Unraveling Molecular Mechanisms Underlying Alzheimer's Disease and its Related Endophenotypes

Sven J. van der Lee

Unraveling Molecular Mechanisms Underlying Alzheimer's disease and its Related Endophenotypes

Sven J. van der Lee

Acknowledgments

The work described in this thesis was conducted at the Department of Epidemiology at the Erasmus Medical Center, Rotterdam, The Netherlands.

The Erasmus Rucphen Family study as a part of EUROSPAN (European Special Populations Research Network) was supported by European Commission FP6 STRP grant number 018947 (LSHG-CT-2006-01947) and also received funding from the European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F4-2007-201413 by the European Commission under the programme "Quality of Life and Management of the Living Resources" of 5th Framework Programme (no. QLG2-CT-2002-01254). The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The study participants, researchers, employees and funders of all other studies used in this thesis are gratefully acknowledged. Their generous contribution of time and devotion to build and enforce scientific studies of high quality are at the basis of unraveling Alzheimer's diseases.

Other financial support leading to this thesis: ADAPTED: Alzheimer's Disease Apolipoprotein Pathology for Treatment Elucidation and Development (number 115975); the CoSTREAM project (www.costream.eu), the European Union's Horizon 2020 research and innovation programme under grant agreement No 667375, Population imaging genetics (ImaGene).

Printing of this thesis was supported by: The Erasmus MC, the department of epidemiology, Erasmus MC, Alzheimer Nederland, Nutricia Research, Chipsoft.

ISBN: 978-90-9030560-8

Layout: Sven J. van der Lee and Gildeprint BV.

Cover: Leaf of *Sinnigia Leucotricha* covered in the genetic code of *PLCG2* p.P522R. design in conjunction with Peter de Bakker

Unraveling Molecular Mechanisms Underlying Alzheimer's disease and its Related Endophenotypes

Het ontrafelen van de moleculaire mechanismen onderliggend aan de ziekten van Alzheimer en gerelateerde endofenotypen

Proefschrift

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

prof.dr. H.A.P. Pols

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op

Vrijdag 13 oktober 2017 om 9:30 uur

Sven Johan van der Lee
geboren te Delft, Nederland

Erasmus University Rotterdam

Promotiecommissie:

Promotoren:	Prof.dr.ir C.M. van Duijn
	Prof.dr. M.A. Ikram
Overige leden	Prof.dr. M.M.B. Breteler
	Prof.dr. T. Hankemeier
	Prof.dr. P. Scheltens
Copromotor:	Dr. N. Amin

Sine Labore Nihil

Table of Contents

Publications and manuscripts in this thesis

Chapter 2.1

Sven J. van der Lee, Frank J. Wolters, M. Kamran Ikram, Albert Hofman, M. Arfan Ikram, Najaf Amin, Cornelia M. van Duijn. *The effect of common genetic variants on the onset of Alzheimer's disease and dementia in carriers of the APOE ε4 genotype.* (Submitted)

Chapter 2.2

Henne Holstege, Sven J. van der Lee, Marc Hulsman, Tsz Hang Wong, Jeroen G.J. van Rooij, Marjan Weiss, Eva Louwersheimer, Frank J. Wolters, Najaf Amin, André G. Uitterlinden, Albert Hofman, M. Arfan Ikram, John C. van Swieten, Hanne Meijers-Heijboer, Wiesje M. van der Flier, Marcel J. T. Reinders, Cornelia M. van Duijn, Philip Scheltens. *Characterization of pathogenic SORL1 genetic variants for association with Alzheimer's disease: A clinical interpretation strategy*. European Journal of Human Genetics. 2017 May 24; 25: 973–981.

Chapter 2.3

Frank J. Wolters, Sven J. van der Lee, Peter J. Koudstaal, Cornelia M. van Duijn, Albert Hofman, M. Kamran Ikram, Meike W. Vernooij, M. Arfan Ikram. *Parental family history of dementia in relation to subclinical brain disease and dementia risk.* Neurology. 2017 April 25; 88(17): 1642-1649.

Chapter 3.1

Adriana I. Iglesias*, Sven J. van der Lee*, Pieter W.M. Bonnemaijer, René Höhn, Abhishek Nag, Puya Gharahkhani, Anthony P. Khawaja, Linda Broer, Paul J. Foster, Christopher J. Hammond, Pirro G. Hysi, Elisabeth M. van Leeuwen, Stuart MacGregor, David A. Mackey, Johanna Mazur, Stefan Nickels, André G. Uitterlinden, Caroline C.W. Klaver, Najaf Amin, Cornelia M. van Duijn. *Haplotype Reference Consortium Panel: Practical implications of imputations with large reference panels.* Human Mutation 2017 August; 38(8): 1025-1032.

* equal contribution of first authors

Chapter 3.2

Sven J. van der Lee, Henne Holstege, Tsz Hang Wong, Johanna Jakobsdottir, Joshua C. Bis, Vincent Chouraki, Jeroen van Rooij, Megan L. Grove, Albert V. Smith, Najaf Amin, Seung-Hoan Choi, Alexa S. Beiser, Melissa E. Garcia, Wilfred F.J. van IJcken, Yolande A.L. Pijnenburg, Eva Louwersheimer, Rutger W.W. Brouwer, Mirjam C.G.N. van den Hout, Edwin Oole, Gudny Eirkisdottir, Daniel Levy, Jerome I. Rotter, Valur Emilsson, Christopher J. O'Donnell, Thor Aspelund, Andre G. Uitterlinden, Lenore J. Launer, Albert Hofman, Alzheimer's Disease Neuroimaging Initiative (ADNI), Eric Boerwinkle, Bruce M. Psaty, Anita L. DeStefano, Philip Scheltens, Sudha Seshadri, John C. van Swieten, Vilmundur Gudnason, Wiesje M. van der Flier, M. Arfan Ikram, Cornelia M. van Duijn. *PLD3-variants in population studies.* Nature. 2015 April 2; 520(7545): E2–E3.

Chapter 3.3

Johanna Jakobsdottir*, Sven J. van der Lee*, Joshua C. Bis*, Vincent Chouraki*, David Li-Kroeger, Shinya Yamamoto, Megan L. Grove, Adam Naj, Maria Vronskaya, Jose L. Salazar, Anita L. DeStefano, Jennifer A. Brody, Albert V. Smith, Najaf Amin, Rebecca Sims, Carla A. Ibrahim-Verbaas, Seung-Hoan Choi, Claudia L. Satizabal, Oscar L. Lopez, Alexa Beiser, M. Arfan Ikram, Melissa E. Garcia, Caroline Hayward, Tibor V. Varga, Samuli Ripatti, Paul W. Franks, Göran Hallmans, Olov Rolandsson, Jan-Håkon Jansson, David J. Porteous, Veikko Salomaa, Gudny Eirkisdottir, Kenneth M. Rice, Hugo J. Bellen, Daniel Levy, Andre G. Uitterlinden, Valur Emilsson, Jerome I. Rotter, Thor Aspelund, Cohorts for Heart and Aging Research in Genomic Epidemiology consortium, Alzheimer's Disease Genetic Consortium, Genetic and Environmental Risk in Alzheimer's Disease consortium, Christopher J. O'Donnell, Annette L. Fitzpatrick, Lenore J. Launer, Albert Hofman, Li-San Wang, Julie Williams, Gerard D. Schellenberg, Eric Boerwinkle**, Bruce M. Psaty**, Sudha Seshadri**, Joshua M. Shulman**, Vilmundur Gudnason**, Cornelia M. van Duijn**. *Rare functional variant in TM2D3 is associated with late-onset Alzheimer's disease.* PLOS Genetics. 2016 October 20: e1006327.

* equal contribution of first authors

** equal contribution of senior authors

Chapter 3.4

Rebecca Sims*, Sven J. van der Lee*~, Adam C. Naj*, Céline Bellenguez*, Nandini Badarinarayan, Johanna Jakobsdottir, Brian W. Kunkle, Anne Boland, Rachel Raybould, Joshua C. Bis, Eden R. Martin, Benjamin Grenier-Boley, Stefanie Heilmann-Heimbach, Vincent Chouraki, Amanda B. Kuzma, Kristel Sleegers, Maria Vronskaya, Agustin Ruiz, Robert R. Graham, Robert Olaso, Per Hoffmann, Megan L. Grove, Badri N. Vardarajan, Mikko Hiltunen, Markus M. Nöthen, Charles C. White, Kara L. Hamilton-Nelson, Jacques Epelbaum, Wolfgang Maier, Seung-Hoan Choi, Gary W. Beecham, Cécile Dulary, Stefan Herms, Albert V. Smith, Cory C. Funk, Céline Derbois, Andreas J. Forstner, Shahzad Ahmad, Hongdong Li, Delphine Bacq, Denise Harold, Claudia L. Satizabal, Otto Valladares, Alessio Squassina, Rhodri Thomas, Jennifer A. Brody, Liming Qu, Pascual Sanchez-Juan, Taniesha Morgan, Frank J. Wolters, Yi Zhao, Florentino Sanchez Garcia, Nicola Denning, Myriam Fornage, John Malamon, Maria Candida Deniz Naranjo, Elisa Majounie, Thomas H. Mosley, Beth Dombroski, David Wallon, Michelle K Lupton, Josée Dupuis, Patrice Whitehead, Laura Fratiglioni, Christopher Medway, Xueqiu Jian, Shubhabrata Mukherjee, Lina Keller, Kristelle Brown, Honghuang Lin, Laura B. Cantwell, Francesco Panza, Bernadette McGuinness, Sonia Moreno-Grau, Jeremy D. Burgess, Vincenzo Solfrizzi, Petra Proitsi, Hieab H. Adams, Mariet Allen, Davide Seripa, Pau Pastor, L. Adrienne Cupples, Nathan D Price, Didier Hannequin, Ana Frank-García, Daniel Levy, Paramita Chakrabarty, Paolo Caffarra, Ina Giegling, Alexa S. Beiser, Vimantas Giedraitis, Harald Hampel, Melissa E. Garcia, Xue Wang, Lars Lannfelt, Patrizia Mecocci, Gudny Eiriksdottir, Paul K. Crane, Florence Pasquier, Virginia Boccardi, Isabel Henández, Robert C. Barber, Martin Scherer, Lluis Tarraga, Perrie M. Adams, Markus Leber, Yuning Chen, Marilyn S. Albert, Steffi Riedel-Heller, Valur Emilsson, Duane Beekly, Anne Braae, Reinhold Schmidt, Deborah Blacker, Carlo Masullo, Helena Schmidt, Rachelle S. Doody, Gianfranco Spalletta, WT Longstreth, Jr, Thomas J. Fairchild, Paola Bossù, Oscar L. Lopez, Matthew P. Frosch, Eleonora Sacchinelli, Bernardino Ghetti, Pascual Sánchez-Juan, Qiong Yang, Ryan M. Huebinger, Frank Jessen, Shuo Li, M. Ilyas Kamboh, John Morris, Oscar Sotolongo-Grau, Mindy J. Katz, Chris Corcoran, Jayanadra J. Himali, C. Dirk Keene, JoAnn Tschanz, Annette L. Fitzpatrick, Walter A. Kukull, Maria Norton, Thor Aspelund, Eric B. Larson, Ron Munger, Jerome I. Rotter, Richard B. Lipton, María J Bullido, Albert Hofman, Thomas J.

Montine, Eliecer Coto, Eric Boerwinkle, Ronald C. Petersen, Victoria Alvarez, Fernando Rivadeneira, Eric M. Reiman, Maura Gallo, Christopher J. O'Donnell, Joan S. Reisch, Amalia Cecilia Bruni, Donald R. Royall, Martin Dichgans, Mary Sano, Daniela Galimberti, Peter St George-Hyslop, Elio Scarpini, Debby W. Tsuang, Michelangelo Mancuso, Ubaldo Bonuccelli, Ashley R. Winslow, Antonio Daniele, Chuang-Kuo Wu, GERAD/PERADES, CHARGE, ADGC, EADI, Oliver Peters, Benedetta Nacmias, Matthias Riemenschneider, Reinhard Heun, Carol Brayne, David C Rubinsztein, Jose Bras, Rita Guerreiro, John Hardy, Ammar Al-Chalabi, Christopher E Shaw, John Collinge, David Mann, Magda Tsolaki, Jordi Clarimón, Rebecca Sussams, Simon Lovestone, Michael C O'Donovan, Michael J Owen, Timothy W. Behrens, Simon Mead, Alison M. Goate[a], Andre G. Uitterlinden, Clive Holmes, Carlos Cruchaga, Martin Ingelsson, David A. Bennett, John Powell, Todd E. Golde, Caroline Graff, Philip L. De Jager, Kevin Morgan, Nilufer Ertekin-Taner, Onofre Combarros, Bruce M. Psaty, Peter Passmore, Steven G Younkin, Claudine Berr, Vilmundur Gudnason, Dan Rujescu, Dennis W. Dickson, Jean-Francois Dartigues, Anita L. DeStefano, Sara Ortega-Cubero, Hakon Hakonarson, Dominique Campion, Merce Boada, John "Keoni" Kauwe, Lindsay A. Farrer, Christine Van Broeckhoven, M. Arfan Ikram, Lesley Jones, Johnathan Haines, Christophe Tzourio, Lenore J. Launer, Valentina Escott-Price, Richard Mayeux, Jean-François Deleuze, Najaf Amin, Peter A Holmans, Margaret A. Pericak-Vance, Philippe Amouyel**, Cornelia M. van Duijn**, Alfredo Ramirez**, Li-San Wang**, Jean-Charles Lambert**, Sudha Seshadri**, Julie Williams**~, Gerard D. Schellenberg**~. *Rare coding variants in PLCG2, ABI3 and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease.* Nature Genetics 2017 July 17 [online publication ahead of print].

* equal contribution of first authors

** equal contribution of senior authors

~ corresponding authors

Chapter 4.1

Sven J. van der Lee*, Gennady V. Roshchupkin*, Hieab H.H. Adams*, Helena Schmidt, Edith Hofer, Yasaman Saba, Reinhold Schmidt, Albert Hofman, Najaf Amin, Cornelia M. van Duijn, Meike W. Vernooij, M. Arfan Ikram, Wiro J. Niessen. *Grey matter*

heritability in family-based and population-based studies using voxel-based morphometry. Human Brain Mapping. 2017 May; 38(5): 2408-2423.
* equal contribution of first authors

Chapter 4.2

Gennady V. Roshchupkin*, Hieab H.H. Adams*, Sven J. van der Lee*, Meike W. Vernooij, Cornelia M. van Duijn, André G. Uitterlinden, Aad van der Lugt, Albert Hofman, Wiro J. Niessen, M. Arfan Ikram. *Fine-mapping the effects of Alzheimer's disease risk loci on brain morphology.* Neurobiology of Aging. 2016 December; 48: 204-211.
* equal contribution of first authors

Chapter 4.3

Sven J. van der Lee*, Hieab H.H. Adams*, Najaf Amin, Lisa R. Yanek, Rasika A. Mathias, Albert Vernon Smith, Tamara B. Harris, Konstantinos Arfanakis, Lei Yu, Saima Hilal, Ching Yu Cheng, Tien Yin Wong, Wan Ting Zhao, Edith Hofer, Yasaman Saba, Derrek P. Hibar, Josh W. Cheung, Neda Jahanshad, Tomas Paus, Manon Bernard, Joshua C. Bis, Oscar L. Lopez, Kent D. Taylor, Ken Rice, WT Longstreth, Claudia Satizabal, Ganesh Chauhan, Joris Deelen, Erik van den Akker, Mariam Beekman, Paul A. Nyquist, Diane M. Becker, Lenore J. Launer, Vilmundur Gudnason, David A. Bennett, Philip L. De Jager, Christopher Chen, M. Kamran Ikram, Reinhold Schmidt, Helena Schmidt, Paul M. Thompson, Zdenka Pausova, Bruce M. Psaty, Sudha Seshadri, Stephanie Debette, Eline Slagboom Cornelia M. van Duijn, Meike W. Vernooij, M. Arfan Ikram, Charles S. Decarli *Novel genetic loci associated with brain lobar volumes.* (In preparation)
* equal contribution of first authors

Chapter 5.1

Vincent Chouraki*, Sven van der Lee*, Carlos Cruchaga, Benjamin Grenier-Boley, Charles DeCarli, Charles White, Christophe Tzourio, Chritiane Reitz, Claudia Satizabal, Claudine Berr, Giuseppe Tosto, Hieab H. Adams, Jean-François Dartigues, Jeannette Simino, Luc Buée, M Arfan Ikram, Najaf Amin, André G. Uitterlinden, Anita DeStefano, Qiong Yang, Shuo Li, Steven Younkin, Susanna Schraen, Thomas Mosley, ADNI, Phil de

Jager, Philippe Amouyel, Alexa Beiser, Jean Charles Lambert**, Sudha Seshadri**, Cornelia M van Duijn** *1000 Genomes-based Genome-wide Association Study Links APOE ε4 and BACE1 Variants with Plasma Amyloid-β Levels.* (Submitted)

* equal contribution of first authors

** equal contribution of senior authors

Chapter 5.2

Juho Tynkkynen*,Vincent Chouraki*, Sven J. van der Lee, Jussi Hernesniemi, Qiong Yang, Shuo Li, Alexa Beiser, Martin G Larson, Katri Sääksjärvi, , Archana Singh-Manoux, Robert E Gerszten, Thomas J. Wang, Aki S. Havulinna, Peter Wùrtz, Krista Fischer, Ayse Demirkan, M Arfan Ikram, Najaf Amin, Markus Perola, Andres Metspalu, Antti J. Kangas, Pasi Soininen, Mika Ala-Korpela, Ramachandran S Vasan, Mika Kivimäki, Cornelia M. van Duijn, Sudha Seshadri**, Veikko Salomaa**. *Association of branched chain amino acids and other circulating metabolites with risk of incident dementia and Alzheimer's disease: a prospective study in eight cohorts.* (Submitted)

* equal contribution first authors

** equal contribution senior authors

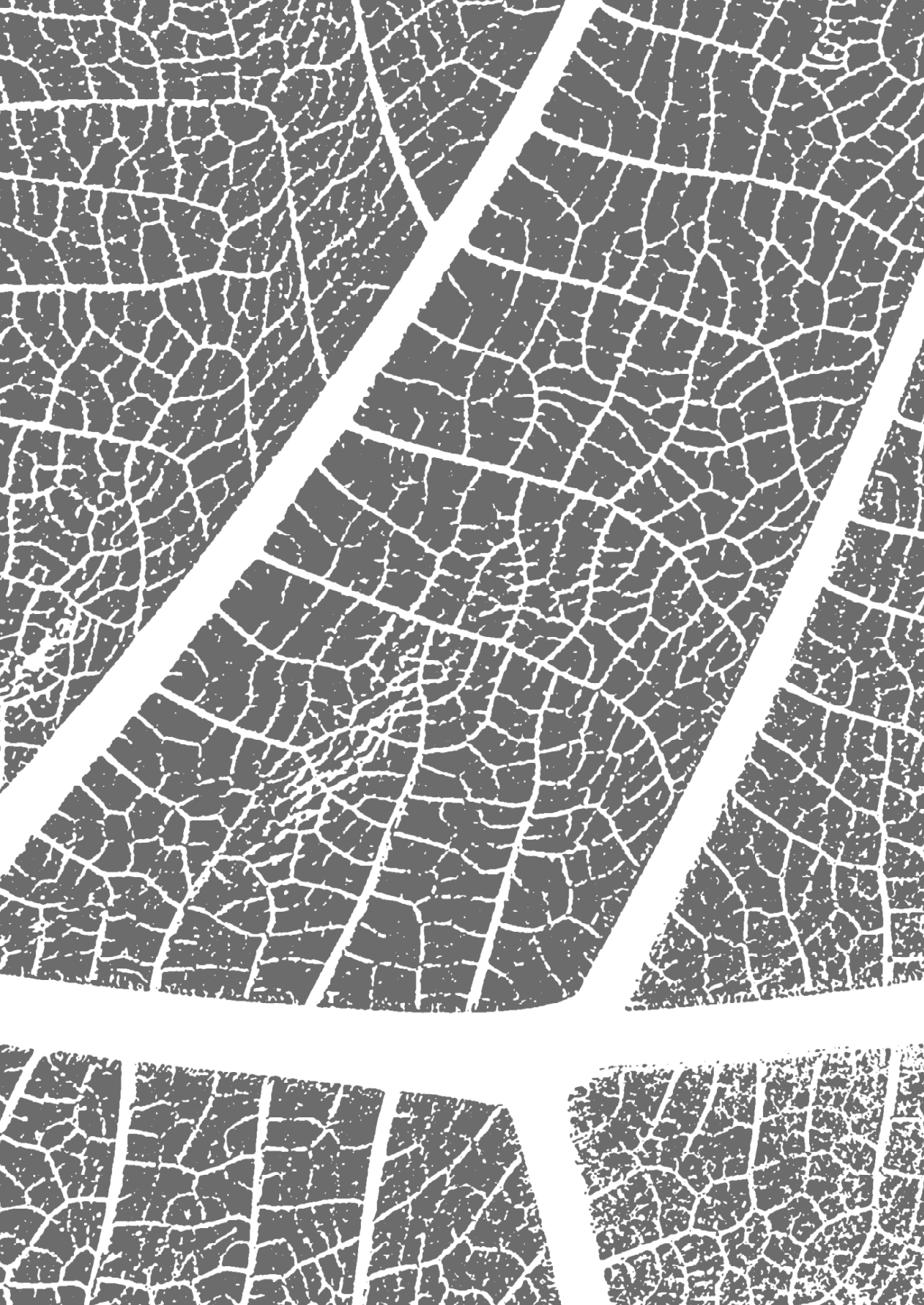
Chapter 5.3

Sven J. van der Lee, Charlotte E. Teunissen, René Pool, Martin J. Shipley, Alexander Teumer, Vincent Chouraki, Debora Melo van Lent, Juho Tynkkynen, Krista Fischer, Jussi Hernesniemi, Andres Metspalu, Archana Singh-Manoux, Aswin Verhoeven, Gonneke Willemsen, Francien A. de Leeuw, Holger Wagner, Jenny van Dongen, Johannes Hertel, Kathrin Budde, Ko Willems van Dijk, Leonie Weinhold, M. Arfan Ikram, Maik Pietzner, Markus Perola, Michael Wagner, Nele Friedrich, P.E. Slagboom, Philip Scheltens, Qiong Yang, Robert E. Gertzen, Sarah Egert, Shuo Li, Thomas Hankemeier, Catharine E.M. van Beijsterveldt, Vasan Ramachandran, Wolfgang Maier, Carel F.W. Peeters, Hans Jörgen Grabe, Alfredo Ramirez, Sudha Seshadri, Toomas Haller, Mika Kivimäki, Veikko Salomaa, Ayşe Demirkan, Dorret Boomsma, Wiesje M. van der Flier, Najaf Amin, Cornelia M. van Duijn. *Circulating metabolites and general cognitive ability and dementia: Evidence from 11 cohort studies.* (Submitted)

Chapter 5.4

Sven J. van der Lee*, Dina Vojinovic*, Cornelia M van Duijn, Meike W. Vernooij, Oscar H. Franco, Najaf Amin, Thomas Hankemeier, Ayşe Demirkan, M. Arfan Ikram, Aad van der Lugt, Daniel Bos. *Metabolic profiling of intracranial arteriosclerosis*. (In preparation)

* equal contribution of first authors



Chapter 1

General introduction

Dementia poses a huge burden on patients, their family, friends, and society. In 2015, 47 million people worldwide were suffering from the disease and this number is expected to rapidly increase in the coming decades due to the aging population.[1,2] The most common type (50-70%) of dementia is Alzheimer's disease (AD).[3] There is currently no effective treatment to prevent AD, or to treat its progressive short-term memory complaints, language problems or the physical decline later in the disease course.[4] The natural course of AD inevitably leads to total dependence on others for care.[4] The immense number of future patients and lack of preventive options make dementia and AD the major health challenge of the coming decades.[1] Unraveling the pathophysiology of AD must be a global research priority to be able to identify an effective treatment.[4]

Clinical epidemiology of AD

At autopsy, Alzheimer's disease is defined by the typical combined presence of amyloid-β plaques and neurofibrillary tangles. Early-onset AD patients (age at onset before 65 years) usually only have the typical neuropathological features of AD,[5] but many patients with a late age of onset show signs of cerebrovascular disease in addition to the "classic" Alzheimer's neuropathology.[6,7] Research in the past decade already led to major progress in understanding the epidemiology of late onset AD and dementia (age at onset after 65 years). These epidemiological studies have confirmed that cardiovascular factors and low education are risk factors of dementia[8,9] and that a quarter to third of dementia cases could potentially be prevented through optimal prevention or treatment of cardiovascular risk factors and diseases and improvement of educational level.[8] Cardiovascular disease prevention and increased educational levels possibly caused the declining trend in the incidence of AD and dementia, as observed in multiple population-based studies.[10,11] However, classical cardiovascular risk factors leave many AD patients unexplained. Part of the disease is explained by genetics as the percentage of AD occurrence explained by genetic factors is estimated to be from 60 to 80%.[12] Studying the genetics of AD is still an important driver of etiological and predictive AD research.

Genetics of Alzheimer's disease

Since previous century the progress in identifying genetic risk factors for AD has been exponential (**Figure 1).** These discoveries have played a pivotal role in our understanding of the pathophysiology of these neuropathological findings of AD. For instance, that the plaques that at autopsy contained aggregated amyloid-β was already known since 1984.[13] However, the discovery of Mendelian mutations in the gene coding for the APP protein (*APP*) and in presenilin 1 and 2 genes (*PSEN1* and *PSEN2*) facilitated the understanding of the molecular mechanism underlying amyloid-B pathology. These genetic findings implicated abnormal amyloid-β processing as a causal mechanism in AD,[14] leading to the formulation of the amyloid cascade hypothesis as a key mechanism. According to this hypothesis, the imbalance in amyloid-β production and/or clearance leads to gradual accumulation and aggregation of the peptide in the brain. This initiates a neurodegenerative cascade that involves amyloid deposition, inflammation, oxidative stress, and neuronal injury and loss.[14,15] Oligomeric and fibrillary forms of amyloid-β cause long-term potentiation impairment, synaptic dysfunction, and accelerate the formation of neurofibrillary tangles that eventually cause synaptic failure and neuronal death.[14,15]

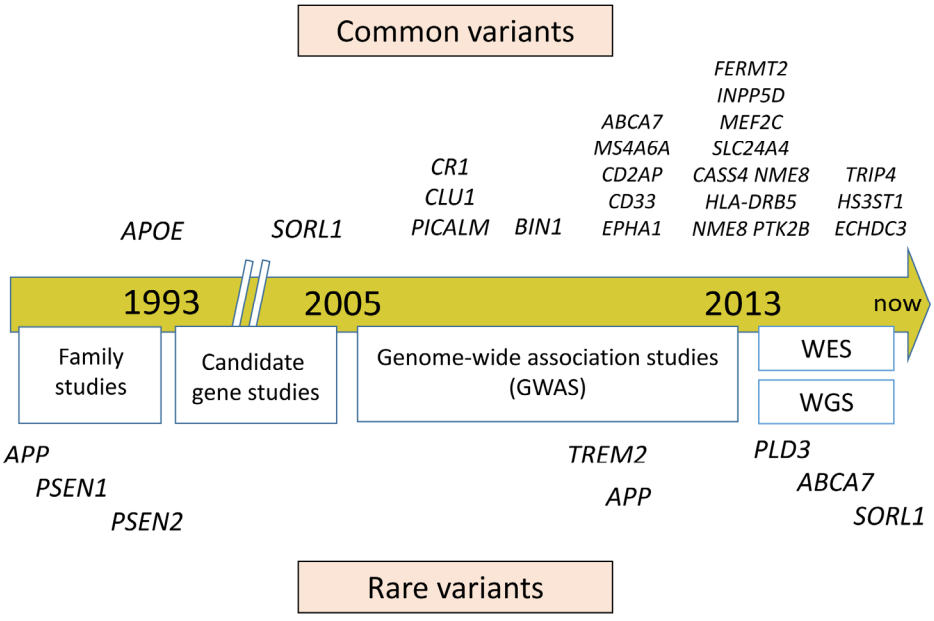


Figure 1: History of AD gene discovery

One of the spinoffs of the genetic discoveries are the development of amyloid-β 40 and 42 markers in cerebral spinal fluid (CSF) and imaging to visualize beta-amyloid plaques in neuronal tissue. These measures are now a part of the clinical diagnosis of AD at least in Europe.[16,17] While the mutations in the *APP, PSEN1* and *PSEN2* genes are known to cause early-onset AD, the first successful genetic discovery for late-onset AD was the ε4 allele of the apolipoprotein E (*APOE*) gene on chromosome 19 using linkage analysis in families with multiple AD cases.[18,19] *APOE* ε4 is the most important genetic variant for both early-onset and late-onset AD as it is found in roughly 25% of the general population and triples the risk for AD with each copy carried.[20,21] *APOE* ε4 homozygotes have a lifetime risk of AD of more than 50% by the age of 85 years, compared to less than 10% for non-carriers by this age.[22,23] Studies in the candidate gene era, which followed these initial successes of family studies, were unable to reliably identify additional genetic variants (**Figure 1**). New successes came with technological developments that led to relatively inexpensive genome-wide genotyping arrays and statistical imputation of common variants (frequency > 5%) based on reference genome panels.[24] These developments paved the way for genome-wide association studies (GWAS), which basically compare the frequencies of common genetic variants between unrelated cases and controls over the whole genome. Individual studies combined efforts to boost sample sizes and thus statistical power to find new genetic loci. In 2009, two of these consortia, the European Alzheimer's Disease Initiative (EADI) and the Genetic and Environmental Risk in Alzheimer's disease (GERAD) reported for the first time associations beyond *APOE* in the *CLU, CR1* and *PICALM* genes.[25,26] In 2010, the Cohorts for Heart and Aging in Genomic Epidemiology (CHARGE) consortium identified variants in the Bridge Integrator1 (*BIN1*) gene and the *EPHA1* gene.[27] This was the first AD GWAS to use genotype imputation. Imputation is a statistical technique that estimates missing genotypes in a population by comparing haplotypes of this population with those of a genotyped reference panel.[28] In 2011, a collaborative effort across these three consortia (the EADI, GERAD, and CHARGE) identified new variants associated with AD in *ABCA7*[29] and at the same time the Alzheimer's Disease Genetics Consortium (ADGC), reported genome-wide significant results in the *MS4A* gene cluster.[30] Collaboration of those two initiatives identified additional signals in or near *EPHA1*, *CD33*, and *CD2AP* by combining their data.[29,30] A cross consortium

collaborative effort combining the data of ADGC, EADI, CHARGE and GERAD was initiated called International Genomics of Alzheimer's Project (IGAP). The resulting GWAS that compared 25,580 cases and 48,466 controls led to the discovery of 11 new associations in loci near the genes *HLA-DRB5/DRB1, PTK2B, SORL1, SLC24A4RIN3, INPP5D, MEF2C, NME8, ZCWPW1, CELF1, FERMT2*, and *CASS4*. Attempts to replicate these novel signals confirmed the signal in *ZCWPW1* and let to the discovery of associated variants in additional loci.[31,32] In total, as a result of these collaborative efforts 23 common genetic variants were found to be associated with AD.[26,27,29,30,32-35] While these common genetic variants all had small effects (odds ratio (OR) ranging from 1.05-1.3)[26,27,29,30,32-35] (**Figure 2**), a major implication was the identification of eight novel biological pathways in AD in addition to the known amyloid cascade (**Figure 2**).[36] These include endocytosis, haemostasis, cholesterol transport, hematopoietic cell lineage, protein folding, clathrin complex, immune response and protein ubiquitination.[36] Despite these discoveries, a large proportion of AD cases remained unexplained as the identified variants were common and had a small impact on disease risk.[37]

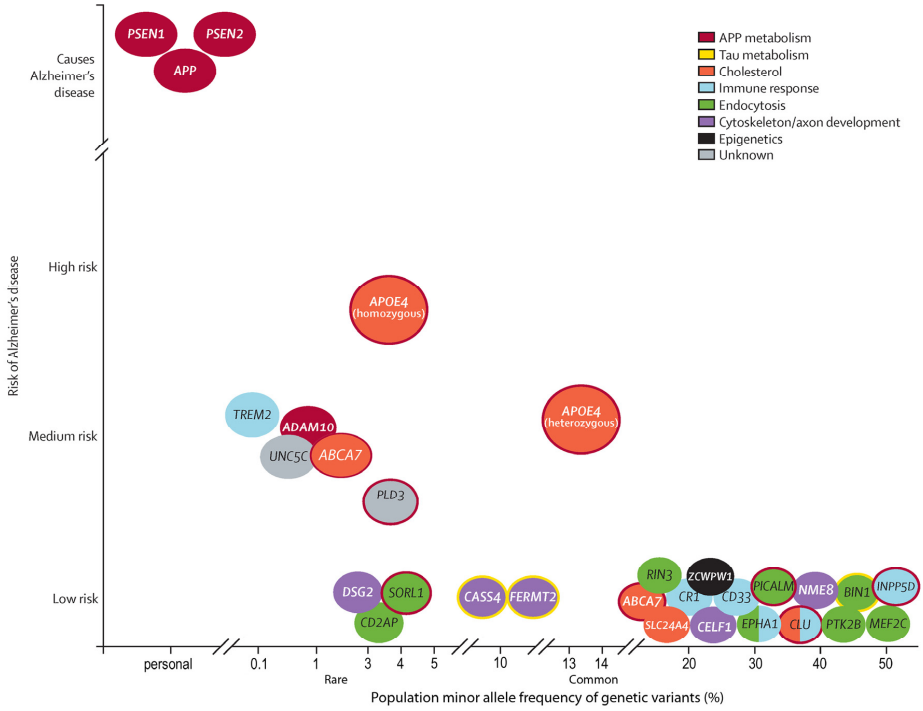


Figure 2: Population frequency, effect size and presumed pathway of effect of AD genetic variants. Modified figure from scheltens *et al.*[49]

The genome-wide genotyping array technology facilitating GWAS was followed by improvements in sequencing techniques, so called next-generation sequencing (NGS). This development made large scale screening of cases and controls for rare coding variants possible. With whole exome sequencing (WES) and whole genome sequencing (WGS) a rare coding variant in *TREM2* was identified.[38-40] The rare variant in *TREM2* has an effect comparable to *APOE* ε4 (**Figure 2**). Further, rare coding variation associated with AD with relatively large effects (OR>2) were also discovered in *SORL1* and *ABCA7,* in which previously common genetic variants were found (**Figure 2**).[41-45] Rare variants were also described in *PLD3* although the scientific debate on this gene is ongoing, as I describe in part in this thesis (**Chapter 3.2**). Despite these successes, the predictive ability of all the identified genetic variants (**Figure 2**) was found to be marginal over age and sex asking for larger studies or other approaches.[46-48]

Another contribution of NGS to the field of genetics was the development of exome-chip – a genotyping array consisting of ~230,000 rare exonic variants and a cheaper alternative to the more expensive NGS. This enabled genotyping of rare variants in large population-based studies. Significant discoveries were made for AD using the exome-chip, which I describe in **chapters 3.3 and 3.4** of this thesis. A limitation of the exome-chip however, is the limited number of genetic variants that can be studied. One way to expand research on rare genetic variants was to improve imputation methods and reference panels based on NGS of populations and patient studies. The HapMap project[50] and the 1000 Genomes Project[51-53] are the most used reference panels to date. Further population specific reference panels like genome of the Netherlands (GoNL)[52] and UK10K[53] are also available. The observation that larger reference panels provide better imputation of missing genetic variants led to the Haplotype Reference Consortium (HRC) reference panel, which combines all the aforementioned panels in addition to other sources of sequence data. The first version of the panel contains the information of over 32,000 individuals enabling in theory more reliable statistical inference of rarer genetic variation than was previously possible.[54]

Endophenotypes of AD in Magnetic resonance imaging (MRI) and blood

AD is known to have a long pre-clinical phase in which there are no apparent clinical signs. Sub-clinical signs start to appear up to twenty years before the onset of clinical symptoms.[55] These subclinical signs (endophenotypes) include amyloid-β depositions and structural brain changes (**Figure 3**)[55] and can be studied in apparently healthy individuals. As these endophenotypes are genetically determined[56,57] they can be powerful tools to study the pathophysiology of AD. A good example is amyloid-β, which is a clinical diagnostic marker of AD and in blood has been found to be a weak predictor of AD.[58] In CSF the genetic background of amyloid-β has been studied,[59] but the genetic background in of Amyloid-β in blood remains obscure.[60] Apart from amyloid-β there is rising interest in newer circulating metabolites in relation with AD. Previous studies have shown circulating metabolites in blood (e.g. lipoproteins, amino acids, fatty acids, and other small molecules) to be associated with cognitive function and conversion from normal cognition to dementia or AD.[61-67] However, these studies were relatively small and their findings have not been replicated,[65,68] emphasizing the need for studies in large well-characterized populations.[69,70]

In the long preclinical phase of AD, during which there are no symptoms of cognitive decline, structural brain changes such as cortical atrophy and localized atrophy of the hippocampus (**Figure 3**) can already be detected.[55,71] Brain imaging allows to address the shape of the brain in large healthy populations prior to the onset of the disease. One may speculate that the effects of genes that cause dementia are likely to affect the brain prior to disease onset. Conversely improvement of our knowledge of the genetics of brain development will provide new insights into the processes of the brain.[72-75] In my thesis I study on the genetics of brain magnetic resonance imaging (MRI) and calcifications in the internal carotid artery, a risk factor for dementia and stroke measured through computer tomography (CT).[76-78]

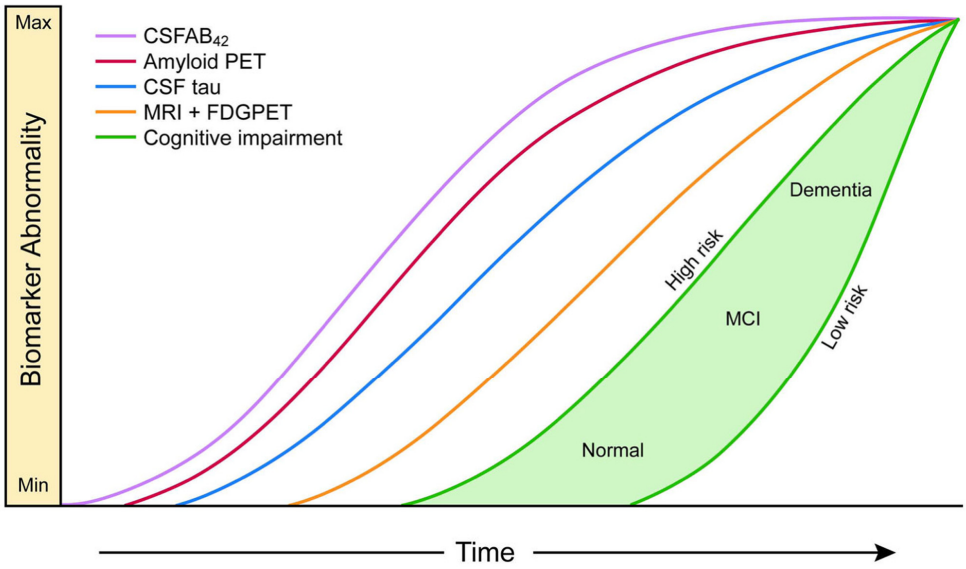


Figure 3: Markers of the AD pathological cascade over time. leading eventually to mild cognitive impairment (MCI) and clinical dementia. Reprinted with persmission from Jack et al.[55]

Aim of this thesis

The overall aim of this thesis is to identify molecular and biological mechanisms underlying AD and its endophenotypes and evaluate the scope of these findings in clinical and preventive studies.

Chapter 2 studies the previously discovered genetic risk factors of AD and determined how they can be translated to clinical use. In **chapter 2.1** I evaluate the combined effect of all common genetic variants associated with AD on AD and dementia risk. In **chapter 2.2,** I study rare genetic variation in the *SORL1* gene and propose a classification for rare variants in the *SORL1* gene that could be used in clinic. In **chapter 2.3,** I study the risk associated with a positive family history, which is the fastest and most used genetic test in clinic.

The next chapters describe association studies of rare genetic variants with AD. **Chapter 3.1** studies the utility of a new large reference panel, the Haplotype Reference Consortium (HRC) reference panel in rare variant association studies. In **chapter 3.2,** I

demonstrate conflicting evidence to a publication proposing *PLD3* as a gene that contains multiple rare risk variants for AD. In **chapter 3.3,** I describe an exome-chip study in the CHARGE consortium and in **chapter 3.4** the largest rare-variant exome chip study to date performed in the IGAP consortium. In this chapter I use the imputation panel for rare genetic variant discovery studied in **chapter 2.1**.

Chapter 4 studies the genetics of endophenotypes of AD determined by brain-imaging. I first search for common genetic factors that influence the lobar volumes of the brain (**chapter 4.1**). I then study the detailed information on grey matter derived from high resolution imaging of the brain, so called voxel based morphometry. Heritability maps, i.e. the genetics contribution, of grey matter voxels are described in **chapter 4.2**. In **chapter 4.3,** I study the effects on grey matter voxels of genetic variants known to associate with AD.

Chapter 5 explores circulating endophenotypes and risk factors of AD**.** In **chapter 5.1,** I performed a genome-wide association study of blood β-amyloid, an endophenotype in blood of AD. In the next chapters, I change my focus to the relation of hundreds of circulating metabolites with AD and dementia (**chapter 5.2**) and general cognitive function (**chapter 5.3).** The same metabolites I used in the last chapter (**chapter 5.4)** to study the association of metabolites with internal carotid artery calcifications, a risk factor for dementia.

In **chapter 6,** I summarize the main findings of this thesis and put them in the broader context of Alzheimer's disease research. I also give potential clinical implications and suggestions for further research.

References:

1. Alzheimer's A. 2015 Alzheimer's disease facts and figures. *Alzheimers Dement* 2015; **11**(3): 332-84.
2. Toledo JB, Arnold SE, Raible K, et al. Contribution of cerebrovascular disease in autopsy confirmed neurodegenerative disease cases in the National Alzheimer's Coordinating Centre. *Brain* 2013; **136**(Pt 9): 2697-706.
3. Alzheimer's A. 2016 Alzheimer's disease facts and figures. *Alzheimers Dement* 2016; **12**(4): 459-509.
4. Winblad B, Amouyel P, Andrieu S, et al. Defeating Alzheimer's disease and other dementias: a priority for European science and society. *Lancet Neurol* 2016; **15**(5): 455-532.

5. Campion D, Dumanchin C, Hannequin D, et al. Early-onset autosomal dominant Alzheimer disease: prevalence, genetic heterogeneity, and mutation spectrum. *Am J Hum Genet* 1999; **65**(3): 664-70.
6. Hyman BT, Phelps CH, Beach TG, et al. National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. *Alzheimers Dement* 2012; **8**(1): 1-13.
7. Montine TJ, Phelps CH, Beach TG, et al. National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease: a practical approach. *Acta Neuropathol* 2012; **123**(1): 1-11.
8. de Bruijn RFAG, Bos MJ, Portegies MLP, et al. The potential for prevention of dementia across two decades: the prospective, population-based Rotterdam Study. *BMC Med* 2015; **13**: 132.
9. Lipnicki DM, Sachdev PS, Crawford J, et al. Risk factors for late-life cognitive decline and variation with age and sex in the Sydney Memory and Ageing Study. *PLoS One* 2013; **8**(6): e65841.
10. Satizabal CL, Beiser AS, Chouraki V, Chene G, Dufouil C, Seshadri S. Incidence of Dementia over Three Decades in the Framingham Heart Study. *New Engl J Med* 2016; **374**(6): 523-32.
11. Schrijvers EM, Verhaaren BF, Koudstaal PJ, Hofman A, Ikram MA, Breteler MM. Is dementia incidence declining?: Trends in dementia incidence since 1990 in the Rotterdam Study. *Neurology* 2012; **78**(19): 1456-63.
12. Gatz M, Reynolds CA, Fratiglioni L, et al. Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry* 2006; **63**(2): 168-74.
13. Glenner GG, Wong CW. Alzheimer's disease: initial report of the purification and characterization of a novel cerebrovascular amyloid protein. 1984. *Biochem Biophys Res Commun* 2012; **425**(3): 534-9.
14. Hardy JA, Higgins GA. Alzheimer's disease: the amyloid cascade hypothesis. *Science* 1992; **256**(5054): 184-5.
15. Lemere CA, Masliah E. Can Alzheimer disease be prevented by amyloid-beta immunotherapy? *Nat Rev Neurol* 2010; **6**(2): 108-19.
16. McKhann GM, Knopman DS, Chertkow H, 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 & Dementia* 2011; **7**(3): 263-9.
17. Mckhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical-Diagnosis of Alzheimers-Disease - Report of the Nincds-Adrda Work Group under the Auspices of Department-of-Health-and-Human-Services Task-Force on Alzheimers-Disease. *Neurology* 1984; **34**(7): 939-44.
18. Saunders AM, Strittmatter WJ, Schmechel D, et al. Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology* 1993; **43**(8): 1467-72.
19. Pericak-Vance MA, Bebout JL, Gaskell PC, Jr., et al. Linkage studies in familial Alzheimer disease: evidence for chromosome 19 linkage. *Am J Hum Genet* 1991; **48**(6): 1034-50.
20. Farrer LA, Cupples LA, Haines JL, et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *JAMA* 1997; **278**(16): 1349-56.
21. Sleegers K, Van Duijn CM. Alzheimer's Disease: Genes, Pathogenesis and Risk Prediction. *Community Genet* 2001; **4**(4): 197-203.
22. Genin E, Hannequin D, Wallon D, et al. APOE and Alzheimer disease: a major gene with semi-dominant inheritance. *Mol Psychiatry* 2011; **16**(9): 903-7.
23. Seshadri S, Drachman DA, Lippa CF. Apolipoprotein E epsilon 4 allele and the lifetime risk of Alzheimer's disease. What physicians know, and what they should know. *Arch Neurol* 1995; **52**(11): 1074-9.
24. Bush WS, Moore JH. Chapter 11: Genome-wide association studies. *PLoS Comput Biol* 2012; **8**(12): e1002822.

25. Harold D, Abraham R, Hollingworth P, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* 2009; **41**(10): 1088-93.
26. Lambert JC, Heath S, Even G, et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet* 2009; **41**(10): 1094-9.
27. Seshadri S, Fitzpatrick AL, Ikram MA, et al. Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA* 2010; **303**(18): 1832-40.
28. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* 2012; **44**(8): 955-9.
29. Hollingworth P, Harold D, Sims R, et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet* 2011; **43**(5): 429-35.
30. Naj AC, Jun G, Beecham GW, 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.
31. Ruiz A, Heilmann S, Becker T, 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; **4**: e358.
32. Desikan RS, Schork AJ, Wang Y, et al. Polygenic Overlap Between C-Reactive Protein, Plasma Lipids, and Alzheimer Disease. *Circulation* 2015; **131**(23): 2061-9.
33. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 2013; **45**(12): 1452-8.
34. Jun G, Ibrahim-Verbaas CA, Vronskaya M, et al. A novel Alzheimer disease locus located near the gene encoding tau protein. *Mol Psychiatry* 2016; **21**(1): 108-17.
35. Miyashita A, Koike A, Jun G, et al. SORL1 is genetically associated with late-onset Alzheimer's disease in Japanese, Koreans and Caucasians. *PLoS One* 2013; **8**(4): e58618.
36. International Genomics of Alzheimer's Disease C. Convergent genetic and expression data implicate immunity in Alzheimer's disease. *Alzheimers Dement* 2015; **11**(6): 658-71.
37. Escott-Price V, Shoai M, Pither R, Williams J, Hardy J. Polygenic score prediction captures nearly all common genetic risk for Alzheimer's disease. *Neurobiol Aging* 2017; **49**: 214 e7- e11.
38. Guerreiro R, Wojtas A, Bras J, et al. TREM2 variants in Alzheimer's disease. *N Engl J Med* 2013; **368**(2): 117-27.
39. Jonsson T, Stefansson H, Steinberg S, et al. Variant of TREM2 associated with the risk of Alzheimer's disease. *N Engl J Med* 2013; **368**(2): 107-16.
40. Ruiz A, Dols-Icardo O, Bullido MJ, 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; **35**(2): 444 e1-4.
41. Nicolas G, Charbonnier C, Wallon D, et al. SORL1 rare variants: a major risk factor for familial early-onset Alzheimer's disease. *Mol Psychiatry* 2016; **21**(6): 831-6.
42. Pottier C, Hannequin D, Coutant S, et al. High frequency of potentially pathogenic SORL1 mutations in autosomal dominant early-onset Alzheimer disease. *Mol Psychiatry* 2012; **17**(9): 875-9.
43. Vardarajan BN, Zhang Y, Lee JH, et al. Coding mutations in SORL1 and Alzheimer disease. *Ann Neurol* 2015; **77**(2): 215-27.
44. Verheijen J, Van den Bossche T, van der Zee J, et al. A comprehensive study of the genetic impact of rare variants in SORL1 in European early-onset Alzheimer's disease. *Acta Neuropathol* 2016; **132**(2): 213-24.
45. Steinberg S, Stefansson H, Jonsson T, et al. Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. *Nat Genet* 2015; **47**(5): 445-7.
46. Escott-Price V, Sims R, Bannister C, et al. Common polygenic variation enhances risk prediction for Alzheimer's disease. *Brain* 2015; **138**(Pt 12): 3673-84.
47. Chouraki V, Reitz C, Maury F, et al. Evaluation of a Genetic Risk Score to Improve Risk Prediction for Alzheimer's Disease. *J Alzheimers Dis* 2016; **53**(3): 921-32.

48. Sleegers K, Bettens K, De Roeck A, et al. A 22-single nucleotide polymorphism Alzheimer's disease risk score correlates with family history, onset age, and cerebrospinal fluid Abeta42. *Alzheimers Dement* 2015; **11**(12): 1452-60.
49. Scheltens P, Blennow K, Breteler MM, et al. Alzheimer's disease. *Lancet* 2016; **388**(10043): 505-17.
50. Altshuler DM, Gibbs RA, Peltonen L, et al. Integrating common and rare genetic variation in diverse human populations. *Nature* 2010; **467**(7311): 52-8.
51. Genomes Project C, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature* 2015; **526**(7571): 68-74.
52. Boomsma DI, Wijmenga C, Slagboom EP, et al. The Genome of the Netherlands: design, and project goals. *Eur J Hum Genet* 2014; **22**(2): 221-7.
53. Consortium UK, Walter K, Min JL, et al. The UK10K project identifies rare variants in health and disease. *Nature* 2015; **526**(7571): 82-90.
54. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016.
55. Jack CR, Jr., Knopman DS, Jagust WJ, et al. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *Lancet Neurol* 2013; **12**(2): 207-16.
56. Rice JP, Saccone NL, Rasmussen E. Definition of the phenotype. *Adv Genet* 2001; **42**: 69-76.
57. Flint J, Munafo MR. The endophenotype concept in psychiatric genetics. *Psychol Med* 2007; **37**(2): 163-80.
58. Chouraki V, Beiser A, Younkin L, et al. Plasma amyloid-beta and risk of Alzheimer's disease in the Framingham Heart Study. *Alzheimers Dement* 2015; **11**(3): 249-57 e1.
59. Deming Y, Li Z, Kapoor M, et al. Genome-wide association study identifies four novel loci associated with Alzheimer's endophenotypes and disease modifiers. *Acta Neuropathol* 2017; **133**(5): 839-56.
60. Chouraki V, De Bruijn RF, Chapuis J, et al. A genome-wide association meta-analysis of plasma Abeta peptides concentrations in the elderly. *Mol Psychiatry* 2014; **19**(12): 1326-35.
61. Hye A, Riddoch-Contreras J, Baird AL, et al. Plasma proteins predict conversion to dementia from prodromal disease. *Alzheimers Dement* 2014; **10**(6): 799-807 e2.
62. Muenchhoff J, Poljak A, Song F, et al. Plasma protein profiling of mild cognitive impairment and Alzheimer's disease across two independent cohorts. *J Alzheimers Dis* 2015; **43**(4): 1355-73.
63. Song F, Poljak A, Crawford J, et al. Plasma apolipoprotein levels are associated with cognitive status and decline in a community cohort of older individuals. *PLoS One* 2012; **7**(6): e34078.
64. Proitsi P, Kim M, Whiley L, et al. Association of blood lipids with Alzheimer's disease: A comprehensive lipidomics analysis. *Alzheimers Dement* 2016.
65. Li D, Misialek JR, Boerwinkle E, et al. Plasma phospholipids and prevalence of mild cognitive impairment and/or dementia in the ARIC Neurocognitive Study (ARIC-NCS). *Alzheimers Dement (Amst)* 2016; **3**: 73-82.
66. Mapstone M, Cheema AK, Fiandaca MS, et al. Plasma phospholipids identify antecedent memory impairment in older adults. *Nat Med* 2014; **20**(4): 415-8.
67. Fiandaca MS, Zhong X, Cheema AK, et al. Plasma 24-metabolite Panel Predicts Preclinical Transition to Clinical Stages of Alzheimer's Disease. *Front Neurol* 2015; **6**: 237.
68. Casanova R, Varma S, Simpson B, et al. Blood metabolite markers of preclinical Alzheimer's disease in two longitudinally followed cohorts of older individuals. *Alzheimers Dement* 2016; **12**(7): 815-22.
69. Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. *Nature* 2014; **505**(7485): 612-3.
70. O'Bryant SE, Gupta V, Henriksen K, et al. Guidelines for the standardization of preanalytic variables for blood-based biomarker studies in Alzheimer's disease research. *Alzheimers Dement* 2015; **11**(5): 549-60.
71. Thompson PM, Cannon TD, Narr KL, et al. Genetic influences on brain structure. *Nat Neurosci* 2001; **4**(12): 1253-8.

