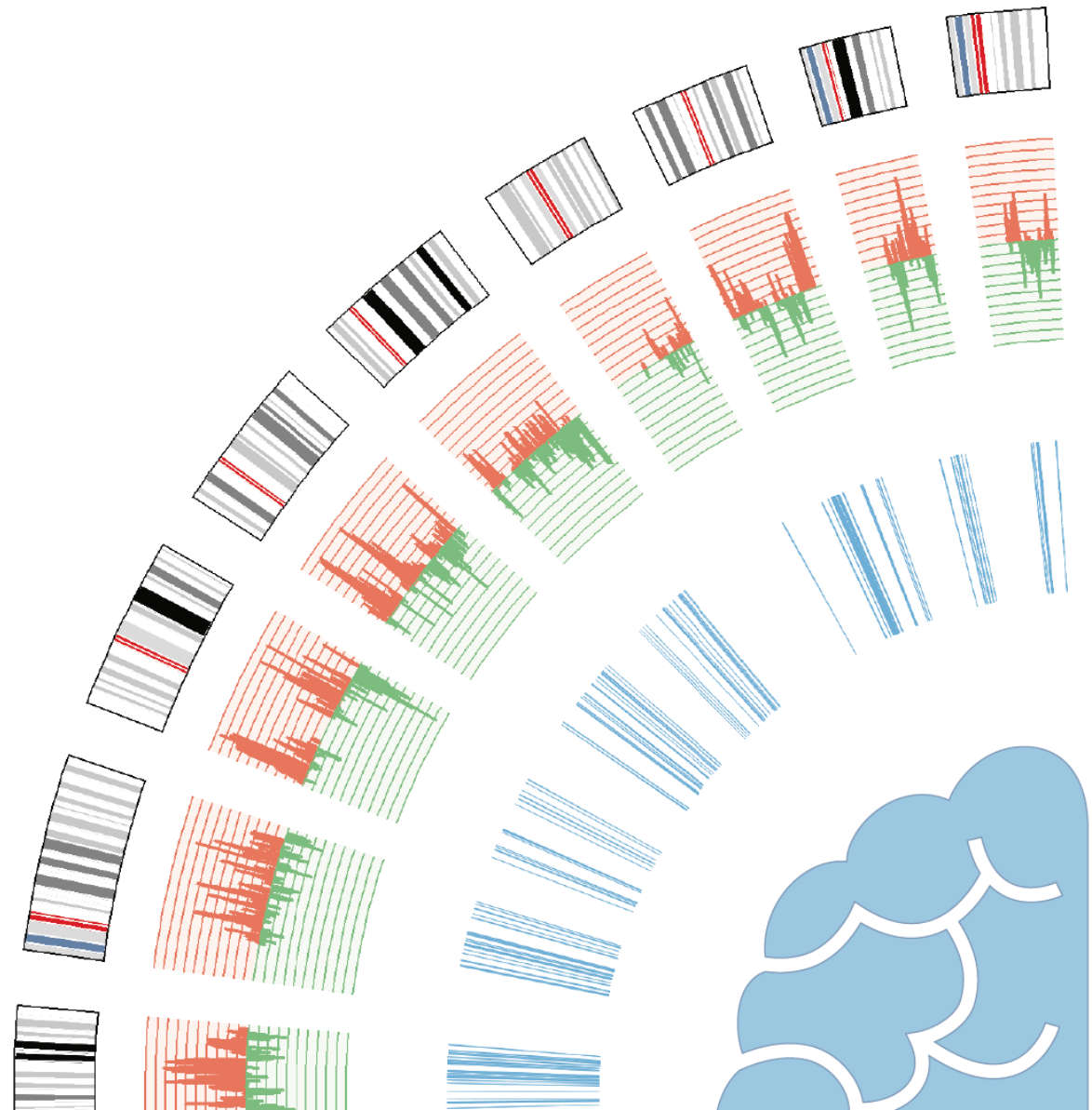
Which biomarker? Could be genetic? In the context of Mendelian AD, genetic testing for *APP, PSEN1* or *PSEN2* mutations is discriminative. Despite that, most AD patients do not present these mutations. The molecular characterization of preclinical stages of AD can be really useful to establish a plasma or CSF biomarker or a composite of biomarkers, where known AD genetic variants or polygenic risk scores can have their space. In fact, the selection of *APOE Ɛ4* carriers is already a very common strategy to enrich preclinical cohorts in AD cases.

How to translate the biological complexity for a disease in bits of useful information? McClellan and King wrote, in an interesting review about the genetic heterogeneity of human disease, (117): "The degree of biological complexity underlying a phenotype is an excellent predictor of locus heterogeneity." The case of *BRCA1* and *BRCA2* genes, where thousands of mutations for breast cancer have been detected, provides a good example (117). AD is probably not different, and the number of common and rare genetic variants mediating disease risk could be overwhelming. In that scenario, searching out common biological pathways for multiple variants might shed light on the disease mechanism.

This thesis contributes to the elucidation of new loci and the proposal of new core pathways. It brings new possibilities to the genetic field of AD, considering the recessive component for this disease. Despite that, giving them a deeper biological significance will require an immense effort.

CHAPTER 5

CONCLUSIONS



1. The generation of clinical endophenotypes of dementia led us to detect three categories of AD genes. These categories allowed us to identify differential biological routes leading to dementia.

2. Vasculature regulation may be an essential part of the causative mechanism in pure forms of AD.

3. The meta-analysis of GR@ACE with additional GWAS datasets revealed two novel AD loci, the *ANKRD31*-rs4704171 (in the *HMGCR* genomic region) and *NDUFAF6*-rs10098778.

4. Our GWAS strategy confirms *SCIMP*-rs7225151 and *CD33*-rs3865444. Both loci have obtained uncontroversial genome-wide statistical significance during this study.

5. Homozygosity component was overrepresented in AD cases respect to controls. Our results suggest that recessive effects may explain a fraction of the AD heritability.

6. We efficiently captured consanguineous AD cases from an outbred population, which supposes a refined method to analyze inbreeding in AD.

7. The *SPON1* gene is a new candidate AD locus with a potential recessive component, as it is a biological partner of APP.

8. Using a homozygosity mapping approach, we detected potential recessive variants in the *HS3ST1* and *TP53INP1/NDUFAF6* loci, previously associated with AD using conventional GWAS strategies.

9. The *APOE Ɛ4* is a genetic risk factor to suffer SCD. The FACEHBI sample presents an enrichment of *APOE* ε4 carriers, suggesting that a pool of AD cases is nested in this subpopulation.

10. The *APOE* gene explains 11% of brain amyloid variance, suggesting the role of additional genetic or epigenetic factors in modulating this trait.

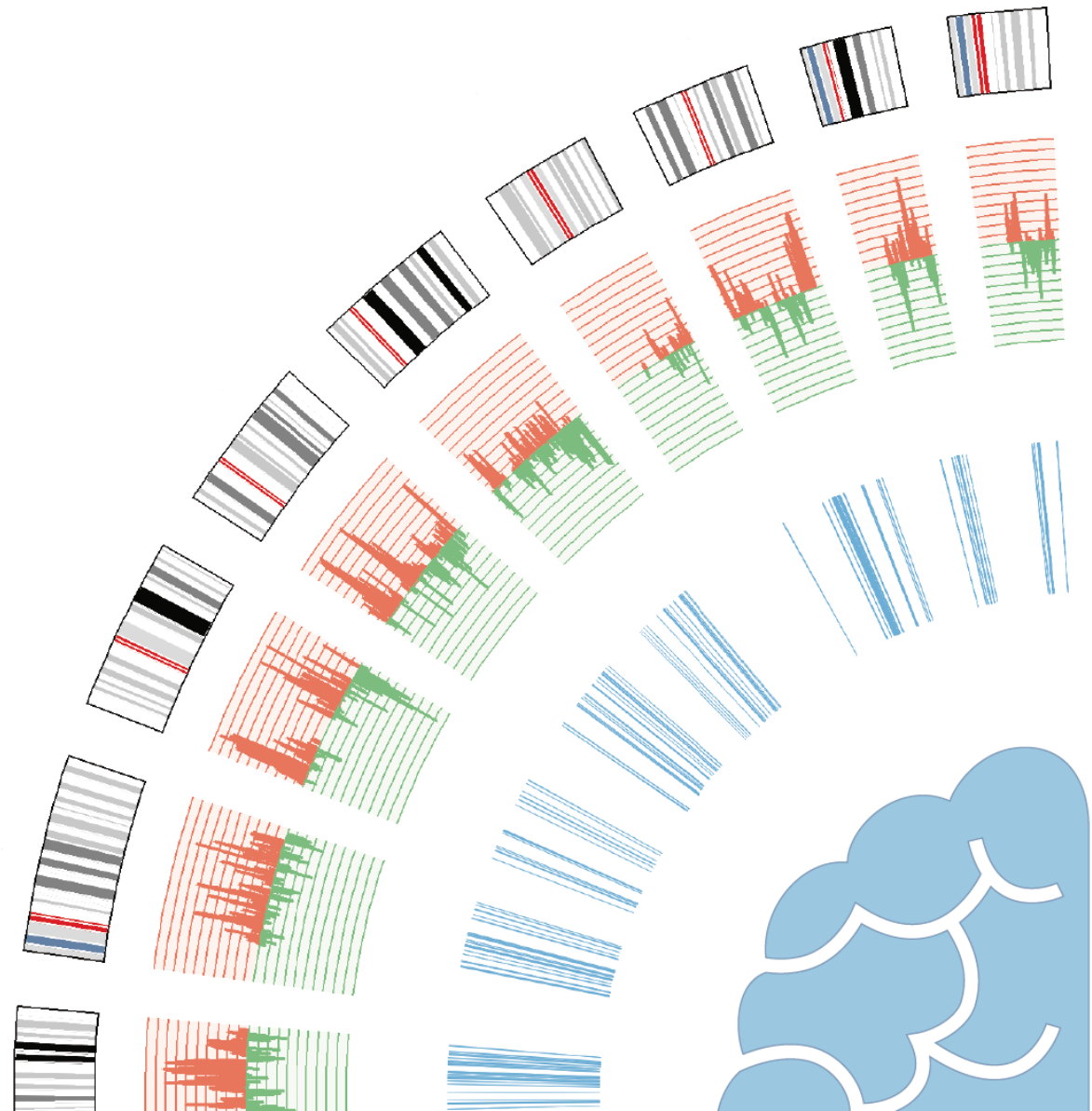
11. Results emerging from this thesis reinforce the involvement of the *ABCA7*-rs4147929 in AD.

12. Non-universal and weak LRLD is present between *APOE*-rs7412 and *CD33*-rs3865444 markers. The forces originating LRLD in this region remain unknown.

13. Our data does not show a link between genes related with neuronal apoptosis and AD.

CHAPTER 6

REFERENCES



1. World Health Organisation. The Global strategy and action plan on ageing and health. WHO. 2018.

2. 2019 Alzheimer's disease facts and figures. Alzheimer's Dement. 2019 Mar 1;15(3):321–87.

3. Alzheimer's Association. Alzheimer's & Dementia: Global Resources. Alzheimer's Dement Glob Resour. 2018.

4. World Health Organisation. World Alzheimer report 2012, A public health priority. 2012.

5. 2018 Alzheimer's disease facts and figures. Alzheimer's Dement. 2018 Mar 1;14(3):367–429.

6. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011 May;7(3):263–9.

7. Alzheimer's Disease International. World Alzheimer Report 2009. 2009.

8. Serrano-Pozo A, Frosch MP, Masliah E, Hyman BT. Neuropathological Alterations in Alzheimer Disease. Cold Spring Harb Perspect Med. 2011 Sep 1;1(1):a006189–a006189.

9. Schneider JA, Arvanitakis Z, Bang W, Bennett DA. Mixed brain pathologies account for most dementia cases in community-dwelling older persons. Neurology. 2007 Dec 11;69(24):2197–204.

10. Schneider JA, Arvanitakis Z, Leurgans SE, Bennett DA. The neuropathology of probable Alzheimer disease and mild cognitive impairment. Ann Neurol. 2009 Aug;66(2):200–8.

11. Toledo JB, Arnold SE, Raible K, Brettschneider J, Xie SX, Grossman M, et al. Contribution of cerebrovascular disease in autopsy confirmed neurodegenerative disease cases in the National Alzheimer's Coordinating Centre. Brain. 2013 Sep 1;136(9):2697–706.

12. Arvanitakis Z, Capuano AW, Leurgans SE, Bennett DA, Schneider JA. Relation of cerebral vessel disease to Alzheimer's disease dementia and cognitive function in elderly people: a cross-sectional study. Lancet Neurol. 2016 Aug;15(9):934–43.

13. Baumgart M, Snyder HM, Carrillo MC, Fazio S, Kim H, Johns H. Summary of the evidence on modifiable risk factors for cognitive decline and dementia: A population-based perspective. Alzheimer's Dement. 2015 Jun;11(6):718–26.

14. Sando SB, Melquist S, Cannon A, Hutton M, Sletvold O, Saltvedt I, et al. Risk-reducing effect of education in Alzheimer's disease. Int J Geriatr Psychiatry. 2008 Nov;23(11):1156–62.

15. Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi HM. Forecasting the global burden of Alzheimer's disease. Alzheimer's Dement. 2007 Jul;3(3):186–91.

16. Jack CR, Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol. 2010 Jan;9(1):119–28.

17. Jessen F, Amariglio RE, van Boxtel M, Breteler M, Ceccaldi M, Chételat G, et al. A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. Alzheimers Dement. 2014 Nov;10(6):844–52.

18. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. Arch Neurol. 1999 Mar;56(3):303–8.

19. Lautenschlager NT, Cupples LA, Rao VS, Auerbach SA, Becker R, Burke J, et al. Risk of dementia among relatives of Alzheimer's disease patients in the MIRAGE study: What is in store for the oldest old? Neurology. 1996 Mar;46(3):641–50.

20. Goldman JS, Hahn SE, Catania JW, LaRusse-Eckert S, Butson MB, Rumbaugh M, et al. Genetic counseling and testing for Alzheimer disease: joint practice guidelines of the American College of Medical Genetics and the National Society of Genetic Counselors. Genet Med. 2011 Jun;13(6):597–605.

21. Jarmolowicz AI, Chen H-Y, Panegyres PK. The patterns of inheritance in early-onset dementia: Alzheimer's disease and frontotemporal dementia. Am J Alzheimers Dis Other Demen . 2015 May 21;30(3):299–306.

22. Campion D, Dumanchin C, Hannequin D, Dubois B, Belliard S, Puel M, et al. Early-Onset Autosomal Dominant Alzheimer Disease: Prevalence, Genetic Heterogeneity, and Mutation Spectrum. Am J Hum Genet . 1999 Sep;65(3):664–70.

23. Van Cauwenberghe C, Van Broeckhoven C, Sleegers K. The genetic landscape of Alzheimer disease: clinical implications and perspectives. Genet Med . 2016 May 27;18(5):421–30.

24. Pottier C, Hannequin D, Coutant S, Rovelet-Lecrux A, Wallon D, Rousseau S, et al. High frequency of potentially pathogenic SORL1 mutations in autosomal dominant early-onset Alzheimer disease. Mol Psychiatry . 2012 Sep 3;17(9):875–9.

25. Nicolas G, Charbonnier C, Wallon D, Quenez O, Bellenguez C, Grenier-Boley B, et al. SORL1 rare variants: a major risk factor for familial early-onset Alzheimer's disease. Mol Psychiatry . 2016 Jun 25;21(6):831–6.

26. Cuyvers E, De Roeck A, Van den Bossche T, Van Cauwenberghe C, Bettens K, Vermeulen S, et al. Mutations in ABCA7 in a Belgian cohort of Alzheimer's disease patients: a targeted resequencing study. Lancet Neurol . 2015 Aug;14(8):814–22.

27. Le Guennec K, Nicolas G, Quenez O, Charbonnier C, Wallon D, Bellenguez C, et al. *ABCA7* rare variants and Alzheimer disease risk. Neurology . 2016 Jun 7;86(23):2134–7.

28. De Roeck A, Van den Bossche T, van der Zee J, Verheijen J, De Coster W, Van Dongen J, et al. Deleterious ABCA7 mutations and transcript rescue mechanisms in early onset Alzheimer's disease. Acta Neuropathol. 2017 Sep 27];134(3):475–87.

29. Wingo TS, Lah JJ, Levey AI, Cutler DJ. Autosomal recessive causes likely in early-onset Alzheimer disease. Arch Neurol . 2012 Jan;69(1):59–64.

30. Bergem AL, Engedal K, Kringlen E. The role of heredity in late-onset Alzheimer disease and vascular dementia. A twin study. Arch Gen Psychiatry . 1997 Mar;54(3):264–70.

31. Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S, et al. Role of genes and environments for explaining Alzheimer disease. Arch Gen Psychiatry . 2006 Feb 1;63(2):168–74.