72. Adams HH, Hibar DP, Chouraki V, et al. Novel genetic loci underlying human intracranial volume identified through genome-wide association. *Nat Neurosci* 2016.
73. Bis JC, DeCarli C, Smith AV, et al. Common variants at 12q14 and 12q24 are associated with hippocampal volume. *Nat Genet* 2012; **44**(5): 545-51.
74. Hibar DP, Stein JL, Renteria ME, et al. Common genetic variants influence human subcortical brain structures. *Nature* 2015; **520**(7546): 224-9.
75. Stein JL, Medland SE, Vasquez AA, et al. Identification of common variants associated with human hippocampal and intracranial volumes. *Nat Genet* 2012; **44**(5): 552-61.
76. Bos D, Portegies MLP, van der Lugt A, et al. Intracranial Carotid Artery Atherosclerosis and the Risk of Stroke in Whites The Rotterdam Study. *Jama Neurology* 2014; **71**(4): 405-11.
77. Bos D, van der Rijk MJM, Geeraedts TEA, et al. Intracranial Carotid Artery Atherosclerosis Prevalence and Risk Factors in the General Population. *Stroke* 2012; **43**(7): 1878-84.
78. Bos D, Vernooij MW, de Bruijn RFAG, et al. Atherosclerotic calcification is related to a higher risk of dementia and cognitive decline. *Alzheimers & Dementia* 2015; **11**(6): 639-47.

Chapter 2

Translation of genetic risk factors of AD to clinic

Chapter 2.1

The effect of common genetic variants on the onset of Alzheimer's disease and dementia in carriers of the *APOE* ε4 genotype

Sven J. van der Lee, Frank J. Wolters, M. Kamran Ikram, Albert Hofman, M. Arfan Ikram, Najaf Amin, Cornelia M. van Duijn.

This chapter is submitted

Abstract

Background: Alzheimer's disease (AD) is one of the most heritable (60-80%) diseases in the elderly and the most common type of dementia. In addition to the major genetic determinant of AD, the apolipoprotein E (*APOE*) gene, twenty-three mostly common genetic variants have been associated with AD. In this population-based study, we determined the effects of these twenty-three variants and *APOE* on risk and onset age of AD and dementia.

Methods: We studied incident dementia in 12,255 cognitively healthy participants (58.5% female) of the community-based Rotterdam Study. During median follow-up of 11.0 years (inter quartile range 4.9-15.9 and 133,123 person years), 1,609 participants developed dementia, of whom 1,262 were classified as AD (78.4%), and 4,590 persons died of causes other than dementia. Cumulative incidence curves up to age 100 years were calculated for AD and dementia, taking into account mortality as a competing event. These risk curves were stratified by *APOE* genotypes and tertiles of a weighted genetic risk score (GRS) of twenty-three AD-associated genetic variants.

Results: Both *APOE* and the GRS significantly modified the risks of AD and dementia. There was evidence for significant interaction between *APOE* genotypes and the GRS in the association with AD (p=0.02) and dementia (p=0.03), such that the risk of *APOE**44 genotype carriers was modified most by the GRS. In *APOE**44 genotype carriers the difference in risk of AD between the high risk tertile and the low risk tertile by age 85 years was 24.6% (34.3% for dementia), translating in a 7 to 10 year difference in age at onset. Comparing the risk extremes, which were *APOE**22/23 carriers in the low risk tertile of the GRS versus *APOE**44 carriers in the high risk tertile of the GRS, the difference in risk by age 85 years was 57.8% (4.4% vs. 62.2%, $p=4.4\times10^{-13}$) for AD, and 69.2% (7.5% vs 76.7%, $p=3.0\times10^{-23}$) for all-cause dementia. These risk differences translate in an 18 to 29 year age at onset difference of AD and a 19 to 22 year age at onset difference of dementia.

Conclusion: In this population-based study common variants with small individual effects jointly significantly modify the risk and onset age of AD and dementia, particularly in *APOE**4 carriers. These findings highlight the potential of common variants in determining AD risk.

Introduction

Alzheimer's disease (AD) is the most common form of dementia. AD is a multi-factorial disease with a considerable genetic component (60-80%)[1] and the apolipoprotein E (*APOE*) gene is the strongest common genetic risk factor for AD.[2] The gene has three common alleles, the protective allele *APOE**2, the neutral or reference allele *APOE**3, and the risk allele *APOE**4.[3] The magnitude of the *APOE**4 effect is best described by looking at the high absolute risk carriers have to get AD and dementia. It has been estimated from case-control studies that *APOE**4 homozygote carriers have a risk of AD of over 50% by the age of 85 years, compared to less than 10% for non-carriers by this age.[4,5] Because of this high risk, there is increasing interest to include *APOE**4 homozygote carriers in pre-symptomatic AD treatment or prevention trials to reduce the necessary duration of these costly studies.[6,7] However, total risk of AD is not the only important factor as the clinical onset of AD and dementia has a wide variability,[8] ranging from mid-life to the ninth decade even within the *APOE**44 carriers.[4]

In addition to APOE, twenty-three other genetic variants have been identified in the past decade that significantly modify risk of AD. Recently, it has been shown that combining the effects of these twenty-three variants[9-18] results in a polygenic risk score that is not only associated with risk of AD,[19,20] neuropathological hallmarks of AD[21] and conversion of MCI to AD[22-24] but also quantifies the age at onset in APOE*4 carries and non-carriers.[21] For research applications and patients it matters a lot at which age one develops AD, so especially the finding that age of onset is in part determined by these common variants is crucial. However, as discussed in the largest conducted study to date,[21] in order to be applicable to practice, these findings await validation in large community-based cohort studies. This is of particular importance as estimates of absolute risk of AD may be overestimated due to competing risk of death,[25] which cannot be fully accounted for in case-control studies.[4,21]

In the present study of a large population-based community cohort followed for up to 25 years for incident dementia, we determine the aggregated effect of common variants

separate and in conjunction with *APOE* on the risk and age at onset of AD and all-cause dementia.

Methods

Study population

This study included 12,255 non-demented participants from the Rotterdam Study (RS), which is a prospective population-based cohort study.[26] In 1990, residents aged 55 and older residing in Ommoord, a district of Rotterdam, the Netherlands, were invited to participate in the study. Of the 10,215 invited inhabitants, 7,983 agreed to participate in the baseline examinations. In 2000, 3,011 participants (out of 4,472 invitees) who had become 55 years of age or moved into the study district since the start of the study were added to the cohort. In 2006 a further extension of the cohort was initiated in which 3,932 subjects, out of 6,057 invited, aged ≥45 years living in the Ommoord district were included.[26] Follow-up examinations take place every 3 to 4 years.[26] We excluded participants who did not contribute follow-up time over the age of 60 years. The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC and by the Ministry of Health, Welfare and Sport of the Netherlands, implementing the Population Studies Act: Rotterdam Study (Erasmus Rotterdam Gezondheid Onderzoek). All participants provided written informed consent to participate in the study and to obtain information from their treating physicians.

Diagnosis of dementia

Participants were screened for dementia at baseline and subsequent center visits with the Mini-Mental State Examination and the Geriatric Mental Schedule organic level.[24] Those with a Mini-Mental State Examination score <26 or Geriatric Mental Schedule score >0 underwent further investigation and informant interview, including the Cambridge Examination for Mental Disorders of the Elderly. In addition, the entire cohort was continuously under surveillance for dementia through electronic linkage of the study database with medical records from general practitioners and the regional institute for outpatient mental health care.[27] Available information on cognitive testing and clinical neuroimaging was used when required for diagnosis of dementia subtype. A consensus panel led by a consultant neurologist established the final diagnosis

according to standard criteria for dementia (DSM-III-R) and Alzheimer's disease (NINCDS–ADRDA).[28] Follow-up for dementia was done until the first of January 2014 for the first Rotterdam cohort, until the first of January 2015 for the first extension and the first of January 2013 for the second extension of the cohorts. Follow-up for dementia was complete for 92% of potential person-years.

Genotyping and imputations

DNA was extracted from blood samples drawn by venepuncture at baseline visit using standard methods. The Rotterdam Study cohorts were genotyped using commercially genotyping arrays and genotyping quality control was done per cohort.[29] Preparation for imputation was done using scripts provided online (HRC or 1000G Imputation preparation and checking: http://www.well.ox.ac.uk/~wrayner/tools/; v4.2.1). Imputation to the Haplotype Reference Consortium (HRC)[30] was facilitated by the "Michigan Imputation server". The server used SHAPEIT2 (v2.r790) to phase the data and imputation to the Haplotype Reference Consortium (HRC) reference panel (v1.0) was performed with Minimac 3. Imputed genotypes were returned by the service. *APOE* genotype was determined using polymerase chain reaction on coded DNA samples in the baseline cohort[28] and with a bi-allelic TaqMan assay (rs7412 and rs429358) in the two extensions of the Rotterdam Study. *APOE* genotype was imputed for 2.8% of individuals in baseline cohort, 0.5% in the first extension of the cohort and 4.6% in the second extension of the cohort using the '*best guess*' imputed genotypes (i.e. rounded dosages) of rs7412 (*APOE**2 determining variant) and rs429358 (*APOE**4 determining variant) as the *APOE* genotype. Using the data of those genotyped, imputation of the *APOE**2 and *APOE**4 allele and direct genotyping were 98.9% and 98.2% concordant respectively. Combining genotyped and imputed data, the *APOE* genotypes were available for 92.8% of all participants.

Statistical analysis

All analysis were performed in R (version 3.2.3).[31] Baseline characteristics were compared across *APOE* genotypes with the genotype *APOE**33 as reference genotype and in the tertiles of the GRS with the low risk genotype as reference group using t-tests (continuous measures) and chi-squared tests (categorical measures).

Alzheimer's diseases genetic risk score (GRS)

We included 23 genetic variants that showed genome-wide significant evidence of association with AD[11,14-16] to calculate a weighted GRS using reported effect estimates as weights. If multiple studies reported the effects of a variant, the effect estimate from the largest study (with respect to sample size) was used. The studies discovering the associations and reported the weights that we used are provided in **Supplementary Table 1**. Because the number of participants that were diagnosed with other types of dementia, for which were genetic evidence is available, was small; dementia with Lewy bodies (N=14), dementia in patients with Parkinson's disease (N=51), frontotemporal dementia (N=6), genetic variants associated with other causes of dementia were not considered. The genetic risk score was calculated as the sum of the products of SNP dosages of the 23 genetic variants (excluding *APOE*) and their respective weights. All 23 variants selected for the calculation of the genetic risk score were well imputed (median imputation score (R^2)> 0.993) (**Supplementary Table 1**). We split the population into a high, middle and low risk based on tertiles of the GRS. The boundaries of the tertiles were determined by those entering the study before the age 60 years, as survival bias is anticipated at old age.

Assessment of cumulative incidence

Cumulative incidence, or risk, of AD and all cause dementia was calculated up to 100 years based on all incident cases occurring in the follow-up (133,123 person years) of the 12,255 non-demented individuals with genotype data available. Participants were censored at the date of dementia diagnosis, death, lost to follow-up, or the administrative censoring date, whichever came first. We calculated cumulative incidence of both AD and all-cause dementia (AD plus non-AD dementia) using the 'etmCIF' function (Cumulative Incidence Function) from package 'etm'.[32,33] When estimating risk of AD we accounted for non-AD dementia and mortality as competing events. When estimating risk of dementia we accounted for mortality as competing event. In short, the function estimates overall survival irrespective of the causes by a modification of the Kaplan–Meier estimate,[34] adapted for left truncation[35] and calculates age and cause-specific risk estimates and corresponding 95% confidence intervals. Risks curves for AD and dementia in carriers of the *APOE**23 and *APOE**22

genotypes showed similar patterns (**Supplementary Figure 1**), as did the risks of *APOE**24 and *APOE**34 genotype carriers (**Supplementary Figure 2**). These genotypes were therefore pooled in analyses, as *APOE**22/23 and *APOE**24/34. Analyses were stratified by 1) *APOE* genotypes, 2) tertiles of the GRS and 3) both *APOE* genotypes and tertiles of the GRS. We calculated the differences between the risk estimates by age 85 years as previously described.[33] Interaction (multiplicative) between *APOE* genotypes (*APOE**22/23, *APOE**33, *APOE**24/34, *APOE**44, coded as 1,2,3,4 respectively) and the GRS as well as the single variants was tested using Cox proportional hazards, adjusting for main genetic effects, sex and age at inclusion.

Results

Baseline characteristics

Characteristics at study entry of in total 12,255 non-demented individuals (58.5% female) are presented in **Table 1**. During follow-up (mean 10.9 ± 6.6 years, inter quartile range 4.9-15.9), 1,609 participants developed dementia (all causes) of which 1,262 were classified as AD (78.4% of all dementia cases) and 4,590 persons died of causes other than dementia. Overall, cumulative incidence, or the lifetime risk, by the age of 100 years of AD was 25.0% (95% CI 23.8-26.3) and of dementia 31.4% (95% CI 30.1-32.8) (**Supplementary Figure 3**).

Effect of *APOE* genotypes and common variants on the risk of AD and dementia

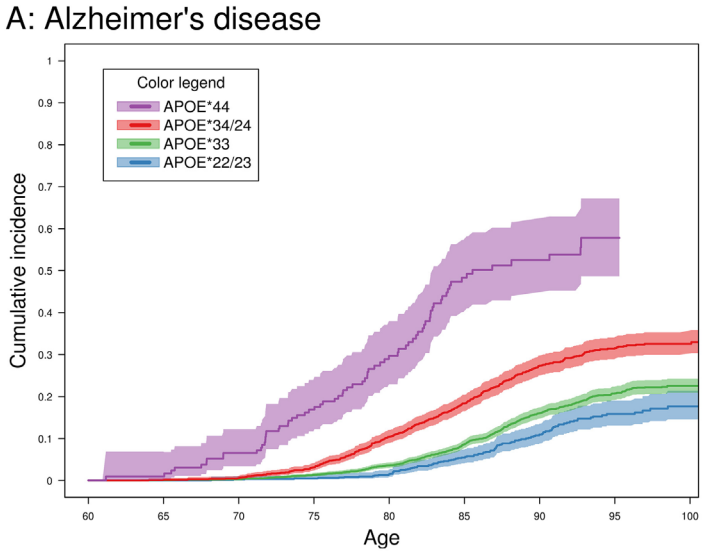
APOE genotypes had a strong effect on risk of AD (**Figure 1A**). By age 85 years, the risk for AD was 48.3% (95% CI 40.1-57.3) for those homozygous for *APOE**4 and 18.4% (95% CI 16.5-20.4) for those heterozygous for *APOE**4. Compared to the *APOE**4 genotype carriers, non-carriers had a lower risk, the risks were 8.6% (95% CI 7.7-9.6) for *APOE**33 carriers and 5.5% (95% CI 4.1-7.4) for *APOE**22/23 carriers. Estimates of the risk curves for all-cause dementia by *APOE* genotypes were larger (**Figure 1B**) but patterns were similar. Stratified by tertiles of the GRS, the risk of AD by age 85 was 15.7% (95% CI 14.1-17.6) for the high risk tertile, 11.3% (95% CI 9.9-12.9) for the middle tertile and 8.2% (95% CI 7.0-9.6) for the low risk tertile (**Figure 2A**, **Supplementary Table 2**).

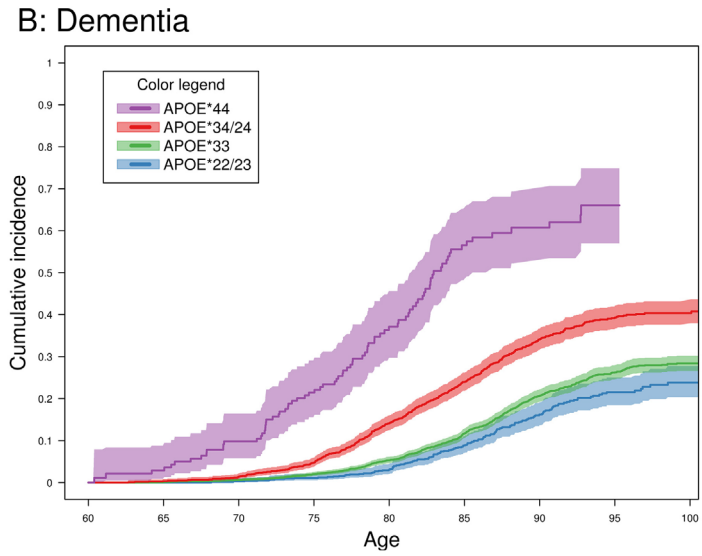
Table 1: Baseline characteristics of the study population by tertiles of the GRS and *APOE* genotype

	All n=12255	low risk tertile n=3402	middle risk tertile n=3292	high risk tertile n=3317	APOE*44 n=261	APOE*34 n=2608	APOE*24 n=312	APOE*33 n=6662	APOE*23 n=1453	APOE*22 n=79
Incident AD cases (%)	1262 (10.3)	287 (8.4)	341 (10.4)	429 (12.9)	72 (27.6)	385 (14.8)	40 (12.8)	585 (8.8)	97 (6.7)	6 (7.6)
Dementia cases other than AD (%)	347 (10.3)	88 (2.6)	90 (2.7)	116 (3.5)	11 (4.2)	95 (3.6)	13 (4.2)	162 (2.4)	38 (2.6)	3 (3.8)
Female sex (%)	7164 (58.5)	1979 (58.2)	1922 (58.4)	1900 (57.3)	139 (53.3)	1518 (58.2)	171 (54.8)	3820 (57.3)	875 (60.2) *	46 (58.2)
Age at entry (years)	67.48 (8.4)	67.2 (8.1)	67 (8)	67 (7.9)	64.8 (6)	# 66.6 (7.6)	# 67.3 (8)	67.3 (8.2)	67.2 (8.2)	69.2 (8.4) *
Follow-up time (years)	10.86 (6.6)	11.2 (6.6)	11.2 (6.6)	11.2 (6.5)	10 (6.2)	* 10.8 (6.5)	* 10.6 (6.5)	11.2 (6.6)	11.8 (6.6)	* 11.7 (6.2)
Smoking (%)						*	*			
Never smoker	4150 (33.9)	1157 (34)	1094 (33.2)	1121 (33.8)	70 (26.8)	830 (31.8)	96 (30.8)	2286 (34.3)	530 (36.5)	28 (35.4)
Former	5250 (42.8)	1470 (43.2)	1453 (44.1)	1423 (42.9)	126 (48.3)	1199 (46)	138 (44.2)	2850 (42.8)	592 (40.7)	32 (40.5)
Smoker at baseline	2479 (20.2)	684 (20.1)	665 (20.2)	696 (21)	60 (23)	512 (19.6)	69 (22.1)	1362 (20.4)	291 (20)	14 (17.7)
Average pack years	16.45 (22.5)	16.7 (23.1)	16.6 (22.3)	16.3 (22.2)	17.1 (22.9)	16.9 (22.4)	17.4 (22.9)	16.3 (22.3)	15.8 (23.1)	13.7 (21.3)
Educational level (%)										
Primary education	2210 (18)	606 (17.8)	589 (17.9)	566 (17.1)	37 (14.2)	471 (18.1)	44 (14.1)	1173 (17.6)	271 (18.7)	15 (19)
Further education	8242 (67.3)	2319 (68.2)	2244 (68.2)	2291 (69.1)	179 (68.6)	1738 (66.6)	212 (67.9)	4562 (68.5)	991 (68.2)	54 (68.4)
Higher education	1582 (12.9)	437 (12.8)	418 (12.7)	420 (12.7)	43 (16.5)	362 (13.9)	51 (16.3)	845 (12.7)	177 (12.2)	10 (12.7)
Diabetes mellitus (%)	1196 (9.8)	336 (9.9)	304 (9.2)	334 (10.1)	25 (9.6)	248 (9.5)	31 (9.9)	669 (10)	134 (9.2)	9 (11.4)
Hypertension (%)	6711 (54.8)	1886 (55.4)	1838 (55.8)	1872 (56.4)	150 (57.5)	1425 (54.6)	178 (57.1)	3737 (56.1)	843 (58)	54 (68.4) *
Systolic blood pressure (mm Hg)	139.5 (21.8)	139.4 (22)	139.1 (21.3)	139.1 (21.8)	138.8 (22.5)	138 (21.2)	# 140.1 (22.5)	139.8 (21.8)	140 (21.8)	145.2 (24.9) *
Diastolic blood pressure (mm Hg)	76.79 (11.9)	76.7 (12)	76.6 (11.8)	76.7 (11.8)	77.5 (13.1)	76.1 (11.5)	* 76.6 (12.1)	76.9 (11.9)	76.8 (12)	77.8 (12.4)
Body mass Index (kg/m2)	26.84 (4)	26.9 (4)	26.7 (3.9)	26.9 (4.1)	26.6 (3.5)	26.6 (3.9)	* 26.8 (3.9)	26.8 (4)	27.1 (4.1)	* 27 (4.4)
Serum cholesterol (mmol/L)	6.24 (1.2)	6.3 (1.2)	6.2 (1.2)	6.2 (1.2)	6.4 (1.2)	* 6.4 (1.2)	# 6.1 (1.1)	6.2 (1.2)	5.9 (1.3)	6.1 (1.8)
Serum high density lipoprotein (mmol/L)	1.37 (0.4)	1.4 (0.4)	1.4 (0.4)	1.4 (0.4)	1.3 (0.4)	* 1.3 (0.4)	# 1.4 (0.4)	1.4 (0.4)	1.4 (0.4)	# 1.4 (0.4)

* p<0.05 # p<0.001

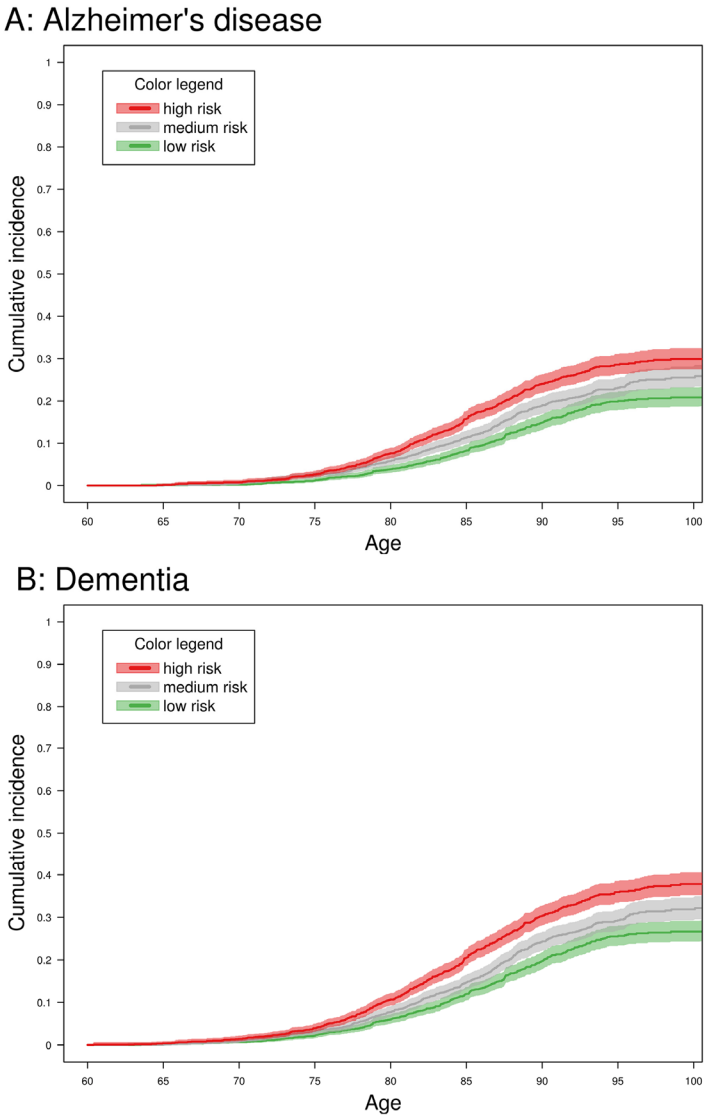
APOE = apolipoprotein E genotypes. AD= Alzheimer's disease. Measurement details of all non-genetic measures are described in detail elsewhere.[35] Reported descriptive measures are measured at entry to the study. Further education= lower/intermediate general education or lower vocational education or intermediate vocational education or higher general education, higher education= higher vocational education or university. Descriptive measures presented as mean (standard deviation) for continuous variables, and percentages (N) for nominal and ordinal variables. APOE genotypes are compared with respect to the APOE*33 genotype. Characteristics of the middle and high GRS tertile did not significantly differ compared to the first tertile.





APOE*44	82	136	128	108	64	23	10	2	0
APOE*34/24	875	1272	1370	1326	987	580	220	49	5
APOE*33	1866	2874	3170	3035	2458	1500	648	178	30
APOE*22/23	437	678	743	730	573	387	183	58	9

Figure 1: Risk curves of Alzheimer's disease (A) and dementia (B) by *APOE* genotypes. The risk curves show the cumulative incidence of Alzheimer's disease (A) and dementia (B). The shaded areas show the upper and lower 95% confidence limits of the corresponding cumulative incidence curve.The number of individuals at risk by age is shown under the graph.



high risk	954	1454	1624	1541	1182	695	286	69	10
medium risk	954	1419	1545	1522	1195	749	320	88	13
low risk	954	1468	1584	1545	1260	780	342	96	16

Figure 2: Risk curves of Alzheimer's disease (A) and dementia (B) by tertiles of the GRS. The risk curves show the cumulative incidence per 100 individuals of Alzheimer's disease (A) and dementia (B). The shaded areas show the upper and lower 95% confidence limits of the corresponding cumulative incidence curve. The number of individuals at risk by age is shown under the graph.

Table 2: Cumulative incidence of AD, dementia and death by other causes.

Age	AD Risk, 95%CI	Dementia Risk, 95%CI	Death by other causes Risk, 95%CI	Alive and not demented (%)	n
60	0.0, 0.0-0.0	0.0, 0.0-0.0	0.0, 0.0-0.0	100	3428
65	0.1, 0.0-0.2	0.2, 0.1-0.4	2.5, 2.1-3.0	97.3	5220
70	0.5, 0.4-0.8	1.0, 0.8-1.3	6.8, 6.1-7.5	92.2	5670
75	2.1, 1.7-2.4	3.1, 2.6-3.5	13.4, 12.5-14.3	83.5	5423
80	5.7, 5.1-6.3	8.1, 7.5-8.8	22.7, 21.6-23.7	69.2	4254
85	11.6, 10.8-12.4	15.6, 14.7-16.5	35.2, 34.0-36.4	49.2	2630
90	19.0, 18.0-20.0	24.4, 23.3-25.6	48.6, 47.3-50.0	27.0	1161
95	23.4, 22.2-24.6	29.6, 28.3-30.9	59.3, 57.9-60.7	11.1	316
100	25.0, 23.8-26.3	31.4, 30.1-32.8	65.5, 64.0-66.9	3.1	48

APOE = apolipoprotein E genotypes. AD= Alzheimer's disease. n = individuals alive and non-demented. Cumulative incidence (per 100 individuals or percentage) of AD takes into account competing non-AD dementia and mortality. Cumulative incidence (per 100 individuals or percentage) of dementia takes into account competing mortality. The percentage of the population alive and not demented is shown as 100% minus dementia risk and risk of death by other causes than dementia.

The 7.5% risk difference by age 85 years between the high and low tertiles was statistically significant (p=5.6×10^{-12}). Observed differences were similar for all-cause dementia (**Figure 2B**, **Supplementary Table 3**).

Effect of common genetic variants on risk by *APOE* genotypes

In **Figure 3** the risk curves of AD and dementia stratified by both *APOE* and GRS risk groups are shown. The corresponding risk estimates and confidence intervals by 5 year increments in age are summarized in **Supplementary Tables 2 and 3**. Carriers of *APOE**22/23 in the low risk tertile had the lowest risk by age 85 years, 4.4% (95%CI, 2.4-8.1) for AD and 7.5% (95%CI 4.7-11.7) for dementia. By age 85 years there was 57.8% risk difference (p=4.4×10^{-13}) for AD and a 69.2% risk difference (p=3.0×10^{-23}) for dementia between genetic risk extremes, i.e. carriers of *APOE**44 in the high GRS risk tertile and *APOE**22/23 carriers in the low GRS tertile. Within the *APOE* genotype categories, the risks of both AD and all-cause dementia were higher in the high risk tertile of the GRS compared to the low risk tertile (**Figure 3**, **Supplementary Tables 2 and 3**).

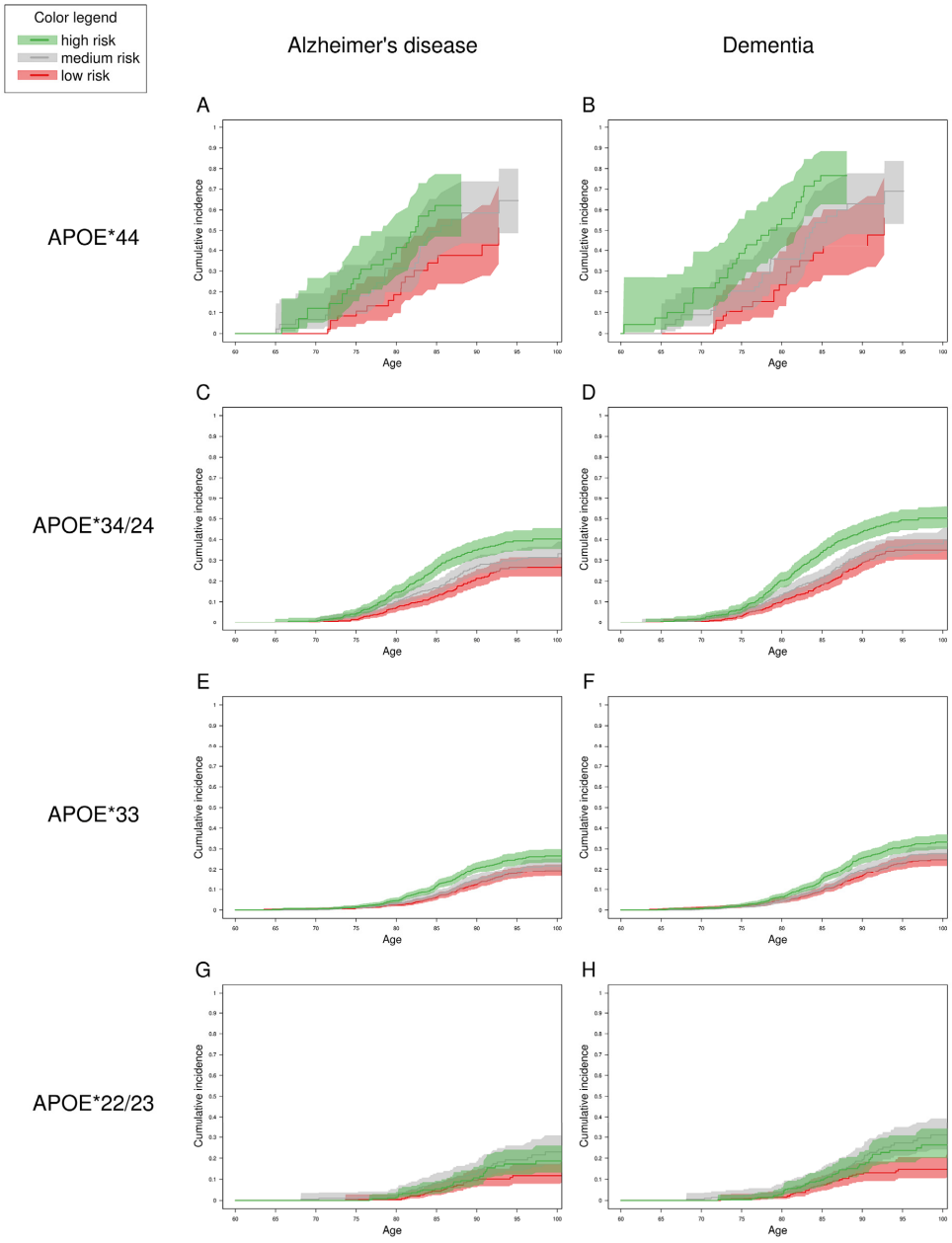


Figure 3: Risk curves of Alzheimer's disease (A) and dementia (B) by *APOE* genotypes and tertiles of the GRS. The risk curves show the cumulative incidence of Alzheimer's disease (A,C,E,G) and dementia (B,D,F,H). The shaded areas show the upper and lower 95% confidence limits of the corresponding cumulative incidence curve.

Figure 3, further suggest the effect of the joint risk score is largest in *APOE**4 carriers. When tested there was significant evidence for interaction between *APOE* genotypes and the GRS for AD (p=0.02) and dementia (p=0.03), that was not driven by single variants (**Supplementary Table 4**). By the age of 85 years, the risk of AD for *APOE**44 carriers in the high GRS risk tertile was 62.2% (95% CI 47.1-77.4) compared to 37.6% (95% CI 23.9-55.7) in the low GRS-tertile, that is a risk difference of 24.6% (p=1.6×10^{-2}). The risk difference for AD by this age for *APOE**34/24 carriers was 13.5% (p=1.3×10^{-7}), 5.4% (p=1.2×10^{-5}) for *APOE**33 carriers, and 1.1% (p=0.30) for *APOE**22/23 carriers. A similar trend was seen for dementia: by age 85, the risk difference between the high risk tertile and the low risk tertile was 34.3% (p=7.3×10^{-4}) for *APOE**44 carriers, 15.6% (p=4.0×10^{-8}) for *APOE**34/24 carriers, 5.6% (p=7.3×10^{-5}) for *APOE**33 carriers, and 2.2% (p=0.19) for *APOE**22/23 carriers.

Figure 4 and 5 illustrate the risk by age, *APOE* and GRS categories for AD and dementia respectively. The colours from green to red show increasing risk in six categories that are marked by 5 lines denoting when the risk increased to over 5%, 12.5%, 25%, 40% and 60%. These figures show that *APOE**44 carriers in the high risk tertile attained 5% risk of AD by age 67 years (64 years for dementia) and 12.5% by age 71 years (67 years for dementia). For comparison, *APOE**22/23 carriers in the low risk tertile attained 5% risk of AD by the age of 85 years (83 years for dementia) and 12.5% by age 100 years (89 years for dementia). This translates into a difference in age at onset in individuals with highest compared to lowest genetic risk of 18 to 29 years for AD and 19 to 22 years for dementia. Also, age at onset differences within *APOE* genotypes can be observed from **Figure 4 and 5**. In *APOE**44 carriers a 40% risk of AD is attained 10 years earlier by those in the high risk tertile (79 years) compared to those in the low risk tertile (89 years). For *APOE**24/34 carriers the age at which 25% risk of AD is attained is 8 years earlier in those in the high risk tertile (92 years) compared to those in the low risk tertile (84 years). For all-cause dementia, the difference in age at which 25% risk is attained is 8 years comparing *APOE**44 carriers in the high (72 years) and low tertile (80 years) and 7 years in *APOE**24/34 carriers (81 years in high and 88 years in low risk category).

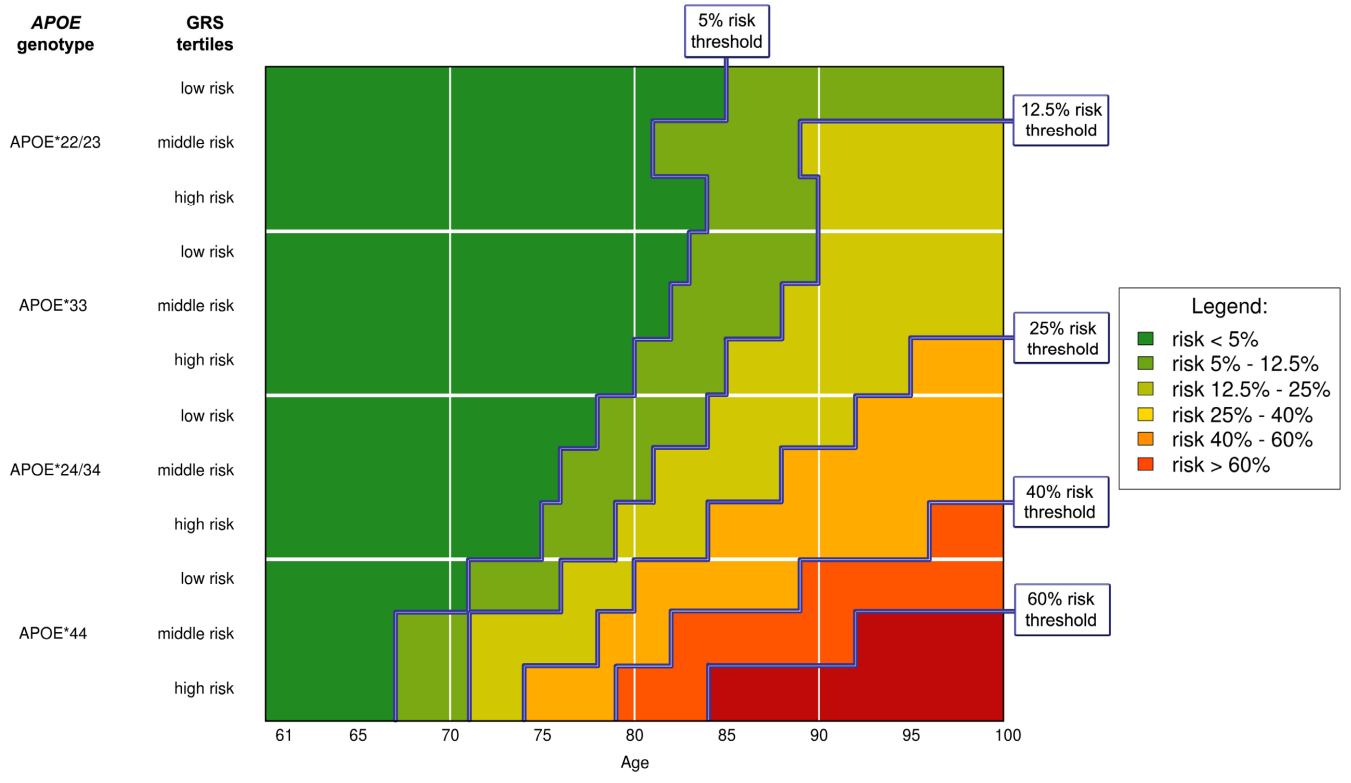


Figure 4: Risk of AD by age, *APOE* genotypes and tertiles of the GRS. The risk curves show the cumulative incidence per 100 individuals of AD by age, *APOE* genotypes and tertiles of the GRS is shown. The risk is categorized and coloured in six risk categories (lower than 5%, between 5% and 12.5%, between 12.5%% and 25%, between 25% and 40%, between 40% and 60% and over 60% risk). The age by which the risks of 5%, 12.5%, 25%, 40% and 60% is attained is marked with connected lines.

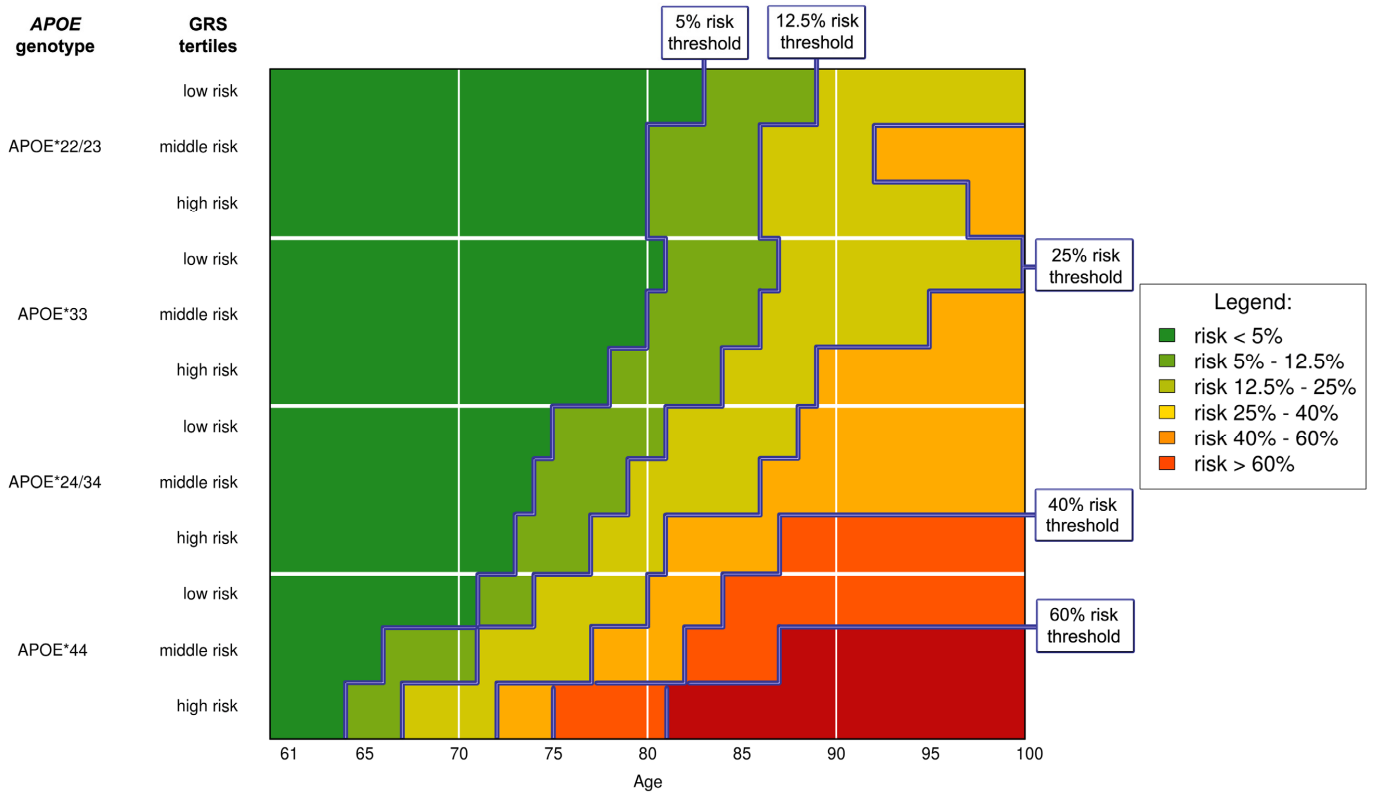


Figure 5: Risk of dementia by age, *APOE* genotypes and tertiles of the GRS. The cumulative incidence per 100 individuals of AD by age, *APOE* genotypes and tertiles of the GRS is shown. The risk is categorized and coloured in six risk categories (lower than 5%, between 5% and 12.5%, between 12.5%% and 25%, between 25% and 40%, between 40% and 60% and over 60% risk). The age by which the risks of 5%, 12.5%, 25%, 40% and 60% is attained is marked with connected lines.

Discussion

In our large population-based study, a GRS of common genetic variants modifies the risk and onset of AD and dementia above and beyond the effect of APOE, when taking into account the competing risk of mortality. The risk modification by the joint effect of common variants is most pronounced in *APOE**44 carriers in whom there is a difference of up to 10 years in onset age between those in the low risk tertile of the GRS and those in the high risk tertile of the GRS. This shift in age of onset translates into a higher risk of AD and dementia at all ages and into significant differences in risk by the age of 85 years. The same set of genetic variants that identify the *APOE**44 carriers at highest risk of AD also identify a risk subgroup that has a very low risk of AD and dementia (*APOE**22/23 carriers with few common risk variants). Between these subgroups with the highest risk and with the lowest risk there is a difference of almost 2 decades in age at onset.

The identification of subgroups at high genetic risk of AD with an earlier onset in the general population has important implications for precision medicine. Pathological changes related to AD begin to develop up to decades before the earliest clinical symptoms[8]. Therefore, preventive interventions are increasingly initiated in the subgroup of individuals with a high genetic risk at a younger age.[6,7,36] An important additional benefit for these costly trials is that selection of only the highest risk subgroups will decrease the duration of the trails,[6,21] and individuals at highest risk should have a chance to access the most promising treatments. On the other end of the risk spectrum are individuals with a very low risk of AD and a generally late onset age. These persons are of interest for inclusion in epidemiological studies aiming to discover protective factors and if these low-risk individuals, against expectation, develop AD at an early age these individuals are of interest for inclusion in genetic studies as they possibly carry high risk rare genetic variants.

An interesting finding in this study is the interaction between *APOE* genotypes and other genetic risk factors. Differential effect estimates of common genetic variants by

*APOE**4 genotypes have been reported previously,[19,20,37] here add to these findings that this significant extends to absolute risk and age at onset. The interaction between *APOE* genotypes and the other common genetic variants associated with AD may be driven by biological pathways that have been identified for AD based on the effects of common genetic variants.[38] These include endocytosis, haemostasis, cholesterol transport, hematopoietic cell lineage, protein folding, clathrin complex, immune response and protein ubiquitination.[38] *APOE* is a part of at least four of these pathways,[38,39] this may explain the observed interaction.

Cumulative incidence of AD and all-cause dementia has been estimated previously, with and without adjusting for competing risks.[4,5,25,40-44] The overall estimates in current study of lifetime risks[25,43] and the risks by 85 years of age stratified by *APOE* genotypes are comparable to previous reports adjusting for competing risks.[43] We found very similar patterns of the risk curves of AD and dementia, which is expected as most all-cause dementia is of the AD type, but may also in part be due to the effects on other types of dementia of *APOE*[39,45,46] and common genetic variants associated with AD.[47] Previous estimates showed significant effects of common genetic variants on age at onset,[19,21,48] we now showed that they modify the absolute risk of AD by shifting risk curves, especially within APOE*4 genotype carriers. The added value of genetic risk scores of common variants with small effects in terms of improved discrimination between AD cases and controls was reported previously as marginal.[19,20,22,49] However, our study and that of Desikan *et al.*[21] showed effects are substantial for risk and age at onset.

Strengths of the present study include the population-based setting, the prospective ascertainment of dementia, the completeness (92%) and the length of follow-up (up to 25 years), and adjustment for competing risk of mortality. We estimated cumulative incidences up to the age of 100 years with a relatively large numbers of subjects in the oldest old, for example at the age of 90 years 1161 participants were included. As the Rotterdam Study studies a population of mainly native Dutch descent and estimates of the genetic variants are based on studies of populations of European ancestry, racial and

ethnic differences need to be explored in other cohorts. Intermittent re-estimations will be necessary if new common and rare genetic variants are discovered. An inherent assumption of correcting for competing risk of mortality is that the genetic variants studied are not influencing mortality independent of AD and all-cause dementia. This assumption is backed by a recent study of the genetics of longevity[50] showing only a 5 month decrease in lifespan per APOE*4 allele (without adjusting for AD effects on mortality), and showed no significant effect for the other AD-associated variants.

In summary, we show that the small effect common genetic variants together significantly modify the risk of AD and dementia and may even partly determine the variability in the age of onset within and across *APOE* genotypes. Our findings contribute towards efficacy improvement of clinical trials and better risk prediction for AD and dementia.

Supplementary Table 1: Genetic variants included in the genetic risk score

Chr	Rsid	Alternate allele	Assigned-gene	Locus discovered in	Effect estimate from	Maf	Weight ALT-HRC	Rsq-RS1	Rsq-RS2	Rsq-RS3
19	rs4147929	G	*ABCA7*	Hollingworth et al. , Naj et al.	Lambert et al. (2013)	0.19	-0.135	0.916	0.917	0.991
2	rs6733839	T	*BIN1*	Seshadri et al.	Lambert et al. (2013)	0.409	0.188	0.960	0.911	0.962
20	rs7274581	C	*CASS4*	Lambert et al. (2013)	Lambert et al. (2013)	0.083	-0.139	0.990	0.989	0.990
6	rs10948363	G	*CD2AP*	Hollingworth et al. , Naj et al.	Lambert et al. (2013)	0.266	0.098	0.998	0.998	0.998
11	rs10838725	C	*CELF1*	Lambert et al. (2013)	Lambert et al. (2013)	0.316	0.075	0.998	0.998	0.998
8	rs9331896	T	*CLU*	Harold et al. , Lambert et al. (2009)	Lambert et al. (2013)	0.379	0.146	0.902	0.974	0.901
1	rs6656401	G	*CR1*	Lambert et al. (2009)	Lambert et al. (2013)	0.197	-0.157	0.953	0.948	0.950
10	rs7920721	G	*ECHDC3*	Desikan et al.	Desikan et al.	0.3867	-0.029	1.000	1.000	1.000
7	rs11771145	A	*EPHA1*	Hollingworth et al. , Naj et al.	Lambert et al. (2013)	0.338	-0.102	0.998	0.998	0.999
14	rs17125944	C	*FERMT2*	Lambert et al. (2013)	Lambert et al. (2013)	0.092	0.122	1.000	1.000	1.000
6	rs111418223	A	*HLA-DRB1/5*	Lambert et al. (2013)	Lambert et al. (2013)	0.276	-0.108	0.314	0.312	0.314
4	rs13113697	G	*HS3ST1*	Desikan et al.	Desikan et al.	0.2825	-0.029	0.999	0.998	0.999
2	rs35349669	T	*INPP5D*	Lambert et al.	Lambert et al. (2013)	0.488	0.066	0.975	0.973	0.976
17	rs118172952	G	*KANSL1*	Jun et al.	Lambert et al. (2013)	0.873	-0.151	0.710	0.700	0.708
5	rs190982	A	*MEF2C*	Lambert et al. (2013)	Lambert et al. (2013)	0.408	0.080	0.979	0.934	0.978
11	rs983392	G	*MS4A6A*	Hollingworth et al. , Naj et al.	Lambert et al. (2013)	0.403	-0.108	0.989	0.990	0.991
7	rs2718058	G	*NME8*	Lambert et al. (2013)	Lambert et al. (2013)	0.373	-0.070	1.000	1.000	1.000
11	rs10792832	G	*PICALM*	Harold et al.	Lambert et al. (2013)	0.358	0.130	0.999	0.999	0.999
8	rs28834970	C	*PTK2B*	Lambert et al. (2013)	Lambert et al. (2013)	0.366	0.096	0.993	0.990	0.994
14	rs10498633	T	*SLC24A4*	Lambert et al. (2013)	Lambert et al. (2013)	0.217	-0.104	0.999	0.999	1.000
11	rs11218343	C	*SORL1*	Lambert et al. (2013)	Lambert et al. (2013)	0.039	-0.270	0.998	0.995	0.998
6	rs75932628	T	*TREM2*	Guerreiro et al. , Jonsson et al.	Ruiz et al.	0.0016	0.889	0.762	0.726	0.668
7	rs1476679	T	*ZCWPW1*	Lambert et al. (2013)	Lambert et al. (2013)	0.287	0.078	0.995	0.996	0.995

Ordered by assigned gene name. References: Harold et al. [51], Seshadri et al.[17], Hollingworth et al.[10], Naj et al.[13] Lambert et al. (2009,2013)[11 9], Jonsson et al. [18] and Ruiz et al.[52]. Minor allel Frequency (Maf) of RS1 is shown and is representative of the MAF in RS2 and RS3. R^2= imputation quality. RS1= baseline Rotterdam study cohort, RS2=first extension Rotterdam study, RS3=second extension. Rotterdam study cohort were imputed separately. Gene names are NCBI gene names assigned to the loci in the corresponding references.

Supplementary Table 2: Cumulative incidence of AD by *APOE* genotypes and tertiles of the GRS

		All			*APOE**44			*APOE**34/24			*APOE**33			*APOE**22/23		
Group	**Age**	**Risk, 95%CI**	**n**	**p**	**Risk, 95%CI**	**n**	**p**	**Risk, 95%CI**	**n**	**p**	**Risk, 95%CI**	**n**	**p**	**Risk, 95%CI**	**n**	**p**
all	*65*	*0.1, 0.0-0.2*	*5220*		*1.0, 0.1-6.9*	*136*		*0.2, 0.0-0.6*	*1272*		*0.1, 0.0-0.3*	*2874*		*0.0, 0.0-0.0*	*678*	
low tertile	65	0.1, 0.0-0.5	1468	ref.	0.0, 0.0-0.0	39	ref.	0.0, 0.0-0.0	353	ref.	0.1, 0.0-0.8	873	ref.	0.0, 0.0-0.0	203	ref.
middle tertile	65	0.0, 0.0-0.0	1419	0.16	0.0, 0.0-0.0	44	NaN	0.0, 0.0-0.0	406	NaN	0.0, 0.0-0.0	779	0.16	0.0, 0.0-0.0	188	NaN
high tertile	65	0.1, 0.0-0.5	1454	0.29	0.0, 0.0-0.0	34	NaN	0.3, 0.0-1.8	369	0.16	0.1, 0.0-0.8	850	0.50	0.0, 0.0-0.0	200	NaN
all	*70*	*0.5, 0.4-0.8*	*5670*		*6.5, 3.4-12.2*	*128*		*0.6, 0.3-1.2*	*1370*		*0.3, 0.2-0.6*	*3170*		*0.3, 0.1-1.1*	*743*	
low tertile	70	0.2, 0.1-0.6	1584	ref.	0.0, 0.0-0.0	44	ref.	0.0, 0.0-0.0	381	ref.	0.3, 0.1-1.0	943	ref.	0.0, 0.0-0.0	216	ref.
middle tertile	70	0.5, 0.3-1.0	1545	6.9×10^{-2}	6.7, 2.2-19.3	39	3.7×10^{-2}	0.7, 0.2-2.0	441	4.1×10^{-2}	0.1, 0.0-0.8	863	0.16	0.5, 0.1-3.5	202	0.16
high tertile	70	0.8, 0.4-1.3	1624	1.1×10^{-2}	12.1, 5.2-26.7	28	8.8×10^{-2}	0.8, 0.3-2.4	400	4.1×10^{-2}	0.4, 0.2-1.2	962	0.34	0.0, 0.0-0.0	234	NaN
all	*75*	*2.1, 1.7-2.4*	*5423*		*17.5, 12.3-24.7*	*108*		*3.2, 2.4-4.2*	*1326*		*1.4, 1.0-1.8*	*3035*		*0.5, 0.2-1.4*	*730*	
low tertile	75	1.2, 0.8-1.8	1545	ref.	10.8, 4.6-24.0	34	ref.	1.6, 0.7-3.2	389	ref.	0.9, 0.5-1.7	909	ref.	0.4, 0.1-2.8	213	ref.
middle tertile	75	2.3, 1.7-3.2	1522	5.1×10^{-3}	18.1, 9.5-33.0	31	0.16	3.8, 2.4-5.9	414	1.8×10^{-2}	1.2, 0.7-2.1	869	0.24	1.0, 0.2-3.8	208	0.24
high tertile	75	2.6, 2.0-3.5	1541	9.8×10^{-4}	26.4, 15.5-42.7	24	3.0×10^{-2}	4.3, 2.8-6.7	392	6.3×10^{-3}	1.5, 0.9-2.5	901	0.11	0.0, 0.0-0.0	223	0.16
all	*80*	*5.7, 5.1-6.3*	*4254*		*29.7, 22.9-38.0*	*64*		*10.4, 9.0-12.0*	*987*		*3.6, 3.0-4.3*	*2458*		*1.5, 0.8-2.6*	*573*	
low tertile	80	3.8, 3.0-4.8	1260	ref.	18.7, 9.7-34.1	22	ref.	7.1, 5.1-9.8	311	ref.	2.6, 1.8-3.8	761	ref.	0.4, 0.1-2.8	166	ref.
middle tertile	80	5.8, 4.8-7.0	1195	3.5×10^{-3}	31.3, 19.9-47.1	19	8.4×10^{-2}	10.2, 7.8-13.3	305	4.3×10^{-2}	3.2, 2.3-4.5	707	0.21	2.3, 1.0-5.4	165	4.2×10^{-2}
high tertile	80	7.6, 6.5-9.0	1182	3.9×10^{-7}	41.3, 27.8-58.2	13	1.1×10^{-2}	14.7, 11.8-18.2	265	8.1×10^{-5}	4.6, 3.5-6.1	729	6.8×10^{-3}	1.6, 0.6-4.1	175	9.3×10^{-2}
all	*85*	*11.6, 10.8-12.4*	*2630*		*48.3, 40.1-57.3*	*23*		*18.4, 16.5-20.4*	*580*		*8.6, 7.7-9.6*	*1500*		*5.5, 4.1-7.4*	*387*	
low tertile	85	8.2, 7.0-9.6	780	ref.	37.6, 23.9-55.7	10	ref.	12.8, 10.1-16.3	192	ref.	6.1, 4.8-7.8	468	ref.	4.4, 2.4-8.1	111	ref.
middle tertile	85	11.3, 9.9-12.9	749	1.4×10^{-3}	49.4, 35.5-65.3	6	0.15	16.7, 13.6-20.5	180	4.9×10^{-2}	7.7, 6.1-9.6	448	8.8×10^{-2}	8.6, 5.5-13.2	114	4.1×10^{-2}
high tertile	85	15.7, 14.1-17.6	695	5.6×10^{-12}	62.2, 47.1-77.4	4	1.6×10^{-2}	26.3, 22.5-30.6	151	1.3×10^{-7}	11.5, 9.7-13.7	428	1.2×10^{-5}	5.5, 3.2-9.3	113	0.30
all	*90*	*19.0, 18.0-20.0*	*1161*		*53.8, 45.2-62.9*	*10*		*27.4, 25.1-29.8*	*220*		*16.0, 14.7-17.4*	*648*		*11.0, 8.8-13.6*	*183*	
low tertile	90	14.8, 13.2-16.7	342	ref.	43.0, 27.7-62.3	4	ref.	21.4, 17.7-25.7	73	ref.	12.5, 10.4-14.8	205	ref.	9.5, 6.2-14.4	60	ref.
middle tertile	90	19.0, 17.1-21.1	320	1.3×10^{-3}	58.5, 43.8-73.8	5	9.8×10^{-2}	27.3, 23.2-32.0	64	2.6×10^{-2}	14.7, 12.5-17.3	199	8.7×10^{-2}	13.0, 9.0-18.6	53	0.13
high tertile	90	24.0, 22.0-26.3	286	6.0×10^{-11}	62.2, 47.1-77.4	1	5.6×10^{-2}	35.0, 30.6-39.7	64	5.7×10^{-6}	20.6, 18.0-23.5	174	2.8×10^{-6}	10.7, 7.1-15.8	49	0.35
all	*95*	*23.4, 22.2-24.6*	*316*		*57.8, 48.7-67.2*	*2*		*31.6, 29.1-34.2*	*49*		*20.8, 19.3-22.4*	*178*		*15.8, 13.1-19.1*	*58*	
low tertile	95	20.0, 17.9-22.2	96	ref.	51.2, 33.5-71.7	1	ref.	26.6, 22.3-31.5	11	ref.	18.0, 15.4-20.9	65	ref.	11.7, 7.9-17.2	20	ref.
middle tertile	95	23.1, 20.9-25.5	88	2.5×10^{-2}	64.5, 48.7-80.0	1	0.15	30.3, 25.9-35.3	20	0.13	18.9, 16.3-22.0	51	0.32	19.2, 14.0-26.1	16	2.6×10^{-2}
high tertile	95	28.6, 26.3-31.1	69	8.2×10^{-8}	62.2, 47.1-77.4	1	0.20	39.4, 34.7-44.4	11	8.4×10^{-5}	24.9, 22.0-28.1	44	5.2×10^{-4}	17.3, 12.2-24.0	15	7.2×10^{-2}
all	*100*	*25.0, 23.8-26.3*	*48*					*33.0, 30.3-35.8*	*5*		*22.5, 20.9-24.3*	*30*		*17.7, 14.6-21.4*	*9*	
low tertile	100	20.8, 18.7-23.2	16	ref.				26.6, 22.3-31.5	2	ref.	19.3, 16.6-22.5	10	ref.	11.7, 7.9-17.2	4	ref.
middle tertile	100	25.5, 23.1-28.1	13	3.1×10^{-3}				33.2, 28.1-39.0	2	3.4×10^{-2}	21.7, 18.7-25.1	8	0.14	23.0, 16.9-30.8	3	4.0×10^{-3}
high tertile	100	29.9, 27.5-32.5	10	6.0×10^{-8}				40.3, 35.5-45.4	2	3.9×10^{-2}	26.5, 23.4-29.9	7	6.5×10^{-4}	18.7, 13.3-26.0	2	4.0×10^{-2}

APOE = apolipoprotein E genotypes. AD= Alzheimer's disease. NaN = difference in risk is zero. n = the number at risk by the corresponding age.

Empty cells show that none of the participants was alive and not demented. Cumulative incidence per 100 individuals (%) or risk of AD at 5 year intervals taking into account non-AD dementia and mortality as competing events is shown. The sum of the n at risk by tertiles does not necessarily add up to all subjects at risk at that age as not all subjects with *APOE* status also had the GRS measured.

Supplementary Table 3: Cumulative incidence of all-cause dementia by *APOE* genotypes and tertiles of the GRS

		All			*APOE**44			*APOE**34/24			*APOE**33			*APOE**22/23		
Group	**Age**	**Risk, 95%CI**	**n**	**p**	**Risk, 95%CI**	**n**	**p**	**Risk, 95%CI**	**n**	**p**	**Risk, 95%CI**	**n**	**P**	**Risk, 95%CI**	**n**	**p**
all	*65*	*0.2, 0.1-0.4*	*5220*		*2.9, 0.9-8.8*	*136*		*0.3, 0.1-0.9*	*1272*		*0.1, 0.0-0.3*	*2874*		*0.0, 0.0-0.0*	*678*	
low tertile	65	0.1, 0.0-0.5	1468	ref.	0.0, 0.0-0.0	39	ref.	0.0, 0.0-0.0	353	ref.	0.1, 0.0-0.8	873	ref.	0.0, 0.0-0.0	203	ref.
middle tertile	65	0.1, 0.0-0.5	1419	0.48	0.0, 0.0-0.0	44	NaN	0.3, 0.0-1.9	406	0.16	0.0, 0.0-0.0	779	0.16	0.0, 0.0-0.0	188	NaN
high tertile	65	0.4, 0.2-0.9	1454	4.8×10^{-2}	7.4, 1.9-26.9	34	7.3×10^{-2}	0.5, 0.1-2.2	369	7.8×10^{-2}	0.1, 0.0-0.8	850	0.50	0.0, 0.0-0.0	200	NaN
all	*70*	*1.0, 0.8-1.3*	*5670*		*9.8, 5.8-16.4*	*128*		*1.4, 0.9-2.2*	*1370*		*0.6, 0.4-1.0*	*3170*		*0.3, 0.1-1.1*	*743*	
low tertile	70	0.7, 0.4-1.3	1584	ref.	0.0, 0.0-0.0	44	ref.	0.5, 0.1-2.1	381	ref.	1.0, 0.5-1.9	943	ref.	0.0, 0.0-0.0	216	ref.
middle tertile	70	0.9, 0.5-1.5	1545	0.25	8.9, 3.5-22.1	39	1.8×10^{-2}	1.4, 0.6-3.1	441	0.10	0.3, 0.1-1.0	863	4.5×10^{-2}	0.5, 0.1-3.5	202	0.16
high tertile	70	1.4, 0.9-2.2	1624	2.6×10^{-2}	21.9, 11.5-39.4	28	7.9×10^{-4}	2.0, 1.0-4.0	400	3.0×10^{-2}	0.6, 0.3-1.4	962	0.22	0.0, 0.0-0.0	234	NaN
all	*75*	*3.1, 2.6-3.5*	*5423*		*22.1, 16.2-29.8*	*108*		*4.8, 3.8-6.0*	*1326*		*2.0, 1.6-2.6*	*3035*		*1.1, 0.6-2.1*	*730*	
low tertile	75	2.2, 1.6-3.0	1545	ref.	12.9, 6.0-26.5	34	ref.	3.2, 1.9-5.3	389	ref.	1.7, 1.1-2.7	909	ref.	0.8, 0.2-3.2	213	ref.
middle tertile	75	3.1, 2.4-4.1	1522	4.2×10^{-2}	20.4, 11.2-35.6	31	0.17	5.1, 3.5-7.5	414	6.9×10^{-2}	1.6, 1.0-2.7	869	0.44	1.8, 0.7-4.8	208	0.17
high tertile	75	3.9, 3.1-5.0	1541	1.5×10^{-3}	38.6, 25.3-55.7	24	2.7×10^{-3}	6.4, 4.5-9.1	392	1.1×10^{-2}	2.3, 1.5-3.4	901	0.19	0.4, 0.1-2.7	223	0.26
all	*80*	*8.1, 7.5-8.8*	*4254*		*37.1, 29.7-45.7*	*64*		*14.1, 12.5-15.9*	*987*		*5.3, 4.6-6.1*	*2458*		*3.1, 2.1-4.5*	*573*	
low tertile	80	6.0, 5.0-7.2	1260	ref.	23.5, 13.3-39.5	22	ref.	9.9, 7.5-13.0	311	ref.	4.5, 3.4-6.0	761	ref.	1.6, 0.6-4.3	166	ref.
middle tertile	80	7.7, 6.5-9.1	1195	2.1×10^{-2}	35.8, 23.7-51.7	19	0.10	13.4, 10.7-16.8	305	4.7×10^{-2}	4.4, 3.3-5.9	707	0.45	4.1, 2.1-7.7	165	5.8×10^{-2}
high tertile	80	10.7, 9.3-12.2	1182	1.6×10^{-7}	55.8, 41.2-71.5	13	8.5×10^{-4}	20.4, 17.0-24.3	265	3.8×10^{-6}	6.4, 5.1-8.1	729	3.2×10^{-2}	3.1, 1.6-6.1	175	0.13
all	*85*	*15.6, 14.7-16.5*	*2630*		*56.6, 48.2-65.2*	*23*		*24.0, 21.9-26.2*	*580*		*11.6, 10.6-12.8*	*1500*		*8.8, 7.0-11.1*	*387*	
low tertile	85	11.9, 10.5-13.6	780	ref.	42.4, 28.0-60.3	10	ref.	18.2, 15.0-22.2	192	ref.	9.2, 7.6-11.2	468	ref.	7.5, 4.7-11.7	111	ref.
middle tertile	85	14.6, 13.0-16.4	749	1.1×10^{-2}	53.9, 39.8-69.4	6	0.16	21.6, 18.1-25.7	180	0.10	10.4, 8.6-12.6	448	0.18	10.8, 7.3-15.9	114	0.11
high tertile	85	20.5, 18.6-22.5	695	6.2×10^{-12}	76.7, 62.6-88.4	4	7.3×10^{-4}	33.8, 29.6-38.4	151	4.0×10^{-8}	14.8, 12.7-17.2	428	7.3×10^{-5}	9.7, 6.5-14.3	113	0.19
all	*90*	*24.4, 23.3-25.6*	*1161*		*62.0, 53.5-70.7*	*10*		*34.2, 31.7-36.8*	*220*		*20.7, 19.3-22.2*	*648*		*16.2, 13.7-19.2*	*183*	
low tertile	90	19.7, 17.8-21.8	342	ref.	47.8, 32.0-66.7	4	ref.	28.8, 24.6-33.5	73	ref.	16.6, 14.3-19.2	205	ref.	12.5, 8.7-17.8	60	ref.
middle tertile	90	24.3, 22.2-26.6	320	1.1×10^{-3}	63.0, 48.2-77.7	5	0.10	33.1, 28.7-38.0	64	9.5×10^{-2}	19.6, 17.1-22.5	199	5.2×10^{-2}	19.3, 14.4-25.6	53	3.1×10^{-2}
high tertile	90	30.3, 28.1-32.7	286	4.4×10^{-12}	76.7, 62.6-88.4	1	5.5×10^{-3}	43.5, 39.0-48.4	64	4.3×10^{-3}	25.6, 22.8-28.6	174	2.4×10^{-6}	17.1, 12.5-23.0	49	9.6×10^{-2}
all	*95*	*29.6, 28.3-30.9*	*316*		*66.0, 57.0-74.9*	*2*		*39.3, 36.7-42.1*	*49*		*26.3, 24.6-28.0*	*178*		*21.5, 18.5-25.0*	*58*	
low tertile	95	25.6, 23.4-28.1	96	ref.	56.0, 37.9-75.8	1	ref.	35.0, 30.3-40.2	11	ref.	23.0, 20.2-26.2	65	ref.	14.7, 10.5-20.6	20	ref.
middle tertile	95	29.3, 26.9-31.9	88	1.7×10^{-2}	69.0, 53.2-83.6	1	0.16	36.9, 32.2-42.1	20	0.30	24.5, 21.6-27.8	51	0.25	27.2, 21.1-34.8	16	1.9×10^{-4}
high tertile	95	35.9, 33.5-38.5	69	2.7×10^{-9}	76.7, 62.6-88.4	1	4.5×10^{-2}	49.4, 44.5-54.5	11	3.0×10^{-5}	30.9, 27.8-34.3	44	2.5×10^{-4}	23.6, 18.0-30.8	15	1.6×10^{-2}
all	*100*	*31.4, 30.1-32.8*	*48*					*40.8, 38.0-43.7*	*5*		*28.4, 26.6-30.2*	*30*		*23.8, 20.3-27.8*	*9*	
low tertile	100	26.6, 24.3-29.1	16	ref.				35.0, 30.3-40.2	2	ref.	24.6, 21.6-28.0	10	ref.	14.7, 10.5-20.6	4	ref.
middle tertile	100	31.9, 29.3-34.6	13	2.2×10^{-3}				39.7, 34.4-45.6	2	0.11	27.6, 24.3-31.2	8	0.11	31.0, 24.2-39.2	3	2.1×10^{-4}
high tertile	100	37.8, 35.2-40.6	10	6.1×10^{-10}				50.3, 45.3-55.5	2	1.3×10^{-5}	33.3, 29.9-36.9	7	1.7×10^{-4}	26.3, 20.0-34.1	2	4.3×10^{-3}

APOE = apolipoprotein E genotypes. NaN = difference in risk is zero. n = the number at risk by the corresponding age.

Empty cells show that none of the participants was alive and not demented. Cumulative incidence per 100 individuals (%) or risk of dementia at 5 year intervals taking into account mortality as competing event is shown. The sum of the n at risk by tertiles does not necessarily add up to all subjects at risk at that age as not all subjects with *APOE* status also had the GRS measured.

Supplementary Table 4: Interaction of single variants in the GRS with *APOE* genotypes

Gene	snp	$HR_{interaction}$	$P_{interaction}$
ABCA7	rs4147929	0.97	0.73
BIN1	rs6733839	1.04	0.57
CASS4	rs7274581	1.03	0.79
CD2AP	rs10948363	1.04	0.56
CELF1	rs10838725	1.15	4.6×10^{-2}
CLU	rs9331896	1.15	3.4×10^{-2}
CR1	rs6656401	1.08	0.37
ECHDC3	rs7920721	1.08	0.27
EPHA1	rs11771145	1.05	0.45
FERMT2	rs17125944	1.02	0.82
HLA-DRB1/5	rs111418223	0.80	7.6×10^{-2}
HS3ST1	rs13113697	0.85	2.4×10^{-2}
INPP5D	rs35349669	0.94	0.38
KANSL1	rs118172952	0.97	0.75
MEF2C	rs190982	0.93	0.32
MS4A6A	rs983392	1.15	3.3×10^{-2}
NME8	rs2718058	1.01	0.84
PICALM	rs10792832	0.99	0.94
PTK2B	rs28834970	1.00	0.99
SLC24A4-RIN3	rs10498633	0.95	0.50
SORL1	rs11218343	1.00	0.99
TREM2	rs75932628	1.87	0.35
ZCWPW1	rs1476679	1.08	0.24

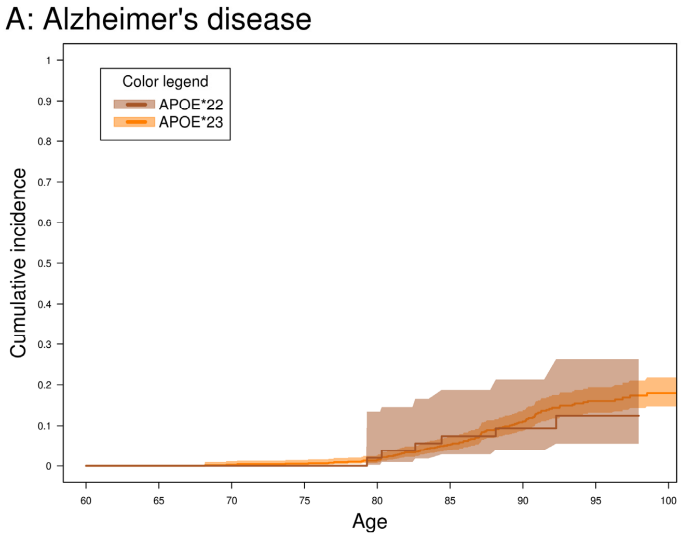
Interaction between *APOE* genotypes (*APOE**22/23, *APOE**33, *APOE**24/34 and *APOE**44, coded as 1,2,3,4 respectively) and the individual genetic variants included in the GRS tested using Cox proportional hazards, adjusting for main genetic effects, sex and age at inclusion.

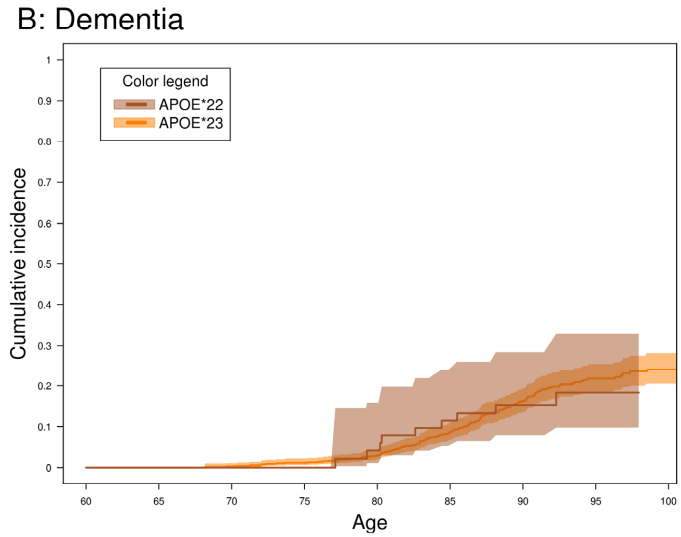
References:

1. Gatz M, Reynolds CA, Fratiglioni L, et al. Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry* 2006; **63**(2): 168-74.
2. Saunders AM, Strittmatter WJ, Schmechel D, et al. Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology* 1993; **43**(8): 1467-72.
3. Farrer LA, Cupples LA, Haines JL, et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *JAMA* 1997; **278**(16): 1349-56.
4. Genin E, Hannequin D, Wallon D, et al. APOE and Alzheimer disease: a major gene with semi-dominant inheritance. *Mol Psychiatry* 2011; **16**(9): 903-7.
5. Seshadri S, Drachman DA, Lippa CF. Apolipoprotein E epsilon 4 allele and the lifetime risk of Alzheimer's disease. What physicians know, and what they should know. *Arch Neurol* 1995; **52**(11): 1074-9.
6. Moulder KL, Snider BJ, Mills SL, et al. Dominantly Inherited Alzheimer Network: facilitating research and clinical trials. *Alzheimers Res Ther* 2013; **5**(5): 48.
7. Reiman EM, Langbaum JB, Fleisher AS, et al. Alzheimer's Prevention Initiative: a plan to accelerate the evaluation of presymptomatic treatments. *J Alzheimers Dis* 2011; **26 Suppl 3**: 321-9.
8. Jack CR, Jr., Knopman DS, Jagust WJ, et al. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *Lancet Neurol* 2013; **12**(2): 207-16.
9. Lambert JC, Heath S, Even G, et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet* 2009; **41**(10): 1094-9.
10. Hollingworth P, Harold D, Sims R, et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet* 2011; **43**(5): 429-35.
11. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 2013; **45**(12): 1452-8.
12. Guerreiro R, Wojtas A, Bras J, et al. TREM2 variants in Alzheimer's disease. *N Engl J Med* 2013; **368**(2): 117-27.
13. Naj AC, Jun G, Beecham GW, 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.
14. Desikan RS, Schork AJ, Wang Y, et al. Polygenic Overlap Between C-Reactive Protein, Plasma Lipids, and Alzheimer Disease. *Circulation* 2015; **131**(23): 2061-9.
15. Jonsson T, Atwal JK, Steinberg S, et al. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* 2012; **488**(7409): 96-9.
16. Jun G, Ibrahim-Verbaas CA, Vronskaya M, et al. A novel Alzheimer disease locus located near the gene encoding tau protein. *Mol Psychiatry* 2016; **21**(1): 108-17.
17. Seshadri S, Fitzpatrick AL, Ikram MA, et al. Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA* 2010; **303**(18): 1832-40.
18. Jonsson T, Stefansson H, Steinberg S, et al. Variant of TREM2 associated with the risk of Alzheimer's disease. *N Engl J Med* 2013; **368**(2): 107-16.
19. Sleegers K, Bettens K, De Roeck A, et al. A 22-single nucleotide polymorphism Alzheimer's disease risk score correlates with family history, onset age, and cerebrospinal fluid Abeta42. *Alzheimers Dement* 2015; **11**(12): 1452-60.
20. Chouraki V, Reitz C, Maury F, et al. Evaluation of a Genetic Risk Score to Improve Risk Prediction for Alzheimer's Disease. *J Alzheimers Dis* 2016; **53**(3): 921-32.
21. Desikan RS, Fan CC, Wang Y, et al. Genetic assessment of age-associated Alzheimer disease risk: Development and validation of a polygenic hazard score. *PLoS Med* 2017; **14**(3): e1002258.

22. Rodriguez-Rodriguez E, Sanchez-Juan P, Vazquez-Higuera JL, et al. Genetic risk score predicting accelerated progression from mild cognitive impairment to Alzheimer's disease. *J Neural Transm (Vienna)* 2013; **120**(5): 807-12.
23. Adams HH, de Bruijn RF, Hofman A, et al. Genetic risk of neurodegenerative diseases is associated with mild cognitive impairment and conversion to dementia. *Alzheimers Dement* 2015; **11**(11): 1277-85.
24. Lacour A, Espinosa A, Louwersheimer E, 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; **22**(1): 153-60.
25. Seshadri S, Wolf PA, Beiser A, et al. Lifetime risk of dementia and Alzheimer's disease. The impact of mortality on risk estimates in the Framingham Study. *Neurology* 1997; **49**(6): 1498-504.
26. Hofman A, Brusselle GG, Darwish Murad S, et al. The Rotterdam Study: 2016 objectives and design update. *Eur J Epidemiol* 2015; **30**(8): 661-708.
27. de Bruijn RFAG, Bos MJ, Portegies MLP, et al. The potential for prevention of dementia across two decades: the prospective, population-based Rotterdam Study. *BMC Med* 2015; **13**: 132.
28. Schrijvers EM, Verhaaren BF, Koudstaal PJ, Hofman A, Ikram MA, Breteler MM. Is dementia incidence declining?: Trends in dementia incidence since 1990 in the Rotterdam Study. *Neurology* 2012; **78**(19): 1456-63.
29. van Leeuwen EM, Kanterakis A, Deelen P, et al. Population-specific genotype imputations using minimac or IMPUTE2. *Nat Protoc* 2015; **10**(9): 1285-96.
30. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016.
31. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; 2015.
32. Allignol A, Schumacher M, Beyersmann J. Empirical Transition Matrix of Multi-State Models: The etm Package. *J Stat Softw* 2011; **38**(4): 1-15.
33. Meister R, Schaefer C. Statistical methods for estimating the probability of spontaneous abortion in observational studies--analyzing pregnancies exposed to coumarin derivatives. *Reprod Toxicol* 2008; **26**(1): 31-5.
34. Kaplan EL. Citation Classic - Nonparametric-Estimation from Incomplete Observations. *Cc/Life Sci* 1983; (24): 14-.
35. Tsai WY, Jewell NP, Wang MC. A Note on the Product-Limit Estimator under Right Censoring and Left Truncation. *Biometrika* 1987; **74**(4): 883-6.
36. NCT02565511 BMNLoMU-. A Study of CAD106 and CNP520 Versus Placebo in Participants at Risk for the Onset of Clinical Symptoms of Alzheimer's Disease. https://ClinicalTrials.gov/show/NCT02565511; 2016 Nov 14.
37. Marden JR, Mayeda ER, Walter S, et al. Using an Alzheimer Disease Polygenic Risk Score to Predict Memory Decline in Black and White Americans Over 14 Years of Follow-up. *Alzheimer disease and associated disorders* 2016; **30**(3): 195-202.
38. International Genomics of Alzheimer's Disease C. Convergent genetic and expression data implicate immunity in Alzheimer's disease. *Alzheimers Dement* 2015; **11**(6): 658-71.
39. Liu CC, Kanekiyo T, Xu H, Bu G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nat Rev Neurol* 2013; **9**(2): 106-18.
40. Myers RH, Schaefer EJ, Wilson PW, et al. Apolipoprotein E epsilon4 association with dementia in a population-based study: The Framingham study. *Neurology* 1996; **46**(3): 673-7.
41. Bickeboller H, Campion D, Brice A, et al. Apolipoprotein E and Alzheimer disease: Genotype-specific risks by age and sex. *Am J Hum Genet* 1997; **60**(2): 439-46.

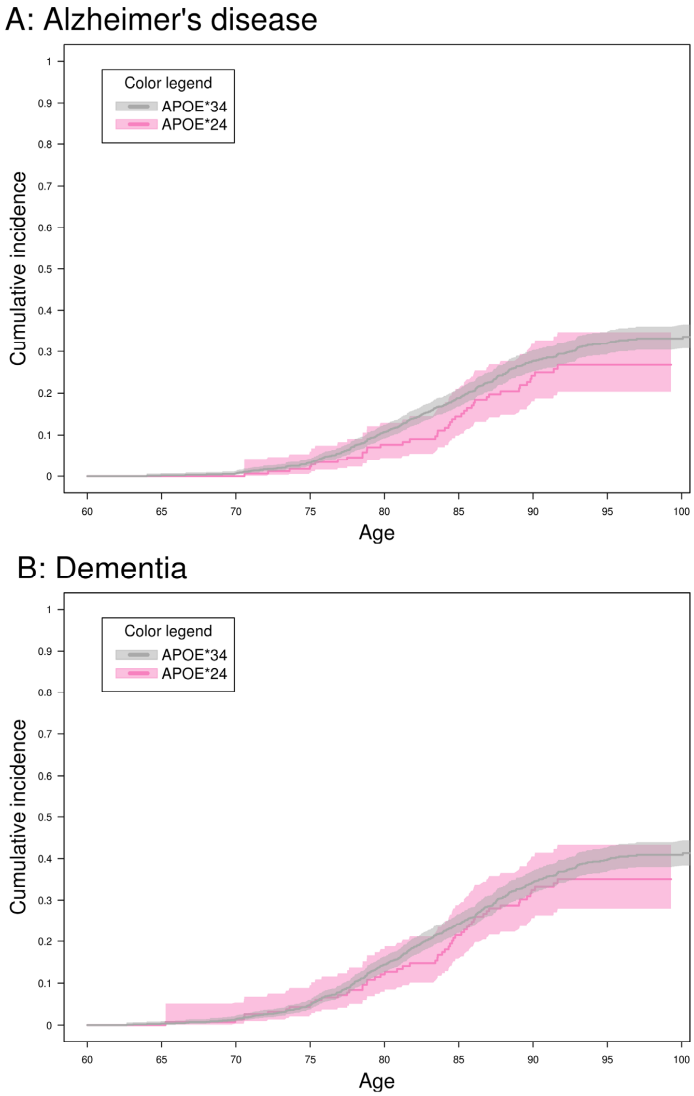
42. Cupples LA, Farrer LA, Sadovnick AD, Relkin N, Whitehouse P, Green RC. Estimating risk curves for first-degree relatives of patients with Alzheimer's disease: the REVEAL study. *Genet Med* 2004; **6**(4): 192-6.
43. Seshadri S, Wolf PA. Lifetime risk of stroke and dementia: current concepts, and estimates from the Framingham Study. *Lancet Neurol* 2007; **6**(12): 1106-14.
44. Seshadri S, Beiser A, Kelly-Hayes M, et al. The lifetime risk of stroke: estimates from the Framingham Study. *Stroke* 2006; **37**(2): 345-50.
45. Bras J, Guerreiro R, Darwent L, et al. Genetic analysis implicates APOE, SNCA and suggests lysosomal dysfunction in the etiology of dementia with Lewy bodies. *Hum Mol Genet* 2014; **23**(23): 6139-46.
46. Khan TA, Shah T, Prieto D, et al. Apolipoprotein E genotype, cardiovascular biomarkers and risk of stroke: systematic review and meta-analysis of 14,015 stroke cases and pooled analysis of primary biomarker data from up to 60,883 individuals. *Int J Epidemiol* 2013; **42**(2): 475-92.
47. Guerreiro R, Escott-Price V, Darwent L, et al. Genome-wide analysis of genetic correlation in dementia with Lewy bodies, Parkinson's and Alzheimer's diseases. *Neurobiol Aging* 2016; **38**: 214.e7-10.
48. Naj AC, Jun G, Reitz C, et al. Effects of multiple genetic loci on age at onset in late-onset Alzheimer disease: a genome-wide association study. *JAMA Neurol* 2014; **71**(11): 1394-404.
49. Escott-Price V, Sims R, Bannister C, et al. Common polygenic variation enhances risk prediction for Alzheimer's disease. *Brain* 2015; **138**(Pt 12): 3673-84.
50. McDaid AF, Joshi PK, Porcu E, et al. Bayesian association scan reveals loci associated with human lifespan and linked biomarkers. *Nature Communications* 2017; **8**.
51. Harold D, Abraham R, Hollingworth P, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* 2009; **41**(10): 1088-93.
52. Ruiz A, Dols-Icardo O, Bullido MJ, 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; **35**(2): 444 e1-4.





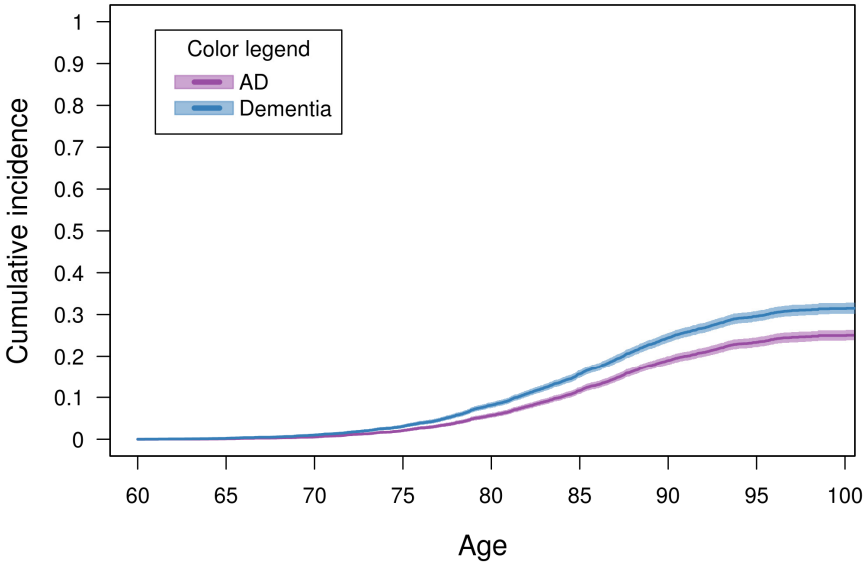
APOE*22	22	24	32	33	36	29	15	4	1
APOE*23	415	654	711	696	537	359	168	55	9

Supplementary Figure 1: Risk curves of Alzheimer's disease (A,C) and dementia (B,D) comparing *APOE22 vs. *APOE**23.** The risk curves show the cumulative incidence of Alzheimer's disease (A,C) and dementia (B,D). The shaded areas show the upper and lower 95% confidence limits of the corresponding cumulative incidence curve. The number of individuals at risk by age is shown in the table under the graph. The curves were compared by age 85 years p_{AD}=0.28 and $p_{dementia}$=0.26.



APOE*34	790	1146	1225	1180	884	517	189	45	5
APOE*24	85	126	146	147	103	62	31	4	1

Supplementary Figure 2: Risk curves of Alzheimer's disease (A,C) and dementia (B,D) comparing *APOE24 vs. *APOE**34.** The risk curves show the cumulative incidence of Alzheimer's disease (A,C) and dementia (B,D). The shaded areas show the upper and lower 95% confidence limits of the corresponding cumulative incidence curve. The number of individuals at risk by age is shown in the table under the graph. The curves were compared by age 85 years p_{AD}=0.08 and $p_{dementia}$=0.21.



Supplementary Figure 3: Risk curves of Alzheimer's disease and dementia. The risk curves show the cumulative incidence of Alzheimer's disease and dementia. The shaded areas show the upper and lower 95% confidence limits of the corresponding cumulative incidence curve.

Chapter 2.2

Characterization of pathogenic *SORL1* genetic variants for association with Alzheimer's disease: A clinical interpretation strategy

Henne Holstege, Sven J. van der Lee, Marc Hulsman, Tsz Hang Wong, Jeroen G.J. van Rooij, Marjan Weiss, Eva Louwersheimer, Frank J. Wolters, Najaf Amin, André G. Uitterlinden, Albert Hofman, M. Arfan Ikram, John C. van Swieten, Hanne Meijers-Heijboer, Wiesje M. van der Flier, Marcel J. T. Reinders, Cornelia M. van Duijn, Philip Scheltens.

European Journal of Human Genetics. 2017 May 24; 25: 973–981.

Abstract

Background: Accumulating evidence suggests that genetic variants in the *SORL1* gene are associated with Alzheimer disease (AD), but a strategy to identify which variants are pathogenic is lacking.

Methods: In a discovery sample of 115 *SORL1* variants detected in 1,908 Dutch AD cases and controls we identified the variant characteristics associated with *SORL1* variant pathogenicity. Findings were replicated in an independent sample of 103 *SORL1* variants detected in 3,193 AD cases and controls. In a combined sample of the discovery and replication samples, comprising 181 unique *SORL1* variants, we developed a strategy to classify *SORL1* variants into five subtypes ranging from pathogenic to benign. We tested this pathogenicity screen in *SORL1* variants reported in two independent published studies.

Results: *SORL1* variant pathogenicity is defined by the Combined Annotation Dependent Depletion (CADD) score and the minor allele frequency (MAF) reported by the Exome Aggregation Consortium (ExAC) database. Variants predicted strongly damaging (CADD score >30), which are extremely rare (ExAC-MAF $<1x10^{-5}$) increased AD risk by 12-fold (95%CI 4.2-34.3; $p=5x10^{-9}$). Protein truncating *SORL1* mutations were all unknown to ExAC and occurred exclusively in AD cases. More common *SORL1* variants (ExAC-MAF $\geq 1x10^{-5}$) were not associated with increased AD risk, even when predicted strongly pathogenic. Findings were independent of gender and the *APOE-ε4* allele. High-risk *SORL1* variants were observed in a substantial proportion of the AD cases analyzed (2%).

Conclusions: Based on their effect size, we propose to consider high-risk *SORL1* variants next to variants in *APOE*, *PSEN1, PSEN2* and *APP* for personalized risk assessments in clinical practice.

Introduction

Approximately one third of the population older than 85 years has Alzheimer's disease (AD).[1] Despite intensive research, the pathophysiology underlying AD is still poorly understood. The risk to develop (non-Mendelian) AD is estimated to be 60-80% heritable,[2] suggesting that the identification of genetic determinants of AD will provide further insights in underlying molecular mechanisms of AD.

Rapidly accumulating genetic and biological evidence suggests that disturbed function of the sortilin-related receptor 1 (SORL1) is associated with Alzheimer's Disease (AD).[3-7] Functional SORL1 reduces the amyloid-β levels in the brain, thereby reducing the load of neurotoxic amyloid-β plaques, a neuropathological hallmark of AD.[8] SORL1 reduces amyloid-β levels by (1) binding the amyloid precursor protein (APP), preventing its processing into amyloid-β; and by (2) binding amyloid-β and directing it to the lysosome for degradation.

To exert its function, the SORL1 protein includes domains from the low density lipoprotein receptor (LDLR)-like family, including complement-type repeats that interact in a 1:1 stoichiometric complex with APP.[9,10] The SORL1 protein also includes a VPS10 domain from the vacuolar protein sorting-10 receptor family, which binds soluble Aβ for endosomal inclusion and sorting for lysosomal degradation.[11,12] Therefore, impaired SORL1 function is associated both with disturbed APP processing and disturbed Aβ degradation, two central events underlying the pathophysiology of AD.[13-16]

Since 2007, a multitude of studies associated both common and rare variants in the *SORL1* gene with AD[17-21] but different variants were associated across studies and risk-effects ranged from small modifying effects to causal effects.[22,23] In genome-wide association studies (GWAS) including thousands of AD cases and controls, common genetic variants near and in the *SORL1* gene were found to associate significantly with AD.[24] However, each of these variants confer only a small increase in AD risk (OR=~1.2), which is comparable to the small changes in AD risk conferred by the other ~20 genetic loci that were also identified in these GWAS studies (reviewed by van Cauwenberghe *et*

al., 2016[25]). In contrast, recent studies reported that rare *SORL1* variants are associated with a 5-fold increased risk for early onset AD.[5,6] These studies suggest that the effect on AD risk of these rare *SORL1* variants is comparable to that of carrying the *ε4* allele of Apolipoprotein-E gene (*APOE*), the most important common risk factor for AD; homozygous and heterozygous *APOE-ε4* carriers are respectively exposed to a 3-5-fold increased AD risk and 10-15-fold increased AD risk compared to non-*APOE-ε4* carriers.[26] Furthermore, recent targeted sequencing studies identified rare pathogenic *SORL1* mutations that segregated with disease in families with familial AD and late onset AD.[3,4] These findings suggest that specific *SORL1* variants are causal for AD, with effects comparable to single mutations in the amyloid precursor protein gene (*APP*),[27] or the presenilin-1 and presenilin-2 genes (*PSEN1*, *PSEN2*),[28,29] that are associated with an autosomal dominant inheritance pattern of AD. However, it is currently unclear which specific *SORL1* variants are major risk factors for AD and which can be considered benign. This raises the need for a strategy to determine *SORL1* variant pathogenicity.

Ideally, one would like to determine penetrance for each detected *SORL1* variant in multiple large and informative families. However, the rareness of the *SORL1* variants that were thus far associated with AD complicates such efforts.[3-6] Therefore, a more feasible approach might be to distinguish between pathogenic and non-pathogenic variants using *independent* variant characteristics that might be associated with *SORL1* variant pathogenicity. In this work we explored the contribution on disease outcome of (i) the functional protein domain affected by the *SORL1* variant; (ii) the minor allele frequency (MAF) of the *SORL1* variant in the population; and (iii) predicted damagingness of the *SORL1* variant on the basis of sequence context.

To predict the damagingness of *SORL1* variants we annotated them with the Combined Annotation Dependent Depletion (CADD) score,[30] a novel functional annotation tool that allows for an unbiased annotation of almost all variants in the human genome. The CADD score reflects the difference between the characteristics of genetic variation that is tolerated (fixed) in the human genome and the characteristics of pathogenic variants (randomly simulated variants enriched with pathogenic variants). Scores are based on

>60 functional prediction tools which include functional annotations, allelic conservation, and regulatory effects.

The minor allele frequency (MAF) of a variant can be derived from the (large) sample in which it was discovered, but also from publically available databases such as the ExAC database,[31] which includes variants detected in a set of 60,706 exomes. Importantly, the MAF per variant is *not* included in determining its CADD score, such that the CADD score and MAF can be used as independent determinants of variant pathogenicity.

We determined the characteristics of *SORL1* variant pathogenicity in a discovery sample consisting of 640 Dutch early- and late onset AD patients and in 1268 older Dutch cognitively healthy controls. We tested whether these characteristics replicated in an independent dataset reported by Verheijen *et al*[6] and we performed a combined analysis including data from both studies. Based on these results, we suggest five *SORL1* variant subtypes according to the five-class system of variant pathogenicity supported by the ACMG[32] and ACGS.[33] Lastly, we validated our classification strategy by applying it to *SORL1* variants described in two additional independent publications.[4,5]

Material and Methods

Detailed Methods are described in the **Supplementary Methods**. For a schematic overview of the analysis setup see **Figure 1.**

Samples

Discovery Sample: The exome collection was assembled from four Dutch studies: (i) the Rotterdam Study[34] contributed 250 AD cases (median age at disease onset 84.5 ± 6.62, 71.2% female) and 1,204 controls (median age at last visit of 82.4 ± 6.8, 54.3% female), (ii) The Amsterdam Dementia Cohort (ADC-VUmc) contributed 320 AD (median age of 58.4 ± 6.5, 51.9% female)[35], (iii) the Alzheimer Centrum Zuidwest Nederland (ACZN) cohort contributed 80 AD cases (median age 59.2 ± 7.2, 57.1% female), and (iv) the 100-plus Study (www.100plus.nl) contributed 64 controls (median age of 101.1 years ± 3.5, 79.7% female).

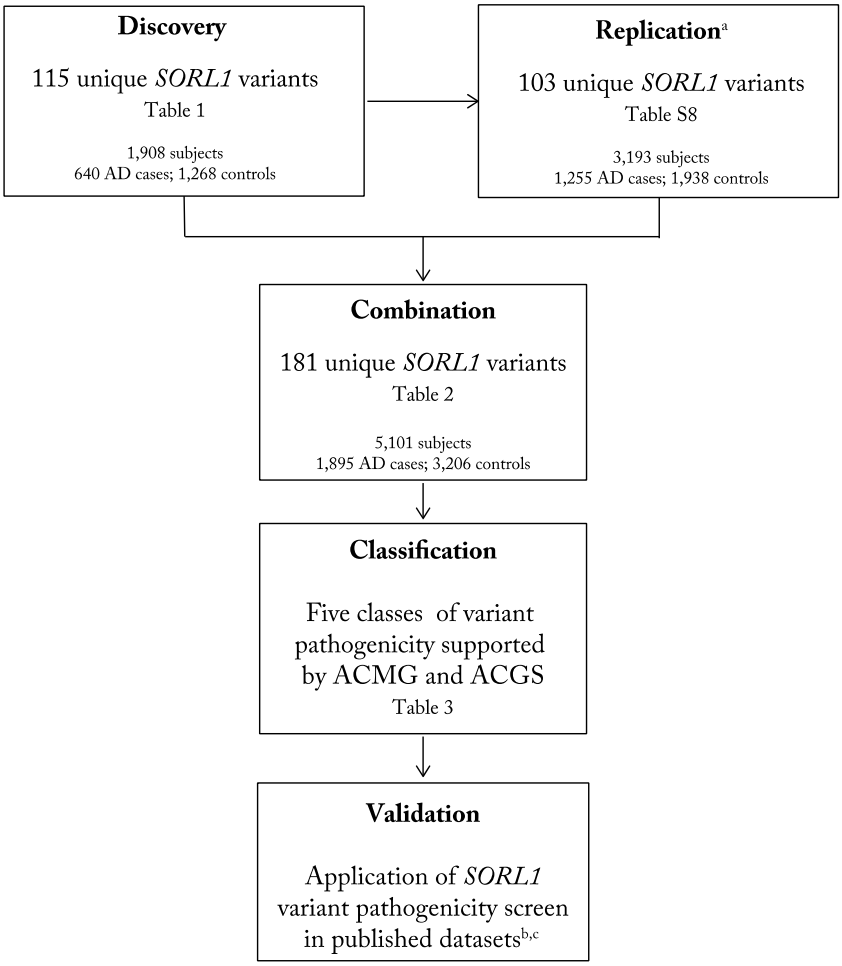


Figure 1. Flowchart of *SORL1* variant pathogenicity analysis. (a) *SORL1* variants in the independent replication dataset were reported by Verheijen *et al.,* 2016. Pathogenicity screen was applied to *SORL1* variants reported by (b) Vardarajan *et al.,* 2015 and (c) Nicolas *et al.,* 2015

In total, 1,908 exomes passed quality control: 640 cases (median age at onset of 64.8 years, IQR: 57.3-82.2, 60% female) and 1,268 controls (median age at last screening of 82.7 years, IQR 78.3-87.6, 55.6% female) (For cohort characteristics see **Table S1**; for age distribution of cases and controls see **Figure S1**) No known AD-causing mutations were detected in the *APP*, *PSEN1* and *PSEN2* genes. Combined data included (i) AD status, (ii) *APOE* status, (iii) gender, and (iv) age at onset for AD and (v) age at last screening for controls.

Replication sample: We performed a replication analysis in an independent dataset recently published by Verheijen *et al.*[6] They reported 103 *SORL1* variants in 1,255 European early onset AD cases and 1,938 age-matched controls. Rare *SORL1* variants that occurred in either cases *or* controls were given per subject (including gender and age, but not *APOE* genotype), and no two variants occurred in the same subject; the number of case- and control-carriers were given for rare *SORL1* variants that occurred in both cases *and* controls; common *SORL1* variants (MAF>0.01 in the sample) were available as sample-MAFs. For an in-depth analysis of independence between the discovery and the replication samples see **Supplementary Data** and **Table S2**.

Exome sequencing and variant detection in discovery sample

Exomes from the Rotterdam Study and the ACZN cohort were captured with the Nimblegen v2 Seqcap EZ Exome capture kit. The exomes from the ADC-VUmc cohort and the 100-plus Study cohort were captured with the Nimblegen SeqCap EZ Exome capture kit v3. For *SORL1* variant calling, we used the intersection between these capture kits in the SORL1 gene: 93.2% of exons 1-47, 2.7% of exon 48. DNA from all samples was prepared with the Illumina TruSeq Paired-End Library Preparation Kit and 100 bp paired-end reads were acquired by sequencing the libraries on a HiSeq 2000 or 2500. We sequenced to at least 40x mean coverage to ensure sufficient read depth for variant calling.

We removed population outliers based on the first two PCA components and those with an identity by decent (IBD) value >0.1. Technical differences between data acquisition commonly introduces 'differential missingness' (i.e. loci may be genotyped in one exome but not in another) which may ultimately result in unwanted bias towards either cases or controls. To overcome this bias we implemented additional quality control (see **Supplementary Methods)** to obtain a set consisting of 115 true positive variant calls with negligible missingness across the sample. Variants are listed in **Table S3,** and they are submitted to the LOVD database (www.LOVD.nl/SORL1). The subset of variants that were detected only once in the sample (singleton variants) were validated by Sanger sequencing (**Table S4**).

Statistical analysis

Variant annotation: Variants were annotated with the Combined Annotation Dependent Depletion score (CADD) version 1.3[30], and using the Variant Effect Predictor (VEP) tool in the Ensembl database[36]. Variants were annotated with SIFT v.5.2.2 /PolyPhen v.2.2.2 prediction scores (**Table S3**).

Discovery analysis: Since the rarity of most detected variants does not allow a per-variant calculation of disease association, we tested the burden of all *SORL1* variants that adhered to a specified set of characteristics in AD cases and controls[37]. For this, variant characteristics were based on combinations of the minor allele frequency (MAF) and CADD scores. The MAF categories included: MAF>0.01, 0.001<MAF<0.01, 0.0005<MAF<0.001, and MAF<0.0005 (singletons). The CADD score categories included: CADD 0-20 (predicted not or mildly pathogenic), 20-30 (predicted moderately pathogenic) and >30 (predicted strongly pathogenic). In addition, variants were stratified according to SORL1 protein domains. We then performed burden tests using an additive genetic model and logistic regression score test with the *burdenMeta* function in the "seqMeta" package v.1.6.0[38] in R (v3.2.2), while including gender as a covariate. *APOE* genotypes were missing for 36 controls and 6 cases, therefore, we performed separate burden tests using both gender and *APOE-ε4* genotype as covariates. Furthermore, we tested for an interaction effect between the burden of *SORL1* variants and the presence of the *APOE-ε4* allele.

Replication analyses were performed with a one tailed Fisher's exact test due to data availability. As rare variants did not overlap between subjects this approach is the same as a burden test. Correction for *APOE* and gender was not possible because *APOE* genotypes were not publically available in the replication sample, and gender only for a sample-subset.

Finally, for a combined analysis of the discovery sample and the replication sample we annotated all variants with their MAFs reported in the publically available ExAC database v.0.3.1[31] (**Table S5**). To more closely distinguish between benign and (possibly)

pathogenic variants, we stratified the CADD tranches into CADD 0-10, 10-20, 20-30 and >30. Combined analyses were performed with a one tailed Fisher's exact. Multiple testing correction was applied to correct for 70 tests performed in discovery, replication and combined analysis. *p*-values lower than 0.05 after Bonferroni were considered statistically significant ($p<7.1x10^{-4}$). We present unadjusted *p*-values.

Results

We detected 115 *SORL1* variants in the exomes of 640 AD cases and 1268 controls: 4 frameshift, 1 stop-gain, 54 missense, 29 synonymous, 4 regulatory, 4 splice site, and 19 intronic variants. Details of the filtering steps and Sanger validation are in the supplementary results. Of these 115 variants, 15 were common (MAF>0.01) and none was significantly associated with AD (**Table S3**). The remaining 100 variants were rare with a MAF<0.01. These variants did not occur more often in cases than in controls (OR=1.2; 95%CI 0.9-1.6; *p*=0.23) (**Table 1A**). There were 36 variants predicted deleterious by SIFT and Polyphen, and these variants were associated with a 2-fold increased AD risk (OR=1.9; 95%CI 1.2-2.9; $p=7.2x10^{-3}$) (**Table 1A**).

Singleton variants with high CADD score are associated with AD

Of the 100 rare variants (MAF<0.01), 26 variants were predicted moderately pathogenic (CADD 20-30) and carrying such a variant is not associated with a significantly increased AD risk (OR=1.3; 95%CI 0.78-2.1; *p*=0.34) (**Table 1A**). In contrast, the 19 variants that were predicted strongly pathogenic (CADD>30) were seen in 15 cases and 8 controls, such that carrying a variant with these characteristics is associated with a 4-fold increased risk for AD: OR=4.0; 95%CI 1.7-9.0; $p=9.9x10^{-4}$ (**Table 1A**).

Interestingly, 16 of these 19 strongly pathogenic variants with MAF<0.01 were seen only once in our sample: singletons. Among the 16 carriers of strongly damaging singletons (CADD>30), 14 had developed AD, such that strongly pathogenic singletons were associated with a >10-fold increase of AD risk (OR=11.3; 95%CI 4.0-32.1; $p=4.9x10^{-6}$). Notably, all five truncating mutations (stop-gain/frameshift) were singletons in our sample, and their carriers all developed AD ($p=1.6x10^{-3}$). In sharp contrast, variants with CADD >30 that occur more than once in this sample were not associated with AD

(**Table 1B**). Gender and carrying the *APOE-ε4* genotype did not influence these findings (**Table S6)** and we detected no evidence for an interaction of a pathogenic *SORL1* variant (CADD 20-30 or CADD>30) and carrying the *APOE-ε4* genotype (**Table S7**, **Supplemental Results**). Of note: no subject carried more than one singleton variant, with the exception of one control subject who carried two intronic singleton variants.

The association of singleton variants that were predicted pathogenic with AD was further illustrated by the finding that the median CADD score for the 30 singletons detected in cases (28.3; IQR 17.6-32.8) was significantly higher than the median CADD score for the 40 singletons in controls (14.9, IQR 8.1-23.1; p=5.2x10^{-5}). The median CADD score of only slightly more common variants did not significantly differ between cases and controls (p=0.23) (**Figure S2**).