32. Brainstorm Consortium V, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, et al. Analysis of

shared heritability in common disorders of the brain. Science . 2018 Jun 22;360(6395):eaap8757.

33. Glenner GG, Wong CW. Alzheimer's disease: initial report of the purification and characterization of a novel cerebrovascular amyloid protein. Biochem Biophys Res Commun . 1984 May 16;120(3):885–90.

34. Glenner GG, Wong CW. Alzheimer's disease and Down's syndrome: sharing of a unique cerebrovascular amyloid fibril protein. Biochem Biophys Res Commun . 1984 Aug 16;122(3):1131–5.

35. Masters CL, Simms G, Weinman NA, Multhaup G, McDonald BL, Beyreuther K. Amyloid plaque core protein in Alzheimer disease and Down syndrome. Proc Natl Acad Sci. 1985 Jun;82(12):4245–9.

36. Tanzi RE, Gusella JF, Watkins PC, Bruns GA, St George-Hyslop P, Van Keuren ML, et al. Amyloid beta protein gene: cDNA, mRNA distribution, and genetic linkage near the Alzheimer locus. Science. 1987 Feb 20;235(4791):880–4.

37. Kang J, Lemaire HG, Unterbeck A, Salbaum JM, Masters CL, Grzeschik KH, et al. The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor. Nature. 1987 Feb;325(6106):733–6.

38. Robakis NK, Wisniewski HM, Jenkins EC, Devine-Gage EA, Houck GE, Yao XL, et al. Chromosome 21q21 sublocalisation of gene encoding beta-amyloid peptide in cerebral vessels and neuritic (senile) plaques of people with Alzheimer disease and Down syndrome. Lancet. 1987 Feb 14;1(8529):384–5.

39. Goldgaber D, Lerman MI, McBride OW, Saffiotti U, Gajdusek DC. Characterization and chromosomal localization of a cDNA encoding brain amyloid of Alzheimer's disease. Science. 1987 Feb 20;235(4791):877–80.

40. Goate A, Chartier-Harlin MC, Mullan M, Brown J, Crawford F, Fidani L, et al. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. Nature. 1991 Feb 21;349(6311):704–6.

41. Chartier-Harlin M-C, Crawford F, Houlden H, Warren A, Hughes D, Fidani L, et al. Early-onset Alzheimer's disease caused by mutations at codon 717 of the β-amyloid precursor protein gene. Nature. 1991 Oct 31;353(6347):844–6.

42. Murrell J, Farlow M, Ghetti B, Benson MD. A mutation in the amyloid precursor protein associated with hereditary Alzheimer's disease. Science. 1991 Oct 4;254(5028):97–9.

43. Mullan M, Crawford F, Axelman K, Houlden H, Lilius L, Winblad B, et al. A pathogenic mutation for probable Alzheimer's disease in the APP gene at the N-terminus of beta-amyloid. Nat Genet. 1992 Aug;1(5):345–7.

44. Tanzi RE, George-Hyslop PHS, Haines JL, Polinsky RJ, Nee L, Foncin J-F, et al. The genetic defect in familial Alzheimer's disease is not tightly linked to the amyloid β-protein gene. Nature. 1987 Sep;329(6135):156–7.

45. Van Broeckhoven C, Genthe AM, Vandenberghe A, Horsthemke B, Backhovens H, Raeymaekers P, et al. Failure of familial Alzheimer's disease to segregate with the A4-amyloid gene in several European families. Nature. 1987 Sep;329(6135):153–5.

46. Schellenberg GD, Bird TD, Wijsman EM, Orr HT, Anderson L, Nemens E, et al. Genetic linkage evidence for a familial Alzheimer's disease locus on chromosome 14. Science. 1992 Oct 23;258(5082):668–71.

47. St George-Hyslop P, Haines J, Rogaev E, Mortilla M, Vaula G, Pericak-Vance M, et al. Genetic evidence for a novel familial Alzheimer's disease locus on chromosome 14. Nat Genet. 1992 Dec;2(4):330–4.

48. Van Broeckhoven C, Backhovens H, Cruts M, De Winter G, Bruyland M, Cras P, et al. Mapping of a gene predisposing to early-onset Alzheimer's disease to chromosome 14q24.3. Nat Genet. 1992 Dec;2(4):335–9.

49. Bird TD, Lampe TH, Nemens EJ, Miner GW, Sumi SM, Schellenberg GD. Familial Alzheimer's disease in American descendants of the Volga Germans: probable genetic founder effect. Ann Neurol. 1988 Jan;23(1):25–31.

50. Levy-Lahad E, Wijsman E, Nemens E. A familial Alzheimer's disease locus on chromosome 1. Science (80-). 1995;411(1978):16–9.

51. Levy-Lahad E, Wasco W, Poorkaj P, Romano DM, Oshima J, Pettingell WH, et al. Candidate gene for the chromosome 1 familial Alzheimer's disease locus. Science. 1995 Aug 18;269(5226):973–7.

52. Cruts M, Theuns J, Van Broeckhoven C. Locus-specific mutation databases for neurodegenerative brain diseases. Hum Mutat. 2012 Sep;33(9):1340–4.

53. Di Fede G, Catania M, Morbin M, Rossi G, Suardi S, Mazzoleni G, et al. A Recessive Mutation in the APP Gene with Dominant-Negative Effect on Amyloidogenesis. Science (80-). 2009 Mar 13;323(5920):1473–7.

54. Tomiyama T, Nagata T, Shimada H, Teraoka R, Fukushima A, Kanemitsu H, et al. A new amyloid β variant favoring oligomerization in Alzheimer's-type dementia. Ann Neurol. 2008 Mar;63(3):377–87.

55. Müller UC, Deller T, Korte M. Not just amyloid: physiological functions of the amyloid precursor protein family. Nat Rev Neurosci. 2017 Mar 31;18(5):281–98.

56. Prox J, Bernreuther C, Altmeppen H, Grendel J, Glatzel M, D'Hooge R, et al. Postnatal disruption of the disintegrin/metalloproteinase ADAM10 in brain causes epileptic seizures, learning deficits, altered spine morphology, and defective synaptic functions. J Neurosci . 2013 Aug 7;33(32):12915–28, 12928a.

57. Cacace R, Sleegers K, Van Broeckhoven C. Molecular genetics of early-onset Alzheimer's disease revisited. Alzheimer's Dement. 2016 Jun 1;12(6):733–48.

58. Dimitrov M, Alattia J-R, Lemmin T, Lehal R, Fligier A, Houacine J, et al. Alzheimer's disease mutations in APP but not γ-secretase modulators affect epsilon-cleavage-dependent AICD production. Nat Commun. 2013 Dec 2;4(1):2246.

59. Hardy J, Selkoe DJ. The Amyloid Hypothesis of Alzheimer's Disease: Progress and Problems on the Road to Therapeutics. Science (80-). 2002 Jul 19;297(5580):353–6.

60. Pericak-Vance MA, Bebout JL, Gaskell PC, Yamaoka LH, Hung WY, Alberts MJ, et al. Linkage studies in familial Alzheimer disease: evidence for chromosome 19 linkage. Am J Hum Genet . 1991 Jun;48(6):1034–50.

61. Strittmatter WJ, Saunders a M, Schmechel D, Pericak-Vance M, Enghild J, Salvesen GS, et al.

Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. Proc Natl Acad Sci. 1993 Mar 1;90(5):1977–81.

62. Saunders a. M, Strittmatter WJ, Schmechel D, St. George-Hyslop PH, Pericak-Vance M a., Joo SH, et al. Association of apolipoprotein E allele 4 with late-onset familial and sporadic Alzheimer's disease. Neurology. 1993 Aug 1;43(8):1467–1467.

63. Zannis VI, Just PW, Breslow JL. Human apolipoprotein E isoprotein subclasses are genetically determined. Am J Hum Genet. 1981 Jan;33(1):11–24.

64. Zhong N, Weisgraber KH. Understanding the association of apolipoprotein E4 with Alzheimer disease: clues from its structure. J Biol Chem. 2009 Mar 6;284(10):6027–31.

65. Fullerton SM, Clark AG, Weiss KM, Nickerson DA, Taylor SL, Stengård JH, et al. Apolipoprotein E Variation at the Sequence Haplotype Level: Implications for the Origin and Maintenance of a Major Human Polymorphism. Am J Hum Genet. 2000 Oct;67(4):881–900.

66. Guerreiro RJ, Gustafson DR, Hardy J. The genetic architecture of Alzheimer's disease: beyond APP, PSENs and APOE. Neurobiol Aging. 2012 Mar;33(3):437–56.

67. Tang MX, Stern Y, Marder K, Bell K, Gurland B, Lantigua R, et al. The APOE-epsilon4 allele and the risk of Alzheimer disease among African Americans, whites, and Hispanics. JAMA. 1998 Mar 11;279(10):751–5.

68. Corbo RM, Scacchi R. Apolipoprotein E (APOE) allele distribution in the world. Is APOE*4 a "thrifty" allele? Ann Hum Genet. 1999 Jul;63(Pt 4):301–10.

69. Corder E, Saunders A. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science (80-). 1993;8(14):41–3.

70. Genin E, Hannequin D, Wallon D, Sleegers K, Hiltunen M, Combarros O, et al. APOE and Alzheimer disease: a major gene with semi-dominant inheritance. Mol Psychiatry. 2011 Sep;16(9):903–7.

71. Poirier J, Davignon J, Bouthillier D, Kogan S, Bertrand P, Gauthier S. Apolipoprotein E polymorphism and Alzheimer's disease. Lancet. 1993 Sep 18;342(8873):697–9.

72. Corder EH, Saunders AM, Risch NJ, Strittmatter WJ, Schmechel DE, Gaskell PC, et al. Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. Nat Genet. 1994 Jun;7(2):180–4.

73. Breslow JL, Zannis VI, SanGiacomo TR, Third JL, Tracy T, Glueck CJ. Studies of familial type III hyperlipoproteinemia using as a genetic marker the apoE phenotype E2/2. J Lipid Res. 1982 Nov;23(8):1224–35.

74. Roses AD, Lutz MW, Amrine-Madsen H, Saunders AM, Crenshaw DG, Sundseth SS, et al. A TOMM40 variable-length polymorphism predicts the age of late-onset Alzheimer's disease. Pharmacogenomics J. 2010 Oct;10(5):375–84.

75. Seshadri S, Fitzpatrick AL, Ikram MA, DeStefano AL, Gudnason V, Boada M, et al. Genome-wide analysis of genetic loci associated with Alzheimer disease. JAMA. 2010 May;303(18):1832–40.

76. Hollingworth P, Harold D, Sims R, Gerrish A, Lambert J-C, Carrasquillo MM, et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nat Genet. 2011 May;43(5):429–35.

77. Cruchaga C, Karch CM, Jin SC, Benitez BA, Cai Y, Guerreiro R, et al. Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer ' s disease. Nature. 2014;505(7484):550–4.

78. Hooli B V, Lill CM, Mullin K, Qiao D, Lange C, Bertram L, et al. PLD3 gene variants and Alzheimer's disease. Nature . 2015 Apr 2;520(7545):E7-8.

79. Heilmann S, Drichel D, Clarimon J, Fernández V, Lacour A, Wagner H, et al. PLD3 in non-familial Alzheimer's disease. Nature . 2015 Apr 1;520(7545):E3–5.

80. Lambert J-C, Grenier-Boley B, Bellenguez C, Pasquier F, Campion D, Dartigues J-F, et al. PLD3 and sporadic Alzheimer's disease risk. Nature. 2015 Apr 1;520(7545):E1–E1.

81. Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. Nat Genet . 2007 Jan;39(1):17–23.

82. Rogaeva E, Meng Y, Lee JH, Gu Y, Kawarai T, Zou F, et al. The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. Nat Genet . 2007 Feb;39(2):168–77.

83. Alvarez R, Alvarez V, Lahoz CH, Martinez C, Pena J, Sanchez JM, et al. Angiotensin converting enzyme and endothelial nitric oxide synthase DNA polymorphisms and late onset Alzheimer's disease. J Neurol Neurosurg Psychiatry . 1999 Dec 1;67(6):733–6.

84. Hu J, Miyatake F, Aizu Y, Nakagawa H, Nakamura S, Tamaoka A, et al. Angiotensin-converting enzyme genotype is associated with Alzheimer disease in the Japanese population. Neurosci Lett. 1999 Dec 17;277(1):65–7.

85. Kim M, Suh J, Romano D, Truong MH, Mullin K, Hooli B, et al. Potential late-onset Alzheimer's disease-associated mutations in the ADAM10 gene attenuate α-secretase activity. Hum Mol Genet . 2009 Oct 15;18(20):3987.

86. Neill D, Curran MD, Middleton D, Mawhinney H, Edwardson JA, McKeith I, et al. Risk for Alzheimer's disease in older late-onset cases is associated with HLA-DRB1*03. Neurosci Lett . 1999 Nov 12 ;275(2):137–40.

87. Aisen PS, Luddy A, Durner M, Reinhard JF, Pasinetti GM. HLA-DR4 influences glial activity in Alzheimer's disease hippocampus. J Neurol Sci . 1998 Nov 26;161(1):66–9.

88. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science . 1996 Sep 13;273(5281):1516–7.

89. Galton F. II. On Blood-Relationship. Proc R Soc London. 1872;20:130–8.

90. Fisher R. The Correlation between Relatives on the Suppositionof Mendelian Inheritance. Trans Real Sociaty Edinbg. 1918;52:399–433.

91. Wright S. Evolution in Mendelian Populations. Genetics . 1931 Mar;16(2):97–159.

92. Falconer DS. The inheritance of liability to certain diseases estimated from the incidence among relatives. Ann Hum Genet. 1965;29:51–76.

93. The Wellcome Trust Case Control Consortium, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A, et al. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007 Jun 7;447(7145):661.

94. The International HapMap Consortium, Altshuler D, Donnelly P. A haplotype map of the human genome. Nature . 2005 Oct 27;437(7063):1299–320.

95. Evangelou E, Ioannidis JPA. Meta-analysis methods for genome-wide association studies and beyond. Nat Rev Genet. 2013 Jun 9;14(6):379–89.

96. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet Epidemiol . 2008 May;32(4):381–5.

97. Dudbridge F, Gusnanto A. Estimation of significance thresholds for genomewide association scans. Genet Epidemiol . 2008 Apr;32(3):227–34.

98. Panagiotou OA, Ioannidis JPA, Genome-Wide Significance Project. What should the genome-wide significance threshold be? Empirical replication of borderline genetic associations. Int J Epidemiol . 2012 Feb;41(1):273–86.

99. Lambert J-C, Heath S, Even G, Campion D, Sleegers K, Hiltunen M, et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. Nat Genet . 2009 Oct;41(10):1094–9.

100. Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet . 2009 Oct;41(10):1088–93.

101. Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nat Rev Genet . 2010 Jul 2;11(7):499–511.

102. Naj AC, Jun G, Beecham GW, Wang L-S, Vardarajan BN, Buros J, et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nat Genet . 2011 May;43(5):436–41.

103. Antúnez C, Boada M, González-Pérez A, Gayán J, Ramírez-Lorca R, Marín J, et al. The membrane-spanning 4-domains, subfamily A (MS4A) gene cluster contains a common variant associated with Alzheimer's disease. Genome Med . 2011 Jan;3(5):33.

104. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet. 2013 Dec ;45(12):1452–8.

105. Boada M, Antúnez C, Ramírez-Lorca R, DeStefano A, González-Pérez A, Gayán J, et al. ATP5H/KCTD2 locus is associated with Alzheimer's disease risk. Mol Psychiatry . 2014 Jun;19(6):682–7.

106. Ruiz A, Heilmann S, Becker T, Hernández I, Wagner H, Thelen M, et al. Follow-up of loci from the

International Genomics of Alzheimer's Disease Project identifies TRIP4 as a novel susceptibility gene. Transl Psychiatry . 2014 Jan;4(January):e358.

107. Witoelar A, Rongve A, Almdahl IS, Ulstein ID, Engvig A, White LR, et al. Meta-analysis of Alzheimer's disease on 9,751 samples from Norway and IGAP study identifies four risk loci. Sci Rep. 2018 Dec 27;8(1):18088.

108. Escott-Price V, Bellenguez C, Wang L-S, Choi S-H, Harold D, Jones L, et al. Gene-wide analysis detects two new susceptibility genes for Alzheimer's disease. PLoS One . 2014 Jan;9(6):e94661.

109. Baker E, Sims R, Leonenko G, Frizzati A, Harwood J, Grozeva D, et al. Gene based Analysis in HRC Imputed Genome Wide Association Data Identifies Three Novel Genes for Alzheimer's Disease. bioRxiv . 2018 Jul 23;374876.

110. Jun G, Ibrahim-Verbaas CA, Vronskaya M, Lambert J-C, Chung J, Naj AC, et al. A novel Alzheimer disease locus located near the gene encoding tau protein. Mol Psychiatry. 2015 Mar 17.