The median age at onset for carriers of strongly pathogenic *SORL1* singletons (CADD >30, n=14) was 58.9 years, compared to 65.1 years for cases without these singletons (p=0.08, one-tailed Mann-Whitney U test). The 5 cases with carriers of stopcodon/frameshift mutations had a median age at onset of 57.7 (**Supplemental Results).**

Next, we analyzed whether there was a differential burden of pathogenic singleton variants in individual SORL1 protein domains **(Table 1C;** for affected amino-acid positions in protein see **Figure S3)**. All 6 subjects who carried a moderately damaging singleton (CADD 20-30) in the VPS10 domain developed AD, which was associated with almost 20-fold increased AD risk (OR=19.3; 95%CI 3.6-105.2; p=6.1x10^{-4}), and 3 out of 4 subjects with a strongly damaging variant in the VPS10 domain (CADD>30) developed AD (OR=6.6; 95%CI 0.82 – 53.1; p=7.6x10^{-2}). In contrast, only 1 out of 12 subjects who carried such a moderately damaging variant in one of the other domains developed AD.

Table 1. Discovery analysis: 115 *SORL1* variants in 640 cases and 1,268 controls.

CADD score	N-variants in burden test	Cases With at least one variant Total 640	Controls With at least one variant Total 1,268	OR (95% CI)	*p*-value[#] (unadjusted)
A. Variant stratification according to sample MAF <0.01 and SIFT/ Polyphen and CADD					
MAF <0.01					
all variants	100	83	140	1.2 (0.9-1.6)	0.23
SIFT/Polyphen damaging	36	40	44	1.9 (1.2-2.9)	7.2×10^{-3}
CADD 0-20	55	42	94	0.9 (0.6-1.3)	0.59
CADD 20-30	26	28	43	1.3 (0.78-2.1)	0.34
CADD >30	19	15	8	4.0 (1.7-9.0)	9.9×10^{-4}
B. Variant stratification according to sample-MAF and CADD					
MAF >0.01 (variants detected in >38 subjects in this sample)					
0-20	13	640	1,268	1.0 (1.0-1.1)	0.54
20-30	1	48	103	1.0 (0.7-1.3)	0.83
>30	1	17	44	0.8 (0.5-1.3)	0.29
0.001<MAF <0.01 (variants detected in 3-38 subjects in this sample					
0-20	7	27	47	1.2 (0.7-1.9)	0.54
20-30	3	18	24	1.5 (0.8-2.9)	0.21
>30	1	1	3	0.7 (0.1-5.5)	0.73
0.0005<MAF <0.001 (variants detected in 2-3 subjects in this sample)					
0-20	12	6	21	0.7 (0.3-1.4)	0.33
20-30	5	3	9	0.7 (0.2-2.4)	0.57
>30	2	1	3	0.7 (0.1-5.2)	0.70
MAF <0.0005 (singletons in this sample)					
0-20	36	9	26*	0.7 (0.4-1.4)	0.31
20-30	18	7	11	1.3 (0.5-3.3)	0.65
>30	**16**	**14**	**2**	**11.3 (4.0-32.1)**	**4.9×10^{-6}**
missense	11	9	2	8.7 (2.5-30.4)	7.2×10^{-4}
stop/frameshift	5	5	0	19.8 (3.1-126.8) [$]	1.7×10^{-3}
C. Variant stratification of singleton variants according to protein domain and CADD					
VPS10					
0-20	5	1	4	0.6 (0.1-3.6)	0.54
20-30	**6**	**6**	**0**	**19.3 (3.6-105.2)** [$]	**6.1×10^{-4}**
>30	4	3	1	6.6 (0.8-53.1)	7.6×10^{-2}
LDL A and B					
0-20	10	3	7	0.9 (0.2-3.2)	0.81
20-30	4	0	4	0.22 (0.0-1.7) [$]	0.14
>30	8	7	1	10.8 (2.5-46.7)	1.5×10^{-3}
Fibronectin					
0-20	3	0	3	0.2 (0.0-2.5) [$]	0.23
20-30	5	0	5	0.2 (0.0-1.4) [$]	0.11
>30	4	4	0	20.5 (2.6-164.5) [$]	4.5×10^{-3}

Variants were categorized according to the associated minor allele frequency in the sample and CADD score. For each category, multiple variants per individual were collapsed into one variant such that an odds ratio and an associated P value could be calculated using the score-based SeqMeta burden test, while using gender as a covariate. OR = Odds Ratio, VPS10 = Vacuolar Protein Sorting domain 10, LDL-Receptor = Low Density lipoprotein receptor A and B. * 1 control had two intronic singletons with CADD 0-20. [#]Associations that are significant after multiple testing correction are shown in bold ($p < 7.1\times10^{-4}$). [$]Results for the odds-ratios can be considered a one-step approximation to the maximum likelihood result, such that OR's of infinite have lower values when controls contribute no variants to the association.

Replication in published data

We replicated our findings in 103 *SORL1* variants reported in an independent published dataset.[6] In this dataset, singleton variants with CADD>30 associated with a 14-fold increased AD risk (OR=14.1; 95%CI 3.3-60.8; p=3.5x10^{-6}) and, all eight stop-gain/frameshift variants were singletons observed exclusively in cases (p=5.6x10^{-4}) (**Table S8**). Singleton missense variants with CADD>30 associated with an 8-fold increased AD risk (OR=7.8; 95%CI 1.7-35.5; p=2.4x10^{-3}). We were not able to replicate our findings in the VPS10 domain: moderately damaging variants in the VPS10 domain were carried only by five young control subjects (aged 55, 60 and 62 years and 2 with unknown ages).

Combined analysis

We substantiated the characterization of *SORL1* variant pathogenicity in the 181 unique *SORL1* variants in the 1,895 cases and 3,206 controls of the combined the discovery and replication samples (**Table 2**); all variants are listed in **Table S5**.

SORL1 variants that were (i) novel or listed only once in the ExAC database, MAF <1x10^{-5}) and (ii) high predicted variant damagingness (CADD>30) had the largest effect on AD risk (OR=12.0; 95%CI 4.2-34.3; p=5x10^{-9}) (**Table 2, Figure 2**). Slightly more common variants observed 2-12x in the ExAC database (ExAC-MAF between 1x10^{-5} and 1x10^{-4}) with CADD>30 associated with an 8.5-fold increased AD risk (95%CI 1.9-38.8; p=1.4x10^{-3}); more common variants with ExAC-MAF>1x10^{-4} and CADD>30 do not associate with a significantly increased AD risk (**Table 2, Figure 2**). Together, the maximum statistical evidence for an effect on AD risk is obtained for *SORL1* variants with CADD>30 and ExAC-MAF<1x10^{-4} (OR=10.9; 95%CI 4.6-25.7; p=1.8x10^{-11}) (**Table 3**). *SORL1* variants with these characteristics occur in 38 from 1,895 cases (2%) and in 6 from 3,206 controls (0.19%).

Table 2. Combined analysis: 181 unique *SORL1* variants in 1,895 cases and 3,206 controls.

CADD score	N-variants in burden test[a]	Cases With at least one variant **Total 1,895**	Controls With at least one variant **Total 3,206**	OR (95% CI)	P value[#] (unadjusted))
ExAC-MAF >0.01 (each variant detected in >1200 subjects in the ExAC database)					
0-10	8[b]	1,895	3,206	1.0	1.0
10-20	4[b]	1,895	3,206	1.0	1.0
20-30	1	203	300	1.2 (1.0-1.4)	$7.3x10^{-2}$
>30	1	67	162	0.7 (0.5-0.9)	1.0
0.001< ExAC-MAF <0.01 (each variant detected in 121-1200 subjects in ExAC database)					
0-10	5[b]	93	158	1.0 (0.8-1.3)	0.52
10-20	4[b]	68	151	0.8 (0.6-1.0)	0.98
20-30	3	37	40	1.6 (1.0-2.5)	$3.2x10^{-2}$
>30	1	4	7	1.0 (0.3-3.3)	0.63
0.0001< ExAC-MAF <0.001 (each variant detected in 13-120 subjects in ExAC database)					
0-10	5	10	12	1.4 (0.6-3.3)	0.27
10-20	5	9	11	1.4 (0.6-3.4)	0.30
20-30	11	15	21	1.2 (0.6-2.4)	0.34
>30	4	5	4	2.1 (0.6-7.9)	0.21
0.00001 < ExAC-MAF <0.0001 (each variant detected in 2-12 subjects in ExAC database)					
0-10	5	7	10	1.2 (0.5-3.1)	0.45
10-20	20	9	21	0.7 (0.3-1.6)	0.84
20-30	22	9	23	0.7 (0.3-1.4)	0.90
>30	9	10	2	8.5 (1.9 – 38.8)	$1.4x10^{-3}$
ExAC-MAF <0.00001 (unknown or singletons in ExAC database)					
0-10	7	2	8	0.4 (0.1-2.0)	0.93
10-20	15	9	7	2.2 (0.8-5.9)	$9.4x10^{-2}$
20-30	22	13	11	2.0 (0.9-4.5)	$6.6x10^{-2}$
>30	**29**	**28**	**4**	**12.0 (4.2-34.3)**	**$5.0x10^{-9}$**

Variants were categorized according to the associated minor allele frequency reported in the ExAC database (ExAC-MAF) and CADD score. [a]In the replication sample, we were not able to account for the overlap when one subject carried multiple variants with the same MAF; [b]For common variants derived form the replication sample, the variant MAF reported in cases and controls was used to estimate the number of case-carriers and control-carriers. [#]Associations that are significant after multiple testing correction are shown in bold ($p < 7.1x10^{-4}$).

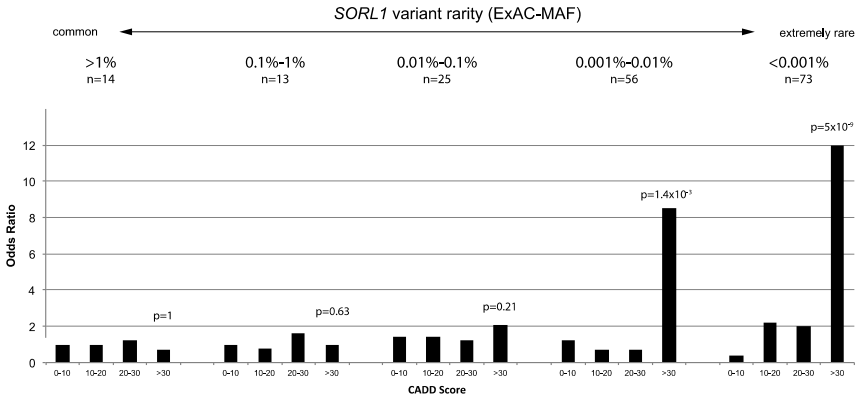


Figure 2: OR by frequency of the variants. Only the rarest variants with the highest CADD scores are associated with increased AD risk. The 181 *SORL1* variants detected in 5,101 AD cases and controls from the combined analysis were first separated by their ExAC-MAF, and then by their CADD values (See also Table 2).

Extremely rare variants (ExAC-MAF <1x10^{-5}) that are moderately damaging (CADD 20-30) and mildly damaging variants (CADD 10-20) are both associated with a 2-fold increased AD risk (OR=2.0; 95%CI 0.9-4.5; p=6.6x10^{-2}) and (OR=2.18, 95%CI 0.8-5.9; p=9x10^{-2}) respectively, whereas variants with CADD 0-10 were not associated with any risk increase (OR=0.4; 95%CI 0.1-2.0; p=0.93) (**Table 2, Figure 2**). In this combined analysis, we found no evidence that moderately pathogenic variants in the VPS10 domain are associated with AD (OR=0.91; 95%CI 0.36-2.29; p=0.66).

Proposed classification of SORL1 variants

SORL1 variants with CADD>30 and ExAC-MAF<1x10^{-4} are associated with a strong increased AD risk (OR=10.9; 95%CI 4.6-25.7; p=1.8x10^{-11}) (**Table 3**), but the subset of protein truncating *SORL1* variants (n=13) occurred exclusively in AD cases, and they were novel to ExAC (OR=inf; 95%CI 5.2-inf; p=2.5x10^{-6}). This suggests that protein truncating *SORL1* variants may be pathogenic (**Table 3,** classification: pathogenic). The subset of non-truncating missense mutations with CADD>30 and ExAC-MAF<1x10^{-4} accounted for a >7-fold increased AD risk (OR=7.1, 95%CI 2.9-17.4, 8.6x10^{-7}), suggesting that variants with these characteristics are strong risk factors for AD (**Table 3,** classification: likely pathogenic). Variants predicted to be mildly-moderately pathogenic (CADD 10-30) with ExAC-MAF<1x10^{-5} were associated with >2-fold

increased AD risk (OR=2.4; 95%CI 1.2-4.6; p=7.7x10^{-3}), suggesting that variants with these characteristics are risk factors for AD and some might be pathogenic (**Table 3**, classification: possibly pathogenic). *SORL1* variants with ExAC-MAF<1x10^{-4}, including those classified pathogenic or likely pathogenic, did not concentrate in specific SORL1 protein domains (**Figure 3**).

When we focus on the more common *SORL1* variants with CADD>30, with ExAC-MAF>10^{-4}, we found that despite their high CADD values, they were not associated with increased AD risk (OR=0.73; 95%CI 0.6-1.0; p=0.99), (**Table 3**, classification: most likely not pathogenic). For *SORL1* variants observed more often in the ExAC database than ExAC-MAF>10^{-5} with CADD values <30, we found no association with AD risk (**Table 3,** classification: likely benign). Likewise, variants with CADD 0-10, regardless of their rarity, were not associated with increase AD risk (**Table 3,** classification: benign**).**

Table 3. Clinical selection criteria of variants

A. Effect on AD risk

Variant selection criteria	**CADD**	**ExAC-MAF**	**N-variants in burden test[a]**	**Cases** With at least one variant **Total 1,895**	**Controls** With at least one variant **Total 3,206**	**OR (95% CI)**	**P value[#] (unadjusted)**
Maximum effect size	>30	<1x10^{-5}	29	28	4	12.0 (4.2-34.3	5.0x10^{-9}
Maximum evidence for effect	>30	<1x10^{-4}	38	38	6	10.9 (4.6-25.9	1.8x10^{-11}

B.Suggested variant subtypes

Variant selection criteria	**CADD**	**ExAC-MAF**	**N-variants in burden test[a]**	**Cases** With at least one variant **Total 1,895**	**Controls** With at least one variant **Total 3,206**	**OR (95% CI)**	**P value[#] (unadjusted)**
Pathogenic	>30 (stop/ frameshift)	<1x10^{-5} *	13	13	0	inf (5.2- inf)	2.5x10^{-6}
Likely pathogenic	>30 (missense)	<1x10^{-4}	25	25	6	7.1 (2.9-17.4)	8.6x10^{-7}
Uncertain significance							
- Possibly pathogenic	10-30	<1x10^{-5}	33	21	15	2.4 (1.2-4.6)	7.7x10^{-3}
- Most likely not pathogenic	>30	>1x10^{-4}	6	76	173	0.73 (0.6-1.0	0.99
Likely benign	10-30	>1x10^{-5}	70	1,895	3,206	1.0	1.0
Benign	0-10	0-1	30	1,895	3,206	1.0	1.0

"Stop/frameshift mutations were all unknown to ExAC. #Associations that are significant after multiple testing correction are shown in bold (p < 7.1×10^{-4}).

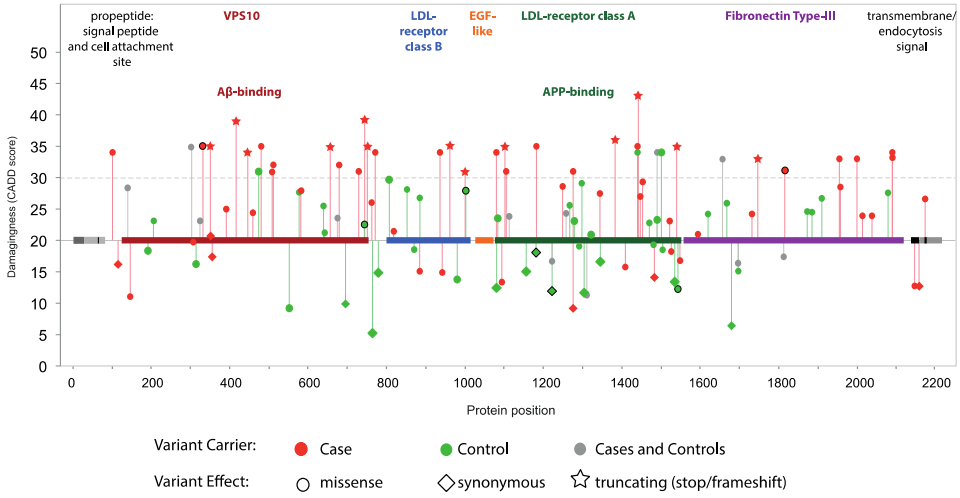


Figure 3. Protein position of *SORL1* variants with MAF <1.10^{-4} in ExAC database. 121 coding variants with ExAC MAF <1.10-4 were detected the combined analysis of 5,101 subjects (1,895 cases and 3,206 controls). Each symbol represents one case carrier (red) or control (green) carrier. Protein domains are depicted on the CADD=20 level, variants with CADD scores between 20-30 are considered 'moderately pathogenic', and variants with CADD scores >30 were considered 'strongly pathogenic'. Markers outlined in black represent variants that were etected in multiple cases or in multiple controls.

Our findings lead us to propose the following five *SORL1* variant subtypes:

Pathogenic: Truncating *SORL1* variants

Likely Pathogenic: *SORL1* variants predicted extremely damaging (CADD>30) and extremely rare (MAF <1.10^{-4} in the publically available ExAC database v.0.3.1[31])

Uncertain significance: — *Possibly pathogenic:* variants predicted mildly to moderately pathogenic (CADD 10-30) which are novel or reported only once in the ExAC database (ExAC-MAF <1.10^{-5}).
— *Most likely not pathogenic:* variants predicted extremely damaging (CADD>30) that are observed more commonly in the ExAC database (ExAC-MAF≥1.10^{-4}).

Likely benign: *SORL1* variants predicted mildly to moderately damaging (CADD 10-30) that are reported more than once in the ExAC database (ExAC-MAF≥1.10^{-5}).

Benign: *SORL1* variants predicted not pathogenic (CADD 0-10) regardless of their rareness.

Application of *SORL1* variant pathogenicity screen

We applied our classification approach to the 17 *SORL1* prioritized variants detected in a family based analysis, Vardarajan *et al.*[4] These variants were enriched in members from 87 Caribbean Hispanic families with late onset AD, compared to 498 age-matched controls. Vardarajan *et al.* identified three truncating deletions unknown to ExAC. Our approach to screen for variant-pathogenicity classified these variants to be 'pathogenic' and indeed, these variants occurred exclusively in affected families. Furthermore, 13 variants are classified 'likely benign', they had CADD scores <30 and occur more than once in ExAC; indeed, these variants were detected both in the affected families and non-affected families. One variant with CADD score 34 and ExAC-MAF >0.01 occurred both in affected families and in unaffected controls; in accordance, our approach classified it to be 'most likely not pathogenic'.

Likewise, we applied our classification strategy to *SORL1* variants reported by Nicolas *et al.*[5] They studied 24 rare variants (sample MAF <0.01) that were predicted deleterious by SIFT and Polyphen in 484 AD cases, mostly with family history of AD, and 498 controls. Of these variants, 15 were novel to the ExAC database: 8 truncating variants and 7 missense variants with CADD >30 that were classified as 'pathogenic' and 'likely pathogenic' variants respectively; indeed, these were seen exclusively in cases. Another case carried a variant with CADD score 35 with ExAC-MAF $6x10^{-5}$ classified to be 'likely pathogenic', indeed, this variant was also detected in 2 cases in the Verheijen sample. Furthermore, Nicolas *et al.* detected a *SORL1* variant unknown to ExAC with CADD score 32, which was located within the VPS10 domain. This 'likely pathogenic', variant was found to disturb the binding of Aβ for lysosomal degradation[12] and Pottier *et al*[3] found that this variant segregated with disease. Nicolas *et al.* also detected a variant with ExAC-MAF 0.003 and CADD score 25.5, which was classified to be 'likely benign': indeed, it occurred equally in cases and controls. Finally, Nicolas et al identified 7 'possibly pathogenic' variants that were novel to ExAC with CADD scores ranging between 26.7-29.4, all occurred in AD cases. The evidence for the association with AD of variants with these characteristics is relatively low, suggesting that further research into the pathogenicity of these variants is necessary.

Discussion

Protein-truncating and rare pathogenic missense variants in the *SORL1* gene associate with AD

It is clear that nonfunctional SORL1 associates with AD, but a comprehensive set of characteristics that defines the associated genetic variants and their impact has been lacking. Therefore, the *"need for pathogenicity assays"* has been raised to aid with the clinical interpretation of *SORL1* variants[6]. Here, we analyzed 181 unique *SORL1* variants detected in a large sample of 1,895 cases and 3,206 controls and we propose that *SORL1* variant- pathogenicity can be classified according to the combination of two independent variant characteristics: the predicted level of variant-damagingness and the level of variant-rareness.

Our findings indicate that stop-gain and frameshift mutations occurred exclusively in cases, suggesting that variants leading to premature disruption of *SORL1* transcription are highly penetrant. This supports previous findings that loss of one copy of *SORL1* (i.e. haploinsufficiency) is causal to AD[6]. Thus far, such high impact on AD is observed only for variants in *PSEN1/2* and *APP*, which are associated with familial AD[25]. Furthermore our findings indicate that variants novel to the ExAC database with a CADD score >30 are associated with a significant 12-fold increased AD risk, which is comparable to the effect of *APOE-ε4* homozygosity[25]. In line with the increased risk we found suggestive evidence that pathogenic *SORL1* variants lead to an earlier age at onset.

Although variants are individually rare, 2% of the AD cases (and <0.2% of the controls) in our analysis carried a *SORL1* variant with these characteristics. By comparison, variants in the classical AD genes *PSEN1*, *PSEN2* and *APP* collectively explain <1% of AD cases[25]. We propose therefore that in clinical practice, rare pathogenic *SORL1* mutations should be considered next to *PSEN1*, *PSEN*2 and *APP*.

Moderately damaging variants in the VPS10 domain might be associated with AD

Pathogenic variants occurred throughout the *SORL1* gene without preference to a specific functional domain. In our discovery analysis, moderately damaging variants in the VPS10 domain were detected only in 4 cases but not in older controls. In contrast, moderately damaging variants in the LDL-receptor A and fibronectin domains occurred only in control subjects. However, we could not confirm these findings in the replication or combined sample, possibly due to the many young control subjects in the replication sample who might still develop disease at a later age. In the future, larger samples will have to clarify whether or not moderately pathogenic variants in the Aβ-binding VPS10 domain might dangerously affect SORL1 function.

Our results indicate that some *SORL1* variants with lower CADD scores may hold some pathogenicity when they are extremely rare, but the effect size is only 2-fold and the evidence for this is not as strong. On the other hand, we found no evidence for pathogenicity for common variants, even variants with CADD scores >30. Risk increases were independent of gender and we detected no evidence for synergy between disrupted *SORL1* function and carrying the *APOE-ε4* genotype.

Five *SORL1* variant subtypes

For the clinical interpretation of *SORL1* variant pathogenicity based on ExAC-MAF and CADD scores, we propose five *SORL1* variant subtypes according to the five-class system of variant pathogenicity supported by the ACMG[32] and ACGS[33]. When we applied our strategy applied to *SORL1* variants reported by the independent studies of Vardarajan *et al*.[4] and Nicholas *et al.*[5] variants were classified according to their occurrence in cases and controls. Even though the classification strategy presented here is based on two large samples, additional research is necessary to determine the exact risk of individual variants. We caution that genetic context might influence variant pathogenicity: for example, one possibly pathogenic *SORL1* variant (CADD score 23.6, ExAC-MAF<1.10^{-5}) was found to segregate with disease and increase AD-risk in a family with several generations of *APOE-ε4* homozygosity[39]. It is likely that classification will be refined as

larger samples with sequencing data become available. Lastly, this classification is based on evidence from populations with European ancestry and should be replicated in populations with other ethnic backgrounds.

Pathogenic *SORL1* variants are rare

We find that truncating SORL1 variants are pathogenic, and that all 24 truncating variants collectively reported across the previous studies by Verheijen *et al*., Nicolas *et al*, Vardarajan *et al* and this present study occurred exclusively in AD cases and were unknown to the ExAC database. Likewise, across these studies, >70% of all likely pathogenic variants (missense variants with CADD score >30) were unknown to ExAC. This suggests that the increased pathogenicity of extremely rare variants may explain part of the *'missing heritability'* that remained undetected in genetic association studies such as GWAS, which test the association of common variants with disease.

In concordance with this, evidence is mounting that especially extremely rare mutations are the major contributors to the development of disease. In a sequencing analysis of 202 drug-targeted genes in 14,002 persons, 74% of the detected mutations occurred only in 1 or 2 subjects, indicating that mutations that associated with disease are abundant, but mostly very rare[40]. Furthermore, Fu et al found that more than half of all SNPs detected in exomes from 6,515 individuals were, in fact, singletons[41]. They found that 86% of detected damaging variants arose very recently in the population, which partly explains the restricted propagation of the variant in the population, i.e. the variant rarity relative to common (old) variants[41].

The rarity of the variants also suggests that natural selection pressure eliminates pathogenic *SORL1* variants from the population. As AD onset occurs well after the reproductive phase, we might expect that variants associated with AD would not be under influence of selection pressure. Therefore it is surprising that pathogenic *SORL1* variants are not propagated in the population. The rarity of harmful *SORL1* variants suggests that *SORL1* function may not be restricted to the maintenance of cells in the brain and that disturbed *SORL1* function might affect reproductive success and/or individual health far before the age of AD onset.

Conclusion

With the increasing availability of whole exome sequencing (WES) in clinical practice, it is possible to detect highly personal exonic variants in *SORL1*. We characterized *SORL1* variants based on variant frequency and damagingness and we suggest five variant subtypes ranging from pathogenic to benign. Our findings suggest that in the clinic, pathogenic *SORL1* variants should be considered in personalized AD risk assessments alongside *APOE, PSEN1, PSEN2, APP*.

References:

1. Alzheimer's Association: 2016 Alzheimer's disease facts and figures. *Alzheimer's & dementia : the journal of the Alzheimer's Association* 2016; **12:** 459-509.
2. Gatz M, Reynolds CA, Fratiglioni L *et al*: Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry* 2006; **63:** 168-174.
3. Pottier C, Hannequin D, Coutant S *et al*: High frequency of potentially pathogenic SORL1 mutations in autosomal dominant early-onset Alzheimer disease. *Molecular psychiatry* 2012; **17:** 875-879.
4. Vardarajan BN, Zhang Y, Lee JH *et al*: Coding mutations in SORL1 and Alzheimer disease. *Annals of neurology* 2015; **77:** 215-227.
5. Nicolas G, Charbonnier C, Wallon D *et al*: SORL1 rare variants: a major risk factor for familial early-onset Alzheimer's disease. *Molecular psychiatry* 2015.
6. Verheijen J, Van den Bossche T, van der Zee J *et al*: A comprehensive study of the genetic impact of rare variants in SORL1 in European early-onset Alzheimer's disease. *Acta neuropathologica* 2016.
7. Andersen OM, Rudolph IM, Willnow TE: Risk factor SORL1: from genetic association to functional validation in Alzheimer's disease. *Acta neuropathologica* 2016; **132:** 653-665.
8. Willnow TE, Andersen OM: Sorting receptor SORLA--a trafficking path to avoid Alzheimer disease. *Journal of cell science* 2013; **126:** 2751-2760.
9. Mehmedbasic A, Christensen SK, Nilsson J *et al*: SorLA complement-type repeat domains protect the amyloid precursor protein against processing. *The Journal of biological chemistry* 2015; **290:** 3359-3376.
10. Andersen OM, Schmidt V, Spoelgen R *et al*: Molecular dissection of the interaction between amyloid precursor protein and its neuronal trafficking receptor SorLA/LR11. *Biochemistry* 2006; **45:** 2618-2628.
11. Hermey G: The Vps10p-domain receptor family. *Cellular and molecular life sciences : CMLS* 2009; **66:** 2677-2689.
12. Caglayan S, Takagi-Niidome S, Liao F *et al*: Lysosomal sorting of amyloid-beta by the SORLA receptor is impaired by a familial Alzheimer's disease mutation. *Science translational medicine* 2014; **6:** 223ra220.
13. Scherzer CR, Offe K, Gearing M *et al*: Loss of apolipoprotein E receptor LR11 in Alzheimer disease. *Archives of neurology* 2004; **61:** 1200-1205.
14. Andersen OM, Reiche J, Schmidt V *et al*: Neuronal sorting protein-related receptor sorLA/LR11 regulates processing of the amyloid precursor protein. *Proceedings of the National Academy of Sciences of the United States of America* 2005; **102:** 13461-13466.
15. Offe K, Dodson SE, Shoemaker JT *et al*: The lipoprotein receptor LR11 regulates amyloid beta production and amyloid precursor protein traffic in endosomal compartments. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 2006; **26:** 1596-1603.

16. Spoelgen R, von Arnim CA, Thomas AV *et al*: Interaction of the cytosolic domains of sorLA/LR11 with the amyloid precursor protein (APP) and beta-secretase beta-site APP-cleaving enzyme. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 2006; **26:** 418-428.
17. Rogaeva E, Meng Y, Lee JH *et al*: The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. *Nature genetics* 2007; **39:** 168-177.
18. Bettens K, Brouwers N, Engelborghs S, De Deyn PP, Van Broeckhoven C, Sleegers K: SORL1 is genetically associated with increased risk for late-onset Alzheimer disease in the Belgian population. *Human mutation* 2008; **29:** 769-770.
19. Caglayan S, Bauerfeind A, Schmidt V *et al*: Identification of Alzheimer disease risk genotype that predicts efficiency of SORL1 expression in the brain. *Archives of neurology* 2012; **69:** 373-379.
20. Tan EK, Lee J, Chen CP, Teo YY, Zhao Y, Lee WL: SORL1 haplotypes modulate risk of Alzheimer's disease in Chinese. *Neurobiology of aging* 2009; **30:** 1048-1051.
21. Kimura R, Yamamoto M, Morihara T *et al*: SORL1 is genetically associated with Alzheimer disease in a Japanese population. *Neuroscience letters* 2009; **461:** 177-180.
22. Liu F, Ikram MA, Janssens AC *et al*: A study of the SORL1 gene in Alzheimer's disease and cognitive function. *Journal of Alzheimer's disease : JAD* 2009; **18:** 51-64.
23. Reitz C, Cheng R, Rogaeva E *et al*: Meta-analysis of the association between variants in SORL1 and Alzheimer disease. *Archives of neurology* 2011; **68:** 99-106.
24. Lambert J, Ibrahim-Verbaas C, Harold D *et al*: Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature genetics* 2013; **45:** 1452-1458.
25. Van Cauwenberghe C, Van Broeckhoven C, Sleegers K: The genetic landscape of Alzheimer disease: clinical implications and perspectives. *Genetics in medicine : official journal of the American College of Medical Genetics* 2016; **18:** 421-430.
26. Corder EH, Saunders AM, Strittmatter WJ *et al*: Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 1993; **261:** 921-923.
27. Goate A, Chartier-Harlin MC, Mullan M *et al*: Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature* 1991; **349:** 704-706.
28. Schellenberg GD, Bird TD, Wijsman EM *et al*: Genetic linkage evidence for a familial Alzheimer's disease locus on chromosome 14. *Science* 1992; **258:** 668-671.
29. Levy-Lahad E, Wasco W, Poorkaj P *et al*: Candidate gene for the chromosome 1 familial Alzheimer's disease locus. *Science* 1995; **269:** 973-977.
30. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J: A general framework for estimating the relative pathogenicity of human genetic variants. *Nature genetics* 2014; **46:** 310-315.
31. Lek M, Karczewski K, Minikel E *et al*: Analysis of protein-coding genetic variation in 60,706 humans. *bioRxiv* 2015.
32. Richards S, Aziz N, Bale S *et al*: Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine* 2015; **17:** 405-423.
33. Wallis Y, Payne S, McAnulty C *et al*: Practice Guidelines for the Evaluation of Pathogenicity and the Reporting of Sequence Variants in Clinical Molecular Genetics: Association for Clinical Genetic Science, 2013.
34. Hofman A, Brusselle GG, Darwish Murad S *et al*: The Rotterdam Study: 2016 objectives and design update. *European journal of epidemiology* 2015; **30:** 661-708.
35. van der Flier WM, Pijnenburg YA, Prins N *et al*: Optimizing patient care and research: the Amsterdam Dementia Cohort. *Journal of Alzheimer's disease : JAD* 2014; **41:** 313-327.
36. McLaren W, Gil L, Hunt SE *et al*: The Ensembl Variant Effect Predictor. *Genome biology* 2016; **17:** 122.
37. Lee S, Abecasis GR, Boehnke M, Lin X: Rare-variant association analysis: study designs and statistical tests. *American journal of human genetics* 2014; **95:** 5-23.

38. R Core Team: *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2013.
39. Louwersheimer E, Cohn-Hokke PE, Pijnenburg YA *et al*: Rare Genetic Variant in SORL1 May Increase Penetrance of Alzheimer's Disease in a Family with Several Generations of APOE-varepsilon4 Homozygosity. *Journal of Alzheimer's disease : JAD* 2017; **56:** 63-74.
40. Nelson MR, Wegmann D, Ehm MG *et al*: An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* 2012; **337:** 100-104.
41. Fu W, O'Connor TD, Jun G *et al*: Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* 2013; **493:** 216-220.

Supplementary material

Supplementary Methods and Supplementary Tables can be accessed by scanning the following code or accessing the journals' website.

Chapter 2.3

Parental family history of dementia in relation to subclinical brain disease and dementia risk

Frank J. Wolters, Sven J. van der Lee, Peter J. Koudstaal, Cornelia M. van Duijn, Albert Hofman, M. Kamran Ikram, Meike W. Vernooij, M. Arfan Ikram.

Neurology. 2017 Apr 25; 88(17): 1642-1649.

Abstract

Objective: To determine the association of parental family history with risk of dementia, by age at onset and sex of affected parent, in a population-based cohort.

Methods: From 2000-2002, we assessed parental history of dementia in non-demented participants of the Rotterdam Study. We investigated associations of parental history with risk of dementia until 2015, adjusting for demographics, cardiovascular risk factors, and known genetic risk variants. Furthermore, we determined the association between parental history and markers of neurodegeneration and vascular disease on MRI.

Results: Of 2,087 participants (mean age 64 years, 55% female), 407 (19.6%) reported a history of dementia in either parent (mean age at diagnosis: 79 years). During a mean follow-up of 12.2 years, 142 participants developed dementia. Parental history was associated with risk of dementia independent of known genetic risk factors (hazard ratio, 95% confidence interval: 1.67, 1.12-2.48), in particular when parents were diagnosed at younger age (HR, 95% CI <80 years: 2.58, 95% CI 1.61-4.15 versus ≥80 years: 1.01, 95% CI 0.58-1.77). Accordingly, age at diagnosis in probands was highly correlated with age at diagnosis in their parents <80 years (r = 0.57, p = 0.001), but not thereafter (r = 0.17, p = 0.55). Among 1161 non-demented participants with brain MRI, parental history related to lower cerebral perfusion, and higher burden of white matter lesions and microbleeds. Dementia risk and MRI markers were similar for paternal versus maternal history.

Conclusion: Parental history of dementia increases risk of dementia, primarily when age at parental diagnosis is <80 years. Unexplained heredity may in part be attributed to cerebral hypoperfusion and small-vessel disease. We found no evidence of preferential maternal compared to paternal transmission.

Introduction

Family history of dementia is an important risk factor for dementia and Alzheimer's disease (AD), independent of known genetic risk factors for AD.[1] Yet, its applicability for clinical risk stratification and research about underlying mechanisms largely depends on the magnitude of the associated risk. For other diseases, such as myocardial infarction, the strength of associations between family history and risk of disease diminishes with increasing age at which family members are affected.[2] Similarly, with regard to dementia, the effect of its major genetic risk factor (*APOE*) as well as the heritability of brain morphology decline with age,[3,4] but prospective studies that quantify associations of family history with risk of dementia by age at onset of affected relatives are lacking.

In search of potential mechanisms that account for the unexplained heredity of dementia, several studies have recently turned to imaging markers of neurodegeneration. These generally explorative studies found that in healthy adults, a family history of dementia is associated with structural brain changes,[5-8] and various other markers of neurodegeneration, including white matter integrity,[9,10] resting state connectivity,[11] glucose metabolism,[12-15] hypoperfusion,[16] and β-amyloid and tau.[12,14,15] Interestingly, several of these studies have suggested a stronger association with maternal compared to paternal family history,[6,12-16] but this was not confirmed in two other reports.[5,7] Sex-specific transmission is plausible in view of findings for ischaemic stroke and myocardial infarction,[17,18] and may relate to chromosome X mutations, mitochondrial DNA, or imprinting.[19] However, no published studies have assessed risk of developing dementia by paternal and maternal history.

We therefore investigated the association of family history, by age at onset and sex of affected parent, with risk of dementia in the general population, and explored underlying imaging abnormalities on structural MRI.

Methods

This study is embedded within the Rotterdam Study, a large ongoing population-based cohort study in the Netherlands among inhabitants aged ≥55 years from the Ommoord area in Rotterdam. For the current study, we included the second wave of invitees, recruited between 2000 and 2002. The Rotterdam Study methods have been described in detail previously.[20] In brief, participants were interviewed at home and subsequently examined at the research center from January 2000 to November 2002, which was used as baseline for this study. Family history of dementia was assessed during baseline interview. Of 3011 eligible participants, 2247 (74.6%) underwent home interview. From August 2005 until July 2013, these participants were randomly invited for magnetic resonance imaging (MRI).

Standard Protocol Approvals, Registrations, and Patient Consents

The Rotterdam Study has been approved by the medical ethics committee according to the Population Study Act Rotterdam Study, executed by the Ministry of Health, Welfare and Sports of the Netherlands. Written informed consent was obtained from all participants.

Assessment of family history

Participants were questioned by trained interviewers about parental family history of dementia, using a structured questionnaire (**Supplementary Table 1**). If this question was answered positively, they were further asked about specific paternal and maternal history of dementia, including age at diagnosis. Vital status of parents and age of death were also recorded.

Genotyping and calculation of genetic risk scores

DNA was extracted from blood samples drawn by venipuncture at baseline. *APOE* genotype was determined with a bi-allelic TaqMan assay (rs7412 and rs429358) in 97.9% of participants. The majority of samples (81.1%) were further genotyped using the Illumina 610K and 660K chip, and imputed to the Haplotype Reference Consortium reference panel (v1.0) with Minimac 3. We included 23 genetic variants that showed genome wide significant evidence of association with Alzheimer's disease to calculate a

weighted genetic risk score (**Supplementary Table 2**). The genetic risk score was calculated as the sum of the products of SNP dosages of the 23 genetic variants (excluding *APOE*) and their respective reported effect estimates. All 23 variants selected for the calculation of the genetic risk score were well imputed (imputation score R2>0.3, median=0.99) (**Supplementary Table 2**).

MRI scan protocol and image processing

Brain MRI was done on a 1.5-T scanner (General Electric Healthcare, Milwaukee, WI, USA), with use of an 8-channel head coil.[21] We acquired a high-resolution axial T1-weighted sequence, proton-density-weighted (PD) sequence, a fluid attenuated inversion recovery (FLAIR) sequence, and a T2*-weighted gradient echo sequence, as described previously.[21] Quantification of cerebrospinal fluid, total parenchymal volume, and volume of white matter hyperintensities was done using an automated tissue segmentation method.[22] All segmentations were inspected and manually corrected if so required. All scans were appraised by trained research physicians, blinded to clinical data, for the presence of cerebral microbleeds (i.e. small round-ovoid areas of signal loss on T2*-weighted images) and lacunar infarcts (i.e. focal lesions ≥3 and <15mm in size with similar signal intensity as cerebrospinal fluid on all sequences and, when located supratentorial, with a hyperintense rim on FLAIR). Cerebral blood flow was determined from 2D phase-contrast images with custom software (Cinetool version 4; General Electric Healthcare).[23] We calculated total brain perfusion (in mL/min per 100 mL brain tissue) by dividing total blood flow (mL/min) by each individual's brain volume (mL) and multiplying the result by 100.

Dementia screening and surveillance

Participants were screened for dementia at baseline and subsequent centre visits using the Mini-Mental State Examination (MMSE) and the Geriatric Mental Schedule (GMS) organic level.[24] Those with MMSE<26 or GMS>0 underwent further investigation and informant interview including the Cambridge Examination for Mental Disorders of the Elderly (CAMDEX). At each centre visit, all participants also underwent routine cognitive assessment, including a verbal fluency test (animal categories), 15-word learning test, letter-digit substitution task, Stroop test, and Purdue pegboard task.

Additionally, the entire cohort was continuously under surveillance for dementia through electronic linkage of the study centre with medical records from general practitioners and the regional institute for outpatient mental healthcare. Available neuroimaging data were reviewed when required for diagnosis of dementia subtype. A consensus panel headed by a consultant neurologist established the final diagnosis according to standard criteria for dementia (DSM-III-R), and Alzheimer's disease (NINCDS-ADRDA). Follow-up until 1st January 2015 was virtually complete (96.8% of potential person years). Within this period, participants were censored at date of dementia diagnosis, death, loss to follow-up, or 1st January 2015, whichever came first.

Other measurements

We assessed educational attainment (lower, further, or higher education), smoking status (never, former, or current), and use of antihypertensive or lipid-lowering medication at baseline by interview. Lipid levels were measured from fasting serum at baseline. Hyperlipidaemia was defined as LDL cholesterol >4.9 mmol/L (190 mg/dL), or use of lipid-lowering medication. Blood pressure was measured twice on the right arm with a random-zero sphygmomanometer. Hypertension was defined as elevated systolic or diastolic blood pressure (>140/90 mmHg) or use of antihypertensive medication. Body mass index was computed from measurements of height and weight (kg/m2). A diagnosis of diabetes mellitus was based on the use of blood glucose-lowering medication or a fasting serum glucose ≥7.0 mmol/L.

Analysis

Analyses included all non-demented participants who provided data on family history at baseline. Missing data on non-genetic covariates (≤1.3%) were imputed using 5-fold multiple imputation, based on determinant, outcome and included covariates. Distribution of covariates was similar in the imputed and non-imputed dataset. We determined the association between parental family history of dementia and risk of dementia and Alzheimer's disease, using Cox proportional hazard models, and stratified results by paternal and maternal family history (or both), sex of proband, and mean age of proband at time of interview. We verified that choice of x-axis (age versus follow-up time) did not affect the results. Subsequently, we determined risk of dementia and

Alzheimer's disease per decade increase in age at onset in parents. To account for potential misclassification of determinant (i.e. parents deceased at young age, or developing dementia after interview) or outcome (i.e. participants who did not yet reach old age at end of follow-up), we performed sensitivity analyses excluding family history of parents who died prematurely (<65 years), excluding participants <70 years at baseline, and excluding non-demented participants censored before age 80.

Next, we compared characteristics of the subset of participants with MRI to those without MRI using age- and sex-adjusted analysis of covariance (ANCOVA) for continuous and logistic regression for dichotomous variables. We then determined the association between family history (overall and stratified by sex of affected parent and age at parental diagnosis) and (standardised values of) total brain parenchymal volume, hippocampal volume, cerebral perfusion, volume of white matter hyperintensities, presence of lacunar infarcts (yes vs. no), and cerebral microbleed count (classified as 0, 1, or ≥2). For continuous outcome variables these analyses were performed using linear regression; for categorical outcomes we used logistic and multinomial regression. Age at parental diagnosis was hereby stratified at 80 years, as this approximates the mean age at diagnosis in the general population (illustrated by a mean age of 80.7 years at diagnosis for our participants, and 78.5 years at time of parental diagnosis).

All analyses were adjusted for age (at time of interview or MRI scan where appropriate) and sex, and additionally in a second model for level of education, smoking habits, history of hypertension, hyperlipidaemia, diabetes mellitus, and body mass index. To account for known genetic risk, in a third and fourth model we additionally adjusted for *APOE* genotype, and *APOE* genotype plus the genetic risk score for Alzheimer's disease, respectively. All imaging analyses were furthermore adjusted for total intracranial volume and interval between interview and MRI scan.

Analyses were done using IBM SPSS Statistics version 23.0 (IBM Corp, Armonk, NY, USA). Alpha-level was set at 0.05.

Results

Of 2,233 eligible participants, 2,078 (93.1%) provided data on parental family history. Family history was positive for dementia in 407 (19.6%) persons. Mean age at diagnosis in affected parents was 78.5 years. Baseline characteristics of participants are presented in **Table 1**. During a mean follow-up of 12.2 years, 142 participants developed dementia, of whom 105 (73.9%) had Alzheimer's disease. Mean age at diagnosis in participants was 80.7 years.

Table 1. Baseline characteristics.

	All participants (n=2,078)	With MRI (n=1,150)	Without MRI (n=928)
Age	64.1 ±7.5	62.0 ±5.5	66.7 ±8.8
Female sex	1,142 (55.0)	614 (53.4)	528 (56.9)
Level of education			
Lower	1,081 (52.7)	552 (48.9)	529 (57.3)
Further	603 (29.4)	349 (30.9)	254 (27.5)
Higher	369 (18.0)	228 (20.2)	141 (15.3)
Smoking history			
Former	1,047 (50.6)	587 (51.4)	460 (49.6)
Current	388 (18.7)	204 (17.8)	184 (19.8)
Hypertension	1,235 (59.5)	599 (52.1)	636 (68.5)
Diabetes mellitus	268 (12.9)	110 (9.6)	158 (17.0)
Body-mass index (kg/m^2)	27.2 ±4.0	26.9 ±3.6	27.5 ±4.5
Hyperlipidaemia	611 (29.4)	330 (28.7)	281 (30.3)
APOE genotype			
ε3/ε3	1,174 (57.7)	663 (58.7)	511 (56.5)
ε2/ε2 or ε2/ε3	292 (14.4)	161 (14.3)	131 (14.5)
ε2/ε4, ε3/ε4, or ε4/ε4	568 (27.9)	305 (27.0)	263 (29.1)
Genetic risk score for Alzheimer's disease	-0.10 ±0.32	-0.09 ±0.32	-0.10 ±0.34
Family history of dementia	407 (19.6)	229 (19.9)	178 (19.2)
Paternal	116 (5.6)	65 (5.7)	51 (5.5)
Maternal	273 (13.1)	156 (13.6)	117 (12.6)
Both	18 (0.9)	8 (0.7)	10 (1.1)
Age at diagnosis in affected parent	78.5 ±8.3	79.2 ±7.5	77.5 ±9.1

Data are presented as frequency (%) for categorical, and mean±standard deviation for continuous variables.

Parental family history of dementia was associated with all-cause dementia and in particular Alzheimer's disease, which was only partly explained by known genetic variants (**Table 2**). These associations were similar for paternal and maternal family history of dementia (**Table 2**), and did not vary significantly by sex of proband (HR 1.82, 95% CI 0.99-3.38 in men vs. 1.43, 95% CI 0.84-2.44 in women; p-value for interaction = 0.44). Results were unaffected by excluding participants whose parents died at young age (before the age of 65: HR 1.95, 1.00-3.82), and grossly similar for participants aged below and above the mean age of 64 years at time of interview (HR 2.45, 95% CI 1.69-3.56 vs. 1.93, 95% CI 1.28-2.93; p-value for interaction = 0.36).

Associations between parental history of dementia and risk of dementia in probands were dependent on age at diagnosis in the parent (**Table 3**). Risk estimates gradually declined per advanced decade of age at diagnosis in parents, such that risk was highest when parents were diagnosed before age 80 (HR, 95% CI, before: 95% CI 2.58, 1.61-4.15 vs. after: 1.01, 95% CI 0.58-1.77). This trend was similar for Alzheimer's disease only (**Supplementary Table 3**). Accordingly, age at diagnosis in probands was highly correlated with age at diagnosis in their parents when parents were diagnosed before age 80 (r = 0.57, p = 0.001), but not thereafter (r = 0.17, p = 0.55). In sensitivity analyses to minimize potential information bias, age trends were similar when excluding non-demented participants censored before they reached age 80, or excluding participants <70 years at baseline (data not shown). Consistent with overall estimates in **Table 2**, known genetic risk factors accounted for only part of the large increased risk with parents affected before age 80.

Of all 2,078 participants who provided family history, 1,150 (55.3%) underwent MRI, a median 5.6 years (IQR 5.1-10.6) after baseline interview. Compared to non-participants, MRI participants were generally younger, and had a more favorable cardiovascular risk profile (**Table 1**). Thirty-four participants who developed dementia between interview and MRI were excluded. Lacunar infarcts were seen in 95 (8.5%) individuals, and at least one cerebral microbleed in 251 (22.5%) individuals (1 in 144, and ≥2 in 107 individuals).

Table 2. Family history and risk of dementia.

	Model I		Model II		Model III		Model IV	
	All dementi (n/N=142/2,078)	Alzheimer's (n/N=105/2,078)	All demen (n/N=142/2,078)	Alzheimer's (n/N=105/2,078)	All dementia (n/N=138/2,034)	Alzheimer's (n/N=101/2,034)	All dementia (n/N=122/1,674)	Alzheimer's (n/N=88/1,674)
	HR, 95% CI	HR, 95% CI	HR, 95% CI	HR, 95% CI	HR, 95% CI	HR, 95% CI	HR, 95% CI	HR, 95% CI
Family history of dementia	2.00, 1.40-2.85	2.37, 1.58-3.54	2.00, 1.40-2.86	2.37, 1.58-3.57	1.82, 1.26-2.63	2.12, 1.39-3.24	1.67, 1.12-2.48	2.01, 1.27-3.18
Paternal (n=116)	2.68, 1.58-4.55	3.19, 1.76-5.79	2.56, 1.51-4.35	3.00, 1.65-5.45	2.24, 1.30-3.88	2.62, 1.41-4.86	2.35, 1.31-4.22	2.68, 1.36-5.30
Maternal (n=273)	1.74, 1.13-2.68	2.02, 1.24-3.29	1.76, 1.14-2.72	2.07, 1.26-3.40	1.67, 1.07-2.62	1.94, 1.16-3.24	1.44, 0.89-2.34	1.75, 1.01-3.06
Both parents (n=18)	1.91, 0.47-7.79	2.78, 0.67-11.45	1.94, 0.46-8.14	2.73, 0.63-11.84	1.30, 0.30-5.64	1.62, 0.35-7.54	1.27, 0.30-5.48	1.81, 0.39-8.31

Table 3. Family history of dementia and risk of dementia by age at diagnosis in parents.

		Model I	Model II	Model III	Model IV
	n/N	HR, 95% CI	HR, 95% CI	HR, 95% CI	HR, 95% CI
No parental family history	98/1,573	REFERENCE	REFERENCE	REFERENCE	REFERENCE
Age at diagnosis in parent*					
<70	6/48	3.90, 1.69-8.98	3.93, 1.73-8.92	2.86, 1.18-6.94	2.71, 1.08-6.82
70-79	22/122	3.11, 1.95-4.94	3.09, 1.93-4.96	2.68, 1.65-4.36	2.54, 1.51-4.29
80-89	13/160	1.34, 0.75-2.39	1.35, 0.75-2.41	1.28, 0.73-2.24	0.97, 0.50-1.89
≥90	2/26	1.09, 0.27-4.41	1.08, 0.29-4.06	1.26, 0.31-5.05	1.25, 0.30-5.19
p-value for trend		<0.0001	<0.0001	<0.0001	0.001

Footnotes Table 2 and 3:

HR=hazard ratio; CI=confidence interval; n=number of dementia cases; N=total sample size.

Model I: adjusted for age and sex

Model II: model I with additional adjustment for level of education, hypertension, diabetes, hyperlipidemia, smoking, and body-mass index.

Model III: model II with additional adjustment for *APOE* genotype

Model IV: model III with additional adjustment for a genetic risk score for Alzheimer's disease

* Youngest age if both parents suffered from dementia

Overall, family history of dementia was not associated with total parenchymal volume or hippocampal volume, or with markers of small-vessel disease. However, after stratification for family history by age at parental diagnosis, we found that participants whose parents were affected at younger age had a larger burden of white matter lesions and cerebral microbleeds (**Figure 1**; for a full table see **Supplementary Table 4**). In addition, those with positive family history had lower cerebral blood flow regardless of age at parental onset (**Figure 1**). Apart from smaller hippocampal volume with paternal family history, results again were similar for paternal and maternal family history (**Figure 2**, for a full table see **Supplementary Table 5**).

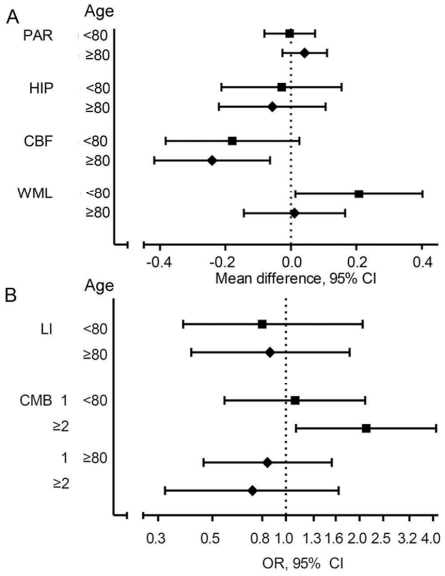


Figure 1. Family history of dementia by age of parental diagnosis in relation to brain MRI measures.