111. Pastor P, Moreno F, Clarimón J, Ruiz A, Combarros O, Calero M, et al. MAPT H1 Haplotype is Associated with Late-Onset Alzheimer's Disease Risk in APOE ε4 Noncarriers: Results from the Dementia Genetics Spanish Consortium. J Alzheimer's Dis . 2015 Sep 29;49(2):343–52.

112. Jun GR, Chung J, Mez J, Barber R, Beecham GW, Bennett DA, et al. Transethnic genome-wide scan identifies novel Alzheimer's disease loci. Alzheimers Dement . 2017 Jul;13(7):727–38.

113. Marioni RE, Harris SE, Zhang Q, McRae AF, Hagenaars SP, Hill WD, et al. GWAS on family history of Alzheimer's disease. Transl Psychiatry . 2018 Dec 18;8(1):99.

114. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. Nat Genet. 2019 Jan 7.

115. Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. Nat Genet. 2019 Mar 28;51(3):414–30.

116. Liu JZ, Erlich Y, Pickrell JK. Case–control association mapping by proxy using family history of disease. Nat Genet . 2017 Mar 16;49(3):325–31.

117. McClellan J, King M-C. Genetic Heterogeneity in Human Disease. Cell. 2010 Apr 16;141(2):210–7.

118. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: An information aesthetic for comparative genomics. Genome Res. 2009 Sep 1;19(9):1639–45.

119. Consortium the HR, McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016 Oct 22;48(10):1279–83.

120. Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogaeva E, Majounie E, et al. TREM2 variants in Alzheimer's disease. N Engl J Med. 2013 Jan 10;368(2):117–27.

121. Jonsson T, Stefansson H, Steinberg S, Jonsdottir I, Jonsson P V, Snaedal J, et al. Variant of TREM2 associated with the risk of Alzheimer's disease. N Engl J Med. 2013 Jan 10;368(2):107–16.

122. Sims R, van der Lee SJ, Naj AC, Bellenguez C, Badarinarayan N, Jakobsdottir J, et al. Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. Nat Genet. 2017 Jul 17;49(9):1373–84.

123. Jonsson T, Atwal JK, Steinberg S, Snaedal J, Jonsson P V, Bjornsson S, et al. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. Nature. 2012 Aug 2;488(7409):96–9.

124. Wetzel-Smith MK, Hunkapiller J, Bhangale TR, Srinivasan K, Maloney JA, Atwal JK, et al. A rare mutation in UNC5C predisposes to late-onset Alzheimer's disease and increases neuronal cell death. Nat Med. 2014 Dec;20(12):1452–7.

125. Bis JC, Jian X, Kunkle BW, Chen Y, Hamilton-Nelson KL, Bush WS, et al. Whole exome sequencing study identifies novel rare and common Alzheimer's-Associated variants involved in immune response and transcriptional regulation. Mol Psychiatry. 2018 Aug 14.

126. Steinberg S, Stefansson H, Jonsson T, Johannsdottir H, Ingason A, Helgason H, et al. Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. Nat Genet. 2015 Mar 25;47(5):445–7.

127. Bellenguez C, Charbonnier C, Grenier-Boley B, Quenez O, Le Guennec K, Nicolas G, et al. Contribution to Alzheimer's disease risk of rare variants in TREM2, SORL1, and ABCA7 in 1779 cases and 1273 controls. Neurobiol Aging. 2017 Nov;59:220.e1-220.e9.

128. Holstege H, van der Lee SJ, Hulsman M, Wong TH, van Rooij JG, Weiss M, et al. Characterization of pathogenic SORL1 genetic variants for association with Alzheimer's disease: a clinical interpretation strategy. Eur J Hum Genet. 2017 Aug 24;25(8):973–81.

129. van der Lee SJ, Holstege H, Wong TH, Jakobsdottir J, Bis JC, Chouraki V, et al. PLD3 variants in population studies. Nature. 2015 Apr 1;520(7545):E2–3.

130. Ruiz A, Dols-Icardo O, Bullido MJ, Pastor P, Rodríguez-Rodríguez E, López de Munain A, et al. Assessing the role of the TREM2 p.R47H variant as a risk factor for Alzheimer's disease and frontotemporal dementia. Neurobiol Aging. 2014 Feb;35(2):444.e1-4.

131. van der Lee SJ, Conway OJ, Jansen I, Carrasquillo MM, Kleineidam L, van den Akker E, et al. A nonsynonymous mutation in PLCG2 reduces the risk of Alzheimer's disease, dementia with Lewy bodies and frontotemporal dementia, and increases the likelihood of longevity. Acta Neuropathol. 2019 May 27.

132. Hirschhorn JN. Genomewide Association Studies — Illuminating Biologic Pathways. N Engl J Med. 2009 Apr 23;360(17):1699–701.

133. Hardy JA, Higgins GA. Alzheimer's disease: the amyloid cascade hypothesis. Science. 1992 Apr 10;256(5054):184–5.

134. Jones L, Holmans PA, Hamshere ML, Harold D, Moskvina V, Ivanov D, et al. Genetic evidence implicates the immune system and cholesterol metabolism in the aetiology of Alzheimer's disease. PLoS One. 2010 Jan;5(11):e13950.

135. International Genomics of Alzheimer's disease Consortium (IGAP). Convergent genetic and expression data implicate immunity in Alzheimer's disease. Alzheimers Dement. 2014 Dec 19.

136. Ridge PG, Hoyt KB, Boehme K, Mukherjee S, Crane PK, Haines JL, et al. Assessment of the genetic variance of late-onset Alzheimer's disease. Neurobiol Aging. 2016 May;41:200.e13-200.e20.

137. Garrod AE. The incidence of alkaptonuria: a study in chemical individuality. Lancet. 1902;1616:1902–11.

138. Lander E, Botstein D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. Science (80-) . 1987 Jun 19;236(4808):1567–70.

139. Jakkula E, Rehnström K, Varilo T, Pietiläinen OPH, Paunio T, Pedersen NL, et al. The Genome-wide Patterns of Variation Expose Significant Substructure in a Founder Population. Am J Hum Genet. 2008 Dec;83(6):787–94.

140. Gross A, Tönjes A, Kovacs P, Veeramah KR, Ahnert P, Roshyara NR, et al. Population-genetic comparison of the Sorbian isolate population in Germany with the German KORA population using genome-wide SNP arrays. BMC Genet. 2011 Jul 28;12(1):67.

141. Farrer LA, Bowirrat A, Friedland RP, Waraska K, Korczyn AD, Baldwin CT. Identification of multiple loci for Alzheimer disease in a consanguineous Israeli-Arab community. Hum Mol Genet. 2003 Feb 15;12(4):415–22.

142. Sherva R, Baldwin CT, Inzelberg R, Vardarajan B, Cupples LA, Lunetta K, et al. Identification of novel candidate genes for Alzheimer's disease by autozygosity mapping using genome wide SNP data. J Alzheimers Dis. 2011;23(2):349–59.

143. Clarimón J, Djaldetti R, Lleó A, Guerreiro RJ, Molinuevo JL, Paisán-Ruiz C, et al. Whole genome analysis in a consanguineous family with early onset Alzheimer's disease. Neurobiol Aging. 2009 Dec;30(12):1986–91.

144. Bras J, Djaldetti R, Alves AM, Mead S, Darwent L, Lleo A, et al. Exome sequencing in a consanguineous family clinically diagnosed with early-onset Alzheimer's disease identifies a homozygous CTSF mutation. Neurobiol Aging. 2016 Oct;46:236.e1-6.

145. McQuillan R, Leutenegger A-L, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al. Runs of Homozygosity in European Populations. Am J Hum Genet. 2008 Sep;83(3):359–72.

146. Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ. Genomic Patterns of Homozygosity in Worldwide Human Populations. Am J Hum Genet . 2012 Aug 10;91(2):275–92.

147. Ghani M, Sato C, Lee JH, Reitz C, Moreno D, Mayeux R, et al. Evidence of recessive Alzheimer disease loci in a Caribbean Hispanic data set: genome-wide survey of runs of homozygosity. JAMA Neurol. 2013 Oct 26;70(10):1261–7.

148. Ghani M, Reitz C, Cheng R, Vardarajan BN, Jun G, Sato C, et al. Association of Long Runs of Homozygosity With Alzheimer Disease Among African American Individuals. JAMA Neurol. 2015 Nov 1;72(11):1313–23.

149. Nalls MA, Guerreiro RJ, Simon-Sanchez J, Bras JT, Traynor BJ, Gibbs JR, et al. Extended tracts of homozygosity identify novel candidate genes associated with late-onset Alzheimer's disease. Neurogenetics. 2009 Jul 7;10(3):183–90.

150. Sims R, Dwyer S, Harold D, Gerrish A, Hollingworth P, Chapman J, et al. No evidence that extended tracts of homozygosity are associated with Alzheimer's disease. Am J Med Genet Part B Neuropsychiatr Genet. 2011 Dec;156(7):764–71.

151. Keller MC, Visscher PM, Goddard ME. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetics. 2011 Sep;189(1):237–49.

152. Combs CK, Karlo JC, Kao SC, Landreth GE. beta-Amyloid stimulation of microglia and monocytes results in TNFalpha-dependent expression of inducible nitric oxide synthase and neuronal apoptosis. J Neurosci. 2001 Feb 15;21(4):1179–88.

153. Zwan MD, Villemagne VL, Doré V, Buckley R, Bourgeat P, Veljanoski R, et al. Subjective Memory Complaints in APOE ε4 Carriers are Associated with High Amyloid-β Burden. J Alzheimer's Dis . 2015 Nov 25;49(4):1115–22.

154. Bangen KJ, Clark AL, Werhane M, Edmonds EC, Nation DA, Evangelista N, et al. Cortical Amyloid Burden Differences Across Empirically-Derived Mild Cognitive Impairment Subtypes and Interaction with APOE ε4 Genotype. J Alzheimers Dis. 2016;52(3):849–61.

155. Ramanan VK, Risacher SL, Nho K, Kim S, Swaminathan S, Shen L, et al. APOE and BCHE as modulators of cerebral amyloid deposition: a florbetapir PET genome-wide association study. Mol Psychiatry. 2014 Mar;19(3):351–7.

156. Ramanan VK, Risacher SL, Nho K, Kim S, Shen L, McDonald BC, et al. GWAS of longitudinal amyloid accumulation on 18F-florbetapir PET in Alzheimer's disease implicates microglial activation gene IL1RAP. Brain. 2015 Aug 11;138(10):3076–88.

157. Rullmann M, Dukart J, Hoffmann K-T, Luthardt J, Tiepolt S, Patt M, et al. Partial-Volume Effect Correction Improves Quantitative Analysis of 18F-Florbetaben -Amyloid PET Scans. J Nucl Med. 2016 Feb 1;57(2):198–203.

158. Seshadri S, Fitzpatrick A, L. Ikram A, DeStefano AL, Gudnason V, Boada M. Genome-wide Analysis of Genetic Loci Associeted with Alzheimer's Disease. JAMA. 2011;303(18):1832–40.

159. Ardlie KG, Kruglyak L, Seielstad M. Patterns of Linkage Disequilibrium in the Human Genome. Nat Rev Genet. 2002;3(April):299–309.

160. Campbell CD, Ogburn EL, Lunetta KL, Lyon HN, Freedman ML, Groop LC, et al. Demonstrating stratification in a European American population. Nat Genet. 2005 Aug 24;37(8):868–72.

161. Lavrik I, Golks A, Krammer PH. Death receptor signaling. J Cell Sci. 2005 Jan 15;118(Pt 2):265–7.

162. Viswanathan A, Rocca WA, Tzourio C. Vascular risk factors and dementia: How to move forward? Neurology. 2009 Jan 27;72(4):368–74.

163. Sakai K, Yamada M. [Cerebral amyloid angiopathy]. Brain Nerve. 2014 Jul;66(7):827–35.

164. Mäkelä M, Kaivola K, Valori M, Paetau A, Polvikoski T, Singleton AB, et al. Alzheimer risk loci and associated neuropathology in a population-based study (Vantaa 85+). Neurol Genet. 2018 Feb;4(1):e211.

165. Zhang S, Reddick R, Piedrahita J, Maeda N. Spontaneous hypercholesterolemia and arterial lesions in mice lacking apolipoprotein E. Science (80-). 1992 Oct 16;258(5081):468–71.

166. Ishibashi S, Brown MS, Goldstein JL, Gerard RD, Hammer RE, Herz J. Hypercholesterolemia in low density lipoprotein receptor knockout mice and its reversal by adenovirus-mediated gene delivery. J Clin Invest. 1993 Aug 1;92(2):883–93.

167. Nakashima Y, Plump AS, Raines EW, Breslow JL, Ross R. ApoE-deficient mice develop lesions of all phases of atherosclerosis throughout the arterial tree. Arterioscler Thromb a J Vasc Biol . 1994 Jan;14(1):133–40.

168. Mahley RW. Apolipoprotein E: cholesterol transport protein with expanding role in cell biology. Science . 1988 Apr 29;240(4852):622–30.

169. Fang L, Choi S-H, Baek JS, Liu C, Almazan F, Ulrich F, et al. Control of angiogenesis by AIBP-mediated cholesterol efflux. Nature. 2013 Jun 29;498(7452):118–22.

170. Robert J, Button EB, Yuen B, Gilmour M, Kang K, Bahrabadi A, et al. Clearance of beta-amyloid is facilitated by apolipoprotein E and circulating high-density lipoproteins in bioengineered human vessels. Elife . 2017 Oct 10;6.

171. Crawford F, Suo Z, Fang C, Mullan M. Characteristics of thein VitroVasoactivity of β-Amyloid Peptides. Exp Neurol. 1998 Mar;150(1):159–68.

172. Jack CR, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimer's Dement. 2018 Apr;14(4):535–62.

173. Lucariello M, Vidal E, Vidal S, Saez M, Roa L, Huertas D, et al. Whole exome sequencing of Rett syndrome-like patients reveals the mutational diversity of the clinical phenotype. Hum Genet. 2016;135(12):1343–54.

174. Deelen J, Evans DS, Arking DE, Tesi N, Nygaard M, Liu X, et al. A meta-analysis of genome-wide association studies identifies multiple longevity genes. Nat Commun. 2019 Dec 14;10(1):3669.

175. Leduc V, De Beaumont L, Théroux L, Dea D, Aisen P, Petersen RC, et al. HMGCR is a genetic modifier for risk, age of onset and MCI conversion to Alzheimer's disease in a three cohorts study. Mol Psychiatry. 2015 Jul 15;20(7):867–73.

176. Broce IJ, Tan CH, Fan CC, Jansen I, Savage JE, Witoelar A, et al. Dissecting the genetic relationship between cardiovascular risk factors and Alzheimer's disease. Acta Neuropathol. 2018 Nov 9.

177. McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al. Runs of Homozygosity in European Populations. Am J Hum Genet. 2008 Sep;83(3):359–72.

178. Desikan RS, Schork AJ, Wang Y, Thompson WK, Dehghan A, Ridker PM, et al. Polygenic Overlap Between C-Reactive Protein, Plasma Lipids and Alzheimer's Disease. Circulation. 2015;131(23):2061.

179. Ho A, Südhof TC. Binding of F-spondin to amyloid-beta precursor protein: a candidate amyloid-beta precursor protein ligand that modulates amyloid-beta precursor protein cleavage. Proc Natl Acad Sci. 2004

Feb 24;101(8):2548–53.

180. Hoe H-S, Wessner D, Beffert U, Becker AG, Matsuoka Y, Rebeck GW. F-Spondin Interaction with the Apolipoprotein E Receptor ApoEr2 Affects Processing of Amyloid Precursor Protein. Mol Cell Biol. 2005 Nov 1;25(21):9259–68.