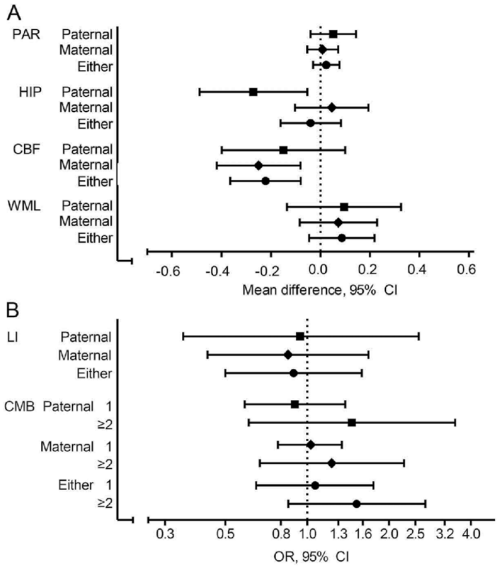


Figure 2. Family history of dementia by affected parent in relation to brain MRI measures.

Legend of Figure 1 and 2: (A) Effect estimates for cerebral microbleeds depict the odds ratio (OR) per age category of having 1 or ≥2 microbleeds with compared to without a positive family history. (B) Model adjusted for age, sex, total intracranial volume, interval between interview and MRI, level of education, hypertension, diabetes mellitus, hyperlipidemia, smoking, body mass index, and APOE genotype. CBF = cerebral blood flow; CI = confidence interval; CMB = cerebral microbleeds; HIP = hippocampal volume; LI = lacunar infarcts; PAR 5 parenchymal tissue volume; WML = white matter lesions.

Discussion

In this prospective population-based study we found an increased risk of dementia with positive family history that is strongly dependent on parental age at diagnosis, but does not differ by paternal or maternal predisposition. Known genetic risk factors accounted for a relatively small share of parental risk. Remaining risk may in part be explained by observed associations of family history with cerebral hypoperfusion and increased burden of small-vessel disease in non-demented participants.

The excess risk of dementia with positive family history in our study is comparable with estimates from prior case-control studies.[25,26] The lack of attenuation after controlling for demographic and lifestyle factors supports family history as a measure of heredity rather than a marker of shared environmental factors. Moreover, known genetic risk factors explained only part of the association in our study, highlighting the important role of unidentified genetic factors involved in the aetiology of dementia.[27] Remaining risk may be due to unidentified epigenetic signatures, low-risk common variants, or high-risk rare variants like *ABCA7* and *SORL1*, but until these are identified, our findings support obtaining family history over genome testing (only).[28] The vast majority of familial excess risk was accounted for by parents diagnosed before age 80, in accordance with estimates modelled in a prior study.[29]

Age at diagnosis correlated well among parents and probands in this group, with correlations similar to those among relatives with early-onset Alzheimer's disease.[30] Thus taking family history is likely much more informative when enquiring about parental onset of dementia before age 80, rather than dementia at any age. This may guide clinical practice, and benefit risk stratification for preventive strategies and selection of participants for research purposes. The increased risk of dementia with parental family history was paralleled by lower cerebral perfusion, and an increased burden of cerebral white matter hyperintensities and microbleeds when parents were affected at younger age. Although one other study did not find an association of white matter hyperintensities with family history,[9] loss of white matter integrity has been associated with family history in two smaller studies.[10,11]

Moreover, cerebral hypoperfusion was associated with family history of dementia in one study,[16] and can also relate to reduced glucose metabolism reported with positive family history previously.[12-15] As hypoperfusion,[31,32] small-vessel disease,[33] and cerebral microbleeds[34] all carry an increased risk of dementia, these may reflect early pathophysiological changes in the brain of those predisposed for developing dementia. Subclinical changes in the brain occur up till decades before onset of dementia symptoms, and neuronal injury is thought to occur years before marked cerebral atrophy is seen on MRI.[35] This might explain why we did not observe an overall association between family history of dementia and total brain volume. Similarly, two other studies reported differences in white matter integrity, as well as amyloid-β42 and tau in cerebrospinal fluid, in the absence of volumetric brain differences.[10,36] As expected, the majority of dementia diagnoses in our study were of the Alzheimer subtype. Yet, these clinical diagnoses may partly reflect other pathology. In fact, mixed pathology is increasingly seen with dementia at higher age, and the risk conferred by a positive family history therefore likely reflects various etiologies, of which we identified perfusion and small vessel disease as contributors.

Although several smaller studies have reported particular or exclusive associations of maternal compared to paternal family history of dementia with biomarkers of neurodegeneration,[6,12-16] other studies did not find such a difference.[5,7] In this population-based study, we did not find evidence of particular maternal transmission with either risk of dementia or imaging biomarkers. Of note, the majority of prior studies did not control for the effects of *APOE*, or even preferentially selected *APOE* ε4 carriers. As *APOE* may have a more profound effect on risk of dementia in women,[3] this might account for part of the associations previously found with maternal family history.

Among the major strengths of our study are its population-based setting with detailed structured questionnaires, meticulous follow-up for dementia, and large sample of participants undergoing MRI. Yet, several limitations need to be discussed. First, albeit structured, interview questions remain susceptible to information bias, in particular

regarding quantitative information such as age at diagnosis. Second, not all of our participants underwent MRI investigation and we cannot rule out selection bias with regard to the imaging analyses. Participants with MRI were generally younger with a favorable cardiovascular risk profile, but as they reported positive family history equally often as participants without MRI, this is unlikely to have affected relative risks. Third, despite similar results in sensitivity analyses, dementia onset after administrative censoring date in younger participants might have caused information bias. Fourth, part of the observed effect of family history may be attributable to identified high-risk rare genetic variants, such as *ABCA7* and *SORL1*. We had no exome sequencing data available, but given the very low prevalence of yet identified variants these are unlikely to explain a large part of the observed effect. Fifth, although we adjusted for many risk factors that probands may share with their parents, some residual confounding with regard to socio-economic status may exist. Finally, the vast majority of participants in our study are of Caucasian descent, potentially limiting generalizability to other ethnicities.

Parental history of dementia is associated with increased risk of dementia, primarily when age at parental diagnosis is below 80 years. This risk increase is only partly explained by known genetic risk factors, and may be further mediated by cerebral hypoperfusion and small-vessel disease. Our findings do not support a preferential risk with maternal compared to paternal family history.

References:

1. Sleegers K, Bettens K, De Roeck A, et al. A 22-single nucleotide polymorphism Alzheimer's disease risk score correlates with family history, onset age, and cerebrospinal fluid Aβ42. *Alzheimer's & dementia : the journal of the Alzheimer's Association* 2015; **11**(12): 1452-60.
2. Stone NJ, Robinson JG, Lichtenstein AH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines.; 2014. p. 2889-934.
3. Farrer LA, Cupples LA, Haines JL, et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *JAMA : the journal of the American Medical Association* 1997; **278**(16): 1349-56.
4. Batouli SAH, Trollor JN, Wen W, Sachdev PS. The heritability of volumes of brain structures and its relationship to age: a review of twin and family studies. *Ageing research reviews* 2014; **13**: 1-9.

5. Ten Kate M, Sanz-Arigita EJ, Tijms BM, et al. Impact of APOE-ε4 and family history of dementia on gray matter atrophy in cognitively healthy middle-aged adults. *Neurobiology of aging* 2016; **38**: 14-20.
6. Honea RA, Swerdlow RH, Vidoni ED, Burns JM. Progressive regional atrophy in normal adults with a maternal history of Alzheimer disease. *Neurology* 2011; **76**(9): 822-9.
7. Okonkwo OC, Xu G, Dowling NM, et al. Family history of Alzheimer disease predicts hippocampal atrophy in healthy middle-aged adults. *Neurology* 2012; **78**(22): 1769-76.
8. Debette S, Wolf PA, Beiser A, et al. Association of parental dementia with cognitive and brain MRI measures in middle-aged adults. *Neurology* 2009; **73**(24): 2071-8.
9. Bendlin BB, Ries ML, Canu E, et al. White matter is altered with parental family history of Alzheimer's disease. *Alzheimer's & dementia : the journal of the Alzheimer's Association* 2010; **6**(5): 394-403.
10. Gold BT, Powell DK, Andersen AH, Smith CD. Alterations in multiple measures of white matter integrity in normal women at high risk for Alzheimer's disease. *Neuroimage* 2010; **52**(4): 1487-94.
11. Wang L, Roe CM, Snyder AZ, et al. Alzheimer disease family history impacts resting state functional connectivity. *Annals of neurology* 2012; **72**(4): 571-7.
12. Mosconi L, Mistur R, Switalski R, et al. Declining brain glucose metabolism in normal individuals with a maternal history of Alzheimer disease. *Neurology* 2009; **72**(6): 513-20.
13. Honea RA, Vidoni ED, Swerdlow RH, Burns JM, Initiative AasDN. Maternal family history is associated with Alzheimer's disease biomarkers. *Journal of Alzheimer's disease : JAD* 2012; **31**(3): 659-68.
14. Mosconi L, Rinne JO, Tsui WH, et al. Amyloid and metabolic positron emission tomography imaging of cognitively normal adults with Alzheimer's parents. *Neurobiology of aging* 2013; **34**(1): 22-34.
15. Maye JE, Betensky RA, Gidicsin CM, et al. Maternal dementia age at onset in relation to amyloid burden in non-demented elderly offspring. *Neurobiology of aging* 2016; **40**: 61-7.
16. Okonkwo OC, Xu G, Oh JM, et al. Cerebral blood flow is diminished in asymptomatic middle-aged adults with maternal history of Alzheimer's disease. *Cerebral cortex (New York, NY : 1991)* 2014; **24**(4): 978-88.
17. Touzé E, Rothwell PM. Sex differences in heritability of ischemic stroke: a systematic review and meta-analysis. *Stroke; a journal of cerebral circulation* 2008; **39**(1): 16-23.
18. Banerjee A, Silver LE, Heneghan C, et al. Sex-specific familial clustering of myocardial infarction in patients with acute coronary syndromes. *Circulation Cardiovascular genetics* 2009; **2**(2): 98-105.
19. Mosconi L, Berti V, Swerdlow RH, Pupi A, Duara R, de Leon M. Maternal transmission of Alzheimer's disease: prodromal metabolic phenotype and the search for genes. *Human genomics* 2010; **4**(3): 170-93.
20. Hofman A, Brusselle GGO, Darwish Murad S, et al. The Rotterdam Study: 2016 objectives and design update. *European journal of epidemiology* 2015; **30**(8): 661-708.
21. Ikram MA, van der Lugt A, Niessen WJ, et al. The Rotterdam Scan Study: design update 2016 and main findings. *European journal of epidemiology* 2015; **30**(12): 1299-315.
22. Vrooman HA, Cocosco CA, van der Lijn F, et al. Multi-spectral brain tissue segmentation using automatically trained k-Nearest-Neighbor classification. *Neuroimage* 2007; **37**(1): 71-81.
23. Vernooij MW, van der Lugt A, Ikram MA, et al. Total cerebral blood flow and total brain perfusion in the general population: the Rotterdam Scan Study. *Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism* 2008; **28**(2): 412-9.
24. de Bruijn RFAG, Bos MJ, Portegies MLP, et al. The potential for prevention of dementia across two decades: the prospective, population-based Rotterdam Study. *BMC medicine* 2015; **13**: 132.

25. Green RC, Cupples LA, Go R, et al. Risk of dementia among white and African American relatives of patients with Alzheimer disease. *JAMA : the journal of the American Medical Association* 2002; **287**(3): 329-36.
26. Fratiglioni L, Ahlbom A, Viitanen M, Winblad B. Risk factors for late-onset Alzheimer's disease: a population-based, case-control study. *Annals of neurology* 1993; **33**(3): 258-66.
27. Lambert JC, Ibrahim-Verbaas CA, Harold D, 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. Aiyar L, Shuman C, Hayeems R, et al. Risk estimates for complex disorders: comparing personal genome testing and family history. *Genetics in medicine : official journal of the American College of Medical Genetics* 2014; **16**(3): 231-7.
29. Silverman JM, Smith CJ, Marin DB, Mohs RC, Propper CB. Familial patterns of risk in very late-onset Alzheimer disease. *Archives of general psychiatry* 2003; **60**(2): 190-7.
30. Ryman DC, Acosta-Baena N, Aisen PS, et al. Symptom onset in autosomal dominant Alzheimer disease: a systematic review and meta-analysis. *Neurology* 2014; **83**(3): 253-60.
31. de la Torre JC. Cerebral hemodynamics and vascular risk factors: setting the stage for Alzheimer's disease. *Journal of Alzheimer's disease : JAD* 2012; **32**(3): 553-67.
32. Mazza M, Marano G, Traversi G, Bria P, Mazza S. Primary cerebral blood flow deficiency and Alzheimer's disease: shadows and lights. *Journal of Alzheimer's disease : JAD* 2011; **23**(3): 375-89.
33. Debette S, Markus HS. The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. *BMJ (Clinical research ed)* 2010; **341**: c3666.
34. Akoudad S, Wolters FJ, Viswanathan A, et al. Association of Cerebral Microbleeds With Cognitive Decline and Dementia. *JAMA neurology* 2016.
35. Jack CR, Knopman DS, Jagust WJ, et al. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *Lancet neurology* 2013; **12**(2): 207-16.
36. Lampert EJ, Roy Choudhury K, Hostage CA, Petrella JR, Doraiswamy PM, Initiative AasDN. Prevalence of Alzheimer's pathologic endophenotypes in asymptomatic and mildly impaired first-degree relatives. *PLoS ONE* 2013; **8**(4): e60747.

Supplementary Tables

Supplementary Table 1. Interview questions (translated from the original Dutch questionnaire).

Is your father still alive?
If not, at what age did he pass away?
Does/did your father have dementia?
If so, at what age was the disease diagnosed?
Is your mother still alive?
If not, at what age did she pass away?
Does/did your mother have dementia?
If so, at what age was the disease diagnosed?

Supplementary Table 2. Included variants in the genetic risk score for Alzheimer's disease.

rsid	ALT-HRC	Assigned-gene	Locus discovered in	Effect estimate from	Maf	Weight ALT-HRC	Imputation quality(R2)
rs41147929	G	*ABCA7*	Hollingworth et al, Naj et al.	Lambert et al. (2013)	0.19	-0.135	0.917
rs6733839	T	*BIN1*	Seshadri et al.	Lambert et al. (2013)	0.409	0.188	0.911
rs7274581	C	*CASS4*	Lambert et al. (2013)	Lambert et al. (2013)	0.083	-0.139	0.989
rs10948363	G	*CD2AP*	Hollingworth et al., Naj et al.	Lambert et al. (2013)	0.266	0.098	0.998
rs10838725	C	*CELF1*	Lambert et al. (2013)	Lambert et al. (2013)	0.316	0.075	0.998
rs9331896	T	*CLU*	Harold et al., Lambert et al. (2009)	Lambert et al. (2013)	0.379	0.146	0.974
rs6656401	G	*CR1*	Lambert et al. (2009)	Lambert et al. (2013)	0.197	-0.157	0.948
rs7920721	G	*ECHDC3*	Desikan et al.	Desikan et al.	0.3867	-0.029	1
rs11771145	A	*EPHA1*	Hollingworth et al. Naj et al.	Lambert et al. (2013)	0.338	-0.102	0.998
rs17125944	C	*FERMT2*	Lambert et al. (2013)	Lambert et al. (2013)	0.092	0.122	1
rs111418223	A	*HLA-DRB1/5*	Lambert et al. (2013)	Lambert et al. (2013)	0.276	-0.108	0.312
rs13113697	G	*HS3ST1*	Desikan et al.	Desikan et al.	0.2825	-0.029	0.998
rs35349669	T	*INPP5D*	Lambert et al.	Lambert et al. (2013)	0.488	0.066	0.973
rs118172952	G	*KANSL1*	Jun et al.	Lambert et al. (2013)	0.873	-0.151	0.7
rs190982	A	*MEF2C*	Lambert et al. (2013)	Lambert et al. (2013)	0.408	0.08	0.934
rs983392	G	*MS4A6A*	Hollingworth et al. Naj et al.	Lambert et al. (2013)	0.403	-0.108	0.99
rs2718058	G	*NME8*	Lambert et al. (2013)	Lambert et al. (2013)	0.373	-0.07	1
rs10792832	G	*PICALM*	Harold et al.	Lambert et al. (2013)	0.358	0.13	0.999
rs28834970	C	*PTK2B*	Lambert et al. (2013)	Lambert et al. (2013)	0.366	0.096	0.99
rs10498633	T	*SLC24A4-RIN3*	Lambert et al. (2013)	Lambert et al. (2013)	0.217	-0.104	0.999
rs11218343	C	*SORL1*	Lambert et al. (2013)	Lambert et al. (2013)	0.039	-0.27	0.995
rs75932628	T	*TREM2*	Guerreiro et al. Jonsson et al.	Ruiz et al.	0.0016	0.889	0.726
rs1476679	T	*ZCWPW1*	Lambert et al. (2013)	Lambert et al. (2013)	0.287	0.078	0.996

Supplementary Table 3. Family history of dementia and risk of Alzheimer's disease by age at diagnosis in parents.

	Model I HR, 95% CI	**Model II** HR, 95% CI	**Model III** HR, 95% CI	**Model IV** HR, 95% CI
No parental family history	REFERENCE	REFERENCE	REFERENCE	REFERENCE
Age at diagnosis in parent				
<70	5.22, 2.07-13.14	5.25, 2.08-13.22	3.65, 1.29-10.29	3.51, 1.23-9.98
70-79	3.81, 2.26-6.42	3.76, 2.20-6.42	3.17, 1.82-5.52	3.27, 1.81-5.91
80-89	1.46, 0.75-2.84	1.51, 0.77-2.95	1.40, 0.71-2.74	0.99, 0.45-2.20
≥90	1.57, 0.38-6.40	1.53, 0.77-2.95	1.80, 0.43-7.51	1.75, 0.41-7.48
p-value for trend	<0.0001	<0.0001	<0.0001	0.0002

HR = hazard ratio; CI = confidence interval.

Model I: adjusted for age and sex

Model II: model I with additional adjustment for level of education, hypertension, diabetes, hyperlipidaemia, smoking, and body-mass index.

Model III: model II with additional adjustment for *APOE* genotype

Model IV: model III with additional adjustment for a genetic risk score for Alzheimer's disease

Supplementary Table 4. Family history of dementia by age of parental diagnosis in relation to brain MRI measures.

	Parenchymal volume β, 95% CI	Hippocampal volume β , 95% CI	Cerebral perfusion β , 95% CI	White-matter lesions β , 95% CI	Lacunar infarcts (yes vs. no) OR, 95% CI	Microbleeds (1 vs. none) OR, 95% CI	Microbleeds (≥2 vs. none) OR, 95% CI
Model I							
No family history	REFERENCE	REFERENCE	REFERENCE	REFERENCE	REFERENCE	REFERENCE	REFERENCE
Parental family history	0.023, -0.029;0.075	-0.052, -0.175;0.071	-0.213, -0.353;-0.073	0.066, -0.066;0.198	0.93, 0.54-1.62	0.96, 0.61-1.50	1.12, 0.68-1.85
Age at diagnosis <80 years	0.008, -.068;0.085	-0.025, -0.204;0.155	-0.166, -0.372;0.039	0.191, -0.002;0.384	0.91, 0.40-2.08	1.12, 0.59-2.15	1.91, 1.01-3.61
Age at diagnosis ≥80 years	0.034, -0.033;0.100	-0.083, -0.241;0.076	-0.233,-0.414;-0.053	-0.007, -0.176;0.162	0.91, 0.45-1.83	0.84, 0.47-1.53	0.71, 0.34-1.49
Model II							
No family history	REFERENCE	REFERENCE	REFERENCE	REFERENCE	REFERENCE	REFERENCE	REFERENCE
Parental family history	0.024, -0.028;0.077	-0.044, -0.169;0.081	-0.209, -0.351;-0.068	0.075, -0.055;0.204	0.91, 0.51-1.61	0.91, 0.57-1.44	1.14, 0.68-1.92
Age at diagnosis <80 years	0.004, -0.072;0.079	-0.018, -0.196;0.160	-0.170, -0.375;0.035	0.206, 0.014;0.400	0.89, 0.38-2.07	1.06, 0.54-2.04	2.00, 1.04-3.86
Age at diagnosis ≥80 years	0.040, -0.023;0.103	-0.071, -.229;0.086	-0.224, -0.403;-0.044	-0.006, -0.174;0.163	0.88, 0.42-1.85	0.83, 0.45-1.51	0.67, 0.31-1.46
Model III							
No family history	REFERENCE	REFERENCE	REFERENCE	REFERENCE	REFERENCE	REFERENCE	REFERENCE
Parental family history	0.023, -0.030;0.076	-0.040, -0.169;0.090	-0.222, -0.365;-0.080	0.086, -0.046;0.218	0.89, 0.50-1.59	0.93, 0.58-1.47	1.22, 0.72-2.06
Age at diagnosis <80 years	-0.004, -0.081;0.074	-0.028, -0.212;0.155	-0.179, -0.383;0.026	0.208, 0.014;0.402	0.88, 0.38-2.06	1.09, 0.56-2.11	2.13, 1.10-4.12
Age at diagnosis ≥80 years	0.042, -0.026;0.110	-0.057, -0.220;0.106	-0.241, -0.418;-0.064	0.011, -0.144;0.166	0.86, 0.41-1.82	0.84, 0.46-1.54	0.73, 0.32-1.64
Model IV							
No family history	REFERENCE	REFERENCE	REFERENCE	REFERENCE	REFERENCE	REFERENCE	REFERENCE
Parental family history	0.008, -.022;0.038	-0.021, -0.160;0.117	-0.258, -0.418;-0.099	0.093, -0.049;0.235	1.03, 0.56-1.89	1.07, 0.65-1.75	1.52, 0.85-2.72
Age at diagnosis <80 years	-0.031, -0.117;0.054	-0.017, -0.216;0.183	-0.247, -0.473;-0.020	0.220, 0.018;0.423	0.92, 0.36-2.33	1.30, 0.64-2.62	2.82, 1.35-5.88
Age at diagnosis ≥80 years	0.041, -0.035;0.117	-0.027, -0.172;0.117	-0.257, -0.455;-0.060	0.022, -0.145;0.189	1.06, 0.49-2.28	0.94, 0.50-1.79	0.87, 0.37-2.06

Betas reflect risk estimates of family history on standardised outcome measure; OR = odds ratio; CI = confidence interval.

Model I: adjusted for age, sex, total intracranial volume, and interval between interview and MRI

Model II: model I with additional adjustment for level of education, hypertension, diabetes, hyperlipidaemia, smoking, and body-mass index

Model III: model II with additional adjustment for APOE genotype

Model IV: model III with additional adjustment for a genetic risk score for Alzheimer's disease

Supplementary Table 5. Paternal and maternal family history of dementia and brain MRI measures.

	Parenchymal volume β , 95% CI	Hippocampal volume β, 95% CI	Cerebral perfusion β, 95% CI	White-matter lesions β, 95% CI	Lacunar infarcts (yes vs. no) OR, 95% CI	Microbleeds (1 vs. none) OR, 95% CI	Microbleeds (≥2 vs. none) OR, 95% CI
Model I							
No family history	REFERENCE	REFERENCE	REFERENCE	REFERENCE	REFERENCE	REFERENCE	REFERENCE
Parental family history	0.023, -0.029;0.075	-0.052, -0.175;0.071	-0.213, -0.353;-0.073	0.066, -0.066;0.198	0.93, 0.54-1.62	0.96, 0.61-1.50	1.12, 0.68-1.85
Paternal	0.060, -0.031;0.151	-0.292, -0.506;-0.078	-0.138, -0.384;0.108	0.004, -0.227;0.235	0.89, 0.34-2.35	0.91, 0.40-2.07	1.27, 0.54-2.98
Maternal	0.006, -0.055;0.066	0.034, -0.110;0.178	-0.246, -0.410;-0.081	0.088, -0.067;0.242	0.94, 0.49-1.79	1.05, 0.63-1.77	1.18, 0.65-2.14
Both parents	0.086, -0.176;0.348	0.250, -0.365;0.865	-0.178, -0.880;0.523	0.154, -0.511;0.820	1.19, 0.14-10.41	N/R	N/R
Model II							
No family history	REFERENCE	REFERENCE	REFERENCE	REFERENCE	REFERENCE	REFERENCE	REFERENCE
Parental family history	0.024, -0.028;0.077	-0.044, -0.169;0.081	-0.209, -0.351;-0.068	0.075, -0.055;0.204	0.91, 0.51-1.61	0.91, 0.57-1.44	1.14, 0.68-1.92
Paternal	0.048, -0.053;0.149	-0.297, -0.513;-0.081	-0.142, -0.376;0.092	0.051, -0.177;0.281	0.96, 0.35-2.60	0.88, 0.39-2.02	1.40, 0.59-3.33
Maternal	0.012, -0.049;0.074	0.050; -0.096;0.197	-0.234, -0.400;-0.068	0.074, -0.080;0.230	0.86, 0.44-1.70	1.01, 0.60-1.71	1.14, 0.61-2.12
Both parents	0.078, -0.185;0.341	0.258; -0.362;0.880	-0.280, -0.638;0.079	0.289, -0.369;0.950	1.41, 0.16-12.93	N/R	N/R
Model III							
No family history	REFERENCE	REFERENCE	REFERENCE	REFERENCE	REFERENCE	REFERENCE	REFERENCE
Parental family history	0.023, -0.030;0.076	-0.040, -0.169;0.090	-0.222, -0.365;-0.080	0.086, -0.046;0.218	0.89, 0.50-1.59	0.93, 0.58-1.47	1.22, 0.72-2.06
Paternal	0.051, -0.041;0.143	-0.271, -0.488;-0.053	-0.150, -0.399;0.100	0.095, -0.135;0.326	0.94, 0.35-2.57	0.90, 0.39-2.07	1.46, 0.61-3.50
Maternal	0.008, -0.054;0.071	0.046, -0.103;0.194	-0.250, -0.419;-0.081	0.072,-0.084;0.229	0.85, 0.43-1.68	1.03, 0.61-1.74	1.23, 0.67-2.27
Both parents	0.075, -0.187;0.337	0.240, -0.380;0.861	-0.277, -0.984;0.429	0.303, -0.354;0.962	1.38, 0.15-12.62	N/R	N/R
Model IV							
No family history	REFERENCE	REFERENCE	REFERENCE	REFERENCE	REFERENCE	REFERENCE	REFERENCE
Parental family history	0.008, -.022;0.038	-0.021, -0.160;0.117	-0.258, -0.418;-0.099	0.093, -0.049;0.235	1.03, 0.56-1.89	1.07, 0.65-1.75	1.52, 0.85-2.72
Paternal	0.038, -0.070;0.147	-0.203, -0.457;0.052	-0.140, -0.430;0.151	0.130, -0.136;0.397	1.09, 0.35-3.38	1.09, 0.43-2.73	2.30, 0.86-6.15
Maternal	-0.005, -0.073;0.063	0.035, -0.126;0.196	-0.307, -0.490;-0.123	0.065,-0.102;0.233	0.97, 0.48-1.97	1.15, 0.66-2.00	1.36, 0.68-2.73
Both parents	0.041, -0.246;0.328	0.171, -0.501;0.844	-0.130, -0.899;0.637	0.411, -0.293;1.118	1.71, 0.17-17.06	N/R	N/R

Betas reflect risk estimates of family history on standardised outcome measure; OR = odds ratio; CI = confidence interval; N/R = not rateable due to small numbers.

Model I: adjusted for age, sex, total intracranial volume, and interval between interview and MRI

Model II: model I with additional adjustment for level of education, hypertension, diabetes, hyperlipidaemia, smoking, and body-mass index

Model III: model II with additional adjustment for *APOE* genotype

Model IV: model III with additional adjustment for a genetic risk score for Alzheimer's disease

Chapter 3

Rare genetic variant discoveries in AD

Chapter 3.1

Haplotype Reference Consortium Panel: Practical implications of imputations with large reference panels

Adriana I. Iglesias*, Sven J. van der Lee*, Pieter W.M. Bonnemaijer, René Höhn, Abhishek Nag, Puya Gharahkhani, Anthony P. Khawaja, Linda Broer, International Glaucoma Genetics Consortium (IGGC) †, Paul J. Foster, Christopher J. Hammond, Pirro G. Hysi, Elisabeth M. van Leeuwen, Stuart MacGregor, David A. Mackey, Johanna Mazur, Stefan Nickels, André G. Uitterlinden, Caroline C.W. Klaver, Najaf Amin, Cornelia M. van Duijn.

* authors contributed equally

† Membership of the IGGC is listed in the Supporting Information

Human Mutation 2017 Aug; 38(8): 1025-1032

Abstract

Recently, the Haplotype Reference Consortium (HRC) released a large imputation panel that allows more accurate imputation of genetic variants. In this study, we compared a set of directly assayed common and rare variants from an exome array to imputed genotypes, i.e., 1000 genomes project (1000GP) and HRC. We showed that imputation using the HRC panel improved the concordance between assayed and imputed genotypes at common, and especially, low-frequency variants. Furthermore, we performed a genome-wide association meta-analysis of vertical cup-disc ratio (VCDR), a highly heritable endophenotype of glaucoma, in four cohorts using 1000GP and HRC imputations. We compared the results of the meta-analysis using 1000GP to the meta-analysis results using HRC. Overall, we found that using HRC imputation significantly improved *p*-values (P = 3.07 x 10^{-61}), particularly for suggestive variants. Both meta-analyses were performed in the same sample size, yet we found eight genome-wide significant loci in the HRC-based meta-analysis versus seven genome-wide significant loci in the 1000GP-based meta-analysis. This study provides supporting evidence of the new avenues for gene discovery and fine mapping that the HRC imputation panel offers.

Introduction

Progress in understanding the genetic aetiology of complex traits is largely determined by the success of genome-wide association studies (GWAS).[1,2] Imputation, i.e. statistical inference of genotypes not directly assayed by arrays,[3] is crucial to the success of GWAS. Imputation increases the number of variants tested per study and allows combination of multiple studies assayed in different arrays, boosting the power of GWAS.[4-6] Recently, the Haplotype Reference Consortium (HRC) released the first version of a panel encompassing 64,976 haplotypes.[7] This resource combines other widely used panels such as the 1000 Genomes Project (1000GP),[8] the Genome of the Netherlands (GoNL),[9,10] the UK10K,[11] and also includes sequencing data from 17 cohorts. [7]

Previous studies have shown that larger reference panels considerably increase imputation accuracy, particularly for low-frequency variants, and that this gain in imputation accuracy results in increased statistical power.[7,12,13] To quantify the advantage of the HRC reference panel in the imputation of common and rare variants, we first assessed the concordance rate between directly assayed genotypes from exome array data and *best-guess* genotypes from 1000GP and HRC imputations. Then, we evaluated any difference in statistical power by comparing the meta-analysis of GWAS of vertical cup-disc ratio (VCDR) using 1000GP imputation versus HRC imputation, in the same samples. VCDR is a well-recognized endophenotype of primary open-angle glaucoma (POAG) [14] and a highly heritable trait ($h^2 = 0.66$) [14] used for clinically assessing glaucoma patients. Here, we assessed the impact of the imputation panel on a meta-analysis of GWAS of VCDR by analyzing four cohorts: the Rotterdam Study (RS-I, RS-II, RS-III) [15] and the Erasmus Rucphen Family (ERF) [16,17] study.

Materials and Methods

Study descriptions

The description of the participating studies can be found in the **Supporting information.** All studies adhered to the tenets of the Declaration of Helsinki and were approved by their local Medical Ethics Committees. Written, informed consent was obtained from all participants.

Genotype and Imputations

The Rotterdam Study cohorts (RS-I, RS-II and RS-III) and the ERF study were genotyped using commercially available genotyping arrays, and genotyping quality control (QC) was done for each cohort individually. Detailed information about genotyping and imputations for each study is presented in **Supplementary Table 1.** The genome-wide arrays were used to impute variants twice, once with the 1000GP [8] and once with HRC.[7] Imputations with the 1000GP were described previously [18]. For HRC, file preparation was done using scripts provided online (HRC Imputation preparation and checking: http://www.well.ox.ac.uk/~wrayner/tools/; v4.2.1). Imputation with HRC was facilitated by the Michigan Imputation server.[19] Filtered genotypes were uploaded. The server uses SHAPEIT2 (v2.r790) to phase the data and Minimac 3 for imputation to the HRC reference panel (v1.0). We used the imputed dosages returned by the service.[7]

Genotyping on the Exome array

In total 3,159 participants from RS-I were genotyped in the HumanExome BeadChip v1.0 from Illumina (Illumina, Inc., San Diego, CA, USA). To increase the quality of the rare variant genotype calls on the exome array, the genotypes were jointly called with 62,266 samples from 11 studies at the University of Texas HSC at Houston (UT Houston).[20] QC procedures for the genotype data were done both centrally at UT Houston and locally. The central QC procedures have been described previously.[20] Locally, additional QC included removal of: 1) individuals with low genotype completion rate (<90%), 2) variants with low genotype call rate (<95%), 3) individuals with sex-mismatches, 4) one individual from duplicate pairs and 5) variants not called in over 5% of the individuals and those that deviated significantly form the expected Hardy-Weinberg Equilibrium proportions ($P < 1 \times 10^{-6}$).

Variant selection and calculation of concordance

A total of 236,756 variants on the exome array passed QC. Of these, we selected variants that fulfilled the following four criteria: 1) were imputed in both the 1000GP and HRC, 2) were polymorphic, 3) had non-ambiguous allele coding and 4) showed no differences in reference allele frequency. In summary, 82,281 variants on the exome array were imputed in both 1000GP and HRC panels. Of these, 31,022 did not fulfil our criteria

(*n* = 29,819 monomorphic, *n* = 1,141 with ambiguous allele coding, and *n* = 62 with large differences [>10%] in reference allele). The final set consisted of 51,259 variants, for which we calculated either the concordance of the alternate allele calls (for the low-frequency and rare variants) or the concordance for both reference and alternate allele (for common variants). Calculation of concordance was performed using the exome array genotypes as benchmark and the *best-guess* genotypes from either 1000GP or HRC. For the common variants, the reported concordance is the percentage of correctly imputed alleles divided by the number of exome array alleles according to the following formula (*n* = number of individuals):

$$\text{Concordance for common variants} = x = 100 \times \frac{(b + d + f + h + 2(a + e + i))}{2n}$$

For the rare variants, the reported concordance is the percentage of correctly imputed alternate alleles divided by the number of exome array alternate alleles according to the following formula:

$$\text{Concordance for rare variants} = x = 100 \times \frac{(e + f + h + 2i)}{(b + e + h + 2(c + f + i))}$$

		Exome array (benchmark) # of alternate alleles		
		0	**1**	**2**
HRC/1000GP # of alternate alleles	**0**	a	b	c
	1	d	e	f
	2	g	h	i

For comparison purposes, assessed variants were divided into nine categories based on their MAF (i.e., common, MAF > 0.05; low-frequency, 0.01 < MAF < 0.05; rare, 0.001 < MAF < 0.01), and R^2_{1000GP} (i.e., $R^2_{1000GP} > 0.8$; R^2_{1000GP} 0.3-0.8; $R^2_{1000GP} < 0.3$). Common variants with $R^2_{1000GP} < 0.3$ that showed a concordance rate < 90% (*n* = 44) were examined in the list of variants located in inaccessible regions of the genome from the GoNL.[10]

Genome-wide association analyses in RS-I-II-III and ERF

We performed a GWAS in four Dutch cohorts: RS-I, RS-II, RS-III and ERF (n = 12,441). GWAS were conducted twice per cohort, once using 1000GP imputations and once using HRC imputations. Association of directly genotyped and imputed dosages with VCDR was tested using linear regression under an additive model. All analyses were adjusted for age, sex and the first five principal components in RS-I, RS-II and RS-III, or family structure in ERF. The 1000GP imputations GWAS analyses were conducted using ProbABEL [21] software. To account for family relationships in ERF, the *mmscore* function implemented in GenABEL [22] was used. Association analyses with HRC imputations were performed in RvTest --*meta score* option; to adjust for familial relationships in ERF, the kinship matrix estimated from the genotyped data was used. Variants with MAF < 0.01 in 1000GP and MAF ≤ 0.001 in HRC, and with low imputation quality score (R^2 < 0.3) were excluded from the analyses. The EasyQC software was used for QC at the study level, as described elsewhere [23]. The inflation factor (λ) of each individual study using 1000GP and HRC panels can be found in the **Supplementary Table 2**.

Meta-analysis of RS-I-II-III and ERF

Meta-analyses for both 1000GP and HRC association results were conducted using the inverse variance-weighted fixed-effect method in METAL.[24] Genomic control was applied in METAL and heterogeneity was calculated. After meta-analyses, we excluded the variants that were not present in at least two of the four cohorts. Given that HRC does not include INDELs, we excluded the INDELs from the 1000GP meta-analysis. The remaining variants were used for comparison of the results between 1000GP and HRC. To calculate the λ in each meta-analysis we used those variants present in both meta-analysis (n = 7,654,440). λ in 1000GP was 1.096, and in HRC 1.102, see QQ plot in **Supplementary Figure 1**.

Statistical analysis

To evaluate the change of p-values in the HRC meta-analysis compared with the 1000GP meta-analysis, variants from 1000GP and HRC meta-analyses were divided into four categories: 1) genome-wide associated in 1000GP or HRC ($P < 5.0 \times 10^{-8}$), 2) suggestively associated in 1000GP or HRC ($5.0 \times 10^{-8} < P \leq 1.0 \times 10^{-4}$), 3) nominally associated (1.0 x

$10^{-4} < P \leq 0.05$), and 4) not associated ($0.05 < P \leq 1$). We then subtracted the chi-square in HRC meta-analysis from the chi-square in 1000GP, and calculated the proportion of variants that showed a positive sign (i.e., larger chi-square in HRC). We then used a proportion Z-test to evaluate if the proportion of variants with a positive sign was significantly different from 0.5.

Results

Comparison 1000GP vs HRC

We first compared the R^2 value derived from the imputation software between 1000GP and HRC imputations (see **Figure 1**). In this study, R^2_{1000GP} refers to the R^2 value in 1000GP imputations and R^2_{HRC} to the value in HRC. As reported by the HRC consortium,[7] we confirmed that the R^2_{HRC} was higher than the R^2_{1000GP}, particularly for rare ($0.001 <$ minor allele frequency (MAF) < 0.01) and low-frequency variants ($0.01 <$ MAF < 0.05) variants.

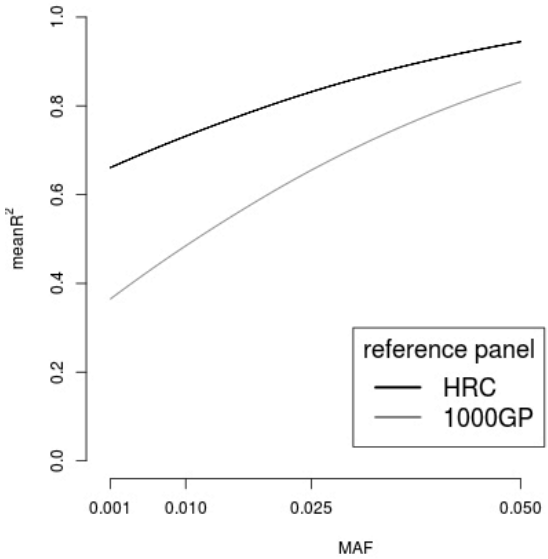


Figure 1. **Imputation accuracy 1000GP vs HRC.** Post-imputation quality metric based on the R^2 value from 1000GP (in blue) and HRC (in red). MAF=Minor allele frequency.

In 3,159 individuals from the Rotterdam Study-I (RS-I) assayed on the HumanExome BeadChip v1.0 (exome array), we determined the concordance between directly assayed genotypes and imputed genotypes. Of the 51,259 studied variants (see Methods section), 21,230 were common (MAF >0.05 in the exome array), 8,006 were low-frequency variants (0.01 < MAF < 0.05) and 11,090 were rare variants (0.001 < MAF < 0.01). A total of 10,933 very rare variants (i.e., MAF < 0.001) were excluded as the minor allele count for the variants were six or less. The concordance rate of very rare variants can be found in **Supplementary Table 3** and **Supplementary Figure 2.**

Figure 2 shows the concordance between exome array and 1000GP in red and the concordance between exome array and HRC in blue (plots are shown per MAF and R^2 bins; for comparison purposes, we used the R^2_{1000GP} to define the bins). **Supplementary Table 4** shows the number of variants per bin and the concordance rates for both 1000GP and HRC. Overall, concordance was better using HRC compared to 1000GP. For common variants with R^2_{1000GP} 0.3-0.8 (**Figure 2B**), the percentage of concordant variants (i.e., concordance rate ≥ 90%) improved from 86% in 1000GP to 99.8% in HRC. In the case of common variants with R^2_{1000GP} < 0.3 (*N variants* = 177, **Figure 2C**), the percentage of concordant variants improved from 46% in 1000GP to 75% in HRC. For low-frequency variants, a remarkable improvement in concordance was observed as shown in **Figures 2D-2F**. Overall (*N variants* = 8,006), the concordance increased from 44% in 1000GP to 66% in HRC. For variants with R^2_{1000GP} > 0.8 (*N variants* = 3,844), the percentage of concordance improved from 85% in 1000GP to 98% in HRC. Improvement was most pronounced for variants with R^2_{1000GP} 0.3-0.8 (**Figure 2E**; *N variants* = 3,301), in which the percentage of concordant variants improved from 7% in 1000GP to 44% in HRC. Despite the overall improvement, the concordance is relatively low for variants with R^2_{1000GP} < 0.3 (**Figure 2F**; *N variants* = 861); concordance improved from 1% in 1000GP to 5% in HRC (see details in **Supplementary Table 4**).

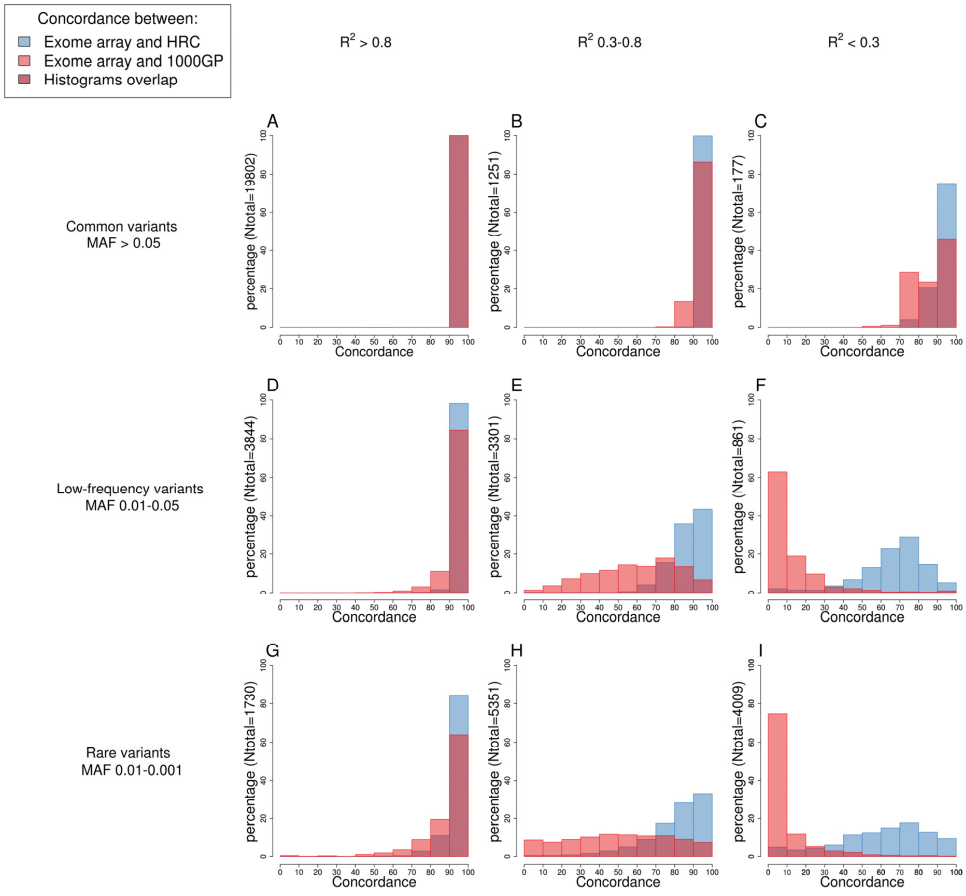


Figure 2. Comparison of concordance between exome-array and imputed data (1000GP and HRC). Histograms are presented by bins selected based on minor allele frequency (MAF) and R^2_{1000GP} values. For each histogram, the concordance rate is on the *x* axis and the percentage of variants (N total of variants) is on the *y* axis. Concordance between exome-array genotypes and *best-guess* genotypes from 1000GP are shown in red, and from HRC in blue. Overlapped regions of the histograms are in dark-red. (**A-C**) show comparison of concordance for common variants; (**D-F**) for low-frequency variants, and (**G-I**) for rare variants. Base on the R^2_{1000GP} threshold (**A-D-G**) show the comparison of concordance for variants with high R^2_{1000GP} (>0.8), (**B-E-H**) for variants with R^2_{1000GP} between 0.3-0.8, and (**C-F-I**) for variants with low R^2_{1000GP} (<0.3).

Using HRC, we found a similar pattern for rare variants as observed for low-frequency variants (**Figure 2G-I**). Overall (*N variants* = 11,090) concordance improved from 14% in 1000GP to 34% using HRC imputations. For rare variants with $R^2_{1000GP} > 0.8$ (**Figure 2G**), concordance improved from 65% in 1000GP to 86% in HRC while for rare variants with R^2_{1000GP} 0.3-0.8, concordance increased from 8% in 1000GP to 35% in HRC (**Figure 2H**).

Impact of HRC imputations on vertical cup-disc ratio GWAS

To evaluate the improvement in statistical power, we compared the results of a meta-analysis of GWAS using 1000GP imputations to a meta-analysis of GWAS using HRC imputations. Both meta-analyses included 12,441 individuals from four independent Dutch cohorts (RS-I, RS-II, RS-III and ERF). **Table 1** describes the baseline characteristics of the study populations. The inflation factor lambda (λ), an indicator of potential population stratification and false positive rate, was comparable in the two meta-analyses with values of 1.096 in the 1000GP meta-analysis and 1.102 in the HRC meta-analysis (**Supplementary Figure 1 and Table 2**), suggesting no increase in false positives in the HRC analyses.

Table 1. Studies characteristics Rotterdam-I-II-III and ERF Studies

Study	*N*	Mean Age	Age SD	Age-range	% Men	Mean VCDR	VCDR range
RS-I	5573	68.0	8.4	55-99	40.90%	0.50	0.05-0.87
RS-II	1987	64.7	7.7	55-96	46.10%	0.50	0.10-0.86
RS-III	2873	57.2	6.6	46-90	43.90%	0.29	0.00-1.00
ERF	2008	48.3	13.8	18-85	43.97%	0.31	0.00-0.87

SD = standard deviation; VCDR = vertical-cup disc ratio

To assess the improvement in *p*-values, we calculated the proportion of variants for which a stronger association was found in the HRC meta-analysis. 62% of variants that showed suggestive association in 1000GP or HRC meta-analyses (i.e., $5.0 \times 10^{-08} < P < 1.0 \times 10^{-04}$; *n variants* = 4,683) showed smaller *p*-values in the HRC meta-analysis ($P = 3.07 \times 10^{-61}$, **Table 2**) suggesting improved sensitivity in the HRC-based meta-analysis. For variants below suggestive significance (i.e., $1.0 \times 10^{-04} < P < 1$) the *p*-values did not improve, suggesting better specificity. No statistically significant differences were observed for the genome-wide significant variants.

Table 2. Improvement in *P*-value in HRC meta-analysis

GWAS *P-value* category	N variants	% of variants with smaller *P*-value in HRC-meta-analysis	Z-score*	*P-value**
Genome-wide significant ($P \leq 5.0 \times 10^{-08}$)	751	52	1.28	0.202
Suggestive ($5.0 \times 10^{-08} < P \leq 1.0 \times 10^{-04}$)	4683	62	16.47	3.07×10^{-61}
Nominally associated ($1.0 \times 10^{-04} < P \leq 0.05$)	534998	49	-16.92	1.53×10^{-64}
No association signal ($0.05 < P \leq 1$)	7321737	48	-119.23	$< 1.0 \times 10^{-230}$

Variants from 1000GP and HRC meta-analyses were divided into four categories, i.e., genome-wide associated, suggestive, nominally associated and not associated. *Z-score and *P*-value from a proportion z-test, in which H_0=50%.

Based on the data of the Rotterdam Study and ERF, we confirmed in our 1000GP meta-analysis seven loci reported by the International Glaucoma Genetics Consortium (IGGC) [25,26], which is 2.6 times the size of this study. Using the HRC imputation, we confirmed one additional common variant (rs6539763, MAF = 0.46, close to *TMTC2*), previously reported by the IGGC [25]. When comparing the top associated variants from both 1000GP and HRC meta-analyses, we found that in three out of the seven confirmed loci, the top associated variant was the same (rs1192415 close to *TGFBR3*, rs7916410 close to *ATOH7*, and rs9613667 in *CHEK2*). For the other four loci, top variants were different in each meta-analysis (**Supplementary Table 5**). In all cases the identified variants were in LD ($0.74 < r^2 < 1$) with the variant previously reported by the IGGC.

Discussion

We showed that the HRC imputations improve the concordance between directly assayed and imputed genotypes over a large range of MAFs. In addition, our VCDR analyses show that the *p*-values obtained with HRC imputations compared to 1000GP imputations are on average significantly smaller for suggestive variants. The improvements were also seen with common variants, predicting that HRC imputations may also be relevant for finding new common variants.

As described by the HRC,[7] the imputation accuracy (measured by the R^2) is higher when using the HRC panel. In our comparisons, 93.27% (*n* = 19,802/21,230) of the common

variants imputed with 1000GP had a $R^2 > 0.8$; this percentage increased to 98.74% in HRC (*n* = 20,962/21,230; **Supplementary Table 6**). Despite the overall improvement, there is still a group of common variants with $R^2 < 0.3$ that showed a low concordance rate (i.e., <90%). Of these, about 18% (8/44) are located in inaccessible regions of the genome that represent a challenge for short-read technologies [9,13,27,28] (**Supplementary Table 7**). Investment in long-read sequencing and single-molecule mapping, along with improvements in the generation of high-quality *de novo* assemblies of genomes, will advance our understanding of the genome and its variation, and will further improve the accuracy of genotype imputations.[29,30]

In this study, we focused on the assessment of genotype concordance rate rather than on the R2 changes as described by the HRC in McCarthy et al. We calculated and assessed the concordance rate between high-quality genotypes [20] from the exome array (enriched in low-frequency and rare variants) and imputed genotypes. The general implication of our study is that for common variants (MAF > 0.05), improvement in concordance supports the view that *best-guess* genotypes, instead of directly assayed data, can be used for replication of GWAS findings but also for Mendelian Randomization studies, particularly when R^2 is >0.8. For low-frequency variants (0.01 < MAF < 0.05), the improvement in concordance is relevant for gene discovery, as previously highlighted by the HRC [7]. A marked improvement was observed for variants with R^2_{1000GP} 0.3-0.8 (**Figure 2E**). Thus, low-frequency variants commonly filtered out in meta-analyses of GWAS at present might be studied reliably using HRC imputations. For rare variants (0.001 < MAF < 0.01) there is also a relevant improvement in concordance that facilitates in silico validation of findings of exome array or sequencing results, and gene discovery. Finally, for rare variation we observed that the concordance of well-imputed variants increased by 20% (65% in 1000GP vs to 86% in HRC).

The HRC panel incorporates 64,976 haplotypes, including 998 Dutch haplotypes from the GoNL [9,18]; previous studies have shown that sample size of the reference panel is often more important than population matching.[4,5] This finding was also observed in a previous study,[13] in which a combined reference panel including GoNL and 1000GP

yielded the best imputation results. This suggests that the improvement in genotype concordance observed in our study is related to the increase sample size of the HRC.

Our analysis of VCDR in the Rotterdam and ERF cohorts used the same group of individuals, thus allowing an informative comparison between meta-analysis results performed with genotypes imputed with 1000GP versus genotypes imputed with HRC. We showed that for over a large range of MAFs, including common variants, a significant improvement in *p*-values can be achieved with HRC imputations. The same trend was observed when *p*-values from both meta-analyses were compared (**Supplementary Figure 3**). **Table 2** shows that the percentage of genome-wide significant, and particularly, suggestive variants in the GWAS performed using the genetic data imputed with the HRC reference panel is larger compared to the GWAS using the 1000GP reference panel. Given the same sample size (n = 12,441), we confirmed in the 1000GP meta-analysis seven previously reported VCDR-loci,[25,26] while in the HRC meta-analysis an additional known VCDR-locus (close to *TMTC2*) was confirmed. The main difference between 1000GP and HRC meta-analysis at the *TMTC2* locus is at the regression coefficient and standard error ($\beta(SE)_{1000GP}$ = -0.010(0.002) vs $\beta(SE)_{HRC}$ = -0.011(0.0019)). These loci have been reported in the HapMap and 1000GP GWAS meta-analyses conducted by the IGGC, which included 27,878 and 32,272 individuals, respectively.[25,26]

A potential limitation in this study is that we used exome array data as benchmark, and genotype calling may be subject to error. However, data were jointly called with around 62,000 other samples, thereby improving the accuracy of our genotype calls.[20] Assessment of concordance using exome-array genotypes allowed us to compare the performance of the both reference panels independently of the post-imputation metric (R^2) derived from imputation software. Our sample size limits the assessment of very rare variants (MAF < 0.001); however, the same trend was observed for these variants (**Supplementary Table 3**). We analysed a genetically homogeneous Dutch population, thus we cannot predict the performance of HRC in non-Europeans. Assessment in other populations will reveal the performance of HRC and 1000GP panels in non-European

samples. Our findings provide further evidence for the gain in power when a large reference panel as released by the HRC is used for imputations. We have shown that concordance between data directly assayed in the exome array and HRC imputations is high, particularly, for variants with an $R^2 > 0.8$ across a wide range of allele frequencies. Furthermore, our analyses of VCDR data in 12,441 subjects showed smaller *p*-values for common and rare variants when using the HRC panel. Thus, imputations with HRC and other larger reference panels improves the statistical power to discover both common and low-frequency variants, opening new avenues for fine-mapping and gene discovery.