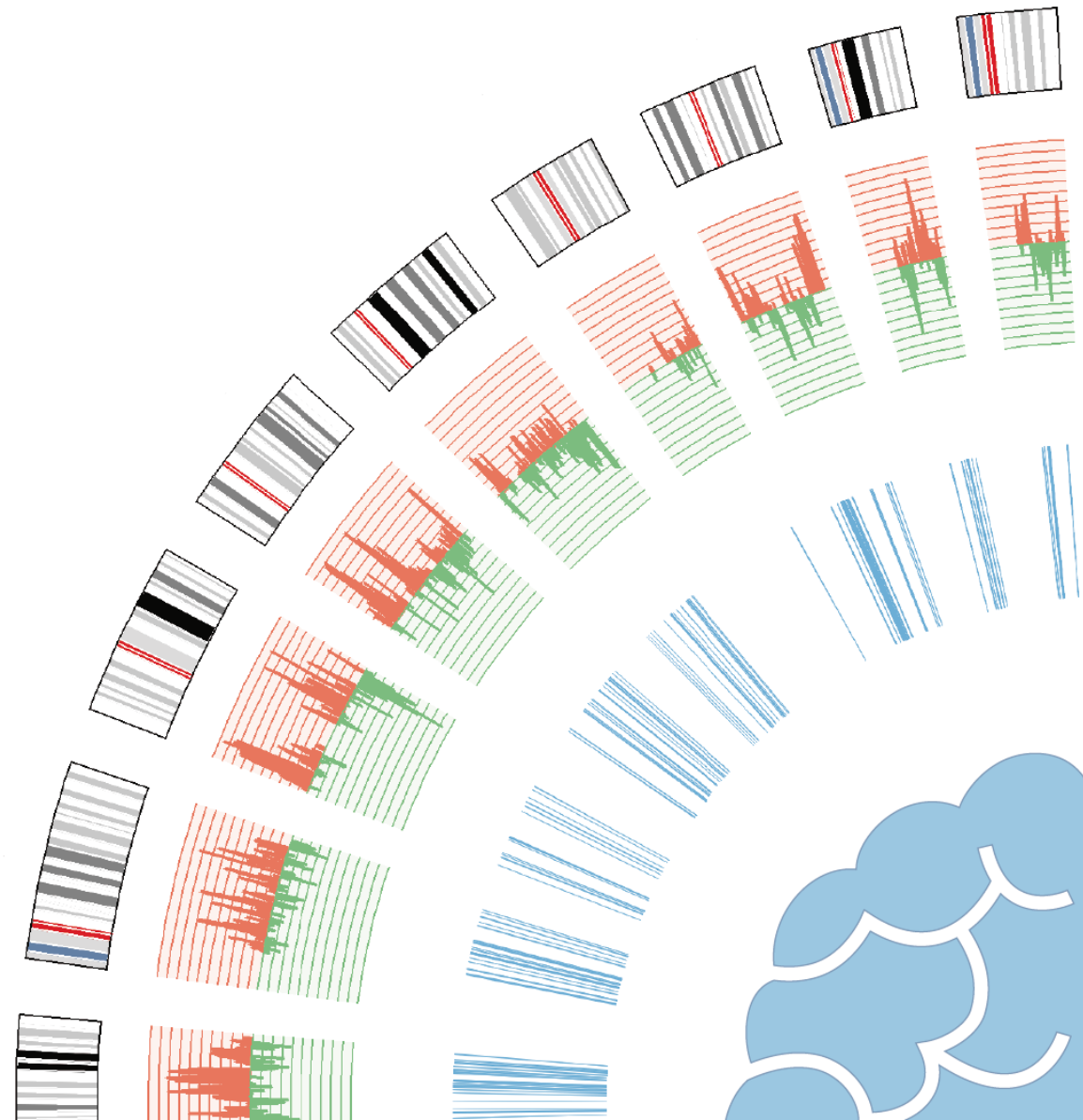
181. Jahanshad N, Rajagopalan P, Hua X, Hibar DP, Nir TM, Toga AW, et al. Genome-wide scan of healthy human connectome discovers SPON1 gene variant influencing dementia severity. Proc Natl Acad Sci . 2013 Mar 19;110(12):4768–73.

182. Laurie CC, Laurie CA, Rice K, Doheny KF, Zelnick LR, McHugh CP, et al. Detectable clonal mosaicism from birth to old age and its relationship to cancer. Nat Genet. 2012 Jun 6;44(6):642–50.

183. Ceballos FC, Hazelhurst S, Ramsay M. Runs of Homozygosity in sub-Saharan African populations provide insights into a complex demographic and health history. bioRxiv. 2018 Nov 14;470583.

CHAPTER 6

APPENDIX



Appendix 1. Supplementary Publication I

Genome research in pre-dementia stages of Alzheimer's disease

Sonia Moreno-Grau.[1,2] and Agustín Ruiz [1,2]

[1]Research Center and Memory clinic Fundació ACE. Institut Català de Neurociències Aplicades. Universitat Internacional de Catalunya. Barcelona. Spain. [2]CIBERNED, Center for Networked Biomedical Research on Neurodegenerative Diseases, National Institute of Health Carlos III, Ministry of Economy and Competitiveness, Spain.

ABSTRACT

Genetic characterization of individuals at risk of Alzheimer's disease (AD), i.e. people having amyloid deposits in the brain without symptoms, people suffering from subjective cognitive decline (SCD) or mild cognitive impairment (MCI), has spurred the interests of researchers. However, their pre-dementia genetic profile remains mostly unexplored. In this study, we reviewed the loci related to phenotypes of AD, MCI and SCD from literature and performed the first meta-analyses evaluating the role of apolipoprotein E (*APOE)* in the risk of conversion from a healthy status to MCI and SCD. For AD dementia risk, an increased number of loci have been identified; to date, 28 genes have been associated with Late Onset AD (LOAD). In MCI syndrome, *APOE* is confirmed as a pheno-conversion factor leading from MCI to AD, and clusterin (*CLU*) is a promising candidate. Additionally, our meta-analyses revealed *APOE* as genetic risk factor to convert from a healthy status to MCI (OR = 1.849 [1.587–2.153]; p = 2.80 x 10^{-15}) and to a lesser extent from healthy status to SCD (OR = 1.151 [1.015 – 1.304]; p = 0.028). Thus, we believe that genetic studies in longitudinal SCD and MCI series may provide new therapeutic targets and improve the existing knowledge of AD. This type of studies must be completed on healthy subjects to better understand the natural disease resistance to brain insults and neurodegeneration.

Keywords: Alzheimer's disease, cognition, genetics, mild cognitive impairment, subjective cognitive decline

Expert Reviews in Molecular Medicine, 2016 May 30;18:e11.
doi: 10.1017/erm.2016.12.

INTRODUCTION

Although Alzheimer's disease (AD) is mainly diagnosed in the elderly, its pathophysiological processes begin several years prior to the onset of symptoms (1).

Clinical AD is preceded by a long asymptomatic period, which has been divided into three stages: 1) an initial preclinical stage, 2) a second mild, but progressive, cognitive impairment (MCI), and 3) the final stage of clinical dementia due to AD (1),(2),(3). Recently, researchers have increasingly focused on the characterization of stages of AD risk, as these provide a critical opportunity for potential intervention (1).

With ageing, there is a natural decline in cognitive skills. Thus, it may be difficult to discriminate between early cognitive changes due to AD and normal ageing process (4). In that context, the first evidence of dementia may be the subjective cognitive decline (SCD), defined as a self-reported memory impairment with normal cognitive performance (5). Complainers present a higher rate of conversion of SCD to either MCI or dementia (Ref 6). Thus, epidemiological studies pointed SCD as a predictor of cognitive decline (7), (8) and as an independent risk factor for dementia (9). In succession, the prodromal stage of dementia, MCI, has been defined as memory impairment beyond that expected for normal ageing (10). Several MCI phenotypes have been associated with AD progression (11); however, amnestic MCI (aMCI) confers a higher risk of conversion (11).

The identification of MCI subjects, or even SCD who will convert to MCI or dementia, puts across an interesting strategy for secondary prevention of AD. In that sense, the biomarkers of β-amyloidosis and tau-mediated neuronal injury are detected in subjects with normal cognition (12). However, these biomarkers are not sufficient to produce the clinical symptoms of MCI and dementia or are not specific to AD either (1). Furthermore, these biomarkers are not sensitive to disease progression (13). Hence, new approaches are required to improve the differentiation of SCD or MCI converters to AD.

AD's genetics has gained much attention since AD presents a heritability of up 70% (14). Recently, researchers have been striving towards the identification of new AD's genetic risk factors. In that sense, the identification of a genetic risk profile for pre-dementia stages may prove to be a powerful approach to select the candidate subjects to prevent or delay the disease progression during the early preclinical stages. If a fraction of SCD and MCI patients are in the pre-AD stages, the identification of an increased number of AD risk alleles as well as that of additional genetic factors specifically influencing SCD or MCI progression can be expected.

In this work, we reviewed the available information in the literature for genome-wide significant variants associated with AD and their involvement in preclinical, prodromal and dementia stages of AD. Additionally, we provide new meta-analyzed data for *APOE* ε4 in preclinical and prodromal stages.

METHODS: META-ANALYSIS

Meta-analysis was performed for exploring the role of *APOE* ε4 in: 1) risk of MCI, and 2) risk of SCD.

Dataset selection

Literature search was conducted in PubMed (http://www.ncbi.nlm.nih.gov/pubmed/) using the following keywords: 1) for MCI: APOE, genetics, risk, mild cognitive impairment and excluding reviews; and 2) for SCD: APOE, subjective cognitive decline. A total of 301 articles were found for MCI and 32 for SCD.

We selected the studies meeting the following criteria: (1) case/control studies or longitudinal studies where it is possible to distinguish a sub-population of cases and a sub-population of controls; (2) studies that provide a complete definition of the participants; (3) studies that evaluated the *APOE* ε4 genotype as a risk factor leading to MCI or SCD, or provided the numbers of *APOE* ε4 genotypes or provided sufficient data to calculate them; and (4) studies that provided an OR with 95% CI as well as the p-value or provide sufficient data to calculate them. Finally, of the 301 articles found for MCI, 207 did not follow inclusion criteria, 29 showed sample overlapping, and 41 had restricted access. A total of 24 articles and 23,668 individuals on MCI were finally included. In the case of SCD, 21 of the 32 articles did not follow the inclusion criteria and 3 showed sample overlapping; finally, a total of 8 articles and 6,824 individuals were included in the meta-analysis.

Meta-analysis

Meta-analysis was conducted using the inverse variant method (fixed-effects model) in Ephisheet Excel application. In the case of heterogeneity, DerSimonian and Liard method (random-effects model) was used. Heterogeneity was considered significant when $I^2 > 50\%$ and $p < 0.05$. Meta-analysis results and forest plots were obtained using OpenMeta.

DEMENTIA STAGE: GENETIC RISK FACTORS OF AD.

AD is a genetically heterogeneous disorder. From a genetic point of view, two patterns of inheritance have been linked to the genomic loci: the autosomal dominant and the polygenic. Traditionally, these patterns have been associated with early and late onset forms of the disease, respectively. However, based on the family history, AD can be subdivided into autosomal dominant, familial and sporadic (15).

Autosomal dominant AD

The familial autosomal dominant pattern in AD represents ~1% of all the AD cases and is found almost exclusively in early onset AD (EOAD) (15). It occurs in at least three individuals in two or more generations, with two of the individuals being first-degree relatives of the third (15).

Linkage and candidate gene studies in EOAD families led to the identification of disease-causing mutations in β-amyloid precursor protein (*APP*), presenilin 1 (*PSEN1*) and presenilin 2 (*PSEN2*) genes (16), (17), (18). Most frequent mutations are shown in *PSEN1* and *APP* loci, respectively, which present complete penetrance in contrast to *PSEN2*, which presents 95% penetrance (15). These identifications promoted the formulation of amyloid cascade hypothesis, which is still considered as a possible disease mechanism. Despite that, there are EOAD families with negative screening for *APP*, *PSEN1* and *PSEN2* mutations supporting the existence of additional causal genes (19). In addition, it is seen that *APOE* ε4 genotype, the major genetic risk factor for Late-Onset Alzheimer's disease (LOAD) (20), also modifies the risk of EOAD (21).

Presently, 262 pathogenic mutations have been identified: 42 in *APP*, 207 in *PSEN1*, and 13 in *PSEN2* (http://www.molgen.ua.ac.be/ADMutations); no other genes have been associated with an autosomal dominant form of AD.

Familial AD (FAD)

FAD occurs in more than one individual and, at least, two of the affected individuals are third-degree relatives or closer (15). Most of the FAD cases are LOAD, but the presence of early onset FAD may be caused by hidden autosomal-dominant AD mutations (15).

Sporadic AD (SAD)

SAD occurs in isolated cases in families or cases separated by more than three degrees of relationship. SAD represents 75% of all AD cases and typically presents a LOAD chart (15).

The commonest AD phenotype, LOAD

The genetic and molecular basis for the commonest AD phenotype, LOAD, remains widely unknown. However, important progress on the isolation of the loci associated with AD has been achieved in the past few years because of the emergence of the genome-wide association (GWAS) and exome studies.

The ε4 allele of the Apolipoprotein E (*APOE*) gene was the first genetic variant associated with LOAD (20), and it remains as the major risk factor for the disease until now.

Behind the *APOE* discovery, the candidate gene approach led to the identification of two clusters of SNPs in *SORL1* gene (22). Recently, this association has been validated by International

Genomics Alzheimer's Project (IGAP) (23). However, candidate gene approach did not show more successful outcomes.

Most discoveries arrived when the GWAS strategy was applied to the large case-controlled datasets. In the GWAS era, common variants located at *CLU*, *PICALM*, *CR1*, *BIN1*, *ABCA7*, *CD2AP*, *CD33*, *EPHA1* and *MS4A6A-MS4A4E* loci were associated with LOAD (24), (25), (26).

To validate the original GWAS findings, replication studies were performed with many independent datasets. Consequently, *CR1*, *PICALM*, *CLU*, and *BIN1* signals have been replicated in Caucasians (27), (28), Caribbean Hispanics (29) or Asian individuals (30), (31). In addition, the *EPHA1* and *CD33* genetic variants were replicated in Caucasian subjects (32), and *ABCA7* in African Americans (33).

A recent meta-analysis developed by IGAP in 74,046 individuals of European ancestry confirmed previously reported GWAS signals (*ABCA7*, *BIN1*, *CLU*, *CR1*, *CD2AP*, *EPHA1*, *MS4A6A-MS4A4E*, and *PICALM*) (23). Nevertheless, the *CD33* locus, previously associated with LOAD, did not reach the genome-wide significance in the replication stages (23). Moreover, the IGAP meta-analysis identified 10 novel genetic regions associated with LOAD: *CASS4*, *SLC24A4-RIN3*, *FERMT2*, *HLA-DRB5-HLA-DRB1*, *INPP5*, *MEF2C*, *PTK2B*, *CELF1*, *NME8*, and *ZCWPW1* and also confirmed a candidate gene, *SORL1* (23). Follow-up studies of the IGAP results revealed additional locus namely *TRIP4* (34) and a novel AD locus within the *MAPT* region at 17q21.31 (35). Finally, the gene-wide analyses of the IGAP dataset identified *TP53INP1* and *IGHV1-67* as the novel AD loci (36), and an independent meta-analysis identified *ATP5H/KCTD2* as the LOAD risk signal (37). Most loci reported after the initial IGAP report would require independent replications confirming its plausibility.

Despite that, the GWAS approach presented a disadvantage, i.e., its inability to detect rare variants, which might be a source of functional variants with larger effects on the LOAD risk (38). This lack was covered by the implementation of genome and exome sequencing technologies. Thus, in the recent years, rare variants with a significant effect on the risk for LOAD have been identified in *APP*, *TREM2*, *PDL3*, and *UNC5C* loci (38), (39), (40), (41). However, more efforts are needed to confirm the original signals. Several studies have confirmed the reported association of *TREM2* with LOAD (42), (43). In addition, the existence of *TREM2* variants associated with the Naso–Hakola disease (44) and frontotemporal dementia (45) supports its role in neurodegeneration. Alternatively, the *PLD3* variants' replication did not replicate the previous effect or overall burden analyses (46), (47). Therefore, prudence is required to define

the genuine signals associated with rare variants.

At present, 28 genetic regions have been associated with LOAD (Table 1), but many of them still require independent validation. These genes can be divided into four major functional clusters: (i) amyloid beta (Aβ) metabolism (*APOE*, *CLU*, *ABCA7*, *CASS4*, *SORL1*, and *APP*), (ii) Tau metabolism (*BIN1*, *SLC24A4-RIN3*, *CASS4*, *FERMT2*, and 17q21.31 *MAPT* region), (iii) synaptic function (*PICALM*, *CD2AP*, *EPHA1*, *SLC24A4-RIN3*, *MEF2C*, and *ZCWPW1*), and (iv) immune response and inflammation (*CLU*, *CR1*, *EPHA1*, *MS4A* cluster, *ABCA7*, *HLA-DRB5-HLA-DRB1*, *INPP5*; *MEF2C*, *TREM2*, and *IGHV1-67*). Seven identified loci do not have a well-established pathway (*PTK2B*, *CELF1*, *NME8*, *TRIP4*, *ATP5H/KCTD2*, *UNC5C*, and *TP53INP1*) (Figure 1).

Along with the identification of single locus, GWAS also permits the genetic confirmation of candidate pathways. Recently, pathway analysis studies have pointed toward the crucial role of the immune system in AD (48), which has been further reinforced by the IGAP results (49). Moreover, the IGAP study also implicates the regulation of endocytosis, cholesterol transport, and protein ubiquitination as prime targets in the etiology of AD (49). The knowledge of the biological pathways involved in disease etiology is crucial in the development of therapeutic strategies to aid in the prevention or treatment of LOAD.

PRODROMAL STAGE: GENETICS OF MILD COGNITIVE IMPAIRMENT SYNDROME

APOE genotype in MCI

Petersen et al (50) were the first to provide evidence that MCI subjects with at least one allele of of *APOE* ε4 presented a higher probability of conversion to dementia. Although subsequent genetic studies supported it (51), (52), they had small sample sizes, which only succeeded in providing an approximate value of the risk effect (53). Thus, the meta-analysis conducted by Elias-Sonnenschein et al. (54) provided the first consistent data corroborating the role of *APOE* ε4 as a genetic risk factor for progression from MCI to AD (Table 2).

The involvement of *APOE* ε4 as a risk factor for MCI remains less explored. Therefore, here we we have explored the risk conferred by *APOE* ε4 genotype to suffer MCI. Our meta-analysis, which includes 23,668 individuals of different ethnic groups, confirmed a significant risk association of *APOE* ε4 genotype and MCI (OR = 1.849 [1.587 – 2.153]) (Figure 2) (Table 2). Our dataset showed high heterogeneity (I^2 = 63%; p-value < 0.001).

TABLE 1. Genetic regions associated with LOAD at the dementia stage from highest to lowest Odds Ratio.

Gene	SNP	OR (CI95%)	P Value	Author
		COMMON VARIANTS		
APOE	rs429358	3.68 (3.30 – 4.11)	9.3×10^{-120}	Corder et al. (20)[b]
ATP5H/KCTD2	rs11870474	1.53 (1.33-1.77)	4.7×10^{-9}	Boada et al. (37)
TRIP4	rs74615166	1.31 (1.17–1.42)	9.7×10^{-9}	Ruiz et al. (34)
BIN1	rs6733839	1.22 (1.18-1.25)	6.9×10^{-44}	Seshadri et al. (25)[c]
ABCA7	rs4147929	1.15 (1.11–1.19)	1.1×10^{-15}	Hollingworth et al. (26)[c]
CR1	rs6656401	1.18 (1.14-1.22)	5.7×10^{-24}	Lambert et al. (119)
FERMT2	rs17125944	1.14 (1.09–1.19)	7.9×10^{-9}	Lambert et al. (23)
HLA-DRB5/DRB1	rs9271192	1.11 [1.08–1.15]	2.9×10^{-12}	Lambert et al. (23)
PTK2B	rs28834970	1.10 (1.08–1.13)	7.4×10^{-14}	Lambert et al. (23)
CD2AP	rs10948363	1.10 (1.07-1.13)	5.2×10^{-11}	Hollingworth et al. (26)[c]
INPP5D	rs35349669	1.08 (1.05–1.11)	3.2×10^{-8}	Lambert et al. (23)
CELF1	rs10838725	1.08 (1.05–1.11)	1.1×10^{-8}	Lambert et al. (23)
NME8	rs2718058	0.93 (0.90–0.95)	4.8×10^{-9}	Lambert et al. (23)
MEF2C	rs190982	0.93 (0.90–0.95)	3.2×10^{-8}	Lambert et al. (23)
ZCWPW1	rs1476679	0.91 (0.89–0.94)	5.6×10^{-10}	Lambert et al. (23)
SLC24A4-RIN3	rs10498633	0.91 (0.88–0.94)	5.5×10^{-9}	Lambert et al. (23)
CD33[a]	rs3865444	0.91 (0.88–0.94)	1.6×10^{-9}	Hollingworth et al. (26)
MS4A	rs983392	0.90 (0.87–0.92)	6.1×10^{-16}	Hollingworth et al. (26)[c]
EPHA1	rs11771145	0.90 (0.88-0.93)	1.1×10^{-13}	Hollingworth et al. (26)[c]
CASS4	rs7274581	0.88 (0.84–0.92)	2.5×10^{-8}	Lambert et al. (23)
PICALM	rs10792832	0.87 (0.85–0.89)	9.3×10^{-26}	Harold et al. (24)[c]
CLU	rs9331896	0.86 (0.84–0.89)	2.8×10^{-25}	Harold et al. (24)[c]

Gene	SNP	OR (CI95%)	P Value	Author
SORL1	rs11218343	0.77 (0.72–0.82)	9.7×10^{-15}	Lambert et al. (23)
IGHV1-67	NA	NA	7.9×10^{-8}	Escott-Price et al. (36)
TP53INP1	NA	NA	1.4×10^{-6}	Escott-Price et al. (36)
17q21.31 region	rs2732703	0.73 (0.65-0.81)	5.8×10^{-9}	Jun et al. (35)
APOE	rs7412	0.62 (0.46 – 0.85)	2.7×10^{-3}	Corder et al. (20)[b]
RARE VARIANTS				
TREM2	rs75932628	5.05 (2.77-9.16)	9.0×10^{-9}	Guerreiro et al. (34)
UNC5C	rs137875858	2.15 (1.21-3.84)	9.5×10^{-3}	Wetzel-Smith et al. (45)
PLD3	rs145999145	2.10 (1.47–2.99)	2.9×10^{-5}	Cruchaga et al. (40)
APP	rs63750847	0.24 (NA)	4.2×10^{-5}	Jonsson et al. (38)

[a] Not replicated in follow-up studies or IGAP. [b] Estimator extracted from Bertram et al. (118). [c] Estimator extracted from Lambert et al. (23). IGAP: International Genomics Alzheimer's Project; LOAD: late onset Alzheimer's disease; NA: not available; OR: odds ratio.

In that sense, sub-population study revealed higher heterogeneity for Caucasian dataset with respect to Asiatic group (I^2 = 68%; p-value < 0.001; I^2 = 51%, p-value = 0.104). It must be considered that the major number of available studies is provided for Caucasians. Additionally, several studies have shown that the risk is higher for aMCI subpopulation in comparison to the rest of the MCI subtypes (55) (Table 2). This makes sense in the context of MCI as the prodromal stage of AD.

Although most of the attention has been focused on the risk allele ε4 of *APOE*, the ε2 allele has also demonstrated its role in AD (56). Some studies have shown that cognitively normal carriers of ε2 allele were less likely to present with cognitive decline (57), (58) and develop AD (relative risk [RR] = 0.76 [0.40–1.44]) (57).

Others showed that the MCI patients with ε2 allele had a better memory in comparison to non-carriers (59). Similarly, a recent study also detected the protective effect of *APOE* ε2 against pheno-conversion from MCI to AD (HR = 0.69 [0.51–0.86], p = 0.004) (60). These observations suggest the role of ε2 allele in the protection against LOAD and its importance as a possible mechanism to reverse the *APOE* ε4 effect (60).

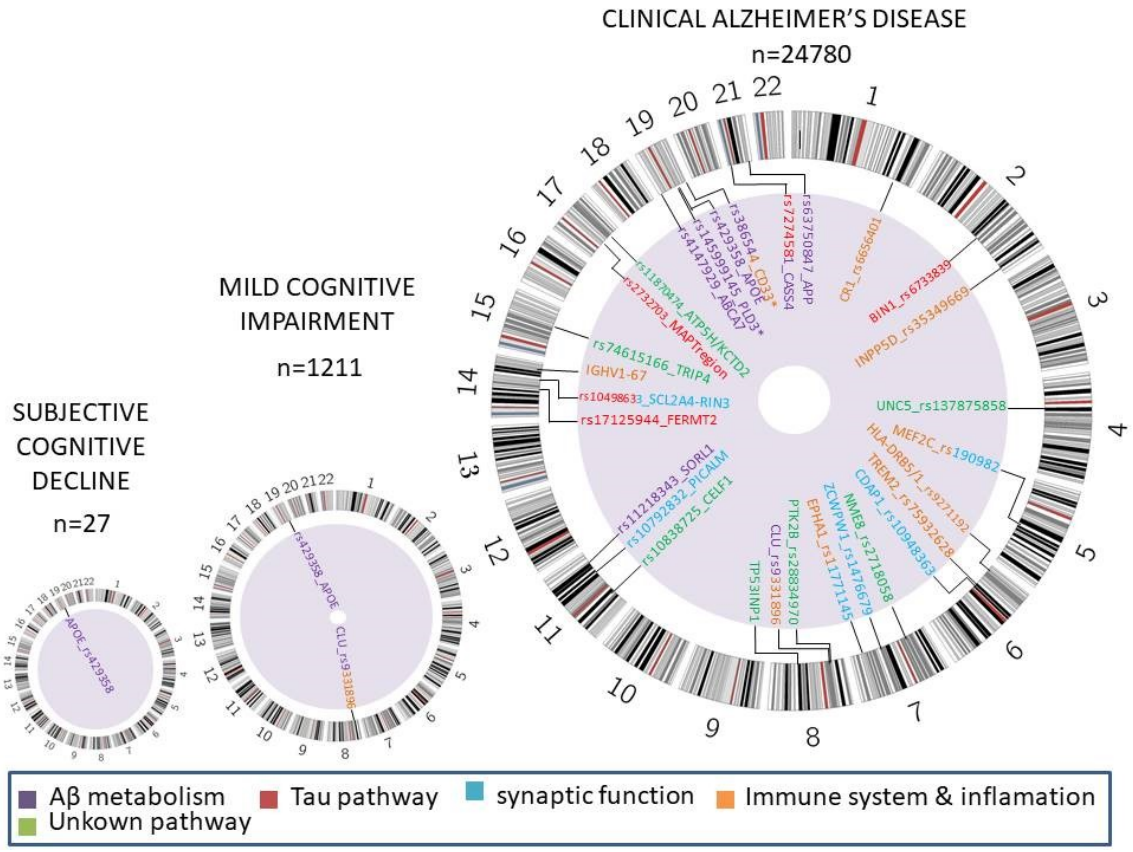


FIGURE 1. Genes by pathways associated with subjective cognitive decline (SCD), mild cognitive impairment (MCI) and clinical Alzheimer's disease (AD) and representation of level of genetic information at each stage

Non-*APOE* LOAD loci in MCI

Apart from *APOE*, the genome-wide significant variants for LOAD in MCI population remain largely unknown. However, in the past few years, several studies have pointed out their influence in cognitive decline.

Thus, *CR1* and *ABCA7* genes have been associated with faster rate of cognitive decline (*CR1* p-value = 0.011; *ABCA7* p-value = 0.013) (61, 62). *EPHA1* and *PICALM* loci were also associated with faster and slower rate of decline (*EPHA1* p = 0.013; *PICALM* p = 0.027), respectively, however they did not support Bonferroni corrections (63).

Following studies for *EPHA1* only replicated it marginally (p = 0.05) (62). Finally, the *CLU* locus was associated with cognitive endophenotypes in several studies. In that sense, the *CLU* risk allele has been associated with a faster rate of decline in some neuropsychological characteristics such as verbal immediate (p = 0.0032) and delayed free recall (p = 0.032) (64) and its protective allele with a

TABLE 2. Genetic variants associated with risk or progression to dementia in mci and SCD subjects

	Gene	Marker	OR (95%, CI)	p-value	Type of Study
MILD COGNITIVE IMPAIRMENT					
Risk Factors	*APOE*	rs42938	1.85 (1.59–2.15) 2.50 (1.13–5.69)[a]	$2.8 \cdot 10^{-15}$ 0.020	Meta-analysis Longitudinal study (55)
Progression Factors	*APOE*	rs429358	2.29 (1.88–2.8)	<0.001	Meta-analysis (54)
	CLU	rs11136000	1.19(0·05–1·32)[b]	0.003	GWAS (60)
	ACOT11	rs12752888	1.78 (1.21–2.64)[b]	<0.001	GWAS (63)
	UBR5-RRM2B	rs7840202	2.26 (1.56–3.26)[b]	<0.001	GWAS (63)
	Unknown	rs11637611	2.15 (1.4–3.31)[b]	<0.001	GWAS (63)
	MAPT	H1 haplotype	2.31 (1.52–3.51)	0.001	Candidate Gene (68)
	5-HTT	S allele	1.73 (1.05–2.84)	0.002	Candidate Gene (74)
	ADRA2B	Deletion (301-303)	0.49 (0.27–0.89)	0.021	Candidate Gene (76)
SUBJECTIVE COGNITIVE DECLINE					
Risk Factor	*APOE*	rs429358	1.15(1.02–1.30)[b]	0.028	Meta-analysis

[a]Amnestic MCI. [b]Parameter of association measure hazard ratio; MCI: mild cognitive impairment; SCD: subjective cognitive decline; OR: odds ratio.

decreased risk of conversion to AD (odds ratio [OR] = 0.25 [0.07–0.84], p = 0.025) (65).

A study that evaluated the progression of normal subjects (n > 2000) to MCI/LOAD pointed the significant effect of *CLU* on logical memory delayed recognition (LMDR) (coefficient for LMDR = −0.51 (−0.92 to −0.11, p = 0.012) (62). It further showed a borderline significant hazard ratio (HR) in the sensitivity analysis between *CLU* risk allele and the risk of progression from MCI to AD (HR = 1.10, p = 0.13; sensitivity HZ = 1.14, p = 0.049) (62). Recently, another experiment conducted on 3,326 MCI subjects of four countries supported the association of the *CLU* locus with the conversion from MCI to AD (HR = 1.19 [0.05–1.32], p = 0.0035) (60) (Table 2).

Most of the identified LOAD loci present small risk effects and are, therefore, not quite informative for risk prediction on their own (65). Consequently, it is apprehended that the use of the genetic risk score (GRS) strategy, where multiple loci with modest effects are combined, might improve the identification of people at risk for common diseases (65). It was observed that MCI carriers of six or more non-*APOE* LOAD risk alleles showed rapid conversion to AD (65). However, in another study, the significant effect was only reached when the *APOE* genotype was considered (HR =

1.29, p = 1.14 × 10^{-9}; sensitivity HR = 1.32, p = 5.73 × 10^{-10}) (62). In a recent study, GRS for 19 LOAD loci with genome-wide significance was associated with MCI (OR = 1.15, p = 0.011) and with progression from MCI to dementia (HR = 1.59, p < 0.001) (66).

With the exception of the *APOE* and *CLU* variants, it seems difficult to arrive at a conclusion about the role of LOAD SNPs in the context of MCI. Studies developed in larger cohorts are needed to check and validate the expected association between the LOAD risk genetic variants and MCI.

New loci associated with MCI

It has been observed that additional loci have been associated with MCI. Recently, a GWAS associated rs12752888 (*ACOT11* gene), rs7840202 (*UBR5* - *RRM2B* region), and rs11637611 (unknown gene) markers with MCI progression (63) (Table 2). However, establishing a relationship with the present pathophysiological hypothesis of AD, at this stage, seems complex. Consequently, these signals require validation if they were to be discarded as false positive and either to be accepted as factors responsible for MCI progression.

Alternatively, candidate gene studies have suggested several aspirant genes associated with the pheno-conversion of MCI to AD. One such example is the microtubule-associated protein tau (*MAPT*) gene. Specifically, a study revealed that the H1/H1 haplotype carriers presented a higher conversion rate of MCI to dementia (67). Moreover, the H1 haplotype has been associated with the risk of aMCI in converters to AD and non-converters (68) (Table 2).

Recently, the region 17q21.31, where *MAPT* was found, has been associated with LOAD reaching GWAS significance in non-*APOE* carriers (35). However, it remains unknown whether the casual variant is located in the *MAPT* or the nearby genes (35). For this reason, the *MAPT* linkage with MCI was explained at this point. Although *MAPT* is expected to be associated with LOAD because of the role of Tau protein in the classical AD hallmarks and the existence of several studies pointing towards this relationship, (69), (70), prudence is required until the location of the causal signal is identified.

Other genes associated with the pheno-conversion of MCI to AD are the vascular endothelial growth factor (71), the brain-derived neurotrophic factor (*BDNF*) (72) or butyrylcholinesterase (*BCHE*) (73). An additional marker in the serotonin transporter (*5HTT*) gene has also been related to MCI (74). It has also been associated with an emotion-induced retrograde amnesia (75), highlighting the role of serotonin in the memory system.

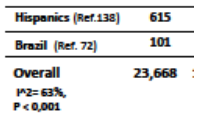


FIGURE 2. Forest plot for APOE ε4 genotype in (a) risk to mild cognitive impairment (MCI) and (b) risk to subjective cognitive decline (SCD).

OR: odds ratio.

The last reported association with MCI is detected in the α2b-adrenergic receptor (*ADRA2B*) (OR = 0.491 [0.268–0.899]; p = 0.021); this association is also identified in AD subjects (OR = 0.463 [0.261–0.822]; p = 0.009) (76) (Table 2). Since none of these genes have been validated for LOAD, it can be said that they may act as genetic progression factors. They are capable of modulating the rate of decline but are not involved in the risk leading to AD.

PRECLINICAL STAGE: SUBJECTIVE COGNITIVE DECLINE

There is limited research on the genetic variants that determine the risk to SCD or the progression of SCD to MCI or AD. Therefore, we have conducted the first meta-analysis exploring the involvement of *APOE* ε4 in the risk to suffer SCD. A significant risk effect was detected (OR = 1.151 [1.015–1.304]) (Figure 2) (Table 2), with a borderline non-significant heterogeneity (I^2=46%, p-value = 0.075), which remains when the analysis is only performed for Caucasians (I^2=48%, p-value = 0.087). However, this significant association disappears (OR = 1.158 [0.933 – 1.437]; p = 0.184) when a random model

is used to conduct the meta-analysis. From our point of view, these results must be taken with prudence. SCD individuals represent a mixed population, where a pool of subjects may develop dementia, not exclusively AD, and others never develop it. Hence, the sample size needed to detect Alzheimer's genuine genes must be larger.