References:

1. McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 2008; **9**(5): 356-69.
2. Price AL, Spencer CC, Donnelly P. Progress and promise in understanding the genetic basis of common diseases. *Proc Biol Sci* 2015; **282**(1821): 20151684.
3. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. *Annu Rev Genomics Hum Genet* 2009; **10**: 387-406.
4. Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet* 2010; **11**(7): 499-511.
5. Huang J, Howie B, McCarthy S, et al. Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat Commun* 2015; **6**: 8111.
6. Hao K, Chudin E, McElwee J, Schadt EE. Accuracy of genome-wide imputation of untyped markers and impacts on statistical power for association studies. *BMC Genet* 2009; **10**: 27.
7. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016; **48**(10): 1279-83.
8. Genomes Project C, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature* 2015; **526**(7571): 68-74.
9. Genome of the Netherlands C. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat Genet* 2014; **46**(8): 818-25.
10. Boomsma DI, Wijmenga C, Slagboom EP, et al. The Genome of the Netherlands: design, and project goals. *Eur J Hum Genet* 2014; **22**(2): 221-7.
11. Consortium UK, Walter K, Min JL, et al. The UK10K project identifies rare variants in health and disease. *Nature* 2015; **526**(7571): 82-90.
12. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet* 2009; **84**(2): 210-23.
13. Deelen P, Menelaou A, van Leeuwen EM, et al. Improved imputation quality of low-frequency and rare variants in European samples using the 'Genome of The Netherlands'. *Eur J Hum Genet* 2014; **22**(11): 1321-6.
14. Charlesworth J, Kramer PL, Dyer T, et al. The path to open-angle glaucoma gene discovery: endophenotypic status of intraocular pressure, cup-to-disc ratio, and central corneal thickness. *Invest Ophthalmol Vis Sci* 2010; **51**(7): 3509-14.
15. Hofman A, Brusselle GG, Darwish Murad S, et al. The Rotterdam Study: 2016 objectives and design update. *Eur J Epidemiol* 2015; **30**(8): 661-708.

16. Pardo LM, MacKay I, Oostra B, van Duijn CM, Aulchenko YS. The effect of genetic drift in a young genetically isolated population. *Ann Hum Genet* 2005; **69**(Pt 3): 288-95.
17. Aulchenko YS, Heutink P, Mackay I, et al. Linkage disequilibrium in young genetically isolated Dutch population. *Eur J Hum Genet* 2004; **12**(7): 527-34.
18. van Leeuwen EM, Kanterakis A, Deelen P, et al. Population-specific genotype imputations using minimac or IMPUTE2. *Nat Protoc* 2015; **10**(9): 1285-96.
19. Das S, Forer L, Schonherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet* 2016; **48**(10): 1284-7.
20. Grove ML, Yu B, Cochran BJ, et al. Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. *PLoS One* 2013; **8**(7): e68095.
21. Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics* 2010; **11**: 134.
22. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 2007; **23**(10): 1294-6.
23. Winkler TW, Day FR, Croteau-Chonka DC, et al. Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc* 2014; **9**(5): 1192-212.
24. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; **26**(17): 2190-1.
25. Springelkamp H, Hohn R, Mishra A, et al. Meta-analysis of genome-wide association studies identifies novel loci that influence cupping and the glaucomatous process. *Nat Commun* 2014; **5**: 4883.
26. Springelkamp H, Iglesias AI, Mishra A, et al. New insights into the genetics of primary open-angle glaucoma based on meta-analyses of intraocular pressure and optic disc characteristics. *Hum Mol Genet* 2016; **Accepted**.
27. Genomes Project C, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012; **491**(7422): 56-65.
28. Genomes Project C, Abecasis GR, Altshuler D, et al. A map of human genome variation from population-scale sequencing. *Nature* 2010; **467**(7319): 1061-73.
29. Chaisson MJ, Wilson RK, Eichler EE. Genetic variation and the de novo assembly of human genomes. *Nat Rev Genet* 2015; **16**(11): 627-40.
30. Berlin K, Koren S, Chin CS, Drake JP, Landolin JM, Phillippy AM. Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. *Nat Biotechnol* 2015; **33**(6): 623-30.

Supplementary material

Supplementary Figures, Tables and IGGC author list can be accessed by scanning the following code or accessing the journals' website.

Chapter 3.2

PLD3-variants in population studies

Sven J. van der Lee, Henne Holstege, Tsz Hang Wong, Johanna Jakobsdottir, Joshua C. Bis, Vincent Chouraki, Jeroen van Rooij, Megan L. Grove, Albert V. Smith, Najaf Amin, Seung-Hoan Choi, Alexa S. Beiser, Melissa E. Garcia, Wilfred F.J. van IJcken, Yolande A.L. Pijnenburg, Eva Louwersheimer, Rutger W.W. Brouwer, Mirjam C.G.N. van den Hout, Edwin Oole, Gudny Eirkisdottir, Daniel Levy, Jerome I. Rotter, Valur Emilsson, Christopher J. O'Donnell, Thor Aspelund, Andre G. Uitterlinden, Lenore J. Launer, Albert Hofman, Alzheimer's Disease Neuroimaging Initiative (ADNI), Eric Boerwinkle, Bruce M. Psaty, Anita L. DeStefano, Philip Scheltens, Sudha Seshadri, John C. van Swieten, Vilmundur Gudnason, Wiesje M. van der Flier, M. Arfan Ikram, Cornelia M. van Duijn.

Nature. 2015 April 2; 520(7545): E2–E3.
Arising from C. Cruchaga et al. Nature. 2014; 505: 550–554

Cruchaga *et al.*[1] report that rare genetic variants in *PLD3* (phospholipase D3) are associated with increased Alzheimer's disease (AD) risk. They showed that *PLD3* is involved in amyloid-β precursor protein (APP) processing and overexpressed in brain tissue from AD patients. However, even the key variant *PLD3*-Val232Met, did not pass genome wide significance. This observation raises the question if the *genetic* association of *PLD3* with AD replicates. We associated *PLD3*-Val232Met with AD in 3 large population-based studies and 3 case-control studies. In total, we meta-analyzed results from 1,914 AD cases and 8,021 controls of European descent. Additionally we searched for other coding *PLD3*-variants in sequence data of 1,067 AD cases and 1,553 controls.

Carrier frequencies of *PLD3*-Val232Met in controls ranged from 0.34-1.42%, consistent with 0-1.17% reported by Cruchaga *et al.*[1] (**Table 1**). Likewise, the frequencies of *PLD3*-Val232Met in cases ranged from 0.66-2.19% compared to 0.7-2.6%.[1] We note that the range of carrier frequencies overlaps between cases and controls, such that in some population based cohorts, the carrier frequency in controls (e.g., 1.28% in FHS) is higher than that of cases in other cohorts (e.g., 0.68% in AGES). Within each cohort frequencies of *PLD3*-Val232Met were higher in cases than controls (**Table 1**), but in none of the populations the case carrier frequency for *PLD3*-Val232Met was significantly increased. However, pooled analyses showed a 1.94 fold increased risk of AD of carriers compared to non-carriers (Odds Ratio [OR] 1.94, adjusted for age and sex, 95% confidence interval=1.05 to 3.57) that was marginally significant (p = 0.03) (**Table 1**). Of note, the crude ORs often differed considerably from the age and sex adjusted estimates. With the exception of ADNI, the ORs were higher after adjustment for age and sex, suggesting that many a-symptomatic carriers were relatively young compared to cases.

Table 1. Association of *PLD3*-Val232Met with Alzheimer's disease.

Cohort	Cases			Controls			Overall carri frequency (%)	Crude OR	OR (95% CI)	*p*
	Carriers (N)	Non-carriers (N)	Carrier frequency (%)	Carriers (N)	Non-carriers (N)	Carrier frequency (%)				
AGES	1	145	0.68	12	2371	0.50	0.51	1.36	3.18 (0.17 - 58.18)	0.44
FHS	5	223	2.19	17	1314	1.28	1.41	1.73	2.63 (0.71 - 9.70)	0.15
RS	6	470	1.26	23	2389	0.95	1.00	1.33	1.43 (0.51 - 4.03)	0.49
GRIP	2	109	1.80	14	975	1.42	1.45	1.28	1.58 (0.27 - 9.91)	0.62
Dutch Alzheimer centers	3	451	0.66	4	609	0.65	0.66	1.01	1.57 (0.25 - 9.38)	0.64
ADNI	7	492	1.40	1	292	0.34	1.01	4.15	2.94 (0.63 - 13.8)	0.17
Combined	24	1,890	1.25	71	7,950	0.89	0.96	1.53	1.94 (1.05 - 3.57)	0.03

The R-package "seqMeta"(version seqMeta_1.4) was used for meta-analysis of single variant score test. Odds Ratios (OR) and *p-values* from score tests are shown adjusted for age and sex, based on a logistic regression model. Crude ORs were calculated using carrier frequencies of cases and controls. Combined crude OR is the *Mantel-Haenszel* estimate of the pooled crude ORs (95% CI 0.91 to 2.57 and two-sided $p = 0.11$). Age, Gene/Environment Susceptibility-Reykjavik Study (AGES). Framingham Heart Study (FHS) and Rotterdam Study (RS) where genotyped on the Illumina exome chip version 1.0. Genetic Research in Isolated Populations (GRIP) subjects were imputed. Dutch Alzheimer centers; encompass whole exome sequence data of AD cases from Amsterdam Dementia Cohort, Alzheimer Center Rotterdam MC and controls from RS (not genotyped on the exome chip). Alzheimer's Disease Neuroimaging Initiative (ADNI) samples are whole genome sequenced.

We further associated other coding variants in *PLD3* with AD and performed a gene-based test using sequence data from two studies encompassing 1,067 AD cases and 1,553 controls. We meta-analyzed results of whole genome sequence data of the ADNI study, 499 AD cases and 293 controls, with results of a combined cohort of 568 Dutch AD cases and 1260 Dutch controls.

We observed 21 rare polymorphic coding variants and 1 splice site variant. Of the 20 observed *PLD3*-variants detected by Cruchaga *et al.*[1], we observed 9 (S63G, P76A, V232M, N284S, C300Y, A442A, G452E, D447G and R488C). Five variants showed the same direction of effect as seen by Cruchaga *et al*.[1] After correcting the -value for multiple testing, none of the variants observed in our study conferred a significant increase in AD risk. Gene-based analysis also did not show significant association of *PLD3*-variants with AD risk (SKAT-O p = 0.61 and burden test OR 1.27 95% confidence interval [CI] =0.85 to 1.9, p = 0.24).

In conclusion, the carrier frequencies of *PLD3*-Val232Met in our data set are consistent with those reported by Cruchaga *et al.*[1] and we showed a nominally significant association of *PLD3*-Val232Met with AD. This is in contrast to findings by Heilmann *et al.*, Lambert *et al.* and Hooli *et al.*[2-4] However, in contrast to Cruchaga *et al.*[1] we found no significant association of other *PLD3*-variants with AD in the single variant or gene based analyses. This is consistent with companion brief communications and hence, our data do not strongly support an important contribution of rare *PLD3*-variants in the etiology of AD. The most notable finding in our study is the high variability of the frequency of *PLD3*-Val232Met across populations, which highlights the need for careful matching of cases and controls for ethnic background when investigating rare variants.

Methods

RS,[5] FHS [6] and AGES [7] were genotyped on the Illumina exome chip.[8] Amsterdam Dementia cohort[9], Alzheimer Center Erasmus MC and RS underwent whole exome sequencing at Center for Biomics, Rotterdam. RS exome sequence and exome chip data partially overlapped, genotypes were concordant and non-overlapping samples were used in *PLD3*-Val232Met analysis. As covariates gender and age at onset for AD cases or

the date of last examination/censoring for cognitively healthy controls were used. GRIP [10] was imputed using the Dutch specific reference panel (Imputation quality [Rsq]=0.74).[11,12] Subjects aged below 55 were excluded. The R-package "seqMeta"(version seqMeta_1.4) was used for meta-analysis of single variant score test and gene-based test.

References:

1 Cruchaga, C. et al. Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. Nature 505, 550-554, doi:10.1038/nature12825 (2014).

2 Hooli, B. V. et al. PLD3 gene variants and Alzheimer's disease. Nature 520, E7-E9, doi:10.1038/nature14040 (2015).

3 Heilmann, S. et al. PLD3 in non-familial Alzheimer's disease. Nature 520, E3-E5, doi:10.1038/nature14039 (2015).

4 Lambert, J. C. et al. PLD3 and sporadic Alzheimer's disease risk. Nature 520, E1-E1, doi:10.1038/nature14036 (2015).

5 Hofman, A. et al. The Rotterdam Study: 2014 objectives and design update. Eur J Epidemiol 28, 889-926, doi:10.1007/s10654-013-9866-z (2013).

6 Splansky, G. L. et al. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. Am J Epidemiol 165, 1328-1335, doi:kwm021 [pii]
10.1093/aje/kwm021 (2007).

7 Harris, T. B. et al. Age, Gene/Environment Susceptibility-Reykjavik Study: multidisciplinary applied phenomics. Am J Epidemiol 165, 1076-1087, doi:kwk115 [pii]
10.1093/aje/kwk115 (2007).

8 Grove, M. L. et al. Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. PLoS One 8, e68095, doi:10.1371/journal.pone.0068095 (2013).

9 van der Flier, W. M. et al. Optimizing patient care and research: the Amsterdam Dementia Cohort. J Alzheimers Dis 41, 313-327, doi:10.3233/JAD-132306 (2014).

10 Liu, F. et al. A genomewide screen for late-onset Alzheimer disease in a genetically isolated Dutch population. Am J Hum Genet 81, 17-31, doi:S0002-9297(07)62813-4 [pii]
10.1086/518720 (2007).

11 Deelen, P. et al. Improved imputation quality of low-frequency and rare variants in European samples using the 'Genome of The Netherlands'. Eur J Hum Genet, doi:ejhg201419 [pii]
10.1038/ejhg.2014.19 (2014).

12 The Genome of the Netherlands, C. Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nat Genet, doi:ng.3021 [pii]
10.1038/ng.3021 (2014).

Chapter 3.3

Rare functional variant in *TM2D3* is associated with late-onset Alzheimer's disease

Johanna Jakobsdottir*, Sven J. van der Lee*, Joshua C. Bis*, Vincent Chouraki*, David Li-Kroeger, Shinya Yamamoto, Megan L. Grove, Adam Naj, Maria Vronskaya, Jose L. Salazar, Anita L. DeStefano, Jennifer A. Brody, Albert V. Smith, Najaf Amin, Rebecca Sims, Carla A. Ibrahim-Verbaas, Seung-Hoan Choi, Claudia L. Satizabal, Oscar L. Lopez, Alexa Beiser, M. Arfan Ikram, Melissa E. Garcia, Caroline Hayward, Tibor V. Varga, Samuli Ripatti, Paul W. Franks, Göran Hallmans, Olov Rolandsson, Jan-Håkon Jansson, David J. Porteous, Veikko Salomaa, Gudny Eirkisdottir, Kenneth M. Rice, Hugo J. Bellen, Daniel Levy, Andre G. Uitterlinden, Valur Emilsson, Jerome I. Rotter, Thor Aspelund, Cohorts for Heart and Aging Research in Genomic Epidemiology consortium, Alzheimer's Disease Genetic Consortium, Genetic and Environmental Risk in Alzheimer's Disease consortium, Christopher J. O'Donnell, Annette L. Fitzpatrick, Lenore J. Launer, Albert Hofman, Li-San Wang, Julie Williams, Gerard D. Schellenberg, Eric Boerwinkle**, Bruce M. Psaty**, Sudha Seshadri**, Joshua M. Shulman**, Vilmundur Gudnason**, Cornelia M. van Duijn**.

* authors contributed equally

** senior authors contributed equally

PLOS Genetics October 20, 2016: e1006327

Abstract

We performed an exome-wide association analysis in 1393 late-onset Alzheimer's disease (LOAD) cases and 8141 controls from the CHARGE consortium. We found that a rare variant (P155L) in *TM2D3* was enriched in Icelanders (~0.5% versus <0.05% in other European populations). In 433 LOAD cases and 3903 controls from the Icelandic AGES sub-study, P155L was associated with increased risk and earlier onset of LOAD [odds ratio (95% CI) = 7.5 (3.5-15.9), p =6.6×10^{-9}]. Mutation in the *Drosophila TM2D3* homolog, *almondex*, causes a phenotype similar to loss of Notch/Presenilin signaling. Human *TM2D3* is capable of rescuing these phenotypes, but this activity is abolished by P155L, establishing it as a functionally damaging allele. Our results establish a rare *TM2D3* variant in association with LOAD susceptibility, and together with prior work suggests possible links to the β-amyloid cascade.

Introduction

Alzheimer's disease (AD), the most common form of dementia, affects more than 10% of those 65 years and older, increasing to 30% in those 85 years and older.[1-4] Pathologically, AD is characterized by extensive brain neurodegenerative cell loss in association with extra-cellular β-amyloid plaques and intra-neuronal tangles consisting of hyper-phosphorylated tau protein. Multiple mutations in the amyloid-β precursor protein (*APP*) and the presenilin-1 and -2 (*PSEN1* and *PSEN2*) genes cause familial early-onset (<65 years) AD.[5,6] Genome-wide association studies (GWAS) have also identified numerous common genetic variants of modest effect sizes for late-onset AD (LOAD).[7-12] However, Apolipoprotein E (*APOE*) still remains the most important known genetic determinant of LOAD susceptibility.[13,14] Recently, rare variants with effect sizes similar to the *APOE*-ε4 allele have also been identified. While the population-attributable risk of such single rare variants is low, their discoveries may have important implications for understanding disease mechanisms and developing novel treatments.[15] Recent examples include a rare protective allele of *APP* [16] as well as variants in the novel genes *TREM2*,[17,18] *PLD3,*[19] *UNC5C*,[20] and *AKAP9*.[21] These findings highlight potentially important, new cellular pathways relevant for disease pathophysiology[15]; for example *TREM2* has spurred intense recent interest in the role of microglia in LOAD.[22,23] Importantly, because of population histories and demography, rare variants may be population-specific. The protective *APP* variant, for example, is found predominantly in Iceland and other Scandinavian populations.[16] The two variants in *AKAP9* have only been reported in African-Americans[21] and *PLD3* variants appear to vary greatly in frequency between populations.[24]

To date, the identification of rare susceptibility variants in LOAD has been hampered by poor representation on genotyping arrays used for large GWAS[25]; moreover, direct sequencing in large numbers of individuals remains costly. Here, using an exome-wide genotyping array (the Illumina HumanExome Beadchip), we report associations with LOAD in four population-based cohorts from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium.[26]

Methods

Studies and participants

Four studies from the CHARGE consortium genotyped a total of 1393 AD cases and 8141 cognitively intact controls for an ExomeChip (EC) genotyping array. All participants in the discovery phase of the analysis were of European or European American descent (**Table 1**). In the follow-up analysis we included Icelandic individuals from the AGES-followup cohort, African-American participants of the Cardiovascular Health Study (CHS), and European or European American individuals from the Alzheimer's Disease Genetics Consortium (ADGC) and Genetic and Environmental Risk in Alzheimer's Disease consortium (GERAD). Further details on phenotyping and other cohort characteristics are in the **Supporting methods**. All participants provided informed consent and all studies were approved by their respective ethics committees.

Genotyping

All four CHARGE cohorts were genotyped for the HumanExome BeadChip v1.0 from Illumina (San Diego, CA, USA). Genotype calling and quality control was performed centrally as described previously.[27] Each study also performed quality control locally (**Supporting methods**). Post-analysis quality control on the top results included visual inspection of cluster plots (**Supplementary Figure 1**) and re-genotyping of *TM2D3* carriers using a TaqMan assay (genotype calls were 100% validated). *TM2D3* variant was genotyped in the AGES-followup cohort using the TaqMan assay. In ADGC and GERAD consortia the *SKAP2* and *TM2D3* variants were genotyped on the Illumina HumanExome BeadChip v1.0 or v1.1 from Illumina (**Supporting methods**).

Statistical analysis

Statistical methods of the discovery phase

In the discovery phase of our analysis we used score tests [28] for the single-variant analysis and the Sequence Kernel Association Test (SKAT[29]) to test for the aggregate effect of multiple low-frequency variants within a gene on LOAD. Each study adjusted for age, sex, and for those principal components associated with LOAD (**Supporting methods**). The discovery analysis was not adjusted for APOE genotypes; however this

important LOAD risk factor was considered in secondary analyses of top findings. Both tests were performed in R with the seqMeta package (http://cran.r-project.org/web/packages/seqMeta/) to meta-analyze study-specific scores and respective variances and covariances (**Supporting methods**).[30] In the single-variant analysis we included variants present in at least two studies, variants with a minor allele frequency (MAF) ≥ 0.5% or assuming that rare damaging alleles could be prevalent predominantly in cases we also included variants with minor allele count (MAC) ≥ 5 in cases. This resulted in 52,026 single-variant tests. In the SKAT analysis we grouped the variants by genes using the start and stop positions as annotated by dbNSFP v2.0 [31] to define gene boundaries. We included only variants of MAF < 5% in the combined sample, and further annotated as missense, stop-gain, stop-loss, or splice-site variants. Then we required that genes have at least two such variants as well as at least two studies having a polymorphic variant. This resulted in 11,303 SKAT tests. After filtering, the Bonferroni correction thresholds that accounted for multiple testing were 9.6×10^{-7} for single-variant tests and 4.4×10^{-6} for SKAT. Two analysts independently performed the meta-analysis, leading to identical results.

Statistical methods in follow-up analysis of TM2D3 in AGES

We used a score test and Fisher's exact test for the follow-up analysis based on an independent Icelandic AGES sample (AGES-followup). We report (**Table 2**) estimates of the ORs based on fitting the full model (AD ~ age + sex + *TM2D3*) instead of the one-step approximation used in seqMeta (see **Supporting methods** for details). For the AGES-discovery cohort we similarly update and report estimates of the ORs (**Table 2**) based on fitting the full model (**Supplementary Table 1A**, **Supplementary Table 2**). In the AGES study, we also used a mixed-model to account for potential confounding effect relatedness might have on the association of P155L in *TM2D3* with LOAD (**Supporting results**). Cox regression was used to test for association of P155L with age-at-onset in the AGES data after adjusting for sex and *APOE* genotype. Individuals with LOAD diagnosis at the baseline visit were excluded so only at-risk individuals were included in the analysis. As detailed in the **Supporting methods**, the survival analysis accounted for left-truncation (i.e. follow-up begins after 65 years) and right-censoring (i.e. censoring that happens if a participant is lost to follow-up before having an event).

Drosophila

Detailed methods for our *Drosophila* experiments can be found in the **Supporting methods**. Briefly, the rescue activity of *amx, TM2D3,* or *TM2D3*^{PI55L} genomic constructs was examined in female flies of the genotype *amx*^{1}*/Df(1)Exel9049; [Rescue Trangene]*/+ or *amx*^{1}*/Df(1)Exel9049; [Rescue Trangene]* when crossed to *amx*^{1} males (with or without the *[Rescue Transgene]*).[32] To visualize the developing nervous system, resulting embryos were stained with anti-Hrp (1:1000),[33] a neuronal membrane marker and anti-ELAV (1:100),[34] a neuronal nuclear marker. For egg hatching, adults were allowed to lay eggs for 5 hours on grape juice agar plates, and larvae were counted 24 hours later.

Results

Exome-wide association study of LOAD in CHARGE

The discovery phase of our analysis examined single-variant associations with LOAD. Meta-analysis was performed across four CHARGE studies including 1393 LOAD cases and 8141 controls (**Table 1**), using logistic regression-based scores[28,30] and adjusting for age and sex (see **Methods** and **Supporting methods**). Single-variant results were filtered using quality control filters that included minimum minor allele frequency (MAF ≥ 0.5%) or minor allele count (MAC ≥ 5) in cases (see also **Methods**) resulting in 52026 total variants tested. Given the rarity of many variants captured on the exome array, we additionally performed association analysis, where we considered rare variants from the same gene in aggregate. For this complementary analysis we used the Sequence Kernel Association Test (SKAT) test,[29] a variance component score test for the association of a set of multiple variants with a trait. SKAT tests the null hypothesis of no variation of effects and therefore the statistical model of SKAT is applicable to test association of variant sets that may have a combination of variants including both risk and protective alleles or those with no effect. Variants were included in the SKAT analysis based on functional annotation and maximum frequency of 5%. Because the SKAT results were filtered by cumulative set-based MAF and MAC, considering all variants in aggregate, and not by minimum MAF of single SNPs, this complementary analysis included many variants that did not meet the criteria for inclusion in the single-variant tests. The results of both the single-variant and SKAT discovery analysis are

presented in **Supplementary Table 1** and quantile-quantile plots are in **Supplementary Figure 1**. Below, we discuss each of the top three results, our efforts to replicate the novel loci in independent cohorts, and for *TM2D3*, subsequent functional validation of the implicated variant in *Drosophila*.

APOE

As expected based on many prior LOAD GWAS,[12] three common intronic variants near *APOE* reached the Bonferroni threshold of exome-wide significance ($p \leq 9.6\times10^{-7}$) in the single-variant analysis: rs769449 in *APOE* ($p = 5.8\times10^{-38}$, $MAF_{case} = 17\%$, $MAF_{control} = 10\%$), rs2075650 in *TOMM40* ($p = 1.2\times10^{-24}$, $MAF_{case} = 18\%$, MAF in controls = 13%), and rs6859 in *PVRL2* ($p = 1.5\times10^{-9}$, $MAF_{case} = 46\%$, $MAF_{control} = 40\%$). Although *APOE* is a well-established LOAD risk locus it did not achieve nominal significance in the SKAT analysis, possible because the established common risk variants at this locus exceed our frequency threshold for consideration in the SKAT analysis (MAF < 0.05) (**Supplementary Table 2**).

SKAP2

Rs17154402 in *SKAP2* on chromosome 7p15 also showed a significant association with LOAD ($p = 2.1\times10^{-7}$, $MAF_{case}=0.22\%$, $MAF_{control}=0\%$). This SNP is predicted to introduce a non-synonymous amino acid change (S253T in the longer isoform) in the SKAP2 protein. The *SKAP2* locus also achieved significant association with LOAD ($p = 4.5\times10^{-7}$) in the SKAT analysis. Although the SKAT analysis included three distinct *SKAP2* missense alleles, the SKAT result was fully explained by the variant identified in the single-variant analysis (**Supplementary Table 2**). rs17154402 was not significantly associated with LOAD risk in the ADGC or GERAD cohorts, which are 2 large European-ancestry case-control datasets (13 carriers in 8,256 controls, MAF 0.08% and 9 carriers in 13,333 cases, MAF 0.03%, $p = 0.05$), and compared to the discovery CHARGE cohort, showed an opposite direction of effect in these samples. We noted that rs17154402 has a higher frequency in person of African compared to European ancestry (MAF~7% in Yorubans and 3% in all African populations combined in the 1000 Genomes Project data, and Grove et al.[27] reported a ~5% MAF in African populations versus 0.22% in European ancestry subjects). We therefore took advantage of an available African-

American subsample of the CHARGE CHS cohort for attempted replication. However, the variant again showed an inconsistent direction of effect and non-significant association (12 carriers in 93 cases; 33 carriers in 208 controls; $p = 0.66$). Based on the non-replication of *SKAP2* in available independent datasets, it was not pursued further.

TM2D3

Our discovery analysis also identified rs139709573, a missense variant in *TM2D3* on chromosome 15q26. Although the association statistic did not exceed the Bonferroni threshold for exome-wide significance in the CHARGE discovery analysis (single-variant $p = 2.0\times10^{-6}$, SKAT $p = 8.3\times10^{-6}$), several important observations led us to consider this variant further. First, we noted that while very rare (0-0.06% MAF in European ancestry populations (**Supplementary Table 3**)) the *TM2D3* variant was enriched nearly 10-fold in the Icelandic AGES cohort (0.45% MAF). Second, rs139709573 was significantly associated ($p = 5.9\times10^{-8}$, **Table 2**) with LOAD when the analysis was restricted to the AGES-discovery sample (143 cases, 2374 controls). Third, prior experimental studies potentially link *TM2D3* to AD-relevant biology. Specifically, *TM2D3* shares homology with the β-amyloid peptide binding protein (BBP or *TM2D1*),[35] and as discussed further below, genetic studies of the conserved *Drosophila* ortholog almondex (*amx*) strongly suggest links to γ-secretase function.[36]

In the SKAT analysis, the association of *TM2D3* with LOAD was fully accounted for by rs139709573; although, 3 other variants were considered in this analysis (**Supplementary Table 2**). In the subsequent analyses and functional follow-up, we therefore restricted our focus on rs139709573. First we evaluated potential inflation in the test statistic of the *TM2D3* rs139709573 result in the AGES-discovery cohort (**Supporting results**) and we further confirmed the genotypes of risk allele carriers by re-genotyping them with a TaqMan assay. We next directly genotyped rs139709573 in an independent Icelandic AGES-followup cohort (290 cases, 1529 controls), consisting of individuals who had not been genotyped previously with the exome array. The rs139709573 association was also detected in this sample ($p = 3.1\times10^{-3}$, **Table 2**). As expected, an analysis combining all of the AGES data on *TM2D3* demonstrated an enhanced association of the *TM2D3* variant with LOAD (OR = 7.45; 95% CI 3.49-15.90;

$p = 6.6\times10^{-9}$). We also observed an association with age-at-onset in the AGES cohorts (hazard ratio = 5.3; 95% CI 2.7-10.5; $p = 1.1\times10^{-6}$, see also non-parametric Kaplan-Meier curves in **Figure 1**). Using the PHASE program [37,38] we estimated that rs139709573 resides on a single haplotype in the Icelandic population (**Supporting results**), and we further excluded potential confounding due to cryptic relatedness (**Supporting results**). Similarly, the association between rs139709573 and LOAD was robust to adjustment for *APOE* genotype (**Table 2**). Given the low frequency of rs139709573 in individuals of European ancestry, there were limited numbers of carriers in GERAD and ADGC (11 carriers in 13333 cases, 0.041% MAF; 4 carriers in 8256 controls, 0.024% MAF). While, the association was not statistically significant ($p = 0.4$), the direction of effect was consistent with that observed in AGES (OR = 1.7, 95% CI 0.5-7.3).

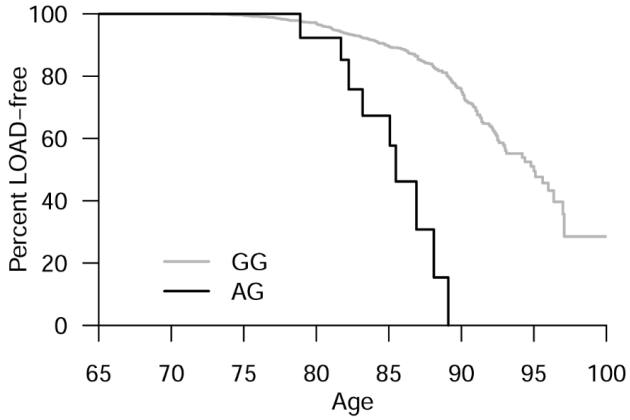


Figure 1. Kaplan-Meier survival curves for *TM2D3*. Curves are based on incident data only. There was no evidence of bias due to competing risk of death (**Supporting results**).

Analysis of *TM2D3* transcripts and expression

The *TM2D3* gene has several alternatively spliced transcripts, including six protein-coding transcripts.[39] rs139709573 falls within an exon common to all six isoforms (**Supplementary Figure 2**) and causes a Proline to Leucine amino acid change (P155L in the longest isoform). Gene and transcript expression estimates were extracted from publically available Genotype-Tissue Expression consortium (GTEx) data,[40], estimated as "transcripts per million" (TPM), using the Toil workflow [41] in the UCSC Xena browser

Table 1. Sample characteristics of discovery cohorts.

	AGES		FHS		CHS		RS	
Characteristics	**Controls**	**Cases**	**Controls**	**Cases**	**Controls**	**Cases**	**Controls**	**Cases**
Study Design	Population-based cohort		Population-based cohort		Population-based cohort		Population-based cohort	
No. participants	2374	143	1338	230	2013	557	2416	463
Women, No (%)	1401 (59)	85 (59)	754 (56)	158 (69)	1134 (56)	343 (62)	1227(51)	319(69)
Age, mean (SD), year	78.89 (4.98)	82.50 (4.94)	79.84 (8.57)	85.06 (6.90)	81.18 (5.15)	82.1 (5.32)	78.2(7.71)	83.3(6.59)
APOE ε4+, No. (%)	638 (27)	66 (46)	258 (20)	74 (32)	397 (20)	176 (32)	609(26)	190(43)

AGES: Age, Gene/Environment Susceptibility study, FHS: Framingham Heart Study, CHS: Cardiovascular Health Study, RS: Rotterdam Study.

Table 2. Sample characteristics and association results for P155L (rs139709573) in *TM2D3* in the two Icelandic AGES cohorts.

Cohort	**Group**	**No (% women)**	**Age, mean (SD), year**	***APOE* ε4+ No (%)**	***TM2D3* carriers, No (%)**	***p* (OR, 95% CI)[a]**	***p*-Fisher's Exact[b]**	**p-conditional[c]**
AGES-discovery	Cases	143 (59)	82.5 (4.9)	66 (46)	7 (4.9)	5.9×10-8 (8.62, 3.43-21.68)	5.6×10-4	8.4×10-8
	Controls	2374 (59)	78.9 (5.0)	638 (27)	20 (0.8)	...	...	...
AGES-follow-up	Cases	290 (59)	84.5 (5.1)	127 (44)	6 (2.1)	3.0×10-3 (5.42, 1.60-18.32)	6.2×10-3	1.1×10-2
	Controls	1529 (57)	79.4 (5.3)	396 (26)	6 (0.4)	...	...	...
Pooled[d]	Cases	433 (59)	83.9 (5.2)	193 (45)	13 (3.0)	6.6×10-9 (7.45, 3.49-15.90)	5.9×10-5	6.8×10-8
	Controls	3903 (58)	79.1 (5.1)	1034 (26)	26 (0.7)	...	...	...

All *p*-values are based on the 2-sided alternative.

[a] Score tests adjusted for age and sex based on a logistic regression model. The unconditional MLE of the OR is reported based on fitting the full model including the SNP. Note, that those point estimates of OR could be inflated (Supplemental results in S1 Text).

[b] *p*-value from the Fisher's exact test for carrier status. For the meta-analysis the data were pooled.

[c] *p*-value conditional on *APOE* ε4 carrier status. Based on a score test after adjusting for age, sex, and *APOE* ε4.

[d] The two cohorts are pooled in a stratified analysis (stratified by cohort).

(http://xena.ucsc.edu/). *TM2D3* is expressed in all GTEx tissues, including from many brain regions (**Supplementary Figure 3**). The GTEx expression data further show that the alternative *TM2D3* transcripts are similarly expressed across diverse human tissues (**Supplementary Figure 4**).

Structure of TM2D3 protein

TM2D3 encodes a predicted double-pass transmembrane protein with evidence of evolutionary conservation (**Figure 2A**). The variant falls within the overall well-conserved C-Type Lectin domain and is adjacent to other invariant residues in the first predicted extracellular domain (**Figure 2B**).[36] While P155L is not predicted to be strongly damaging by PolyPhen,[42], SIFT,[43] or CADD (C-Score=7.3),[44]and cross-species alignments show that the Proline residue is not conserved (**Figure 2A**), the two amino-acids do have important property differences, consistent with a potentially non-conservative substitution. Proline, the only cyclic amino-acid, frequently resides on the surface of folded proteins, and can underlie structural "kinks" or bends. Proline is unique in its inability to form hydrogen bonds that stabilize alpha-helices and beta-sheets. By contrast with Proline, Leucine favors alpha-helical secondary structure, and is more commonly buried in the interior of folded protein structures.[45]

Functional validation of *TM2D3* (P155L) in Drosophila

In order to investigate the potential functional consequences of P155L and its possible link to AD, we turned to the fruit fly, *Drosophila melanogaster*. Human *TM2D3* and the homologous fly gene, *amx*, are 51% identical and 64% similar (**Figure 2A**). Mutations in *amx* cause a strong maternal effect neurogenic phenotype, characteristic of defective Notch signaling, and lead to embryonic lethality (**Figure 3A-B**).[32,46] Similar to the cleavage of APP to generate the pathogenic ß-amyloid peptide, Notch receptor signaling requires homologous cleavage by γ-secretase for receptor activation. In *Drosophila,* loss of *Notch* or its regulators (e.g. *presenilin*) cause similar neurogenic phenotypes, due to impaired lateral inhibition and the inappropriate differentiation of ectodermal tissue toward a neural fate.[47-49] Moreover, prior genetic epistasis experiments suggest that *amx* may function at the γ-secretase cleavage step.[36]

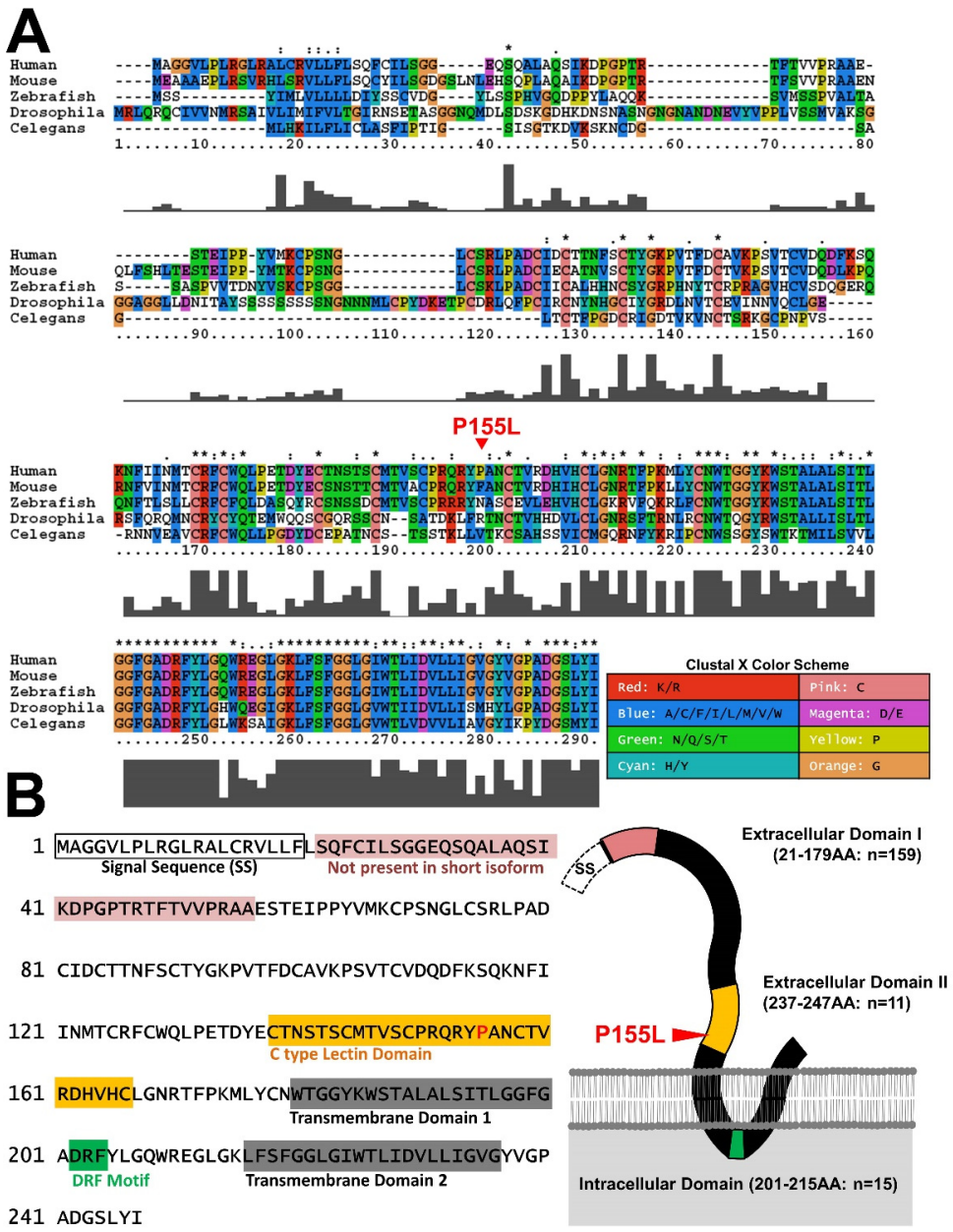


Figure 2. Sequence alignment and domain structure of TM2D3. (A) Sequence alignment of TM2D3 homologs shows overall strong conservation but lack of conservation for P155L. Sequence alignment of human (TM2D3), mouse (Tm2d3), zebrafish (tm2d3), *Drosophila* (amx) and *C. elegans* (C4IDI1.9) using Clustal X2.1 is shown. The conserved amino acids are highlighted according to the standard color scheme of Clustal X. (B) Primary sequence and schematic diagram of the domain structure of human TM2D3.

To determine whether *TM2D3* and *amx* have conserved molecular functions, we "humanized" the *Drosophila amx* gene by replacing its coding sequence with that of human *TM2D3* in the context of a genomic rescue construct (**Figure 3C**). Following establishment of stable transgenic lines, this construct was crossed into an *amx* mutant genetic background (*amx*^{1}*/Df(1)Exel9049)*. Consistent with prior reports,[32,46] when crossed to *amx* mutant males, all embryos laid by *amx* mutant females exhibited a strong neurogenic phenotype, failing to hatch from their eggs (**Figure 3B, 3D**). This embryonic defect and the resulting female sterility can be fully rescued by a genomic rescue construct with the wild-type (+) fly *amx* gene (**Figure 3D, 3E**). Due to the lethality of *Df(1)Exel9049/Y* hemizygous male progeny, complete rescue of the *amx* phenotype is predicted to lead to a maximum of ~75% egg hatching based on Mendelian expectations (**Figure 3D, Supplementary Figure 5**). Introduction of wild-type human *TM2D3*, under control of endogenous *amx* regulatory sequences, demonstrated significant rescue activity compared to *amx* mutant females (34.2% vs. 0% egg hatching $p<0.001$). Overall, based on the egg hatching assay (**Figure 3D**), the activity of the human *TM2D3* construct was estimated to be roughly half that observed for fly *amx*. Consistent with this, we observed a range in the severity of the neurogenic phenotype severity (**Supplementary Figure 6**), including embryos exhibiting a complete rescue (**Figure 3G**) and producing viable progeny that can develop into adult flies with no obvious morphological phenotypes. By contrast, an otherwise identical *TM2D3*^{P155L} genomic construct, but harboring the AD-associated P155L variant, was unable to rescue the neurogenic phenotype (**Figure 3H**) or the associated female sterility (**Figure 3D**). No animals hatched out of more than 900 eggs laid from *TM2D3*^{P155L} female flies. Consistent results were also seen in complementary assessments of rescue activity for the *amx* peripheral nervous system neurogenic phenotype (**Supplementary Figure 6**). In sum, our results demonstrate that human *TM2D3* can functionally substitute for fly *amx* in the context of embryonic Notch signaling, and that the P155L variant causes a loss-of-function in this context.

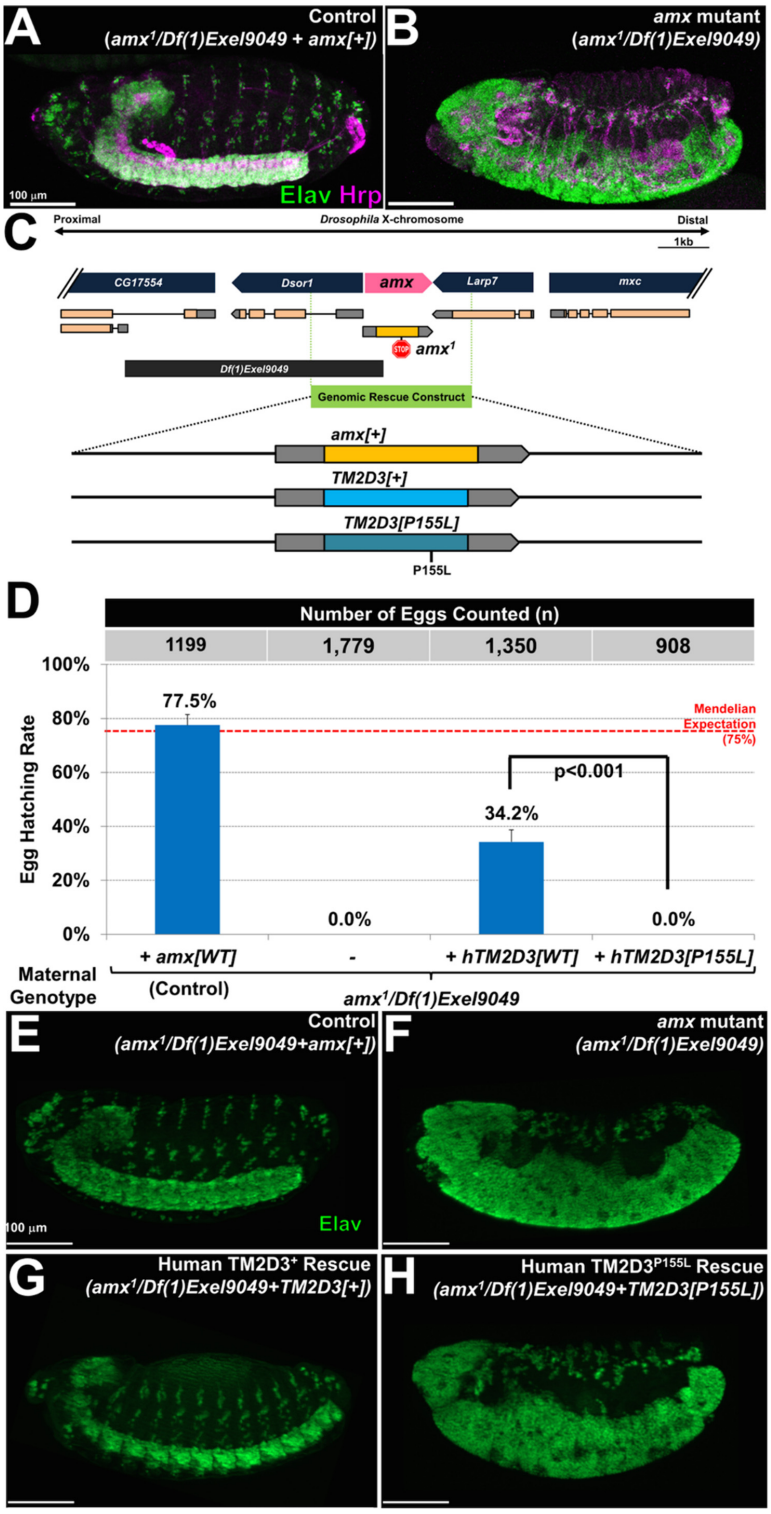

A
Control
(amx1/Df(1)Exel9049 + amx[+])
100 µm
Elav Hrp
B
amx mutant
(amx1/Df(1)Exel9049)
C
Proximal
Drosophila X-chromosome
Distal
1kb
CG17554
Dsor1
amx
Larp7
mxc
STOP
amx1
Df(1)Exel9049
Genomic Rescue Construct
amx[+]
TM2D3[+]
TM2D3[P155L]
P155L
D
Number of Eggs Counted (n)
1199
1,779
1,350
908
100%
80%
60%
40%
20%
0%
Egg Hatching Rate
77.5%
0.0%
34.2%
0.0%
Mendelian Expectation (75%)
p<0.001
+ amx[WT]
(Control)
-
+ hTM2D3[WT]
+ hTM2D3[P155L]
amx1/Df(1)Exel9049
Maternal Genotype
E
Control
(amx1/Df(1)Exel9049+amx[+])
100 µm
Elav
F
amx mutant
(amx1/Df(1)Exel9049)
G
Human TM2D3+ Rescue
(amx1/Df(1)Exel9049+TM2D3[+])
H
Human TM2D3P155L Rescue
(amx1/Df(1)Exel9049+TM2D3[P155L])

Figure 3. *TM2D3^{P155L}* is a loss-of-function allele. (A-B) Embryos laid by *amx* mutant females exhibit a strong neurogenic phenotype. (**A**) Lateral view of a late stage control embryo shows properly patterned and organized central and peripheral nervous system structures. (**B**) Embryos laid by amx mutant females show dramatic increase in the number of neurons, labeled by Elav (neuronal nucleus, green) and Hrp (neuronal membrane, magenta). (**C**) Schematic diagram of the *amx* locus and genomic rescue constructs generated for this study. *amx^{1}* contains an 8 nucleotide deletion that introduces a frameshift followed by a stop codon after residue 184 (Michellod and Randsholt, 2008). *Df(1)Exel9049* is a molecularly defined deletion that covers *amx* and two neighboring genes. The genomic rescue construct contains a ~3.3 kb fragment that can fully rescue the sterility of *amx^{1}/Df(1)Exel9049* mutant females and the neurogenic defects seen in the progeny (**A, E**). Coding region of Amx has been replaced by TM2D3 to "humanize" the fly *amx* gene. (D) Egg hatching assay reveals that *hTM2D3[+]* can partially suppress the female sterility of *amx^{1}/Df(1)Exel9049* mutant females, while *hTM2D3[P155L]* cannot. Due to the lethality of *Df(1)Exel9049/Y* hemizygous male progeny, complete rescue of the *amx* phenotype is expected to lead to a maximum of ~75% egg hatching, as denoted by the Mendelian Expectation line (also see Figure S3). (**E-H**) The developing nervous system is shown for embryos laid by *amx* mutant females with and without *amx[+]*, *TM2D3[+]*, and *TM2D3[P155L]* genomic rescue constructs. *TM2D3[+]* is capable of complete rescue of the *amx* neurogenic phenotype in some embryos (G), whereas all embryos with the *hTM2D3[P155L]* construct exhibit strong neurogenic phenotypes (H). Also see Figure S5 for assessment of rescue of the *amx* peripheral nervous system neurogenic phenotype. Scale bars = 100μm.

Discussion

We identified a novel LOAD-associated gene, *TM2D3*, harboring a rare missense mutation (P155L) that is associated with increased risk and earlier onset of LOAD diagnosis in an Icelandic population. Our experiments in *Drosophila* further suggest that this LOAD-associated P155L variant leads to a loss-of-function of TM2D3 in the context of Notch signaling during embryogenesis. While we could neither replicate nor negate the *TM2D3* finding in available European ancestry populations – possibly because of the low P155L allele frequency, a consistent association was demonstrated in an Icelandic cohort. Although the *TM2D3* variant failed to achieve the Bonferroni significance threshold in our CHARGE-wide meta-analysis our observation of increased frequency and enhanced association within the Icelandic AGES-discovery cohort highlighted the TM2D3 variant for further consideration. Significant association in the AGES follow-up sample coupled with variant functional validation in Drosophila suggests a role for TM2D3 in AD susceptibility. Our investigation thus exemplifies the importance not only for statistical rigor and transparency, but also flexibility in study design, when performing rare variant genetic association studies, in particular when statistical results are accompanied by relevant functional experiments.

TM2D3 has not previously been directly linked to AD. The encoded protein is predicted to contain two transmembrane regions, and additional domains with homology to C-Type lectins and G-protein coupled receptors (DRF motif) (**Figure 2B**). Since the P155L variant resides within the well-conserved C-Type lectin domain, the variant could plausibly disrupt the interaction of TM2D3 with a putative glycosylated binding partner. Alternatively, as suggested above, distinct structural properties of Leucine versus Proline may disrupt protein folding.[45] Interestingly, *TM2D3* shares homology with the β-amyloid peptide binding protein (BBP or *TM2D1*), which avidly binds to and sensitizes cells to toxicity caused by the β-amyloid peptide.[35] However, overexpression of TM2D3 (referred to as BLP2 in Kajkowski et al.[35]) was not associated with similar toxicity in cell culture. In mammals, TM2D3 it is highly expressed in the brain, including in the hippocampus, hypothalamus, and amygdala, potentially consistent with a role in AD pathogenesis (http://biogps.org/#goto=genereport&id=80213 and **Supplementary Figure 3**).[35,50,51] Additionally, micro-array analysis suggests that the expression of *TM2D3* is down-regulated in the hippocampus of AD cases compared to controls.[52]

While incompletely studied in mammals, loss-of-function mutants in the well-conserved *Drosophila* homolog of *TM2D3*, *amx,* cause embryonic phenotypes associated with disruptions in Notch signaling.[53] Further, genetic epistasis experiments support a potential role for *amx* in Notch receptor intramembranous proteolysis,[36] a process that is mediated by the presenilin/γ-secretase complex. One speculative hypothesis is that TM2D3/Amx participates either directly or indirectly in the intramembranous cleavage of Notch as well as in the homologous proteolytic processing of APP by γ-secretase. In this model, the P155L variant might potentially influence the generation of toxic β-amyloid peptides, similar to well-established disease susceptibility variants in *APP*, *PSEN1*, and *PSEN2*.[54]

Taking advantage of the potential functional conservation, we found that human TM2D3 expressed under the control of endogenous *Drosophila amx* regulatory elements is capable of rescuing the embryonic neurogenic phenotype, indicating that Amx and TM2D3 are orthologs and have conserved molecular functions *in vivo*. Importantly, we found that introduction of the LOAD-associated P155L variant into an otherwise

identical *TM2D3* genomic rescue construct abolishes the rescue activity. This result is consistent with either a partial or complete loss-of-function mechanism of the P155L variant, albeit within the heterologous context of Notch signaling in the *Drosophila* embryo. It is notable that the P155L variant was neither well-conserved nor predicted by bioinformatics as damaging; nevertheless, our cross-species approach convincingly demonstrates its functional impact. Our overall strategy of first rescuing the loss-of-function phenotype of the predicted homologous gene in flies, followed by examining the potential consequences of an implicated variant can be a powerful approach for the follow-up of many similar findings emerging from human genomic studies, including others using exome genotyping arrays or next generation sequencing approaches.[55,56]

In summary, we have identified a missense mutation in the *TM2D3* gene with a strong impact on LOAD risk. The *TM2D3* variant is enriched ~10-fold and associated with both risk and age-at-onset of LOAD in the Icelandic population. We further show that P155L is associated with a loss-of-function in the heterologous but potentially relevant context of Notch signaling in *Drosophila* embryos. We therefore speculate that TM2D3 may participate in the proteolytic processing of both Notch and APP, linking it to the amyloid cascade like other well-established AD susceptibility variants. Although we have demonstrated an association of the *TM2D3* variant only in the Icelandic population, our findings may thus have broader implications for understanding LOAD.

References:

1. Barker WW, Luis CA, Kashuba A, et al. Relative frequencies of Alzheimer disease, Lewy body, vascular and frontotemporal dementia, and hippocampal sclerosis in the State of Florida Brain Bank. *Alzheimer Dis Assoc Disord* 2002; **16**(4): 203-12.
2. Hebert LE, Weuve J, Scherr PA, Evans DA. Alzheimer disease in the United States (2010-2050) estimated using the 2010 census. *Neurology* 2013; **80**(19): 1778-83.
3. Alzheimer's A. 2015 Alzheimer's disease facts and figures. *Alzheimers Dement* 2015; **11**(3): 332-84.
4. James BD, Leurgans SE, Hebert LE, Scherr PA, Yaffe K, Bennett DA. Contribution of Alzheimer disease to mortality in the United States. *Neurology* 2014; **82**(12): 1045-50.
5. Cruts M, Theuns J, Van Broeckhoven C. Locus-specific mutation databases for neurodegenerative brain diseases. *Hum Mutat* 2012; **33**(9): 1340-4.
6. Bekris LM, Yu CE, Bird TD, Tsuang DW. Genetics of Alzheimer disease. *J Geriatr Psychiatry Neurol* 2010; **23**(4): 213-27.
7. Naj AC, Jun G, Beecham GW, 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.

8. Hollingworth P, Harold D, Sims R, et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet* 2011; **43**(5): 429-35.
9. Lambert JC, Heath S, Even G, et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet* 2009; **41**(10): 1094-9.
10. Harold D, Abraham R, Hollingworth P, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* 2009; **41**(10): 1088-93.
11. Seshadri S, Fitzpatrick AL, Ikram MA, et al. Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA* 2010; **303**(18): 1832-40.
12. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 2013; **45**(12): 1452-8.
13. Saunders AM, Strittmatter WJ, Schmechel D, et al. Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology* 1993; **43**(8): 1467-72.
14. Strittmatter WJ, Saunders AM, Schmechel D, 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 U S A* 1993; **90**(5): 1977-81.
15. Chouraki V, Seshadri S. Genetics of Alzheimer's disease. *Adv Genet* 2014; **87**: 245-94.
16. Jonsson T, Atwal JK, Steinberg S, et al. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* 2012; **488**(7409): 96-9.
17. Jonsson T, Stefansson H, Steinberg S, et al. Variant of TREM2 associated with the risk of Alzheimer's disease. *N Engl J Med* 2013; **368**(2): 107-16.
18. Guerreiro R, Wojtas A, Bras J, et al. TREM2 variants in Alzheimer's disease. *N Engl J Med* 2013; **368**(2): 117-27.
19. Cruchaga C, Karch CM, Jin SC, et al. Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. *Nature* 2014; **505**(7484): 550-4.
20. Wetzel-Smith MK, Hunkapiller J, Bhangale TR, et al. A rare mutation in UNC5C predisposes to late-onset Alzheimer's disease and increases neuronal cell death. *Nat Med* 2014; **20**(12): 1452-7.
21. Logue MW, Schu M, Vardarajan BN, et al. Two rare AKAP9 variants are associated with Alzheimer's disease in African Americans. *Alzheimers Dement* 2014; **10**(6): 609-18 e11.
22. Zhang B, Gaiteri C, Bodea LG, et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* 2013; **153**(3): 707-20.
23. Ulrich JD, Finn MB, Wang Y, et al. Altered microglial response to Abeta plaques in APPPS1-21 mice heterozygous for TREM2. *Mol Neurodegener* 2014; **9**: 20.
24. van der Lee SJ, Holstege H, Wong TH, et al. PLD3 variants in population studies. *Nature* 2015; **520**(7545): E2-E3.
25. Kiezun A, Garimella K, Do R, et al. Exome sequencing and the genetic basis of complex traits. *Nat Genet* 2012; **44**(6): 623-30.
26. Psaty BM, O'Donnell CJ, Gudnason V, et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ Cardiovasc Genet* 2009; **2**(1): 73-80.
27. Grove ML, Yu B, Cochran BJ, et al. Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. *PLoS One* 2013; **8**(7): e68095.
28. C. RR. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Mathematical Proceedings of the Cambridge Philosophical Society* 1948; **Volume 44**(Issue 1): 50-7.
29. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* 2011; **89**(1): 82-93.
30. Lumley T, Brody J, Dupuis J, Cupples A. Meta-analysis of a rare-variant association test. . 2012.
31. Liu X, Jian X, Boerwinkle E. dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. *Hum Mutat* 2013; **34**(9): E2393-402.

32. Shannon MP. Characterization of the female-sterile mutant almondex of Drosophila melanogaster. *Genetica* 1972; **43**(2): 244-56.
33. Snow PM, Patel NH, Harrelson AL, Goodman CS. Neural-specific carbohydrate moiety shared by many surface glycoproteins in Drosophila and grasshopper embryos. *J Neurosci* 1987; **7**(12): 4137-44.
34. Robinow S, White K. Characterization and spatial distribution of the ELAV protein during Drosophila melanogaster development. *J Neurobiol* 1991; **22**(5): 443-61.
35. Kajkowski EM, Lo CF, Ning X, et al. beta -Amyloid peptide-induced apoptosis regulated by a novel protein containing a g protein activation module. *J Biol Chem* 2001; **276**(22): 18748-56.
36. Michellod MA, Randsholt NB. Implication of the Drosophila beta-amyloid peptide binding-like protein AMX in Notch signaling during early neurogenesis. *Brain Res Bull* 2008; **75**(2-4): 305-9.
37. Stephens M, Scheet P. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *Am J Hum Genet* 2005; **76**(3): 449-62.
38. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001; **68**(4): 978-89.
39. Yates A, Akanni W, Amode MR, et al. Ensembl 2016. *Nucleic Acids Res* 2016; **44**(D1): D710-6.
40. Consortium GT, Ardlie KG, Deluca DS, et al. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science (New York, N Y)* 2015; **348**(6235): 648-60.
41. Vivian J, Rao A, Nothaft FA, et al. Rapid and efficient analysis of 20,000 RNA-seq samples with Toil. *bioRxiv Cold Spring Harbor Labs Journals* 2016.
42. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods* 2010; **7**(4): 248-9.
43. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009; **4**(7): 1073-81.
44. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014; **46**(3): 310-5.
45. Barnes MB, Gray IC. Bioinformatics for Geneticists; 2007.
46. Lehmann R DU, Jimnez F, Campos-Ortega JA. . Mutations of early neurogenesis in Drosophila. . *Wilhelm Roux's Arch Dev Biol* 1981; (190): 226–9.
47. Poulson DF. Chromosomal Deficiencies and the Embryonic Development of Drosophila Melanogaster. *Proc Natl Acad Sci U S A* 1937; **23**(3): 133-7.
48. Struhl G, Greenwald I. Presenilin is required for activity and nuclear access of Notch in Drosophila. *Nature* 1999; **398**(6727): 522-5.
49. Goriely A, Dumont N, Dambly-Chaudiere C, Ghysen A. The determination of sense organs in Drosophila: effect of the neurogenic mutations in the embryo. *Development* 1991; **113**(4): 1395-404.
50. Wu C, Orozco C, Boyer J, et al. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol* 2009; **10**(11): R130.
51. Su AI, Wiltshire T, Batalov S, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci U S A* 2004; **101**(16): 6062-7.
52. Zhang L, Guo XQ, Chu JF, Zhang X, Yan ZR, Li YZ. Potential hippocampal genes and pathways involved in Alzheimer's disease: a bioinformatic analysis. *Genet Mol Res* 2015; **14**(2): 7218-32.
53. Michellod MA, Forquignon F, Santamaria P, Randsholt NB. Differential requirements for the neurogenic gene almondex during Drosophila melanogaster development. *Genesis* 2003; **37**(3): 113-22.
54. Karch CM, Cruchaga C, Goate AM. Alzheimer's disease genetics: from the bench to the clinic. *Neuron* 2014; **83**(1): 11-26.
55. Shulman JM. Drosophila and experimental neurology in the post-genomic era. *Exp Neurol* 2015; **274**(Pt A): 4-13.

56. Bellen HJ, Yamamoto S. Morgan's legacy: fruit flies and the functional annotation of conserved genes. *Cell* 2015; **163**(1): 12-4.
57. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 2011; **12**: 323.
58. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 2016; **34**(5): 525-7.

Supporting methods and results

Supplementary Methods and Supplementary Tables can be accessed by scanning the following code or accessing the journals' website

Chapter 3.4

Rare coding variants in *PLCG2*, *ABI3* and *TREM2* implicate microglial-mediated innate immunity in Alzheimer's disease

Rebecca Sims*, Sven J. van der Lee*~, Adam C. Naj*, Céline Bellenguez*, Nandini Badarinarayan, Johanna Jakobsdottir, Brian W. Kunkle, Anne Boland, Rachel Raybould, Joshua C. Bis, Eden R. Martin, Benjamin Grenier-Boley, Stefanie Heilmann-Heimbach, Vincent Chouraki, Amanda B. Kuzma, Kristel Sleegers, Maria Vronskaya, Agustin Ruiz, Robert R. Graham, Robert Olaso, Per Hoffmann, Megan L. Grove, Badri N. Vardarajan, Mikko Hiltunen, Markus M. Nöthen, Charles C. White, Kara L. Hamilton-Nelson, Jacques Epelbaum, Wolfgang Maier, Seung-Hoan Choi, Gary W. Beecham, Cécile Dulary, Stefan Herms, Albert V. Smith, Cory C. Funk, Céline Derbois, Andreas J. Forstner, Shahzad Ahmad, Hongdong Li, Delphine Bacq, Denise Harold, Claudia L. Satizabal, Otto Valladares, Alessio Squassina, Rhodri Thomas, Jennifer A. Brody, Liming Qu, Pascual Sanchez-Juan, Taniesha Morgan, Frank J. Wolters, Yi Zhao, Florentino Sanchez Garcia, Nicola Denning, Myriam Fornage, John Malamon, Maria Candida Deniz Naranjo, Elisa Majounie, Thomas H. Mosley, Beth Dombroski, David Wallon, Michelle K Lupton, Josée Dupuis, Patrice Whitehead, Laura Fratiglioni, Christopher Medway, Xueqiu Jian, Shubhabrata Mukherjee, Lina Keller, Kristelle Brown, Honghuang Lin, Laura B. Cantwell, Francesco Panza, Bernadette McGuinness, Sonia Moreno-Grau, Jeremy D. Burgess, Vincenzo Solfrizzi, Petra Proitsi, Hieab H. Adams, Mariet Allen, Davide Seripa, Pau Pastor, L. Adrienne Cupples, Nathan D Price, Didier Hannequin, Ana Frank-García, Daniel Levy, Paramita Chakrabarty, Paolo Caffarra, Ina Giegling, Alexa S. Beiser, Vimantas Giedraitis, Harald Hampel, Melissa E. Garcia, Xue Wang, Lars Lannfelt, Patrizia Mecocci, Gudny Eiriksdottir, Paul K. Crane, Florence Pasquier, Virginia Boccardi, Isabel Henández, Robert C. Barber, Martin Scherer, Lluis Tarraga, Perrie M. Adams, Markus Leber, Yuning Chen, Marilyn S. Albert, Steffi Riedel-Heller, Valur Emilsson, Duane Beekly, Anne Braae, Reinhold Schmidt, Deborah Blacker, Carlo Masullo, Helena Schmidt, Rachelle S. Doody, Gianfranco Spalletta, WT Longstreth, Jr, Thomas J. Fairchild, Paola Bossù, Oscar L. Lopez, Matthew P. Frosch, Eleonora Sacchinelli, Bernardino Ghetti, Pascual Sánchez-Juan, Qiong Yang, Ryan M. Huebinger,

Frank Jessen, Shuo Li, M. Ilyas Kamboh, John Morris, Oscar Sotolongo-Grau, Mindy J. Katz, Chris Corcoran, Jayanadra J. Himali, C. Dirk Keene, JoAnn Tschanz, Annette L. Fitzpatrick, Walter A. Kukull, Maria Norton, Thor Aspelund, Eric B. Larson, Ron Munger, Jerome I. Rotter, Richard B. Lipton, María J Bullido, Albert Hofman, Thomas J. Montine, Eliecer Coto, Eric Boerwinkle, Ronald C. Petersen, Victoria Alvarez, Fernando Rivadeneira, Eric M. Reiman, Maura Gallo, Christopher J. O'Donnell, Joan S. Reisch, Amalia Cecilia Bruni, Donald R. Royall, Martin Dichgans, Mary Sano, Daniela Galimberti, Peter St George-Hyslop, Elio Scarpini, Debby W. Tsuang, Michelangelo Mancuso, Ubaldo Bonuccelli, Ashley R. Winslow, Antonio Daniele, Chuang-Kuo Wu, GERAD/PERADES, CHARGE, ADGC, EADI, Oliver Peters, Benedetta Nacmias, Matthias Riemenschneider, Reinhard Heun, Carol Brayne, David C Rubinsztein, Jose Bras, Rita Guerreiro, John Hardy, Ammar Al-Chalabi, Christopher E Shaw, John Collinge, David Mann, Magda Tsolaki, Jordi Clarimón, Rebecca Sussams, Simon Lovestone, Michael C O'Donovan, Michael J Owen, Timothy W. Behrens, Simon Mead, Alison M. Goate[a], Andre G. Uitterlinden, Clive Holmes, Carlos Cruchaga, Martin Ingelsson, David A. Bennett, John Powell, Todd E. Golde, Caroline Graff, Philip L. De Jager, Kevin Morgan, Nilufer Ertekin-Taner, Onofre Combarros, Bruce M. Psaty, Peter Passmore, Steven G Younkin, Claudine Berr, Vilmundur Gudnason, Dan Rujescu, Dennis W. Dickson, Jean-Francois Dartigues, Anita L. DeStefano, Sara Ortega-Cubero, Hakon Hakonarson, Dominique Campion, Merce Boada, John "Keoni" Kauwe, Lindsay A. Farrer, Christine Van Broeckhoven, M. Arfan Ikram, Lesley Jones, Johnathan Haines, Christophe Tzourio, Lenore J. Launer, Valentina Escott-Price, Richard Mayeux, Jean-François Deleuze, Najaf Amin, Peter A Holmans, Margaret A. Pericak-Vance, Philippe Amouyel**, Cornelia M. van Duijn**, Alfredo Ramirez**, Li-San Wang**, Jean-Charles Lambert**, Sudha Seshadri**, Julie Williams**~, Gerard D. Schellenberg**~.

* equal contribution first author

** equal contribution senior author

~ corresponding author

Nature Genetics 2017 July 17 [online publication ahead of print]

Abstract

We identified rare coding variants associated with Alzheimer's disease (AD) in a 3-stage case-control study of 85,133 subjects. In stage 1, 34,174 samples were genotyped using a whole-exome microarray. In stage 2, we tested associated variants ($p < 1\times10^{-4}$) in 35,962 independent samples using *de novo* genotyping and imputed genotypes. In stage 3, an additional 14,997 samples were used to test the most significant stage 2 associations ($P<5\times10^{-8}$) using imputed genotypes. We observed 3 novel genome-wide significant (GWS) AD associated non-synonymous variants; a protective variant in *PLCG2* (rs72824905/p.P522R, $p = 5.38\times10^{-10}$, OR = 0.68, $MAF_{cases} = 0.0059$, $MAF_{controls} = 0.0093$), a risk variant in *ABI3* (rs616338/p.S209F, $p = 4.56\times10^{-10}$, OR = 1.43, $MAF_{cases} = 0.011$, $MAF_{controls} = 0.008$), and a novel GWS variant in *TREM2* (rs143332484/p.R62H, $p = 1.55\times10^{-14}$, OR = 1.67, $MAF_{cases} = 0.0143$, $MAF_{controls} = 0.0089$), a known AD susceptibility gene. These protein-coding changes are in genes highly expressed in microglia and highlight an immune-related protein-protein interaction network enriched for previously identified AD risk genes. These genetic findings provide additional evidence that the microglia-mediated innate immune response contributes directly to AD development.

Late-onset AD (LOAD) has a significant genetic component (h^2 = 58-79%[1]). Nearly 30 LOAD susceptibility loci[2–12] are known, and risk is significantly polygenic[13]. However, these loci explain only a proportion of disease heritability. Rare variants also contribute to disease risk.[14–17] Recent sequencing studies identified a number of genes that have rare variants associated with AD.[9–11,18–24] Our approach to rare-variant discovery is to genotype a large sample with micro-arrays targeting known exome variants with follow-up using genotyping and imputed genotypes in a large independent sample. This is a cost-effective alternative to *de novo* sequencing.[25–29]

We applied a 3-stage design (**Supplementary Figure 1**) using subjects from the International Genomics of Alzheimer's Project (IGAP)(**Table 1**, **Supplementary Tables 1 & 2**). In stage 1, 16,097 LOAD cases and 18,077 cognitively normal elderly controls were genotyped using the Illumina HumanExome microarray. Data from multiple consortia were combined in a single variant meta-analysis (**Supplementary Methods**) assuming an additive model. In total, 241,551 variants passed quality-control (**Supplementary Table 3**). Of these 203,902 were polymorphic, 26,947 were common (minor allele frequency (MAF) ≥ 5%), and 176,955 were low frequency or rare (MAF < 5%). We analyzed common variants using a logistic regression model in each sample cohort and combined data using METAL.[30] Rare and low frequency variants were analyzed using the score test and data combined with SeqMeta[31] (**Supplementary Figure 2**).

We reviewed cluster plots for variants showing association ($p < 1\times10^{-4}$) and identified 43 candidate variants (**Supplementary Table 4**) exclusive of known risk loci (**Supplementary Table 5**). Stage 2 tested these for association in 14,041 LOAD cases and 21,921 controls, using *de novo* and imputation derived genotypes (**Supplementary Methods**). We carried forward single nucleotide variants (SNVs) with GWS associations and consistent directions of effect to stage 3 where genotypes for 6,652 independent cases and 8,345 controls were imputed using the Haplotype Reference Consortium resource[32,33] (**Supplementary Methods**, **Supplementary Table 6**).

Table 1. Summary of the consortium data sets used for stages 1, 2 and stage 3.

	Consortium	N Controls	N Cases	N Total
Stage 1	GERAD/PERADES	2,974	6,000	8,974
	ADGC	7,002	8,706	15,708
	CHARGE	8,101	1,391	9,492
Total		18,077	16,097	**34,174**
Stage 2	GERAD/PERADES genotype	5,049	4,049	9,098
	CHARGE-genotype	1,839	1,434	3,273
	CHARGE-*in silico*	3,246	722	3,968
	EADI-genotype	11,787	7,836	19,623
Total		21,921	14,041	**35,962**
Stage 3	ADGC-*in silico*	8,345	6,652	**14,997**
Stage 1 + 2 + 3 Total		**48,402**	**37,022**	**85,133**

Data are from the Genetic and Environmental Risk for Alzheimer's Disease (GERAD)/Defining Genetic, Polygenic and Environmental Risk for Alzheimer's Disease (PERADES) Consortium, the Alzheimer's Disease Genetic Consortium (ADGC), the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) and the European Alzheimer's disease Initiative (EADI)(Supplement 1).

We identified four rare coding variants with GWS association signals with LOAD ($p < 5\times10^{-8}$)(**Table 2**, **Supplementary Tables 7 & 8**). The first is a missense variant p.P522R ($p = 5.38\times10^{-10}$, OR = 0.68) in *Phospholipase C Gamma 2* (*PLCG2*)(**Table 2**, **Figure 1A**, **Supplementary Table 9**, **Supplementary Figure 3**). This variant is associated with decreased risk of LOAD, showing a MAF of 0.0059 in cases and 0.0093 in controls. The reference allele (p.P522) is conserved across several species (**Supplementary Figure 4**). Gene-wide analysis showed nominal evidence for association at $p = 1.52\times10^{-4}$ (**Supplementary Tables 10 & 11**) and we found no other independent association at this gene (**Supplementary Figure 5**).

The second novel association is a missense change p.S209F ($p = 4.56\times10^{-10}$, OR = 1.43) in *ABI Family Member 3* (*ABI3*). The p.F209 variant shows consistent evidence for increasing LOAD risk across all stages, with a MAF of 0.011 in cases and 0.008 in controls (**Table 2**, **Figure 1B**, **Supplementary Table 12**, **Supplementary Figure 6**). The reference allele is conserved across multiple species (**Supplementary Figure 7**). Gene-wide analysis showed nominal evidence of association ($p = 5.22\times10^{-5}$)(**Supplementary Tables 10 & 11**). The *B4GALNT2* gene, adjacent to *ABI3,* contained

an independent suggestive association (**Supplementary Figure 8**), but this failed to replicate in subsequent stages ($p_{combined} = 1.68\times10^{-4}$) (**Supplementary Table 7**).

Table 2. Summary of stage 1, 2, 3 and combined meta-analysis results for SNVs at $P<5\times10^{-8}$. Data includes p-values, odds ratios (OR), minor allele frequency (MAF) in cases and controls and number of subjects included in each analytical stage. For OR 95% confidence intervals see **Supplementary Table 7**.

SNV	**rs75932628**	**rs143332484**	**rs72824905**	**rs616338**
Chr	6	6	16	17
Position	41129252	41129207	81942028	47297297
Protein Variation	R47H	R62H	P522R	S209F
Gene	*TREM2*	*TREM2*	*PLCG2*	*ABI3*
Effect Allele	T	T	G	T
Stage 1				
P	3.02×10^{-12}	3.48×10^{-9}	1.19×10^{-5}	2.16×10^{-5}
OR	2.46	1.58	0.65	1.42
MAF Cases	0.003	0.015	0.006	0.013
MAF Controls	0.001	0.010	0.011	0.010
N	30018	33786	33786	33786
Stage 2				
P	$4.38E\times10^{-8}$	3.66×10^{-7}	1.35×10^{-4}	8.37×10^{-5}
OR	2.37	3.97	0.70	1.41
MAF Cases	0.004	0.014	0.006	0.010
MAF Controls	0.002	0.006	0.008	0.008
N	35831	3968	35831	35831
Stage 3				
P	1.23×10^{-6}	2.45×10^{-3}	2.48×10^{-2}	1.75×10^{-2}
OR	2.58	1.55	0.69	1.58
MAF Cases	0.006	0.012	0.006	0.010
MAF Controls	0.003	0.008	0.007	0.008
N	14884	15288	15288	14876
Stage1, 2 and 3 Meta-Analysis				
P	5.38×10^{-24}	1.55×10^{-14}	5.38×10^{-10}	4.56×10^{-10}
OR	2.46	1.67	0.68	1.43
MAF Cases	0.004	0.014	0.006	0.011
MAF Controls	0.002	0.009	0.009	0.008
N	80733	53042	84905	84493

Note: Concordance for alternate allele carrier genotypes between imputed versus called SNPs in Stage 3 was 75.2% for rs75932628, 91.1% for rs143332484, 95.7% for rs72824905, and 81.9% for rs616338 (Supplementary Methods and **Supplementary Table** 6).

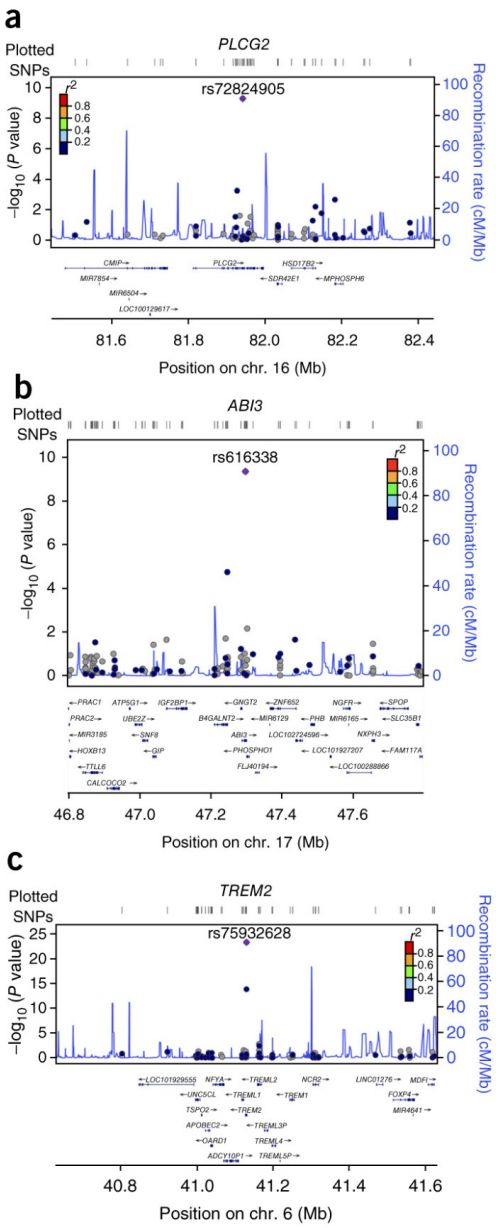


Figure 1. Association plots of *PLCG2*, *ABI3*, and *TREM2*. **(A)** Regional plot of identified association at the *PLCG2* locus. Top hit rs72824905 indicated in purple. Data presented for rs72824905 includes stage 1, stage 2 and stage 3 (N=84,905). **(B)** Regional plot of identified association at the *ABI3* locus. Top hit rs616338 indicated in purple. Data presented for rs616338 includes stage 1, stage 2 and stage 3 (N=84,493). **(C)** Regional plot of identified association at the *TREM2* locus. Top hit rs75932628 indicated in purple. Data presented for rs75932628 and rs143332484 includes stage 1, stage 2 and stage 3 (N=80,733 and 53,042, respectively). SNVs with missing LD information are shown in grey.

Following reports of suggestive association with LOAD,[34,35] we report the first evidence for GWS association at *TREM2* coding variant p.R62H ($p = 1.55\times10^{-14}$, OR = 1.67), with a MAF of 0.0143 in cases and 0.0089 in controls (**Table 2, Figure 1C, Supplementary Table 13**, **Supplementary Figures 9 & 10**). We also observed evidence for the previously reported[9,11] *TREM2* rare variant p.R47H (**Table 2**). These variants are not in linkage disequilibrium (**Supplementary Table 14**) and conditional analyses confirmed that p.R62H and p.R47H are independent risk variants (Supplementary Figure 11). Gene-wide analysis of *TREM2* showed a GWS association ($P_{SKAT} = 1.42\times10^{-15}$) (**Supplementary Tables 10 & 11**). Removal of p.R47H and p.R62H variants from the analysis diminished the gene-wide association but the signal remains interesting ($P_{SKAT\text{-}O}=6.3\times10^{-3}$, $P_{Burden}=4.1\times10^{-3}$). No single SNV was responsible for the remaining gene-wide association (**Supplementary Table 13**, **Supplementary Figure 11**) suggesting that there are additional *TREM2* risk variants in *TREM2*. We previously reported a common variant LOAD association near *TREM2*, in a GWAS of cerebrospinal fluid tau and P-tau.[36] We also observed a different suggestive common variant signal in another LOAD case-control study ($p = 6.3\times10^{-7}$).[2]

We previously identified 8 gene pathway clusters significantly enriched in AD-associated common variants.[36] To test whether biological enrichments observed in common variants are also present in rare variants we used the rare-variant data (MAF < 1%) to reanalyze these eight AD-associated pathway clusters (**Supplementary Methods, Supplementary Table 15**). We used Fisher's method to combine gene-wide p-values for all genes in each cluster. After correction for multiple testing, we observed enrichment for immune response ($p = 8.64\times10^{-3}$), cholesterol transport ($p = 3.84\times10^{-5}$), hemostasis ($p = 2.10\times10^{-3}$), Clathrin/AP2 adaptor complex ($p = 9.20\times10^{-4}$) and protein folding ($p = 0.02$). We also performed pathway analyses on the rare variant data presented here using all 9,816 pathways used previously. The top pathways are related to lipoprotein particles, cholesterol efflux, B-cell differentiation and immune response, areas of biology also enriched when common variants are analyzed[37] (**Supplementary Table 16**).

Previous analysis of normal brain co-expression networks identified 4 gene modules that are enriched for common variants associated with LOAD risk.[2-11] These 4 modules are enriched for immune response genes. We identified 151 genes present in 2 or more of these 4 modules and these showed a strong enrichment for LOAD-associated common variants ($p = 4.0\times10^{-6}$)[36] and for rare variants described here (MAF<1%) (**Supplementary Table 15**, $p = 1.17\times10^{-6}$). We then used a set of high-quality protein-protein interactions[37] to construct, from these 151 genes, an interaction network containing 56 genes, including *PLCG2*, *ABI3* and *TREM2* (**Figure 2**)(**Supplementary Methods**). This subset is strongly enriched for association signals from both the previous common variant analysis ($p = 5.0\times10^{-6}$, **Supplementary Table 17**) and this rare variant gene-set analysis ($p = 1.08\times10^{-7}$, **Supplementary Table 15**). The remaining 95 genes only have nominally-significant enrichment for either common or rare variants (**Supplementary Tables 15 & 17**), suggesting that the 56-gene (**Supplementary Table 18**) network is driving the enrichment.

TREM2, *ABI3* and *PLCG2* have a common expression pattern in human brain cortex, with high expression in microglia cells and limited expression in neurons, oligodendrocytes, astrocytes and endothelial cells (Supplementary Figure 12)[38]. Other known LOAD loci with the same expression pattern include *SORL1,* the *MS4A* gene cluster, and *HLA-DRB1*. *PLCG2*, *ABI3*, and *TREM2* are up-regulated in LOAD human cortex and in two APP mouse models. However, when corrected for levels of other microglia genes, these changes in expression appear to be related to microgliosis (**Supplementary Tables 19 & 20**).

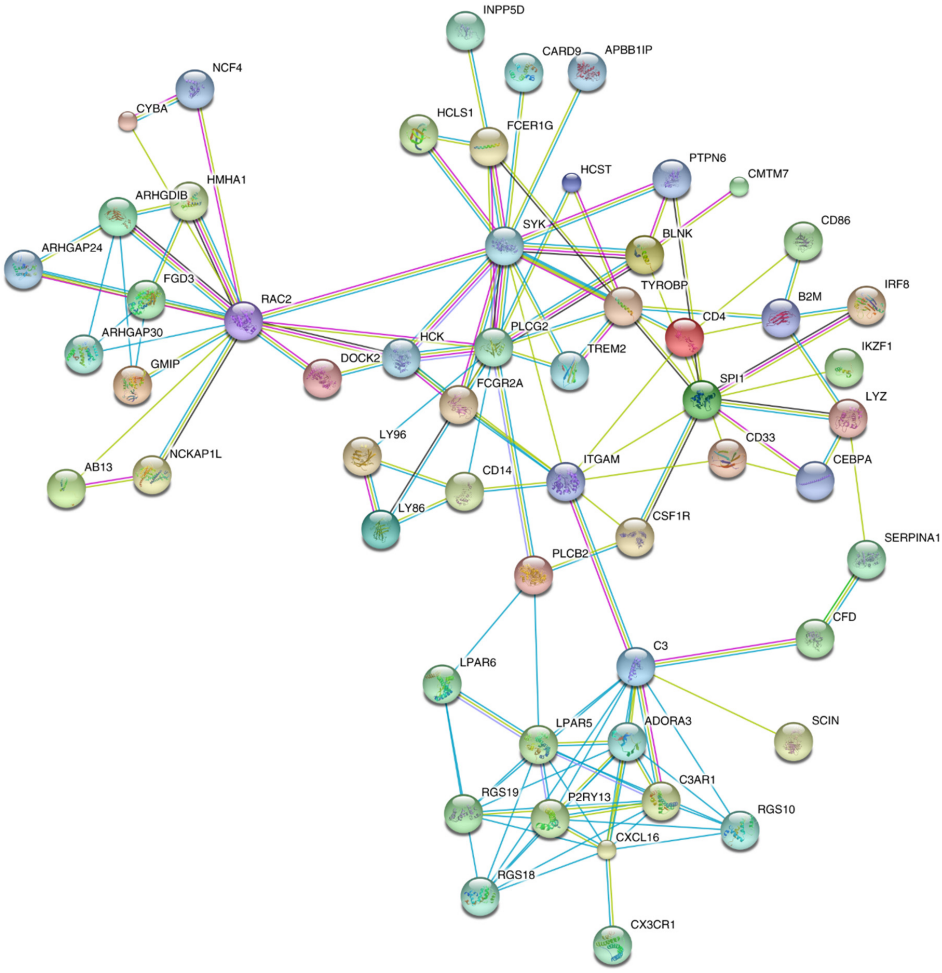


Figure 2. Protein-protein interaction network (using high-confidence human interactions from the STRING database) of 56 genes enriched for both common and rare variants associated with AD risk. Colours of edges refer to the type of evidence linking the corresponding proteins: red=gene fusion, dark blue = co-occurrence, black = co-expression, magenta = experiments, cyan=databases, light green = text mining, mauve = homology. *TREM2*, *PLCG2* and *ABI3* highlighted by red circles, *SYK*, *CSF1R* and *TYROBP* highlighted by blue circles, and *INPP5D*, *SPII* and *CD33* identified as common variant risk loci[2,5–7], highlighted by black circles.

PLCG2 (**Supplementary Figure 13**) encodes a transmembrane signaling enzyme (PLCγ2) that hydrolyses the membrane phospholipid PIP2 (1-phosphatidyl-1D-myo-inositol 4,5-bisphosphate) to secondary messengers IP3 (myo-inositol 1,4,5-trisphosphate) and DAG (diacylglycerol). IP3 is released into the cytosol and acts at the endoplasmic reticulum where it binds to ligand-gated ion channels to increase cytoplasmic Ca^{2+}. DAG remains bound to the plasma membrane where it activates two major signaling molecules, protein kinase C (PKC) and Ras guanyl nucleotide-releasing proteins (RasGRPs), which initiate the NF-κB and mitogen-activated protein kinase (MAPK) pathways. While the IP3/DAG/Ca+2 signaling pathway is active in many cells and tissues, in brain, *PLCG2* is primarily expressed in microglial cells. *PLCG2* variants also cause Antibody Deficiency and Immune Dysregulation (PLAID) and Autoinflammation and PLAID (APLAID).[39] Genomic deletions (PLAID) and missense mutations (APLAID) affect the cSH2 autoinhibitory regulatory region. The result is a complex mix of loss and gain of function in cellular signalling.[39]

Functional annotation (**Supplementary Table 21**) suggests *ABI3* (**Supplementary Figure 14**) plays a role in the innate immune response via interferon-mediated signaling.[40] *ABI3* is co-expressed with *INPP5D* (p = 2.2×10^{-10}), a gene previously implicated in LOAD risk.[2] ABI3 plays a significant role in actin cytoskeleton organization through participation in the WAVE2 complex,[41] a complex that regulates multiple pathways leading to T-cell activation.[42]

TREM2 encodes a transmembrane receptor present in the plasma membrane of brain microglia (**Supplementary Figure 15**). TREM2 protein forms an immune-receptor-signaling complex with DAP12. Receptor activation results in activation of Syk and ZAP70 signaling which in turn activates PI3K activity and influences PLCγ2 activity.[43] In microglia, TREM2-DAP12 induces an M2-like activation[44] and participates in recognition of membrane debris and amyloid deposits resulting in microglial activation and proliferation.[45–47] When *TREM2* knockout (KO) or *TREM2* heterozygous KO mice are crossed with *APP*-transgenics that develop plaques, the size and number of microglia associated with plaques are markedly reduced.[46,47] *TREM2* risk variants are located within exon 2, which is predicted to encode the conserved ligand binding

extracellular region of the protein. Any disruption in this region may attenuate or abolish TREM2 signaling, resulting in the loss or decrease in TREM2 function.[47] The 56-gene interaction network identified here is enriched in immune response genes and includes *TREM2*, *PLCG2*, *ABI3*, *SPI1*, *INPP5D*, *CSF1R*, *SYK* and *TYROBP* (**Figure 2**). *SPI1* is a central transcription factor in microglial activation state that has a significant gene-wide association with AD[5] and is in the proximity of GWS signals identified by IGAP.[2] Loss-of function mutations in *CSF1R* cause hereditary diffuse leukoencephalopathy with spheroids, a white matter disease related to microglial dysfunction[48]. Activated microglial cells surround plaques,[49,50] a finding consistently observed in AD brain and AD transgenic mouse models.[51] In AD mouse model brain, synaptic pruning associates with activated microglial signaling.[52] Pharmacological targeting of CSF1R inhibits microglial proliferation and shifts the microglialf inflammatory profile to an anti-inflammatory phenotype in murine models.[53] SYK regulates Aβ production and tau hyperphosphorylation,[54] is affected by the INPP5D/CD2AP complex[55] encoded by two LOAD associated genes,[2] and mediates phosphorylation of PLCG2.[56] Notably, the anti-hypertensive drug Nilvadipine, currently in a phase III AD clinical trial, targets *SYK* as well as *TYROBP,* a hub gene in an AD-related brain expression network,[38] that encodes the TREM2 complex protein DAP12.

We identified three rare coding variants in *PLCG2, ABI3* and *TREM2* with GWS associations with LOAD that are part of a common innate immune response. This work provides additional evidence that the microglial response in LOAD is directly part of a causal pathway leading to disease and is not simply a downstream consequence of neurodegeneration.[46,47,57,58] Our network analysis supports this conclusion. In addition, PLCγG2, as an enzyme, represents the first classically drug-able target to emerge from LOAD genetic studies. The variants described here account for a small portion of the 'missing heritability of AD'. The remaining heritability may be due to a large number of common variants of small effect size. For rare variants, there may be additional exonic sites with lower MAF or effect size, and/or intronic and intergenic sites. Complete resolution of AD heritability will be facilitated by larger sample sizes and more comprehensive sequence data.

References:

1. Gatz, M. *et al.* Role of genes and environments for explaining Alzheimer disease. *Arch. Gen. Psychiatry* **63,** 168–174 (2006).
2. Lambert, J. C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45,** 1452–1458 (2013).
3. Harold, D. *et al.* Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat. Genet.* **41,** 1088–1093 (2009).
4. Lambert, J.-C. *et al.* Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat. Genet.* **41,** 1094–1099 (2009).
5. Escott-Price, V. *et al.* Gene-wide analysis detects two new susceptibility genes for Alzheimer's disease. *PloS One* **9,** e94661 (2014).
6. Hollingworth, P. *et al.* Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat. Genet.* **43,** 429–435 (2011).
7. Naj, A. C. *et al.* Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat. Genet.* **43,** 436–441 (2011).
8. Ruiz, A. *et al.* TOWARD FINE MAPPING AND FUNCTIONAL CHARACTERIZATION OF GENOME-WIDE ASSOCIATION STUDY-IDENTIFIED LOCUS RS74615166 (TRIP4) FOR ALZHEIMER'S DISEASE. *Alzheimers Dement. J. Alzheimers Assoc.* **10,** P257–P258 (2014).
9. Jonsson, T. *et al.* Variant of TREM2 Associated with the Risk of Alzheimer's Disease. *N. Engl. J. Med.* **368,** 107–116 (2013).
10. Jonsson, T. *et al.* A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* **488,** 96–99 (2012).
11. Guerreiro, R. *et al.* TREM2 Variants in Alzheimer's Disease. *N. Engl. J. Med.* **368,** 117–127 (2013).
12. Seshadri, S. *et al.* Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA* **303,** 1832–1840 (2010).
13. Escott-Price, V. *et al.* Common polygenic variation enhances risk prediction for Alzheimer's disease. *Brain J. Neurol.* **138,** 3673–3684 (2015).
14. Bodmer, W. & Bonilla, C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat. Genet.* **40,** 695–701 (2008).
15. Pritchard, J. K. Are rare variants responsible for susceptibility to complex diseases? *Am. J. Hum. Genet.* **69,** 124–137 (2001).
16. Schork, N. J., Murray, S. S., Frazer, K. A. & Topol, E. J. Common vs. rare allele hypotheses for complex diseases. *Curr. Opin. Genet. Dev.* **19,** 212–219 (2009).
17. Surakka, I. *et al.* The impact of low-frequency and rare variants on lipid levels. *Nat. Genet.* **47,** 589–597 (2015).
18. Vardarajan, B. N. *et al.* Coding mutations in SORL1 and Alzheimer disease. *Ann. Neurol.* **77,** 215–227 (2015).
19. Vardarajan, B. N. *et al.* Rare coding mutations identified by sequencing of Alzheimer disease genome-wide association studies loci. *Ann. Neurol.* **78,** 487–498 (2015).
20. Steinberg, S. *et al.* Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. *Nat. Genet.* **47,** 445–447 (2015).
21. Logue, M. W. *et al.* Two rare AKAP9 variants are associated with Alzheimer's disease in African Americans. *Alzheimers Dement. J. Alzheimers Assoc.* **10,** 609–618.e11 (2014).
22. Jun, G. *et al.* PLXNA4 is associated with Alzheimer disease and modulates tau phosphorylation. *Ann. Neurol.* **76,** 379–392 (2014).
23. Hunkapiller, J. *et al.* A rare coding variant alters UNC5C function and predisposes to Alzheimer's disease. *Alzheimers Dement. J. Alzheimers Assoc.* **9,** P853 (2013).

24. Wetzel-Smith, M. K. *et al.* A rare mutation in UNC5C predisposes to late-onset Alzheimer's disease and increases neuronal cell death. *Nat. Med.* **20,** 1452–1457 (2014).
25. Richards, A. L. *et al.* Exome arrays capture polygenic rare variant contributions to schizophrenia. *Hum. Mol. Genet.* (2016). doi:10.1093/hmg/ddv620
26. Wessel, J. *et al.* Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. *Nat. Commun.* **6,** 5897 (2015).
27. Igartua, C. *et al.* Ethnic-specific associations of rare and low-frequency DNA sequence variants with asthma. *Nat. Commun.* **6,** 5965 (2015).
28. Tachmazidou, I. *et al.* A rare functional cardioprotective APOC3 variant has risen in frequency in distinct population isolates. *Nat. Commun.* **4,** 2872 (2013).
29. Huyghe, J. R. *et al.* Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. *Nat. Genet.* **45,** 197–201 (2013).
30. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26,** 2190–2191 (2010).
31. R Development Core Team. *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing).
32. Das, S. *et al.* Imputation server: next generation genotype imputation service. *Nat. Genet.*
33. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *bioRxiv* 35170 (2015). doi:10.1101/035170
34. Jin, S. C. *et al.* Coding variants in TREM2 increase risk for Alzheimer's disease. *Hum. Mol. Genet.* **23,** 5838–5846 (2014).
35. Lu, Y., Liu, W. & Wang, X. TREM2 variants and risk of Alzheimer's disease: a meta-analysis. *Neurol. Sci.* **36,** 1881–1888 (2015).
36. Cruchaga, C. *et al.* GWAS of cerebrospinal fluid tau levels identifies risk variants for Alzheimer's disease. *Neuron* **78,** 256–268 (2013).
37. International Genomics of Alzheimer's Disease Consortium (IGAP). Convergent genetic and expression data implicate immunity in Alzheimer's disease. *Alzheimers Dement. J. Alzheimers Assoc.* **11,** 658–671 (2015).
38. Zhang, Y. *et al.* Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. *Neuron* **89,** 37–53 (2016).
39. Milner, J. D. PLAID: a Syndrome of Complex Patterns of Disease and Unique Phenotypes. *J. Clin. Immunol.* **35,** 527–530 (2015).
40. Fairfax, B. P. *et al.* Innate Immune Activity Conditions the Effect of Regulatory Variants upon Monocyte Gene Expression. *Science* **343,** 1246949 (2014).
41. Sekino, S. *et al.* The NESH/Abi-3-based WAVE2 complex is functionally distinct from the Abi-1-based WAVE2 complex. *Cell Commun. Signal. CCS* **13,** (2015).
42. Nolz, J. C. *et al.* The WAVE2 Complex Regulates Actin Cytoskeletal Reorganization and CRAC-Mediated Calcium Entry during T Cell Activation. *Curr. Biol. CB* **16,** 24–34 (2006).
43. Xing, J., Titus, A. R. & Humphrey, M. B. The TREM2-DAP12 signaling pathway in Nasu-Hakola disease: a molecular genetics perspective. *Res. Rep. Biochem.* **5,** 89–100 (2015).
44. Neumann, H. & Takahashi, K. Essential role of the microglial triggering receptor expressed on myeloid cells-2 (TREM2) for central nervous tissue immune homeostasis. *J. Neuroimmunol.* **184,** 92–99 (2007).
45. Painter, M. M. *et al.* TREM2 in CNS homeostasis and neurodegenerative disease. *Mol. Neurodegener.* **10,** 43 (2015).
46. Ulrich, J. D. *et al.* In vivo measurement of apolipoprotein E from the brain interstitial fluid using microdialysis. *Mol. Neurodegener.* **8,** 13 (2013).
47. Wang, Y. *et al.* TREM2 lipid sensing sustains the microglial response in an Alzheimer's disease model. *Cell* **160,** 1061–1071 (2015).

48. Rademakers, R. *et al.* Mutations in the colony stimulating factor 1 receptor (CSF1R) gene cause hereditary diffuse leukoencephalopathy with spheroids. *Nat. Genet.* **44,** 200–205 (2012).
49. Perlmutter, L. S., Barron, E. & Chui, H. C. Morphologic association between microglia and senile plaque amyloid in Alzheimer's disease. *Neurosci. Lett.* **119,** 32–36 (1990).
50. Wisniewski, H. M., Wegiel, J., Wang, K. C. & Lach, B. Ultrastructural studies of the cells forming amyloid in the cortical vessel wall in Alzheimer's disease. *Acta Neuropathol. (Berl.)* **84,** 117–127 (1992).
51. Schwab, C., Klegeris, A. & McGeer, P. L. Inflammation in transgenic mouse models of neurodegenerative disorders. *Biochim. Biophys. Acta BBA - Mol. Basis Dis.* **1802,** 889–902 (2010).
52. Hong, S. *et al.* Complement and microglia mediate early synapse loss in Alzheimer mouse models. *Science* aad8373 (2016). doi:10.1126/science.aad8373
53. Olmos-Alonso, A. *et al.* Pharmacological targeting of CSF1R inhibits microglial proliferation and prevents the progression of Alzheimer's-like pathology. *Brain* awv379 (2016). doi:10.1093/brain/awv379
54. Paris, D. *et al.* The Spleen Tyrosine Kinase (syk) Regulates Alzheimer's Aβ Production and Tau Hyperphosphorylation. *J. Biol. Chem.* jbc.M114.608091 (2014). doi:10.1074/jbc.M114.608091
55. Bao, M. *et al.* CD2AP/SHIP1 complex positively regulates plasmacytoid dendritic cell receptor signaling by inhibiting the E3 ubiquitin ligase Cbl. *J. Immunol. Baltim. Md 1950* **189,** 786–792 (2012).
56. Kurosaki, T. & Tsukada, S. BLNK: Connecting Syk and Btk to Calcium Signals. *Immunity* **12,** 1–5 (2000).
57. Wang, Y. *et al.* TREM2-mediated early microglial response limits diffusion and toxicity of amyloid plaques. *J. Exp. Med.* **213,** 667–675 (2016).
58. Yuan, P. *et al.* TREM2 Haplodeficiency in Mice and Humans Impairs the Microglia Barrier Function Leading to Decreased Amyloid Compaction and Severe Axonal Dystrophy. *Neuron* **90,** 724–739 (2016).

Methods

Genotyping and Quality Control

Stage 1

GERAD/PERADES: Genotyping was performed at Life and Brain, Bonn, Germany, with the Illumina HumanExome BeadChip v1.0 (N = 247,870 variants) or v1.1 (N = 242,901 variants). Illumina's GenTrain version 2.0 clustering algorithm in GenomeStudio or zCall[1] was used for genotype calling. Quality control (QC) filters were implemented for sample call rate excluding samples with > 1% missingness, excess autosomal heterozygosity excluding outliers based on < 1% and > 1% minor allele frequency (MAF) separately, gender discordance, relatedness excluding one of each pair related with IBD ≥ 0.125 (the level expected for first cousins), and population outliers (i.e. non European ancestry). Variants were filtered based on call rate excluding variants with >1% missingness, genotype cluster separation excluding variants with a separation score <

0.4 and Hardy-Weinberg equilibrium (HWE) excluding variants with P_{HWE} <1×10^{-4}. Ten principal components (PCs) were extracted using EIGENSTRAT, including the first three PCs as covariates had the maximum impact on the genomic control inflation factor, λ^2. After QC 6,000 LOAD cases and 2,974 elderly controls (version 1.0; 4,093 LOAD cases and 1,599 controls, version 1.1; 1,907 LOAD cases and 1,375 controls) remained. The version 1.0 array had 244,412 variants available for analysis and 239,814 remained for the version 1.1 array.

CHARGE: All four CHARGE cohorts were genotyped for the Illumina HumanExome BeadChip v1.0. To increase the quality of the rare variant genotype calls, the genotypes for all four studies were jointly called with 62,266 samples from 11 studies at the University of Texas HSC at Houston.[3] Quality control (QC) procedures for the genotype data were performed both centrally at UT Houston and at each study. The central QC procedures have been described previously.[3] Minimum QC included: 1) Concordance checking with GWAS data and removal of problematic samples, 2) Removal of individuals with low genotype completion rate (< 90%), 3) Removal of variants with low genotype call rate (< 95%), 4) Removal of individuals with sex-mismatches, 5) Removal of one individual from duplicate pairs, 6) Removal of first-degree relatives based on genetically calculated relatedness (IBS > 0.45), with cases retained over controls, 7) Removal of variants not called in over 5% of the individuals and those that deviated significantly form the expected Hardy-Weinberg Equilibrium proportions (p < 1×10^{-6}).

ADGC: Genotyping was performed in subsets at four centers: NorthShore, Miami, WashU, and CHOP ("CHOP" and "ADC7" datasets) on the Illumina HumanExome BeadChip v1.0. One variant rs75932628 (p.R47H) in *TREM2* clustered poorly across all ADGC cohorts, and was therefore re-genotyped using a Taqman assay. Data on all samples underwent standard quality control procedures applied to genome-wide association studies (GWAS), including excluding variants with call rates < 95%, and then filtering samples with call rate < 95%. Variants with MAF > 0.01 were evaluated for departure from HWE and any variants for P_{HWE} < 10^{-6} were excluded. Population substructure within each of the five subsets (NorthShore, Miami, WashU, CHOP, and ADC7) was examined using PC analysis in EIGENSTRAT,[4] and population outliers (> 6

SD) were excluded from further analyses; the first three PCs were adjusted for as covariates in association testing. Prior to analysis we harmonized the alternate and reference alleles over all datasets. See **Supplementary Table 3** for an overview of cohort genotype calling and quality control procedures. All sample genotyping and quality control was performed blind to participant's disease status.

Stage 2

Twenty-two variants successfully designed for replication genotyping on the Agena Bioscience MassARRAY® platform. Genotyping was performed at Life and Brain, Bonn, Germany, and the Centre National de Génotypage (CNG), Paris, France. Twenty-one variants were successfully genotyped, with one variant (rs147163004 in *ASTN2*) failing visual cluster plot inspection. An additional nine variants were successfully genotyped using the Agena Bioscience MassARRAY® platform or Thermo FisherTaqMan® assay at the CNG, Paris, France in a subset of the replication samples N = 16,850 (7,755 cases, 9,095 controls).

GERAD/PERADES and ACE QC: Filters were implemented for sample call rate, excluding samples with > 10% missingness, and excess autosomal heterozygosity via visual inspection. Variants were filtered based on call rate excluding variants with > 10% missingness and HWE excluding variants with $P_{HWE} < 1\times10^{-5}$ in either cases or controls. *IGAP and EADI QC:* Variants were genotyped in 3 different panels and QC was performed in each panel separately. Samples with more than 3 missing genotypes were excluded, as were males heterozygous for X-Chromosome variants present within the genotyped panels. Variants were excluded based on missingness > 5%, HWE (in cases and controls separately) < 1×10^{-5}, and differential missingness between cases and controls < 1×10^{-5}, for each Country cohort. All variants passed quality control. PCs were determined using previously described methods.[19]

Stage 3

Replication was performed using genotypes from 23 ADGC datasets as described above. Genotyping arrays used have been described in detail before for most datasets, except for the CHAP, NBB, TARCC, and WHICAP datasets. CHAP and WHICAP datasets were genotyped on the Illumina OmniExpress-24 array, while NBB was genotyped on the

Illumina 1M platform. TARCC first wave subjects were genotyped using the Affymetrix 6.0 microarray chip, while subjects in the second wave (172 cases and 74 controls) were genotyped using the Illumina HumanOmniExpress-24 beadchip. Second wave TARCC subjects (TARCC2) were genotyped together with 84 cases and 115 controls from second wave samples ascertained at the University of Miami and Vanderbilt University. All samples used in stage 3 were imputed to the HRC haplotype reference panel,[5,6] which includes 64,976 haplotypes with 39,235,157 SNPs that allows imputation down to an unprecedented MAF=0.00008.