Apart from the *APOE* ε4 polymorphism, other markers have been investigated to assess their possible association with SCD, such as alpha-2 macroglobulin gene (77), presenilin-1 mutation Glu318Gly (78), gene polymorphisms involved in vascular alterations (49), and inflammatory genes (80). However, all these studies did not report any association with SCD.

The genetic profile of the SCD subjects is unexplored in spite of the fact that its analysis could provide new ways to manage the disease. The generation of large SCD datasets integrating genomic information with follow-up data would be an essential step in identifying genetic elements responsible for the progression of SCD to MCI and AD.

OTHER APPROACHES: ENDOPHENOTYPE-BASED APPROACH

The use of quantitative traits closely related to the disease state, namely, endophenotypes, has been proposed as a simpler way to deal with genetic testing of LOAD. Thus, several endophenotypes have emerged across the cognitive spectrum of AD.

Differential amyloid burden and brain volume as endophenotype

Greater amyloid positron emission tomographic (PET) uptake is detected in AD, MCI and SCD cases as well as healthy controls who are carriers of the ε4 allele of *APOE* gene (81), (82). *APOE* ε4 carriers become positive for amyloid PET imaging earlier (81) and show a higher cognitive decline (83). Moreover, signals in *APOE* locus have been detected by GWAS of longitudinal studies for change in amyloid burden (84). Thus, in the past few years, *APOE* contribution to the determination of AD dementia converters has been reinforced. In addition, recent GWAS of longitudinal studies has also provided novel genetic correlations with the amyloid burden, such as *BCHE*, *TREM1*, and *ILR1RAP* (83), (85), 86).

Findings around differential brain volumes have also involved LOAD loci showing that a reduced hippocampal volume (HCV) is associated with *SORL1* in AD patients (87) and *CLU* gene in young healthy controls (88). *HLA-DRB1* locus was correlated with a decrease in total brain volume along large longitudinal cohorts (89). Putamen volume was also associated with genetic variants, involved in apoptosis, axon guidance, and vesicle transport (90). In that context, axon guidance pathway was also associated with reduced HCV, as well

as calcium and ErbB signaling (91). In addition, GRS for LOAD risk variants was associated with cortical thickness (92) and reduced HCV in cognitively normal subjects, although HCV association disappears after removing *APOE* locus (89).

Amyloid-β and tau levels in cerebrospinal fluid (CSF)

CSF levels of Aβ42 and pTau181 have also been used as LOAD endophenotypes. *APOE* locus has been associated with both Aβ42 (93), (94), (95) and pTau181 levels (94). However, Elias-Sonnenschein et al. (93) showed the correlations of *APOE* with Aβ42 but not with CSF tau biomarkers. Correlations with other LOAD loci remain scarce. Kauwe et al. (96) did not find any association for *BIN1*, *CR1*, *CLU*, and *PICALM*, although, recently, *CLU* and *MSA4A* have been associated with Aβ42 levels (93). In addition, GRS for AD loci did not provide any association (97). Novel identifications have pointed to the 3q28 region, *GLIS3* gene, and TREM cluster association with tau biomarkers (94) and *SUCLG2* association with Aβ42 levels (95). For further information, refer to Cruchaga et al. (98).

There is an inverse correlation between brain amyloid burden and Aβ CSF levels (99). From our understanding, both techniques are dealing with the same pathological process, Aβ deregulation. In that scenario, the identification of the following two might be expected: (1) the same genetic factors independently of the analyzed quantitative trait and (2) the reported LOAD loci associated with Aβ metabolism. Available data seem to be too far of these requisites, with the exception of *APOE*, the most consistent across studies and across AD stages. Therefore, from our view, an unsuited sample size affecting statistical power or the use of the incorrect endophenotypes of AD could be preventing new discoveries.

BRAIN GENETIC RESISTANCE FACTORS: STUDIES IN HEALTHY PEOPLE

The presence of Alzheimer-type pathology in healthy elderly people at death (100) evidenced the existence of compensatory mechanisms avoiding a cognitive decline in populations. A GWAS developed in this group of subjects suggested the involvement of the *RELN* in the compensatory mechanism for AD (101) and illustrated that studies on non-demented subjects with AD neuropathology are an interesting starting point to identify brain genetic resistance factors.

The state of resistance to brain insults, where the neuropathological hallmarks without clinical AD existing, has been defined as the cognitive reserve (CR). Individuals with higher CR tolerate the pathology for a longer duration and show signs of cognitive decline later in life (102). Environmental and genetic factors are also

believed to influence CR. It has been observed that the educational level (103), work complexity (104), engagement in leisure (105), or social activity (106) result in a reduced risk of dementia, contributing to CR. However, neuroplastic processes form the base of the above-mentioned factors. A study showed that the years of education is associated with genes involved in synaptic plasticity (107), and not surprisingly, cognition and neuroplasticity seem to be driven by shared genes (108).

The human cognition has a heritable component (109), (110). Cognition status is associated with several genetic variants, such as genes involved in oxidative stress (111), biosynthesis of neurotransmitters (112), ubiquitin metabolism, and immune system (113). It is to be noted that the immune system is highlighted as a prime pathway in LOAD (49) and is expected to be linked with cognition. The association of memory with *BDNF*, *5HTT*, and catechol-O-methyl transferase (*COMT*) genes remains more controversial (114). However, as indicated, *BDNF* and *5HTT* genes have also been associated with the conversion of MCI to AD (72), (74). These make them suitable candidates for further studies in genetics with MCI or SCD subjects.

Recently, two studies conducted on non-demented elderly subjects have showed that genes related to AD (*TOMM40*, *APOE*, *MEF2C*, and *ABCG1*) are significantly associated with the cognitive function (113), (115). This suggests that genes involved in the normal and pathological cognitions somehow overlap (115) and highlight the applicability of the studies performed on healthy people. In addition, the cognition status has also been related to differential brain volumes (116), thus, it seems that the hippocampal volume is a key component of the neuroanatomical basis of CR against memory in multiple sclerosis (117). Thus, although the existence of a genetic component influencing cognition is evident, its relevance in the health and disease processes remains unclear. However, it cannot be denied that its knowledge can bring new insights.

Either way, the investigation of genetic variants affecting cognition and brain structure in healthy people with and without AD neuropathology could be a starting point to determine the intrinsic genetic resistance to dementia. The information obtained through these studies must be comprehensively translated to evaluate its clinical utility in the preclinical stages of AD.

CONCLUSION

There exists an increasing interest in the characterization of the stages of pre-dementia. Taking into account the high genetic component of AD (14), the identification of genetic variants influencing MCI and SCD can provide a new perspective in tackling the disease.

Recent technological improvements have promoted the identification of 28 genetic variants for LOAD. Despite that, data concerning pre-dementia stages remain scarce. At present, *APOE* gene is the most consistent association with risk to MCI and progression from MCI to AD (54). The CLU locus has also showed promising results (60). There are more inconsistent data for SCD. To the best of our knowledge, we are the first to show meta-analyzed data evidencing the role of APOE as a risk factor for SCD (Figure 2).

There is highly pronounced absence of genetic data for pre-dementia stage (Figure 1). In addition, there is a high degree of heterogeneity between available studies in pre-dementia stage. In an extended way, the MCI studies show a statistical correlation between the genotype and neuropsychological test scores, which mainly provides informative data. In most cases, studies with SCD and MCI individuals have small sample sizes. Moreover, the consideration of population characteristics seems pertinent. SCD and MCI individuals comprise highly heterogeneous population, where converters to AD dementia coexist with converters to other forms of dementia and non-converters. In that scenario, the identification of novel and expected LOAD loci may be hampered by the effect size of the true AD group. That could explain the reduction in the effect size of *APOE* along stages. On the other hand, the identification of progression factors that are not previously reported must be considered with prudence until their validation. This limitation points to the necessity of using larger cohorts in studies involving a pre-dementia stage.

Efforts are required to provide useful data, which can help in designing strategies to stop or modulate the course of the disease. In that sense, a GWAS in the MCI population seems mandatory. Moreover, the identification of the genetic factors conferring resilience to dementia in non-demented people could provide a good opportunity in uncovering the compensatory mechanisms that may prevent the disease progression. Therefore, a genome-wide approach for endophenotypes involved in CR also seems affordable and advisable.

In conclusion, genetic research in the pre-dementia stages of non-demented people must be potentiated to obtain advances in AD and design prevention strategies.

REFERENCIES

1. Sperling, R.A. et al. (2011) Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's and Dementia* **7**, 280–92.

2. Albert, M.S. et al. (2011) The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's and Dementia* **7**, 270–9.

3. McKhann, G.M. et al. (2011) The diagnosis of dementia due to Alzheimer's disease:

recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's and Dementia* **7**, 263–9.

4. Caselli, R.J. et al. (2009) Longitudinal modeling of age-related memory decline and the APOE epsilon4 effect. *The New England Journal of Medicine* **361**, 255–63.

5. Jessen, F. et al. (2014) A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. *Alzheimer's and Dementia* **10**, 844–52.

6. Reisberg, B. et al. (2010) Outcome over seven years of healthy adults with and without subjective cognitive impairment. *Alzheimer's and Dementia* **6**, 11–24.

7. Amieva, H. (2008) Prodromal Alzheimer's disease: successive emergence of the clinical symptoms. *Annals of Neurology* **64**, 492–8.

8. Jonker, C., Geerlings, M.I. and Schmand, B. (2000) Are memory complaints predictive for dementia? A review of clinical and population-based studies. *International Journal of Geriatrics Psychiatry* **15**, 983–91.

9. Jessen, F. (2010) Prediction of Dementia by Subjective Memory Impairment. *Archive of General Psychiatry* **67**, 414–422.

10. Petersen, R.C. et al. (1999) Mild cognitive impairment: clinical characterization and outcome. *Archives of Neurology* **56**, 303–8.

11. Espinosa, A. et al. (2013) A longitudinal follow-up of 550 mild cognitive impairment patients: evidence for large conversion to dementia rates and detection of major risk factors involved. *Journal of Alzheimer's Disease* **34**, 769–80.

12. Jack, C.R. et al. (2010) Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology* **9**, 119–28.

13. Drago, V. et al. (2011) Disease tracking markers for Alzheimer's disease at the prodromal (MCI) stage. *Journal of Alzheimer's Disease* **26** Suppl 3, 159–99.

14. Avramopoulos, D. (2009) Genetics of Alzheimer's disease: recent advances. Genome Medicine 1, 34.

15. Goldman, Jill S. et al. (2011) Genetic counseling and testing for Alzheimer Disease: Joint practice guidelines of the American College of Medical Genetics and the National Society of Genetic Counselors. *Genetics in medicine official journal of the American College of Medical Genetics* **13**, 597–605.

16. Goate, A. et al. (1991) Segregation of a missense mutation in the amyloid precursor protein gene with familial alzheimer's disease. *Nature* **349**, 704–6.

17. Sherrington, R. et al. (1995) Cloning of a gene bearing missense mutations in Early-Onset Familial Alzheimer's Disease. *Nature* **375**, 754–60.

18. Levy-Lahad, E., Wijsman E.M and Nemens, E. (1995) A Familial Alzheimer's Disease locus on chromosome 1. *Science* **411**, 16–19.

19. Raux, G. et al. (2005) Molecular diagnosis of autosomal dominant early onset Alzheimer's Disease: An Update. *Journal of Medical Genetics* **42**, 793–95.

20. Corder E, Saunders A (1993) Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **261**, 921–3.

21. Genin, E. et al. (2011) APOE and Alzheimer Disease: A major gene with semi-dominant inheritance. *Molecular psychiatry* **16**, 903–7.

22. Rogaeva, E. et al. (2007) The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. *Nature Genetics* **39**, 168–77.

23. Lambert, J.C. et al. (2013) Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature Genetics* **45**, 1452–8.

24. Harold, D. et al. (2009) Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nature Genetics* **41**, 1088–93.

25. Seshadri, S. et al. (2011) Genome-wide Analysis of Genetic Loci Associated with Alzheimer's Disease. *JAMA*. **303**, 1832–1840.

26. Hollingworth, P. et al. (2011) Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nature Genetics* **43**, 429–35.

27. Jun, G. et al. (2010) Meta-analysis confirms CR1, CLU, and PICALM as Alzheimer disease risk loci and reveals interactions with APOE genotypes. *Archives of Neurology* **67**, 1473–84.

28. Carrasquillo, M. et al. (2011) Replication of BIN1 association with Alzheimer's disease and evaluation of genetic interactions. *Journal of Alzheimer's Disease* **24**, 751–8.

29. Lee, J.H. et al. (2011) Identification of novel loci for Alzheimer disease and replication of CLU, PICALM, and BIN1 in Caribbean Hispanic individuals. *Archives of Neurology* **68**, 320–8.

30. Yu, J-T. (2013) Genetic variation in clusterin gene and Alzheimer's disease risk in Han Chinese. *Neurobiology of Aging* **34**, 1921.e17–23.

31. Jin, C. et al. (2012) Association of the CR1 polymorphism with late-onset Alzheimer's disease

in Chinese Han populations: a meta-analysis. *Neuroscience Letters* **527**, 46–9.

32. Carrasquillo, M.M. et al. (2011) Replication of EPHA1 and CD33 associations with late-onset Alzheimer's disease: a multi-centre case-control study. *Molecular Neurodegeneration* **6**, 54.

33. Reitz, C. et al. (2013) Variants in the ATP-binding cassette transporter (ABCA7), apolipoprotein E ϵ4,and the risk of late-onset Alzheimer disease in African Americans. *JAMA* **309**, 1483–92.

34. Ruiz, A. et al. (2014) Follow-up of loci from the International Genomics of Alzheimer's Disease Project identifies TRIP4 as a novel susceptibility gene. *Translational Psychiatry* **4**, e358.

35. Jun, G. et al. (2015) A novel Alzheimer disease locus located near the gene encoding tau protein. *Molecular Psychiatry*. doi: 10.1038/mp.2015.23. [Epub ahead of print]

36. Escott-Price, V. (2014) Gene-wide analysis detects two new susceptibility genes for Alzheimer's disease. *PLoS One* **9**, e94661.

37. Boada, M. et al. (2014) ATP5H/KCTD2 locus is associated with Alzheimer's disease risk. *Molecular Psychiatry* **19**, 682–7.

38. Jonsson, T. et al. (2012) A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* **488**, 96–9.

39. Guerreiro, R. et al. (2013) TREM2 variants in Alzheimer's disease. *The New England Journal of Medicine* **368**, 117–27.

40. Cruchaga, C. et al. (2014) Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer ' s disease. *Nature* **505**, 550–554.

41. Wetzel-Smith, M.K. et al. (2014) A rare mutation in UNC5C predisposes to late-onset Alzheimer's disease and increases neuronal cell death. *Nature Medicine* **20**, 1452–7.

42. Benitez, B.A. et al. (2013) TREM2 is associated with the risk of Alzheimer's disease in Spanish population. *Neurobiology of Aging* **34**, 1711.e15–7.

43. Ruiz, A. et al. (2014) Assessing the role of the TREM2 p.R47H variant as a risk factor for Alzheimer's disease and frontotemporal dementia. *Neurobiology of Aging* **35**, 444.e1–4.

44. Paloneva, J. (2002) Mutations in two genes encoding different subunits of a receptor signaling complex result in an identical disease phenotype. *American Journal of Human Genetics* **71**, 656–62.

45. Thelen, M. (2014) Investigation of the role of rare TREM2 variants in frontotemporal dementia subtypes. *Neurobiology of Aging* **35**, 2657.e13–9.

46. Hooli, B.V. et al. (2015) PLD3 gene variants and Alzheimer's disease. *Nature* **520**, E7–8.

47. Heilmann, S. et al. (2015) PLD3 in non-familial Alzheimer's disease. *Nature* **520**, E3–E5.

48. Lambert, J-C. et al. (2010) Implication of the immune system in Alzheimer's disease: evidence from genome-wide pathway analysis. *Journal of Alzheimer's Disease* **20**, 1107–18.

49. International Genomics of Alzheimer's disease Consortium (IGAP) (2014) Convergent genetic and expression data implicate immunity in Alzheimer's disease. *Alzheimer's and Dementia* **14**, 02492-3.

50. Petersen, R.C. et al. (1995) Apolipoprotein E status as a predictor of the development of Alzheimer's disease in memory-impaired individuals. *JAMA* **273**, 1274–8.

51. Fleisher, A.S. et al. (2007) Clinical predictors of progression to Alzheimer disease in amnestic mild cognitive impairment. *Neurology* **68**, 1588–95.

52. Artero, S. et al. (2008) Risk profiles for mild cognitive impairment and progression to dementia are gender specific. *Journal of Neurology Neurosurgery & Psychiatry* **79**, 979–84.

53. Fei, M. and Jianhua, W. (2013) Apolipoprotein ε4-allele as a significant risk factor for conversion from mild cognitive impairment to Alzheimer's disease: a meta-analysis of prospective studies. *Journal of Molecular Neuroscience* **50**, 257–63.

54. Elias-Sonnenschein, L.S. et al. (2011) Predictive value of APOE-ε4 allele for progression from MCI to AD-type dementia: a meta-analysis. *Journal of Neurology Neurosurgery & Psychiatry* **82**, 1149–56.

55. Lopez, O.L. et al. (2003) Risk factors for mild cognitive impairment in the Cardiovascular Health Study Cognition Study: part 2. *Archives of Neurology* **60**, 1394–9.

56. Corder, E.H. et al. (1994) Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. *Nature Genetics* **7**, 180–4.

57. Wilson, R.S. et al. (2002) The apolipoprotein E epsilon 2 allele and decline in episodic memory. *Journal of Neurology Neurosurgery & Psychiatry* **73**, 672–7.

58. Blacker, D. (2007) Neuropsychological measures in normal individuals that predict subsequent cognitive decline. *Archives of Neurology* **64**, 862–71.

59. Bonner-Jackson, A., Okonkwo, O. and Tremont, G. (2012) Apolipoprotein E ε2 and functional decline in amnestic mild cognitive impairment and Alzheimer disease. *American Journal of Geriatrics Psychiatry* **20**, 584–93.

60. Lacour, A. et al. (2015) The role of well-established Alzheimer's disease. Genetic risk factors in the process of Mild Cognitive Impairment phenol-conversion to AD. Personal Communication: 9th Panhellenic Conference on Alzheimer's disease & 1st Mediterranean on Neurodegenerative Disease. Thessaloniki, 2015.

61. Chibnik, L.B. et al. (2011) CR1 is associated with amyloid plaque burden and age-related cognitive decline. *Annals of Neurology* **69**, 560–9.

62. Carrasquillo, M.M. et al. (2015) Late-onset Alzheimer's risk variants in memory decline, incident mild cognitive impairment, and Alzheimer's disease. *Neurobiology of Aging* **36**, 60–7.

63. Hu, X. et al. (2011) Genome-wide association study identifies multiple novel loci associated with disease progression in subjects with mild cognitive impairment. *Translational Psychiatry* **1**, e54.

64. Thambisetty, M. et al. (2013) Alzheimer risk variant CLU and brain function during aging. *Biological Psychiatry* **73**, 399–405.

65. Rodríguez-Rodríguez, E. et al (2013) Genetic risk score predicting accelerated progression from mild cognitive impairment to Alzheimer's disease. *Journal of Neuronal Transmission* **120**, 807–12.

66. Adams, H. et al. (2015) Genetic risk of neurodegenerative disease is associated with mild cognitive impairment and conversion to dementia. *Alzheimer's and Dementia* **15**, 118-1.

67. Samaranch, L. et al. (2010) The effect of MAPT H1 and APOE ε4 on transition from mild cognitive impairment to dementia. *Journal of Alzheimer's Disease* **22**, 1065–71.

68. Di Maria, E. et al. (2010) The H1 haplotype of the tau gene (MAPT) is associated with mild cognitive impairment. *Journal of Alzheimer's Disease* **19**, 909-14.

69. Cruchaga, C. et al. (2010) SNPs associated with cerebrospinal fluid phospho-tau levels influence rate of decline in Alzheimer's disease. *PLoS Genetics* **6**, e1001101.

70. Myers, A.J. et al. (2005) The H1c haplotype at the MAPT locus is associated with Alzheimer's disease. *Human Molecular Genetics* **14**, 2399–404.

71. Chiappelli, M. et al. (2006) VEGF gene and phenotype relation with Alzheimer's disease and mild cognitive impairment. *Rejuvenation Research* **9**, 485–93.

72. Forlenza, O.V. et al. (2010) Effect of brain-derived neurotrophic factor Val66Met polymorphism and serum levels on the progression of mild cognitive impairment. *World Journal Biological Psychiatry* **11**, 774–80.

73. Ferris, S. et al. (2009) Progression from mild cognitive impairment to Alzheimer's disease: effects of sex, butyrylcholinesterase genotype, and rivastigmine treatment. *Pharmacogenetics and Genomics* **19**, 635–46.

74. Marini, S. et al. (2011) Implication of serotonin-transporter (5-HTT) gene polymorphism in subjective memory complaints and mild cognitive impairment (MCI). *Archives of Gerontology and Geriatrics* **52**, e71–4.

75. Strange, B.A. et al. (2008) Emotion-induced retrograde amnesia is determined by a 5-HTT genetic polymorphism. *Journal of Neuroscience* **28**, 7036–9.

76. Koutroumani, M. et al. (2013) The deletion variant of α2b-adrenergic receptor is associated with decreased risk in Alzheimer's disease and mild cognitive impairment. *Journal of Neurological Sciences* **328**, 19–23.

77. Zill, P. et al. (2000) Polymorphisms in the alpha-2 macroglobulin gene in psychogeriatric patients. *Neurosciense Letters* **294**, 69–72.

78. Laws, S.M. et al. (2002) Association between the presenilin-1 mutation Glu318Gly and complaints of memory impairment. *Neurobiology of Aging* **23**, 55–8.

79. Watfa, G. et al. (2010) Association study of gene polymorphisms involved in vascular alterations in elderly hypertensives with subjective memory complaints. *Dementia Geriatrics Cognitive Disorders* **30**, 440–8.

80. Lau, S. et al. (2012) Functional effects of genetic polymorphism in inflammatory genes in subjective memory complainers. *Neurobiology of Aging* **33**, 1054–6.

81. Fleisher, A.S. et al. (2013) Apolipoprotein E ε4 and age effects on florbetapir positron emission tomography in healthy aging and Alzheimer disease. *Neurobiology of Aging* **34**, 1–12.

82. Risacher, S.L. et al. (2015) APOE effect on Alzheimer's disease biomarkers in older adults with significant memory concern. *Alzheimers and Dementia*.

83. Mormino, E.C. et al. (2014) Amyloid and APOE ε4 interact to influence short-term decline in preclinical Alzheimer disease. *Neurology* **82**, 1760–7.

84. Ramanan, V.K. et al. (2014) APOE and BCHE as modulators of cerebral amyloid deposition: a florbetapir PET genome-wide association study. *Molecular Psychiatry* **19**, 351–7.

85. Replogle, J. et al. (2015) A TREM1 variant alters the accumulation of Alzheimer-related amyloid pathology. *Annals of Neurology* **77**, 469-77.

86. Ramanan, V.K. et al. (2015) GWAS of longitudinal amyloid accumulation on 18F-

florbetapir PET in Alzheimer's disease implicates microglial activation gene IL1RAP. *Brain* **138**, 3076-88.

87. Louwersheimer, E. et al. (2015) The influence of genetic variants in SORL1 gene on the manifestation of Alzheimer's disease. *Neurobiology of aging* 36, 1605.e13-20.

88. Lancaster, T.M. et al. (2015) Alzheimer's disease risk variant in CLU is associated with neural inefficiency in healthy individuals. *Alzheimers and Dementia* **11**, 1144–52.

89. Chauhan, G. et al. (2015) Association of Alzheimer's disease GWAS loci with MRI markers of brain aging. *Neurobiology of Aging* **36**, 1765.e7–16.

90. Hibar, D.P. et al. (2015) Common genetic variants influence human subcortical brain structures. *Nature* **520**, 224-9

91. Meda, S. et al. (2013) Genetic interactions associated with 12-month atrophy in hippocampus and entorhinal cortex in Alzheimer's Disease Neuroimaging Initiative. *Neurobiology of Aging* **34,** 1518.e9-18.

92. Sabuncu, M. (2011) The Association between a Polygenic Alzheimer Score and Cortical Thickness in Clinically Normal Subjects. *Cerebral Cortex* **22**, 2653-2661.

93. Elias-Sonneschein, L. et al. (2013) Genetic loci associated with Alzheimer's disease and cerebrospinal fluid biomarkers in a Finnish case-control cohort. *PloS One* **8**, e59676.

94. Cruchaga, C. et al. (2013) GWAS of cerebrospinal fluid tau levels identifies risk variants for Alzheimer's disease. *Neuron* **78**, 256-68.

95. Ramirez, A. et al. (2014) SUCLG2 identified as both a determinator of CSF Aβ1-42 levels and an attenuator of cognitive decline in Alzheimer's disease. *Human Molecular Genetics* **23**, 6644-58.

96. Kauwe, J. et al. (2011) Fine mapping of genetic variants in BIN1, CLU, CR1 and PICALM for association with cerebrospinal fluid biomarkers for Alzheimer's disease. *PloS One* **6**, e15918.

97. Martinskainen, H. et al. (2015) Effects of Alzheimer's disease-associated risk loci on cerebrospinal fluid biomarkers and disease progression: a polygenic risk score approach. *Journal Of Alzeimer's Disease* **43**, 565-73.

98. Cruchaga, C., Ebbert, M., Kauwe, J. (2014) Genetic discoveries in AD using CSF amyloid and tau. *Current genetic medicine reports* **2**, 23-29.

99. Fagan, A.M. et al. (2006) Inverse correlation between in vivo amyloid imaging load and cerebrospinal fluid Abeta42 in humans. *Annals of Neurology* 59, 512-9.

100. Neuropathology Group. MedicalResearch Council Cognitive Function and Aging Study. (2001) Pathological correlates of late-onset dementia in a multicentre, community-based population in England and Wales. Neuropathology Group of the Medical Research Council Cognitive Function and Ageing Study (MRC CFAS). *The Lancet* **357**, 169–75.

101. Kramer, P.L. et al. (2011) Alzheimer disease pathology in cognitively healthy elderly: a genome-wide study. *Neurobiology of Aging* **32**, 2113–22.

102. Stern, Y. (2012) Cognitive reserve in ageing and Alzheimer's disease. *The Lancet Neurology* **11**, 1006–12.

103. Sando, S.B. et al. (2008) Risk-reducing effect of education in Alzheimer's disease. *International Journal of Geriatric Psychiatry* **23**, 1156–62.

104. Andel, R. et al. (2005) Complexity of work and risk of Alzheimer's disease: a population-based study of Swedish twins. *The Journals of Gerontology Series B* **60**, P251–8.

105. Scarmeas, N. et al. (2001) Influence of leisure activity on the incidence of Alzheimer's disease. *Neurology* **57**, 2236–42.

106. Wilson, R.S. (2009) Elderly women with larger social networks are less likely to develop dementia. *Evidence-Based Mental Health* **12**, 22.

107. Rietveld, C. et al. (2014) Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 13790-4.

108. Van Veluw, S.J. et al. (2012) Prefrontal cortex cytoarchitecture in normal aging and Alzheimer's disease: a relationship with IQ. *Brain structure & function* **217**, 797-808

109. Deary, I.J., Johnson, W. and Houlihan L.M. (2009) Genetic foundations of human intelligence. *Human Genetics* **126**, 215–32.

110. Glahn, D.C. (2013) Genetic basis of neurocognitive decline and reduced white-matter integrity in normal human brain aging. *Proceedings of National Academy of Sciences of the United States of America* **110**, 19006–11.

111. Harris, S.E. et al. (2007) A genetic association analysis of cognitive ability and cognitive ageing using 325 markers for 109 genes associated with oxidative stress or cognition. *BMC Genetics* **8**, 43.

112. Barral, S. et al. (2014) Common genetic variants on 6q24 associated with exceptional episodic memory performance in the elderly. *The Journal of the American Medical Association Neurology* **71**, 1514–9.

113. Debette, S. et al. (2014) Genome-wide studies of verbal declarative memory in non-demented older

people: the CHARGE consortium. *Biological Psychiatry* **8**, 749-73.

114. Sabb, F.W. et al. (2009) Challenges in phenotype definition in the whole-genome era: multivariate models of memory and intelligence. *Neuroscience* **164**, 88–107.

115. Davies, G. et al. (2015) Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53 949). *Molecular Psychiatry* **20**, 183–92.

116. Dowling, N.M. et al. (2011) Neuropathological associates of multiple cognitive functions in two community-based cohorts of older adults. *Journal of the International Neuropsychological Society* **17**, 602-14.

117. Sumowski, J.F. et al. (2015) Searching for the neural basis of reserve against memory decline: intellectual enrichment linked to larger hippocampal volume in multiple sclerosis. *European Journal of Neurology*. doi: 10.1111/ene.12662.

118. Bertram, L. et al. (2007) Systematic meta-analyses of Alzheimer disease genetic

119. Lambert, J. (2009) Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nature Genetics* **41**, 1094-9.

120. Krzywinski, M. et al. (2009) Circos: an information aesthetic for comparative genomics. *Genome Research* **19**, 1639–1645.

121. Wang, X. et al. (2014) Frequency of the apolipoprotein E ε4 allele in a memoryclinic cohort in Beijing: a naturalistic descriptive study. *Plos One* 9, e99130.

122. Jin, C. et al. (2014) The *SORL1* polymorphism rs985421 may confer the risk for amnestic mild cognitive impairment and Alzheimer's disease in the Han Chinese population. *Neuroscience Letter* 563, 80-81.

123. Sasaki, M. et al. (2009) Prevalence of four subtypes of mild cognitive impairment and APOE in a Japanese community. *International Journal of Geriatric Psychiatry* **24**, 1119-1126.

124. Kim, S.M. et al. (2013) Regional cerebral perfusion in patients with Alzheimer's disease and mild cognitive impairment: effect of APOE Epsilon4 allele. *Neuroradiology* 55, 25-34.

125. Foster J.K. et al. (2013) Lack of reliable evidence for a distinctive ε4-related cognitive phenotype that is independent from clinical diagnostic status: findings from the Australian Imaging, Biomarkers and Lifestyle Study. *Brain* **136**, 2201–16

126. Troyer A.K. et al. (2012) Associative recognition in mild cognitive impairment: relationship to hippocampal volume and apolipoprotein E. *Neuropsychologia* **50**, 3721–8.

127. Traykov L, et al. (1999) Apolipoprotein E phenotypes in demented and cognitively impaired patients with and without cerebrovascular disease. *European Journal of Neurology* 6, 415–21.

128. Luck, T. et al. (2010) Risk factors for incident mild cognitive impairment--results from the German Study on Ageing, Cognition and Dementia in Primary Care Patients (AgeCoDe). *Acta Psychiatrica Scandinavica* **121**, 260–72.

129. Chico L, et al. (2013) Oxidative stress and APOE polymorphisms in Alzheimer's disease and in mild cognitive impairment. *Free Radical Research* **47**, 569–76.

130. Siuda J, et al. (2009) From mild cognitive impairment to Alzheimer's disease - influence of homocysteine, vitamin B12 and folate on cognition over time: results from one-year follow-up. *Neurologia i neurogirurgia Polska* **43**, 321–9.

131. Cervantes S, et al. (2011) Genetic variation in APOE cluster region and Alzheimer's disease risk. *Neurobiology of Aging* **32**, 2107.e7–17.

132. Bartrés-Faz D. et al. (2001) Neuropsychological and genetic differences between age-associated memory impairment and mild cognitive impairment entities. *Journal of the American Geriatrics Society* **49**, 985–90.

133. Buerger K, et al. (2005) Increased levels of CSF phosphorylated tau in apolipoprotein E epsilon4 carriers with mild cognitive impairment. *Neuroscience Letter* **391**(1-2), 48–50.

134. Zekry D, et al.(2010) Telomere length and ApoE polymorphism in mild cognitive impairment, degenerative and vascular dementia. *Journal of Neurological Science* **299**, 108–11.

135. Jefferson, A.L. et al. (2014) APOE and mild cognitive impairment: the Framingham Heart Study. *Age and Ageing* **2**, 307-11.

136. Altmann A, et al. (2014) Sex modifies the APOE-related risk of developing Alzheimer disease. *Annals of Neurology* **75**, 563–73.

137. Tyas S.L. et al. (2007) Transitions to mild cognitive impairments, dementia, and death: findings from the Nun Study. *American Journal of Epidemiology* **165**, 1231–8.

138. O'Bryant S.E. et al. (2013) Risk factors for mild cognitive impairment among Mexican Americans. *Alzheimers & Dementia.* **9**, 622–631.

139. Boyle P.A. et al (2010) The APOE ε4 Allele Is Associated with Incident Mild Cognitive Impairment among Community-Dwelling Older Persons. *Neuroepidemiology* **34**, 43–49.

140. Stewart R, et al. (2001) Depression, APOE genotype and subjective memory impairment: a cross-sectional study in an African-Caribbean population. *Psychological Medicine* **31**, 431–40.

141. Laws S.M. et al. (2002) APOE-epsilon4 and APOE -491A polymorphisms in individuals with subjective memory loss. *Molecular Psychiatry* **7**, 768–75.

142. Verdile G, et al. (2012) Associations between gonadotropins, testosterone and β amyloid in men at risk of Alzheimer's disease. *Molecular Psychiatry* **19,** 69–75.

143. Jessen F, et al. (2007) Patterns of subjective memory impairment in the elderly: association with memory performance. *Psychological Medicine* **37**, 1753–62.

144. Risacher S.L. *et al.* (2015) APOE effect on Alzheimer's disease biomarkers in older adults with significant memory concern. *Alzheimers & Dementia* **11***,* 1417-29.

145. Caselli R.J, et al. (2014) Subjective cognitive decline: self and informant comparisons. *Alzheimers & Dementia* **10**, 93–8.

146. Samieri, C. et al. (2014) Subjective cognitive concerns, episodic memory, and the APOE ε4 allele. *Alzheimer's and Dementia* **10**, 752–759.

ACKNOWLEDGEMENTS

We are indebted to Trinitat Port-Carbó and her family for their support to the Fundació ACE research programs. Fundació ACE collaborates with the Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED, Spain) and is one of the participating centers of the Dementia Genetics Spanish Consortium (DEGESCO). **The present work has been performed as part of the Biochemistry, Molecular Biology and Biomedicine doctoral program of S. Moreno-Grau at Universitat Autònoma de Barcelona (Barcelona, Spain).**

FINANCIAL SUPPORT

This work is supported by Grant PI13/02434 (Acción Estratégica en Salud. Instituto de Salud Carlos III (ISCIII). Ministerio de Economía y Competitividad, Spain), and Obra Social 'La Caixa' (Barcelona, Spain).

Conflict of interest: None.

Comparison of two new commercial genotyping platforms in a cohort of Alzheimer's disease patients

Sonia Moreno-Grau [1], Isabel Hernández [1], Begoña Hernández-Olasagarre [1], Laura Montrreal[1], tziar de Rojas.[1], GR@ACE & DEGESCO consortia, Gemma Monté-Rubio [1], Sergi Valero [1], Adelina Orellana [1], Lluís Tárraga [1], Mercè Boada [1], Agustín Ruiz [1]

[1]Research Center and Memory clinic Fundació ACE. Institut Català de Neurociències Aplicades. Universitat Internacional de Catalunya. Barcelona. Spain

Poster Presentation

Genomics of Brain disorders, 23-24 April 2018

Wellcome Genome Campus, Hinxton, Cambridge, UK

INTRODUCTION

Genome-wide studies have become a common tool to identify genetic variants associated to complex traits. A new generation of genome-wide arrays has been introduced in the field. We sought to explore the performance of two commercial high-throughput genotyping arrays, the 815K Axiom Spain Biobank array (Axiom BA, Affymetrix) and the Infinium global screening array (GSA, Illumina), in 443 Spanish individuals of the Alzheimer's disease cohort from Fundacio ACE, Barcelona, Spain.

METHODS

Two independent Spanish cohorts of Alzheimer's disease were used. The Gr@ACE cohort is composed by 4,124 AD cases and 3,290 controls, which were genotyped using the Axiom BA array in the Spanish National Genotyping Center (Santiago de Compostela, Spain). The second dataset, the FACE EADB cohort (925 cases, 826 MCI and 1,507 controls), was genotyped using GSA array in Life and Brain Center (Bonn, Germany). A total of 443 individuals were genotyped using both platforms.

Wet genotyping was performed according to manufacturer's instruction in each case. SNP calling was conducted using APT software for Axiom BA array and

TABLE 1. Marker failure rate after GWAS Quality Control

GWAS Quality Control	815K Axiom Affymetrix	GSA Illumina
No. Markers	814 923	696 375
Internal probe sets	56 183 (6.9%)	53 551 (7.8%)
Markers failing Standard Quality Control	111 754 (13.7%)	103 112 (14.8%)
Markers MAF < 1%	54 597 (6.7%)	108 195 (15.5%)

Genome Studio for GSA. The same GWAS quality control was applied for each dataset using PLINK software 1.9 (https://www.cog-genomics.org/plink2). Briefly, samples with heterozygosity rate >3SD were excluded. Gender discrepancies were removed. Relatedness, re-sampling and potential undetected contamination were evaluated conducting Identity by Descent (IBD) analysis (PiHat > 0.1875).

Principal component analysis was performed to detect individual ancestry. Markers with different missing rate between cases and controls, failing HWE ($p<0.000001$) and rare variants (MAF<0.01) were excluded. Call rate per sample and per marker was established in 97% and 95%, respectively.Four quality datasets were imputed. The experiment A comprised entire Gr@ACE and FACE EADB cohorts (N > 3,000) and the experiment B includes individuals genotyped using both platforms (N = 443). Imputation was performed using HRC reference panel in Michigan Server (https://imputationserver.sph.umich.edu/start.html).

IBD analysis was conducted to evaluate concordance rate (COR) between platforms using PLINK 1.9 software

RESULTS

After GWAS QC, both platforms present similar proportion of failure rate but GSA had less common variants available (Table 1). Axiom BA array imputed more variants with acceptable quality for further research (R2>0.3; MAF<0.01), independently from dataset sample size (Table 2).

The mean quality for the acceptable variants was higher for Axiom BA platform across different allelic frequencies (Figure 1). The concordance of common variants between both platforms in genotyped data was higher than 0.99 (SD;0.016) (Figure 2). The samples size and the quality of

TABLE 2. Number of markers after imputation procedure

IMPUTATION	Experiment A		Experiment B	
	815K Axiom Affymetrix	GSA Illumina	815K Axiom Affymetrix	GSA Illumina
No. Individuals	7 414	3 258	443	443
Markers R2>0.3	25 206 529	20 248 594	14 724 604	13 677 810
Markers R2>0.3 MAF > 0.01	7 734 408	7 653 310	7 720 363	7 614 477

imputed genotypes are also impacting the concordance rate (COR) between platforms (Figure 2).

CONCLUSIONS

Axiom BA array genotyped correctly more SNPs and imputed a higher number of high quality variants. The higher number of common variants presents in this array could be driving these findings. Both methods provided similar genotype and imputation accuracies displaying also an excellent concordance among genotyped SNPs.

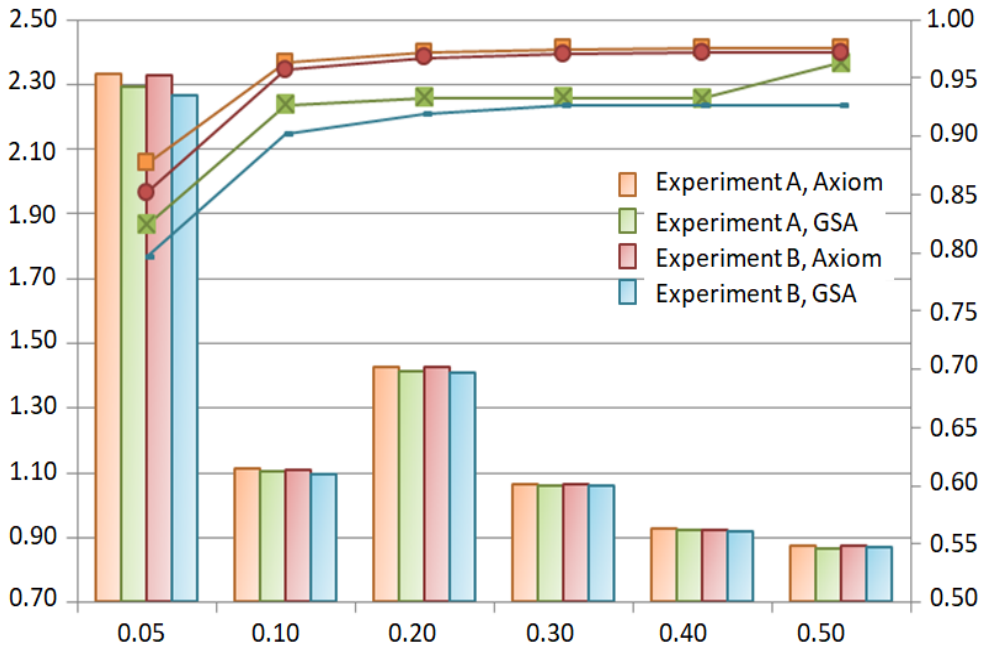


FIGURE 1. Number of imputed variants and imputation quality according to allelic frequency per quality dataset.

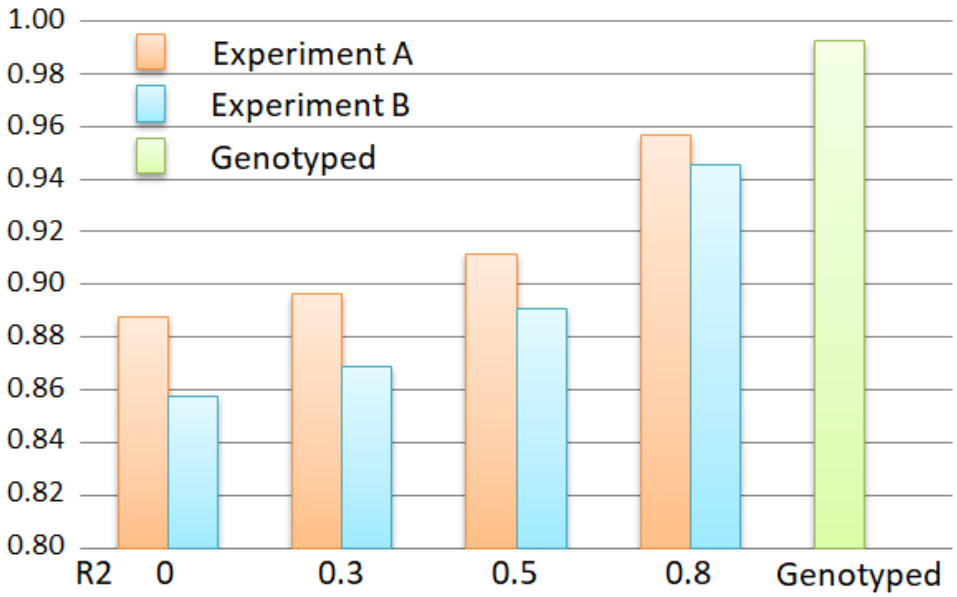


FIGURE 2. Concordance Rate (Cor) between Axiom BA and GSA platforms by imputation quality.

Other publication

Saez ME, Gonzalez-Perez A, Hernandez-Olasagarre B, Bea A, Moreno-Grau S, Rojas I de, et al. Genome Wide Meta-Analysis identifies new loci associated with cardiac phenotypes and uncovers a common genetic signature shared by heart function and Alzheimer's disease. Sci Rep. Accepted 2019 Sep 23.

van der Lee SJ, Conway OJ, Jansen I, Carrasquillo MM, Kleineidam L, van den Akker E, et al. A nonsynonymous mutation in PLCG2 reduces the risk of Alzheimer's disease, dementia with Lewy bodies and frontotemporal dementia, and increases the likelihood of longevity. Acta Neuropathol. 2019 Aug;138(2):237-250.

Real LM, Macías J, Rivero-Juárez A, Téllez F, Merino D, Moreno-Grau S, et al. Genetic markers of lipid metabolism genes associated with low susceptibility to HCV infection. Sci Rep. 2019 Jun 21;9(1):9054.

Marquié M, Castilla-Martí M, Valero S, Martínez J, Sánchez D, Hernández I, et al. Visual impairment in aging and cognitive decline: experience in a Memory Clinic. Sci Rep. 2019 Jun 18;9(1):8698.

Real LM, Fernández-Fuertes M, Sáez ME, Rivero-Juárez A, Frías M, Téllez F, et al. A genome-wide association study on low susceptibility to hepatitis C virus infection (GEHEP012 study). Liver Int. 2019 Jun 17.

Rongve A, Witoelar A, Ruiz A, Athanasiu L, Abdelnour C, Clarimon J, et al. GBA and APOE ε4 associate with sporadic dementia with Lewy bodies in European genome wide association study. Sci Rep. 2019 May 7;9(1):7013.

Franzmeier N, Ren J, Damm A, Monté-Rubio G, Boada M, Ruiz A, et al. The BDNF (Val66Met) SNP modulates the association between beta-amyloid and hippocampal disconnection in Alzheimer's disease. Mol Psychiatry. 2019 Mar 21.

Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. Nat Genet. 2019 Mar;51(3):414-430.

de Rojas I, Romero J, Rodríguez-Gomez O, Pesini P, Sanabria A, Pérez-Cordon A, et al. Correlations between plasma and PET beta-amyloid levels in individuals with subjective cognitive decline: the Fundació ACE Healthy Brain Initiative (FACEHBI). Alzheimers Res Ther. 2018 Nov 29;10(1):119.

Sánchez D, Castilla-Marti M, Rodríguez-Gómez O, Valero S, Piferrer A, Martínez G, et al. Usefulness of peripapillary nerve fiber layer thickness assessed by optical coherencetomography as a biomarker for Alzheimer's disease. Sci Rep. 2018 Nov 5;8(1):16345.

Espinosa A, Hernández-Olasagarre B, Moreno-Grau S, Kleineidam L, Heilmann-Heimbach S, Hernández I, Exploring Genetic Associations of Alzheimer's Disease Loci With Mild Cognitive Impairment Neurocognitive Endophenotypes. Front Aging Neurosci. 2018 Oct 30;10:340.

Sanabria A, Alegret M, Rodriguez-Gomez O, Valero S, Sotolongo-Grau O, Monté-Rubio G, et al. The Spanish version of Face-Name Associative Memory Exam (S-FNAME) performance is related to amyloid burden in Subjective Cognitive Decline. Sci Rep. 2018 Feb 28;8(1):3828.

Sims R, van der Lee SJ, Naj AC, Bellenguez C, Badarinarayan N, Jakobsdottir J, et al. Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. Nat Genet. 2017 Sep;49(9):1373-1384.

DeMichele-Sweet MAA, Weamer EA, Klei L, Vrana DT, Hollingshead DJ, Seltman HJ, et al. Genetic risk for schizophrenia and psychosis in Alzheimer disease. Mol Psychiatry. 2018 Apr;23(4):963-972.

Rodriguez-Gomez O, Sanabria A, Perez-Cordon A, Sanchez-Ruiz D, Abdelnour C, Valero S, et al. FACEHBI: A Prospective Study of Risk Factors, Biomarkers and Cognition in a Cohort of Individuals with Subjective Cognitive Decline. Study Rationale and Research Protocols. J Prev Alzheimers Dis. 2017;4(2):100-108.

Abdelnour C, Rodríguez-Gómez O, Alegret M, Valero S, Moreno-Grau S, Sanabria Á, Hernández I, et al. Impact of Recruitment Methods in Subjective Cognitive Decline. J Alzheimers Dis. 2017;57(2):625-632.

Lacour A, Espinosa A, Louwersheimer E, Heilmann S, Hernández I, Wolfsgruber S, et al. Genome-wide significant risk factors for Alzheimer's disease: role in progression to dementia due to Alzheimer's disease among subjects with mild cognitive impairment. Mol Psychiatry. 2017 Jan;22(1):153-160.

Dols-Icardo O, Nebot I, Gorostidi A, Ortega-Cubero S, Hernández I, Rojas-García R, et al. Analysis of the CHCHD10 gene in patients with frontotemporal dementia and amyotrophic lateral sclerosis from Spain. Brain. 2015 Dec;138(Pt 12):e400.

About the author

Sonia Moreno Grau was born in Valencia, Spain, on the 23rd of November of 1990. After finishing her pre-scientific education, she started her studies in Pharmacy at University of Valencia, Valencia, Spain (2008–2013). In 2013, she came to Barcelona, where she obtained her Master of Biochemistry (Biochemistry, Molecular Biology and Biomedicine) from Autonomous University of Barcelona, Barcelona, Spain. During her master's, she worked in the apoptosis and cell signaling lab, led by Dr. Joan Xavier Comella Carnicé, at the Vall d'Hebron Research Intitute (VHIR). After a year, she started working on this thesis under the supervision of Dr. Agustín Ruiz Laza at the genomics lab of Fundació ACE. During this research, she worked as part of the international consortia Cohorts of Heart and Aging Research in Genomic Epidemiology (CHARGE) and the International Genomics Alzheimer's Project (IGAP). To improve her knowledge of large-scale genotyping technologies, she visited the Genomic unit of the Life and Brain Center in Bonn, Germany in 2016. She was supervised by Dr. Alfredo Ramírez and and Dr. Stefanie Heilmann. To complete her training, one year later, she was in the Spanish National Center for Genotyping (CeGEN) led by Dr. Ángel Carracedo, at Santiago de Compostela, Spain. She was supervised by Dr. Inés Quintela. In 2018, she visited the neurogenomics and informatics lab of Dr. Carlos Cruchaga, in the Department of Psychiatry, Washington University School of St. Louis, St. Louis, Missouri, United States for a three-month research stay. She was supervised by Dr. Maria Victoria Fernández. Her research interests include the genetics of Alzheimer's disease, bioinformatics and integrative genomics, as well as population genetics to understand human evolutionary history. Specifically, she is focused on the exploration of long runs of homozygosity as strategy to study the recessive component of Alzheimer's disease.