Prior to imputation, all genotype data underwent QC procedures that have been described extensively elsewhere.[7,8] Imputation was performed on the Michigan Imputation Server (https://imputationserver.sph.umich.edu/) running MiniMac3.[9,10] Genotypes from genome-wide, high-density SNP genotyping arrays for 16,175 AD cases and 17,176 cognitive-normal individuals were imputed. Across all samples 39,235,157 SNPs were imputed, with the actual number of SNPs imputed for each individual varying based on the regional density of array genotypes available. As a subset of these samples had also been genotyped as part of stage 1, we examined the imputation quality for critical variants by comparing imputed genotypes to those directly genotyped by the exome array; overall concordance was > 99%, while concordance among alternate allele genotypes (heterozygotes and alternate allele homozygotes) was >88.5% on average (N = 13,000 samples). Concordance between Stage 3 imputed genotypes and exome chip genotypes for replicated SNPs is reported in **Supplementary Table 6**.

Analysis

Stage 1

We tested association with LOAD using logistic regression modelling for common and low frequency variants (MAF>1%) and implementing maximum likelihood estimation using the score test and 'seqMeta' package for rare variation (MAF ≤ 1%). Analyses were conducted globally in the GERAD/PERADES consortium, and for each contributing centre in the CHARGE and ADGC consortia under two models (1) an 'unadjusted' model, which included minimal adjustment for possible population stratification, using Country of origin and the first three principal components from PCA, and (2) an

'adjusted' model, which included covariates for age, and sex, as well as Country of origin and the first three principal components. Age was defined as the age at onset of clinical symptoms for cases, and the age at last interview for cognitively normal controls.
Meta-analysis for common and low frequency variants were undertaken in METAL using a fixed-effects inverse variance-weighted meta-analysis. Rare variants were meta-analysed in the SeqMeta R package. In the SeqMeta pipeline, cohort-level analyses generated score statistics through the function 'prepScores()' which were captured in *. Rdata objects. These *. Rdata objects contain the necessary information to meta-analyse SKAT analyses: the individual SNP scores, MAF, and a covariance matrix for each unit of aggregation. Using the 'singlesnpMeta()' and 'skatOmeta()' functions of SeqMeta, the *. Rdata objects for individual studies were meta-analysed. The seqMeta coefficients and standard errors can be interpreted as a 'one-step' approximation to the maximum likelihood estimates. Monomorphic variants in individual studies were not excluded as they contribute to the minor allele frequency information. Three independent analysts confirmed the meta-analysis results.

In the GERAD/PERADES consortium 1,740 participants (888 LOAD cases and 852 controls) did not have age information available and were excluded from the adjusted analyses. Therefore, 16,160 cases and 17,967 controls were included in the unadjusted analyses and 15,272 cases and 17,115 controls were included in the adjusted analyses. The primary analysis utilized the unadjusted model given the larger sample size this provided. See **Supplementary Figure 2** for QQ plots of unadjusted and adjusted analyses.

Stage 2

We tested association with LOAD using the score test and 'seqMeta' package. Analyses were conducted under the two models described above, in the analysis groups indicated in **Supplementary Table 2**. Analyses were undertaken globally in the GERAD/PERADES cohort and by Country in the IGAP cohorts, with the EADI1 cohort only including French participants and the ACE cohort including only Spanish participants. Following the format of the IGAP mega meta-analysis,[7] four PCs were

included for the EADI1 dataset, and one in the Italian and Swedish IGAP clusters. Meta-analysis was undertaken in the SeqMeta R package.

Stage 3

Association analyses performed followed Stage 1 and Stage 2 analytical procedures described below, and only variants in *ABI3*, *PLCG2* and *TREM2* were examined. For gene-based testing, 10 variants in *ABI3*, 35 in *PLCG2*, and 13 in *TREM2* were examined.

Pathway/Gene-set Enrichment Analysis

The eight biological pathway clusters previously identified as enriched for association in the IGAP dataset[11] were tested for enrichment in this rare variation study (**Supplementary Table 15**) in order to test whether the biological enrichments observed in common variants also apply to rare variants. Genes were defined without surrounding genomic sequence, as this yielded the most significant excess of enriched pathways in the common variation dataset.[11] Gene-wide SKAT-O *p*-values for the variants of interest were combined using the Fisher's combined probability test. Given the low degree of LD[12] between rare variants our primary analyses did not control for LD between pathway genes. However, as a secondary analysis, the *APOE* region was removed, and for each pair of pathway genes within 1Mb of each other, the gene with the more significant SKAT-O *p*-value was removed. This highly conservative procedure removes any potential bias in the enrichment test both from LD between the genes, and also from dropping less significant genes from the analysis.

We also performed pathway analyses on the rare variant data presented here using all 9,816 pathways used previously. The top pathways are related to lipoprotein particles, cholesterol efflux, B-cell differentiation and immune response, and closely parallel the common variant results (**Supplementary Table 16**).

Protein interaction Analysis

Previous analysis of normal brain co-expression networks identified 4 gene modules that were enriched for common variants associated with AD risk in the IGAP GWAS. Each of these 4 modules was also found to be enriched for immune-related genes. The 151 genes present in 2 or more of these 4 modules were particularly strongly enriched

for IGAP GWAS association.[41] This set of 151 co-expressed genes thus contains genes of relevance to AD aetiology. To identify these genes, and clarify biological relationships between them for future study, protein interaction analysis was performed. First, a list of high-confidence (confidence score >0.7) human protein-protein interactions was downloaded from the latest version (v10) of the STRING database (http://string-db.org). Then, protein interaction networks were generated as follows:

1. Choose a gene to start the network (the "seed" gene)
2. For each remaining gene in the set of 151 genes, add it to the network if its corresponding protein shows a high-confidence protein interaction with a protein corresponding to any gene already in the network.
3. Repeat step 2 until no more genes can be added
4. Note the number of genes in the network
5. Repeat, choosing each of the 151 genes in turn as the seed gene.

The largest protein interaction network resulting from this procedure resulted in a network of 56 genes connected by high-confidence protein interactions. To test whether this network was larger than expected by chance, given the total number of protein-protein interactions for each gene, random sets of 151 genes were generated, with each gene chosen to have the same total number of protein-protein interactions as the corresponding gene in the actual data. Protein networks were generated for each gene as described above, and the size of the largest such network compared to the observed 56-gene network. 1000 random gene sets were generated, and none of them yielded a protein interaction network as large as 56 genes. Note that the procedure for generating the protein interaction network relies only on protein interaction data, and is agnostic to the strength of GWAS or rare-variant association for each gene. Thus the strength of genetic association in the set of 56 network genes can be tested relative to that in the original set of 151 genes without bias.

Gene-set enrichment analysis of the protein network

The set of 56 network genes was tested for association enrichment in the IGAP GWAS using ALIGATOR,[13] as was done in the original pathway analysis, using a range of

p-value thresholds for defining significant SNPs (and thus the genes containing those SNPs). The same analysis was also performed on the 95 genes in the module overlap but not the protein interaction network (**Supplementary Table 17**). It can be seen that the 56 network genes account for most of the enrichment signal observed in the set of 151 module overlap genes.

The set of 56 network genes, the set of 151 module overlap genes, and the set of 95 genes in the module overlap but not the network were tested for enrichment of association signal in variants with MAF<1% using the gene set enrichment method described above in section 11. Both the set of 151 genes ($p = 1.17\times10^{-6}$) and the subset of 56 genes ($p = 1.08\times10^{-7}$) show highly significant enrichment for association in the rare variants with MAF<1%. It can be seen that the 56 network genes account for most of the enrichment signal observed in the set of 151 module overlap genes (**Supplementary Table 17**). Again, the subset of 56 genes accounts for most of the enrichment signal observed in the set of 151 genes, as the remaining 95 genes have only nominally-significant enrichment ($p = 0.043$). Both the set of 151 genes ($p = 5.15\times10^{-5}$) and the subset of 56 genes ($p = 2.98\times10^{-7}$) show significant enrichment under a conservative analysis excluding the *APOE* region and correcting for possible LD between the genes (**Supplementary Table 17**). Thus, the rare variants show convincing replication of the biological signal observed in the common variant GWAS, and furthermore, the protein network analysis has refined this signal to a set of 56 interacting genes. Given that *TREM2* has a highly significant gene-wide p-value ($p = 1.01\times10^{-13}$) among variants with MAF<1%, enrichment analyses were run omitting it. Both the set of 151 genes ($p = 2.78\times10^{-3}$) and the subset of 56 genes ($p = 0.010$) (**Supplementary Table 18**) still showed significant enrichment of signal, suggesting that the contribution of rare variants to disease susceptibility in these networks is not restricted to *TREM2*. Biological follow-up of genetic results is labour-intensive and expensive. It is therefore important to concentrate such work on the genes that are most important to AD susceptibility. Thus, the rationale for reducing the gene set is that it defines a network of genes that are not only related through co-expression and protein interaction, but also show enrichment for genetic association signal. These genes are therefore strong candidates for future biological study.

Gene expression

We examined mRNA expression of the novel genes *PLCG2* and *ABI3* in neuropathologically characterized brain post-mortem tissue (508 persons): they are expressed at low levels in the dorsolateral prefrontal cortex of subjects from two studies of aging with prospective autopsy (ranked 12,965th out of 13,484 expressed genes).[14] However, *ABI3* and *PLCG2* were more highly expressed in purified microglia/macrophage from the cortex of 11 subjects from these cohorts (1740th and 2600th respectively out of the 11,500 expressed genes)(*unpublished data*). These findings are consistent with the high levels of expression of both PLCG2 and ABI3 in peripheral monocytes, spleen, and whole blood reported by the ROADmap project and in microglia as reported by Zhang *et al.*[15] From the same brain tissue, we examined methylation (n=714)[16] and H3K9ac acetylation (n=676) data and found differential methylation at four CpG sites and lower acetylation at two H3K9ac sites adjacent to *PLCG2* and *ABI3* in relation to increased global neuritic plaque and tangle burden (FDR < 0.05). Similarly, high TREM2 expression has been shown to correlate with increasing neuritic plaque burden.[17]

AMP-AD Gene Expression Data: RNA sequencing was used to measure gene expression levels in the temporal cortex of 80 subjects with pathologically confirmed AD and 76 controls without any neurodegenerative pathologies obtained from the Mayo Clinic Brain Bank and the Banner Sun Health Institute. The human RNA sequencing data is deposited in the Accelerating Medicines Partnership-AD (AMP-AD) knowledge portal housed in Synapse (https://www.synapse.org/#!Synapse:syn2580853/wiki/66722). After QC, our postmortem human cohort has 80 subjects with pathologically confirmed AD and 76 controls without any neurodegenerative pathologies. Assuming two samples of 100 per group, two-sample t-test, same standard deviation, we will have 80% power to detect effect sizes of 0.40, 0.49 and 0.59 at $p < 0.05$, 0.01 and 0.001, respectively, where effect size is the difference in means between two groups divided by the within-group standard deviation. The human RNA sequencing data overview, QC and analytic methods are available at the following Synapse pages, respectively: syn3163039, syn6126114, syn6090802. Multivariable linear regression was used to test for association of gene expression levels with AD diagnosis (Dx) using two different models: In the

Simple model, we adjust for age at death, sex, RNA integrity number (RIN), tissue source, and RNAseq flowcell. In the Comprehensive model, we adjust for all these covariates, and brain cell type markers for five cell-specific genes (*CD68* (microglia), *CD34* (endothelial), *OLIG2* (oligodendroglia), *GFAP* (astrocyte), *ENO2* (neuron)) to account for cell number changes that occur with AD neuropathology. *TREM2*, *PLCG2* and *ABI3* are significantly higher in AD temporal cortex prior to correcting for cell types (Simple model), but this significance is abolished after adjusting for cell-specific gene counts (Comprehensive model). This suggests that these elevations are likely a consequence of changes in cell types that occur with AD, most likely microgliosis given that *TREM2*, *PLCG2* and *ABI3* are microglia-enriched genes[15] (**Supplementary Table 19, Supplementary Figure 12**).

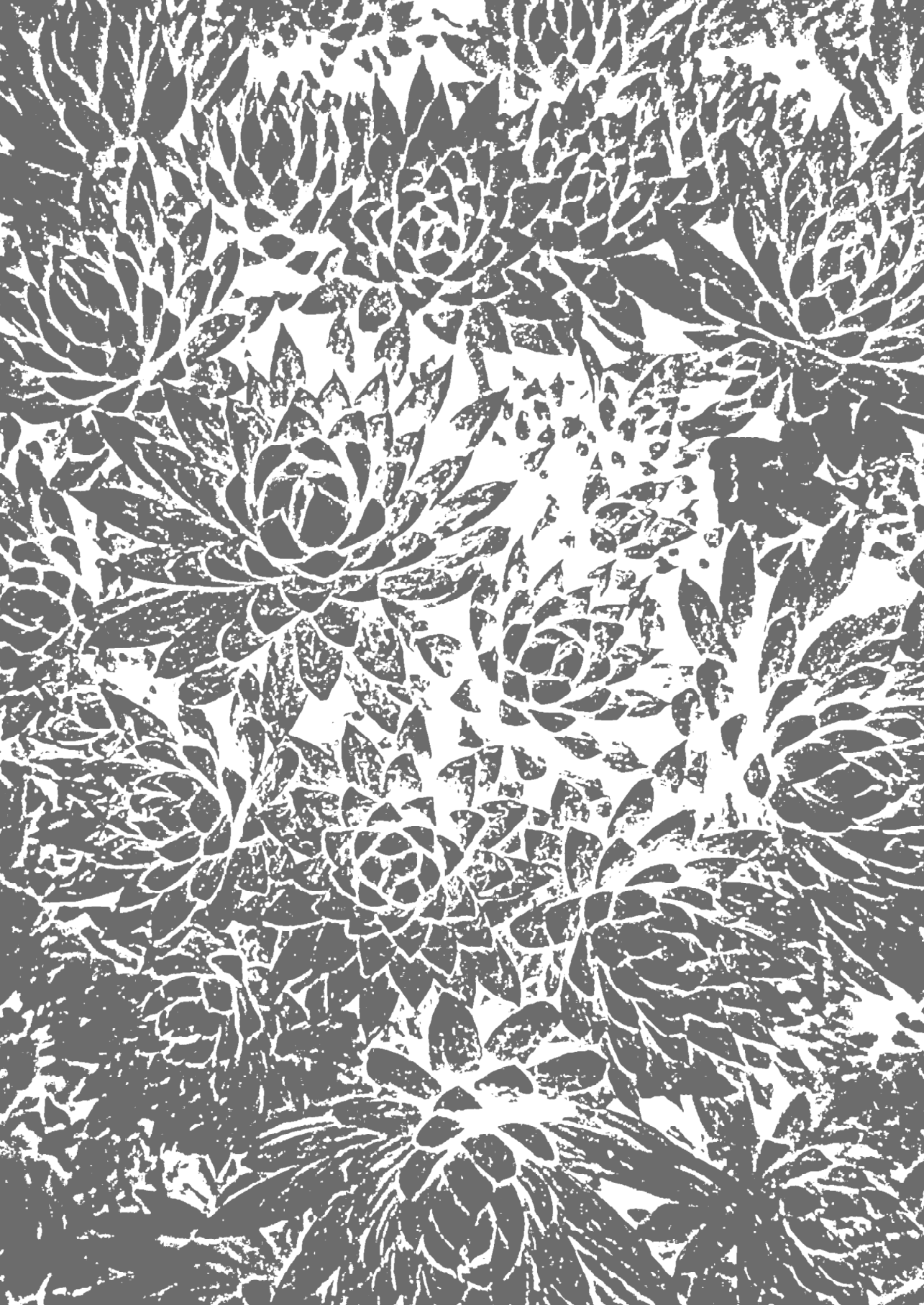
Methods only References

1. Goldstein, J. I. *et al.* zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinforma. Oxf. Engl.* **28,** 2543–2545 (2012).
2. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55,** 997–1004 (1999).
3. Grove, M. L. *et al.* Best Practices and Joint Calling of the HumanExome BeadChip: The CHARGE Consortium. *PLoS ONE* **8,** e68095 (2013).
4. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2,** e190 (2006).
5. Das, S. *et al.* Imputation server: next generation genotype imputation service. *Nat. Genet.*
6. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *bioRxiv* 35170 (2015). doi:10.1101/035170
7. Lambert, J. C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45,** 1452–1458 (2013).
8. Naj, A. C. *et al.* Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat. Genet.* **43,** 436–441 (2011).
9. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44,** 955–959 (2012).
10. Fuchsberger, C., Abecasis, G. R. & Hinds, D. A. minimac2: faster genotype imputation. *Bioinforma. Oxf. Engl.* **31,** 782–784 (2015).
11. International Genomics of Alzheimer's Disease Consortium (IGAP). Convergent genetic and expression data implicate immunity in Alzheimer's disease. *Alzheimers Dement. J. Alzheimers Assoc.* **11,** 658–671 (2015).
12. Talluri, R. & Shete, S. A linkage disequilibrium-based approach to selecting disease-associated rare variants. *PloS One* **8,** e69226 (2013).
13. Holmans, P. *et al.* Gene ontology analysis of GWA study data sets provides insights into the biology of bipolar disorder. *Am. J. Hum. Genet.* **85,** 13–24 (2009).
14. Lim, A. S. P. *et al.* 24-hour rhythms of DNA methylation and their relation with rhythms of RNA expression in the human dorsolateral prefrontal cortex. *PLoS Genet.* **10,** e1004792 (2014).

15. Zhang, Y. *et al.* Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. *Neuron* **89,** 37–53 (2016).
16. De Jager, P. L. *et al.* Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat. Neurosci.* **17,** 1156–1163 (2014).
17. Chan, G. *et al.* CD33 modulates TREM2: convergence of Alzheimer loci. *Nat. Neurosci.* **18,** 1556–1558 (2015).

Supplementary tables and figures

Supplementary tables and figures for this chapter can be accessed by scanning the following code or accessing the journals' website.



Chapter 4

Genetics of Magnetic Resonance Imaging endophenotypes of Alzheimer's disease

Chapter 4.1

Grey matter heritability in family-based and population-based studies using voxel-based morphometry

Sven J. van der Lee*, Gennady V. Roshchupkin*, Hieab H.H. Adams*, Helena Schmidt, Edith Hofer, Yasaman Saba, Reinhold Schmidt, Albert Hofman, Najaf Amin, Cornelia M. van Duijn, Meike W. Vernooij, M. Arfan Ikram, Wiro J. Niessen.

*Authors contributed equally

Human Brain Mapping. 2017 May; 38(5): 2408-2423.

Abstract

Background: The combination of genetics and imaging has improved our understanding of the brain through studies of aggregate measures obtained from high-resolution structural imaging. Voxel-wise analyses have the potential to provide more detailed information of genetic influences on the brain. Here we report a large-scale study of the heritability of grey matter at voxel resolution (1×1×1mm).

Methods: Validated voxel-based morphometry (VBM) protocols were applied to process magnetic resonance imaging data of 3239 unrelated subjects from a population-based study and 491 subjects from two family-based studies. Genome-wide genetic data was used to estimate voxel-wise gray matter heritability of the unrelated subjects and pedigree-structure was used to estimate heritability in families. We subsequently associated two genetic variants, known to be linked with subcortical brain volume, with most heritable voxels to determine **i**f this would enhance their association signals.

Results: Voxels significantly heritable in both estimates mapped to subcortical structures, but also voxels in the language areas of the left hemisphere were found significantly heritable. When comparing regional patterns of heritability, family-based estimates were higher than population-based estimates. However, regional consistency of the heritability measures across study designs was high (Pearson's correlation coefficient = 0.73, $p = 2.6\times10^{-13}$). We further show enhancement of the association signal of two previously discovered genetic loci with subcortical volume by using only the most heritable voxels.

Conclusion: Grey matter voxel-wise heritability can be reliably estimated with different methods. Combining heritability estimates from multiple studies is feasible to construct reliable heritability maps of grey matter voxels.

Introduction

The human brain shows large inter-individual variation, which could be explained by genetic and environmental influences. Studying these influences is essential in better understanding brain structure and function. The degree to which genetics explains phenotypic variation, in other words heritability, depends on many factors: the actual genetic contribution to the trait, environmental effects, measurement error, study design and sample characteristics.[1-3] Recently an overview was published of fifty years of worldwide heritability research in twins encompassing thousands of traits, showing heritability studies are highly informative on how large the genetic contribution to a trait is.[4] Heritability studies could aid future genetic research to focus on particular regions of interest in the brain. For example, large scale genetic studies of brain structures with the highest heritability typically yield the most findings.[5] When studying the multitude of measures of voxel based magnetic resonance imaging (MRI), limiting genetic studies to the most heritable traits could be feasible in light of multiple testing. Recent studies have focused on heritability of detailed MRI measures at a voxel level.[6-12] Different study designs showed comparably high estimates for white matter tract heritability in twin pairs,[9,12] sib-pairs,[7] and extended pedigrees (heritability = 50-90%).[10] The heritability of grey matter was studied by voxel-based morphometry (VBM) previously,[13-15] but the studies were relatively small and relatively large voxels were studied. Moreover, heritability of grey matter VBM has not been estimated in population-based studies.

Here, we perform a large multi-site study to estimate the voxel-wise heritability of grey matter. We calculate pedigree-based heritability in two family-based studies and heritability based on genome-wide genetic data in a large population-based study of unrelated subjects. Using these approaches, we created two grey matter heritability maps and described which regions contain significantly heritable voxels in both designs. We furthermore estimated overall regional consistency of the heritability measures across study designs and explored if usage of our heritability maps could potentially enhance association signals of two genetic variations, previously discovered by genome-wide association studies.[5,16,17]

Methods

Study subjects and imaging protocol

Rotterdam Study – The Rotterdam Study is a population-based cohort study among inhabitants of a district of Rotterdam (Ommoord), The Netherlands, and aims to examine the determinants of disease and health in the elderly with a focus on neurogeriatric, cardiovascular, bone, and eye disease.[18] In 1990 to 1993, 7983 persons participated and were re-examined every 3 to 4 years (RS-I). In 2000 to 2001 the cohort was expanded by 3011 persons who had not yet been part of the Rotterdam Study (RS-II). All participants had DNA extracted from blood at their first visit. In 2006-2008 a second expansion (RS-III) of 3,932 persons aged 45 and over was realized. Genotyping was performed at the Human Genotyping Facility, Genetic Laboratory Department of Internal Medicine, Erasmus MC, Rotterdam. Genotypes were imputed to the 1000 genomes phase I version 3 reference panel, using standard methods and software.[19] From 2005 onwards MRI is part of the core protocol of the Rotterdam study.[20] For this study a total of 4071 unique study participants had both MRI and genetic data and were available for analysis. The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC and by the Ministry of Health, Welfare and Sport of the Netherlands, implementing the Wet Bevolkingsonderzoek: ERGO (Population Studies Act: Rotterdam Study). All participants provided written informed consent to participate in the study and to obtain information from their treating physicians.

Erasmus Rucphen Family (ERF) – The ERF study is a family-based cohort study in a genetically isolated population from a community in the South-West of the Netherlands (Rucphen municipality) including 3,000 participants. Participants are all descendants of a limited number of founders living in the 19th century, and all of Caucasian European descent. Extensive genealogical data is available for this population. The study population is described in detail elsewhere.[21] In a follow-up analysis, non-demented hypertensive (systolic blood pressure ≥ 160, diastolic blood pressure ≥ 100 or use of antihypertensive medication) subjects aged 55-75 years were included for a new battery of tests, including MRI scanning.[22] These 122 participants from the ERF were related to each other in one large pedigree. This large pedigree was split into multiple small

pedigrees for heritability calculations (pedcut version 1.19 http://mga.bionet.nsc.ru/soft/). Participants related to each other in 27 families with in total 880 relatives. The average size of the pedigrees was 32.6 relatives (range 20-44) with on average 4.5 participants with MRI per family. All participants gave informed consent to participate in the study and to obtain information from their treating physicians. The study was approved by the medical ethics committee at Erasmus MC University Medical Center, Rotterdam, The Netherlands.

MRI scanning for ERF and the Rotterdam Study was done on the same 1.5 T MRI unit (GE Healthcare, Milwaukee, USA, Signa Excite software version 11×) fitted with a dedicated 8-channel head coil. The T1-weighted, proton density-weighted (PDw) and fluid-attenuated inversion recovery (FLAIR) sequences were used.[20] For the purpose of segmentation, the T1w scan is acquired in 3D at high in-plane resolution and with thin slices (voxel size < 1 mm3).[20]

Austrian Stroke Prevention Study (ASPS) – The ASPS study is a single-center, prospective follow-up study on the effects of vascular risk factors on brain structure and function in the normal elderly population of the city of Graz, Austria. The procedure of recruitment and diagnostic work-up of study participants has been described previously.[23,24] Between 2006 and 2013 the study was extended for the Austrian Stroke Prevention Family Study (ASPS-Fam).[25] Study participants of the ASPS and their first grade relatives were invited to enter ASPS-Fam. Inclusion criteria were no history of previous stroke or dementia and a normal neurological examination. In total 176 families connecting a total of 719 relatives, among which 369 were study participants with brain-MRI. The average size of the pedigrees was 4 (range 1-10) relatives with on average 2.4 participants with MRI per family. The diagnostic work-up was identical to the original study. The study protocol was approved by the ethics committee of the Medical University of Graz, Austria, and written and informed consent was obtained from all subjects. MRI scanning of the ASPS-Fam was done on a 3.0 T Tim Trio (Siemens, Erlangen). T1-MPRAGE 1×1×1mm was used for image processing.[25]

Image processing

Prior to analysis, a number of pre-processing steps were performed. For multispectral image analysis, the different scans were spatially registered using rigid registration [20]. Subsequently, the brain was extracted from the scan. Hereto a manually segmented brain mask, which excludes cerebellum, eyes and skull, was non-rigidly registered to the T1-weighted image using Elastix.[20] Finally, scans were corrected for intensity non-uniformity using the N3 method; non-uniformity correction was carried out within the brain mask.[20] All T1-weighted images were segmented into supra-tentorial grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF). For the Rotterdam Study and ERF, we used a previously described k-Nearest-Neighbor (kNN) algorithm, which was trained on six manually labeled atlases.[26] For the ASPS-Fam study a Quantib BV tissue segmentation tool was applied (www.quantib.org). Quantib® software implements the same algorithm, which we then used for tissue segmentation in the Rotterdam Study and ERF. There are thus no methodological differences between the methods, both of them based on kNN-based segmentation training on manually labeled subjects for segmenting GM, WM and CSF.

Voxel-based morphometry (VBM) was performed by the same optimized VBM protocol in all three studies[27] and previously described.[28] FSL software[29] was used for VBM data processing. First, all GM density maps were non-linearly registered to the standard GM probability template. For this study we chose the MNI152 GM template (Montreal Neurological Institute) with a 1×1×1 mm voxel resolution.[30] The MNI152 standard-space T1-weighted average structural template image is derived from 152 structural images, which have been warped and averaged into the common MNI152 coordinate system after high-dimensional nonlinear registration. A spatial modulation procedure was used to avoid differences in absolute grey matter volume due to the registration. This involved multiplying voxel density values by the Jacobian determinants estimated during spatial normalization. To decrease signal to noise ratio, all images were smoothed using a 3 mm (FWHM 8 mm) isotropic Gaussian kernel. Thus all results are in MNI space. Brain regions were segmented using atlas-based segmentation based on the Hammer atlas.[31] The modulation step in the VBM pipeline preserves the volume of

a particular tissue within a voxel. The multiplication of the voxel values in the segmented images by the Jacobian determinants derived from the spatial normalization step allows us to calculate volumes by aggregating voxels. In total we estimated heritability for 1,405,508 grey matter voxels in all three studies.

Reproducibility VBM measures

We investigated the test-retest reliability of the VBM measures in a subset of 83 persons who were scanned twice within 1-9 weeks. We quantified the reproducibility by calculating the intraclass correlation (ICC) of the gray matter density measures for every voxel (Online viewer, **Supplementary Figure 1**).[32]

Heritability analysis

Population-based heritability estimates were calculated using Genome-wide Complex Trait Analysis (GCTA v1.24)[33] (http://cnsgenomics.com/software/gcta/) in the population-based Rotterdam Study. GCTA implements REML (restricted maximum likelihood) analysis, this method compares genotypic similarity between individuals to their phenotypic similarity. Formula's underlying the GCTA method to determine heritability estimates are described elsewhere[3] and thoroughly explained in a commentary by the authors.[34] The 1000 Genomes imputed genotypes (Imputation quality (Rsq) > 0.5 and minor allele frequency (MAF) > 0.01) were used to create a genetic relationship matrix (GRM) in GCTA.[35] The power of GCTA analysis is determined by pair-wise genetic relationships in the studied population.[3,34] Therefore the three cohorts of the Rotterdam study were combined and analyzed as one in the voxel-wise heritability analysis. Pairwise genetic relatedness between all individuals (N=4,071) was calculated and for pairs with more than 0.02 genotype similarity[35] one person was removed ($N_{removed}$ = 832). REML analysis was then performed in the remaining 3,239 unrelated subjects using the GRM correcting for age and sex. All grey matter heritability was estimated once**.**

Family-based heritability was estimated using maximum-likelihood variance components methods implemented in the SOLAR (version 6.6.2) software.[36] Formulas for the calculation of heritability estimates are described in detail elsewhere.[36] Briefly,

the algorithms in SOLAR employ maximum likelihood variance decomposition methods. The covariance matrix Ω for a pedigree of individuals is given by:

$$\Omega = 2.\Phi.\sigma_g^2 + I.\sigma_e^2$$

where σ^2_g is the genetic variance due to the additive genetic factors, Φ is the kinship matrix representing the pair-wise kinship coefficients among all individuals, σ^2_e is the variance due to individual-specific environmental effects, and I is an identity matrix (under the assumption that all environmental effects are uncorrelated among family members). Narrow sense heritability is defined as the fraction of phenotypic variance σ^2_P attributable to additive genetic factors:

$$h^2 = \frac{\sigma_g^2}{\sigma_P^2}.$$

The variance parameters are estimated by comparing the observed phenotypic covariance matrix with the covariance matrix predicted by kinship (Almasy and Blangero, 1998). Significance of heritability is tested by comparing the likelihood of the model in which σ^2_g is constrained to zero with that of a model in which σ^2_g is estimated. Twice the difference between the $\log_e$ likelihoods of these models yields a test statistic, which is asymptotically distributed as a ½:½ mixture of a χ^2 variable with 1 degree-of-freedom and a point mass at zero. If the algorithm converges SOLAR outputs the heritability value, the significance value (p), and the standard error for each voxel.[8,36]
ERF study and ASPS-Fam were not jointly analysed because ERF subjects were scanned on a 1.5T MRI and ASPS-Fam subjects on a 3.0T MRI. Instead inverse variance meta-analysis using heritability and heritability standard errors was performed in METAL [37] to boost power and improve stability of heritability estimates [6]. Heritability estimates were calculated in both studies with age and sex as covariates. Variance component methods implemented in SOLAR are vulnerable for inflation if phenotypes have a leptokurtic to distribution. Therefore we applied inverse normal transformations in SOLAR to all voxels, but some voxels still violated the distribution of too high residual kurtosis (kurtosis >0.9) and were therefore excluded.[38] Non converging heritability estimates of 0 without standard errors were also excluded from the meta-analysis. In the family-based studies some voxels had valid p-values and a heritability of 1, but missing standard errors. These voxels were located in the middle of voxel-clusters with

high heritability (online viewer reference) (close to 1). Therefore standard errors for such voxels were imputed to retain these voxels for meta-analysis. This resulted in imputation of the standard error for 6.4% of voxels in the ERF study and a negligible percentage of voxels in ASPS-Fam (<0.001%).

Enhancement of association signal

We explored whether voxel heritability information could enhance the association of genetic variants with brain structures. The genetic variants most significantly associated with hippocampal volume (rs77956314 on 12q24.22, near the gene *HRK*) and putamen volume (rs945270 on 14q22.3, downstream of the gene *KTN1*) were selected from a recently published genome-wide association study on subcortical structures [5]. To select the most heritable voxels in the hippocampus and putamen, we ordered them using three approaches. First, we ranked the voxels from low to high family-based heritability estimates. Second, we ranked them from low to high population-based heritability estimates. In the third approach we summed the ranks obtained from both the family- and population-based estimates and used the sum of the ranks to prioritize the voxels. Using these three approaches we excluded the voxels in a step-wise manner by removing the 5% least heritable voxels. For each step we computed the volume by summing the values of the remaining voxels. As a voxel represents grey matter density in 1 mm^3, the sum of voxels gives the volume of grey matter. We determined the association of the two genetic variants in an additive model with the volumes in linear regression analyses (adjusted for age, sex, and the first three principal components) and compared this to association of the volume derived from all voxels mapped to the structure (i.e. the total VBM-volume of the hippocampus or putamen). The p-value of the association of the genetic variants with the subsets of voxels divided by the p-value of the association of the genetic variants with the total VBM-volume was calculated to measure change in the strength of the association. Genetic effects were calculated in the three cohorts of the Rotterdam study separately (RS-I = 844, RS-II = 1003, RS-III = 2190) and were combined using an inverse variance weighted meta-analysis in METAL.[37]

Statistical analysis

Descriptive statistics were compared using one-way ANOVA and chi-squared tests. To correct for multiple comparisons we applied FDR p-value thresholds[39] for both population and family heritability separately to declare which voxels are significantly heritable.

Results

Population characteristics

Characteristics of the study population are shown in **Table 1.** The spread of the age of subjects in the ERF study (age range 55-76) was smaller compared to ASPS-Fam (38-86) and the Rotterdam Study (46-98) due to the fact that inclusion criteria for scanning was restricted to midlife **(Table 1)**. However, the average age at the time of MRI-scanning of the cohorts was very similar, ranging from 64.3 (± 4.5) years in the ERF study, 64.9 (± 10.7) years in ASPS and 64.9 (± 10.7) in the Rotterdam Study (p = 0.86). The percentage of women was 52.5% in ERF, 60.4% in ASPS-Fam and 55.3% in the Rotterdam study, these differences were non-significant (p = 0.13) (**Table 1**).

Heritability estimates

In total 454,184 (33.3% of all voxels) were FDR-significant in the family-based estimates. Mean heritability of significant voxels was 0.44 ± 0.12 SD (all voxels 0.29 ± 0.17 SD), with heritability estimates ranging from 0.23 to 1. In total 68,616 (4.9% of all voxels) were FDR-significant in the population-based estimates. Mean heritability of the significant voxels was 0.34 ± 0.04 SD (all voxels 0.11 ± 0.10), with heritability estimates ranging from 0.25 to 0.56. We found heritability of 44,349 voxels (3.2% of all voxels) to be FDR significant in the family- as well as the population-based heritability estimates**.** These significantly heritable voxels were clustered, mostly within subcortical brain structures (**Figure 1). Table 2** shows the percentage of voxels that were significantly heritable of the total of voxels in a structure in both estimates, as well as the average regional heritability, considering all voxel-wise heritability estimates. Highest percentage of significantly heritable in both estimates voxels were located in the caudate nucleus (right 72.4% and left 68.6%) followed by the putamen (right 57.5% and

left 32.6%). Other subcortical structures with a large percentage of significantly heritable voxels were; left pallidum (32.2%), left nucleus accumbens (29.7%), right pallidum (28.5%), left amygdala (21.4%), left hippocampus (17.9%), left thalamus (14.4%), right amygdala (12.8%) and the right insula (11.4%). Apart from the subcortical structures, parts of the right lateral occipitotemporal gyrus (gyrus fusiformis) (10.4%), left straight gyrus (gyrus rectus) (10.4%), left subcallosal area (8.0%) and the left lingual gyrus (7.9%) harbored a proportion significantly heritable voxels (**Figure 1 and Table 2**).

Table 1. Descriptive statistics

	ERF	**ASPS-Fam**	**RS**	***p***
Country	Netherlands	Austrian	Netherlands	
Study type	Family-based	Family-based	Population-based	
Field strength	1.5T	3.0T	1.5T	
Sequence	T1-weighted	T1-MPRAGE	T1-weighted	
TR/TE (ms)	13.8/2.8	1900/2.19	13.8/2.8	
TI (ms)	400	900	400	
Flip angle (degrees)	20	9	20	
Voxel size	1x1x1mm	1x1x1mm	1x1x1mm	
Minimum-maximum age	56-76	38-86	46-98	
Age (± SD)	64.3 ± 4.5	64.9 ± 10.7	64.7 ± 10.7	0.86‡
Minimum-maximum age	56-76	38-86	46-98	
Percentage females	52.50%	60.40%	55.30%	0.13†
N participants with MRI-scan	122	369	3239*	
Total relatives in pedigrees	880	718	-	

Descriptive statistics of the included studies. ‡ p calculated with one-way ANOVA, † p chi-squared test. *The total number of participants with brain magnetic imaging and genetics data in the Rotterdam study was 4071, but for pairs with more than 0.02 genotype similarity one person was removed ($N_{removed}$ = 832). MRI = Magnetic Resonance Imaging, ERF = Erasmus Rucphen Family study, ASPS-Fam = Austrian Stroke Prevention Family Study, RS = Rotterdam Study, SD = standard deviation. T1-MPRAGE = T1 weighted 3D sequences with magnetization preparation, TR = repetition time, TE = echo time, TI inversion time.

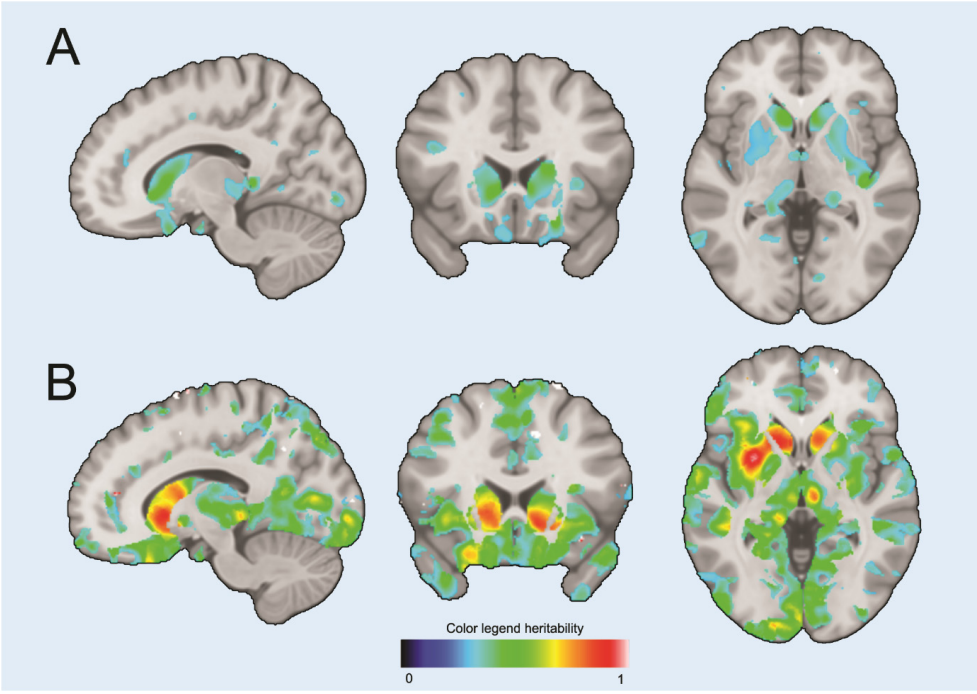


Figure 1. Example of FDR-Significant voxels in both population-based (**A**) and family-based (**B**) estimates. Significant voxels cluster in subcortical structures, such as the caudate nucleus. All results can be interactive accessed (www.imagene.nl/heritability) and downloaded from the website.

When comparing regional heritability, estimates calculated in families was always higher than the population-based estimates ($p<0.001$) **(Figure 2A)** and the difference in heritability between family-based estimates and population-based estimates was relatively stable (mean difference of regional heritability = 0.21 ± 0.08) **(Table 2)**. Therefore, the regional heritability pattern of the family-based estimates significantly predicted the regional pattern of heritability in the population-based study (Pearson's correlation coefficient = 0.73, $p = 2.6\times10^{-13}$) (**Figure 2B**).

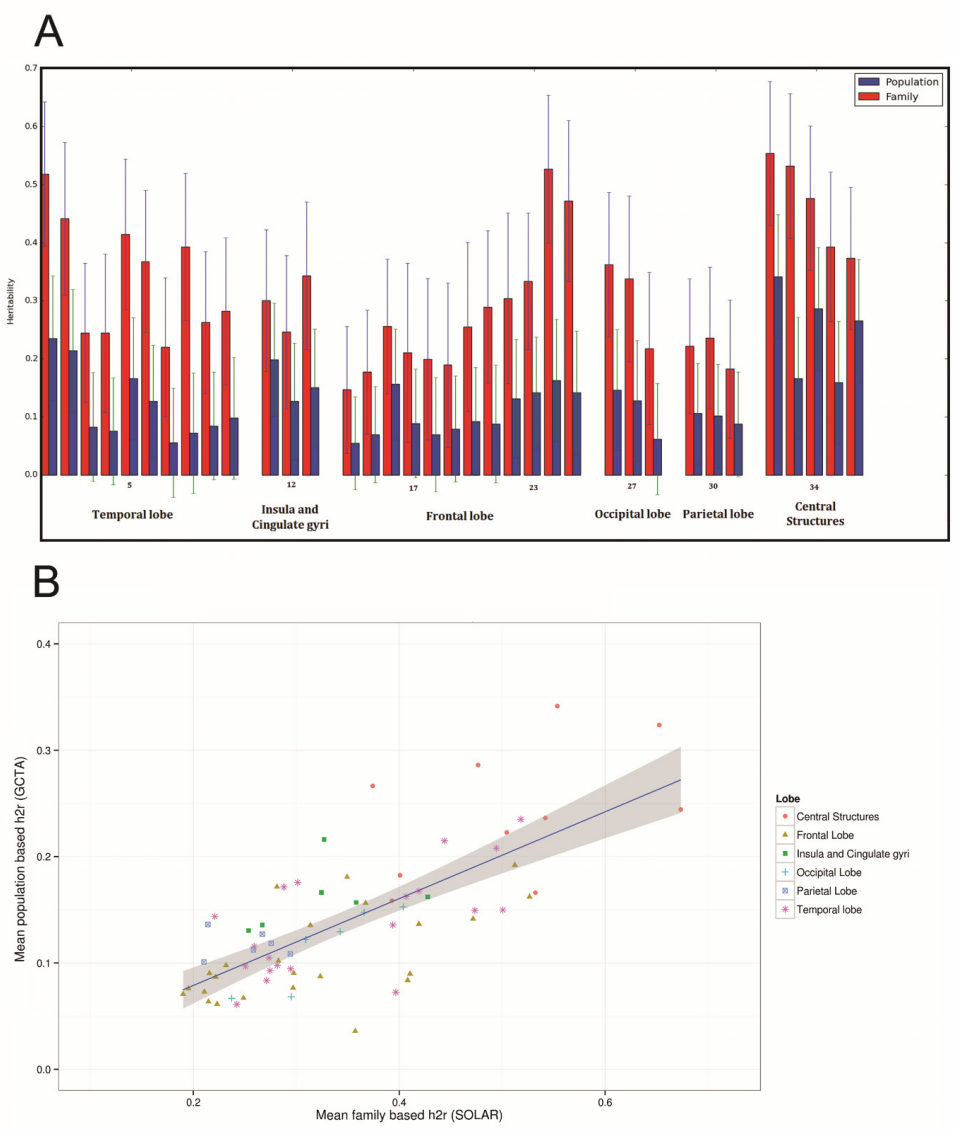


Figure 2. A Barplot showing regional brain heritability. Structures that are in both the left as well as the right hemisphere were averaged for this figure. It can clearly be seen that the heritability from family-based studies is higher than heritability form the unrelated population (P<0.001). **B** Scatter plot of the average regional heritability of all brain structures. The correlation of the family-based and population-based estimates was high (Pearson's correlation coefficient = 0.73, $p = 2.6 \times 10^{-13}$). Data points per structure correspond to family and population heritability in table 2.

Table 2: Regional heritability estimates

Lobe	Structure	N-voxels	Population significant (%)	Family significant (%)	Significant both (%)	Population (±SD)	Family (±SD)	Δ h2
Temporal lobe	Amygdala left	2328	21.4	79.6	21.4	0.21 (±0.09)	0.44 (±0.11)	0.23
Temporal lobe	Amygdala right	2443	13	88.2	12.8	0.21 (±0.09)	0.49 (±0.10)	0.29
Temporal lobe	Anterior temporal lobe, lateral part left	5169	0.1	21.9	0.1	0.08 (±0.08)	0.27 (±0.16)	0.19
Temporal lobe	Anterior temporal lobe, lateral part right	5535	1.2	34.6	0.4	0.10 (±0.08)	0.27 (±0.14)	0.17
Temporal lobe	Anterior temporal lobe, medial part left	10427	3.3	33.9	0.9	0.09 (±0.09)	0.27 (±0.15)	0.18
Temporal lobe	Anterior temporal lobe, medial part right	11157	7.4	34.1	3	0.17 (±0.08)	0.29 (±0.10)	0.12
Temporal lobe	Gyri parahippocampalis et ambiens left	7411	10.7	66.5	6.8	0.17 (±0.10)	0.42 (±0.17)	0.25
Temporal lobe	Gyri parahippocampalis et ambiens right	7548	4.5	87.9	4.3	0.15 (±0.10)	0.50 (±0.12)	0.35
Temporal lobe	Hippocampus left	3482	18.1	93.7	17.9	0.24 (±0.07)	0.52 (±0.15)	0.28
Temporal lobe	Hippocampus right	3725	8.1	83.1	7.7	0.15 (±0.11)	0.47 (±0.16)	0.32
Temporal lobe	Lateral occipitotemporal gyrus (gyrus fusiformis) left	6603	0	69.5	0	0.07 (±0.07)	0.40 (±0.14)	0.32
Temporal lobe	Lateral occipitotemporal gyrus (gyrus fusiformis) right	6631	10.9	68.9	10.4	0.16 (±0.09)	0.41 (±0.15)	0.24
Temporal lobe	Medial and inferior temporal gyri left	23978	0	26.7	0	0.06 (±0.06)	0.24 (±0.14)	0.18
Temporal lobe	Medial and inferior temporal gyri right	24547	1.9	33.8	1.3	0.12 (±0.08)	0.26 (±0.17)	0.14
Temporal lobe	Posterior temporal lobe left	64214	3.5	28.5	2.3	0.10 (±0.10)	0.25 (±0.15)	0.15
Temporal lobe	Posterior temporal lobe right	64529	2.6	36.9	0.6	0.09 (±0.09)	0.29 (±0.16)	0.2
Temporal lobe	Superior temporal gyrus, anterior part left	6618	4.9	49.1	2.3	0.18 (±0.08)	0.30 (±0.09)	0.13
Temporal lobe	Superior temporal gyrus, anterior part right	6874	3	32.8	0.4	0.10 (±0.08)	0.28 (±0.12)	0.18
Temporal lobe	Superior temporal gyrus, central part left	17666	8.1	66.3	5.7	0.14 (±0.10)	0.39 (±0.12)	0.26
Temporal lobe	Superior temporal gyrus, central part right	18329	11.8	20.7	1.5	0.14 (±0.11)	0.22 (±0.13)	0.08
Parietal Lobe	Postcentral gyrus left	39052	4	35	2.3	0.11 (±0.09)	0.29 (±0.17)	0.19
Parietal Lobe	Postcentral gyrus right	34055	4.7	30	1.7	0.13 (±0.10)	0.27 (±0.15)	0.14
Parietal Lobe	Remainder of parietal lobe left (including supramarginal and angular gyrus)	56090	8.3	21	2.4	0.14 (±0.11)	0.21 (±0.15)	0.08
Parietal Lobe	Remainder of parietal lobe right (including supramarginal and angular gyrus)	55452	4	18	1.1	0.10 (±0.10)	0.21 (±0.15)	0.11
Parietal Lobe	Superior parietal gyrus left	53683	3.6	28.9	2.3	0.11 (±0.09)	0.26 (±0.16)	0.15
Parietal Lobe	Superior parietal gyrus right	59456	2.8	32.1	1.6	0.12 (±0.09)	0.28 (±0.16)	0.16
Occipital Lobe	Cuneus left	15193	3.2	40.6	2.5	0.12 (±0.09)	0.31 (±0.17)	0.19
Occipital Lobe	Cuneus right	14886	6.6	51.6	5.8	0.13 (±0.10)	0.34 (±0.15)	0.21

Lobe	Structure	N-voxels	Population significant (%)	Family significant (%)	Significant both (%)	Population (±SD)	Family (±SD)	Δ h2
Occipital Lobe	Lateral remainder of occipital lobe left	61105	0.6	35.3	0.3	0.07 (±0.07)	0.30 (±0.16)	0.23
Occipital Lobe	Lateral remainder of occipital lobe right	61419	0.4	21.9	0.3	0.07 (±0.08)	0.24 (±0.15)	0.17
Occipital Lobe	Lingual gyrus left	18574	8.3	65.7	7.9	0.15 (±0.10)	0.40 (±0.14)	0.25
Occipital Lobe	Lingual gyrus right	19412	10.5	60.2	6.1	0.15 (±0.11)	0.37 (±0.14)	0.22
Insula and Cingulate gyri	Cingulate gyrus, anterior (supragenual) part left	12284	6.3	27.6	1.3	0.14 (±0.10)	0.27 (±0.09)	0.13
Insula and Cingulate gyri	Cingulate gyrus, anterior (supragenual) part right	13770	1.1	23.4	0.1	0.13 (±0.08)	0.25 (±0.11)	0.12
Insula and Cingulate gyri	Cingulate gyrus, posterior part left	12016	13.2	40.1	6.9	0.17 (±0.12)	0.32 (±0.14)	0.16
Insula and Cingulate gyri	Cingulate gyrus, posterior part right	12643	6.7	51.9	5.6	0.16 (±0.09)	0.36 (±0.14)	0.2
Insula and Cingulate gyri	Insula left	20040	8.5	62.5	4.5	0.16 (±0.10)	0.43 (±0.16)	0.27
Insula and Cingulate gyri	Insula right	20796	20.6	46.3	11.4	0.22 (±0.10)	0.33 (±0.15)	0.11
Frontal Lobe	Anterior orbital gyrus left	9240	0.1	25	0.1	0.07 (±0.07)	0.25 (±0.14)	0.18
Frontal Lobe	Anterior orbital gyrus right	8652	0	11.4	0	0.10 (±0.08)	0.23 (±0.20)	0.13
Frontal Lobe	Inferior frontal gyrus left	26296	1.3	33.6	0.3	0.08 (±0.08)	0.30 (±0.19)	0.22
Frontal Lobe	Inferior frontal gyrus right	24054	0.7	11.3	0.2	0.07 (±0.08)	0.21 (±0.16)	0.14
Frontal Lobe	Lateral orbital gyrus left	5264	0	59.3	0	0.04 (±0.05)	0.36 (±0.11)	0.32
Frontal Lobe	Lateral orbital gyrus right	5348	0.3	42	0.1	0.09 (±0.10)	0.30 (±0.15)	0.21
Frontal Lobe	Medial orbital gyrus left	8408	3.4	40.2	3.3	0.09 (±0.10)	0.32 (±0.15)	0.24
Frontal Lobe	Medial orbital gyrus right	8486	4.6	27.2	3.9	0.10 (±0.10)	0.28 (±0.21)	0.18
Frontal Lobe	Middle frontal gyrus left	53570	0.9	12.4	0.2	0.06 (±0.08)	0.21 (±0.17)	0.15
Frontal Lobe	Middle frontal gyrus right	56150	1.4	9.5	0.2	0.07 (±0.08)	0.19 (±0.15)	0.12
Frontal Lobe	Posterior orbital gyrus left	7098	0	58.8	0	0.09 (±0.07)	0.41 (±0.20)	0.32
Frontal Lobe	Posterior orbital gyrus right	7469	8.5	38.6	5.5	0.14 (±0.11)	0.31 (±0.20)	0.18
Frontal Lobe	Precentral gyrus left	44289	0.3	18.1	0	0.06 (±0.07)	0.22 (±0.14)	0.16
Frontal Lobe	Precentral gyrus right	45058	1.3	24.1	0.8	0.09 (±0.09)	0.22 (±0.16)	0.13
Frontal Lobe	Pre-subgenual anterior cingulate gyrus left	1888	0	84.3	0	0.08 (±0.05)	0.41 (±0.08)	0.32
Frontal Lobe	Pre-subgenual anterior cingulate gyrus right	1166	2.6	90	0.3	0.14 (±0.07)	0.47 (±0.10)	0.33
Frontal Lobe	Straight gyrus (gyrus rectus) left	5397	16.6	69.7	10.4	0.18 (±0.11)	0.35 (±0.10)	0.17
Frontal Lobe	Straight gyrus (gyrus rectus) right	6467	6.7	41.5	3.5	0.17 (±0.11)	0.28 (±0.12)	0.11
Frontal Lobe	Subcallosal area left	527	8	87.7	8	0.19 (±0.07)	0.51 (±0.11)	0.32
Frontal Lobe	Subcallosal area right	496	4.2	92.5	4.2	0.16 (±0.09)	0.53 (±0.08)	0.36
Frontal Lobe	Subgenual anterior cingulate gyrus left	1932	1.1	73.3	1.1	0.14 (±0.08)	0.42 (±0.10)	0.28
Frontal Lobe	Subgenual anterior cingulate gyrus right	2044	1.9	60.5	1.9	0.16 (±0.09)	0.37 (±0.13)	0.21

Lobe	Structure	N-voxels	Population significant (%)	Family significant (%)	Significant both (%)	Population (±SD)	Family (±SD)	Δ h2
Frontal Lobe	Superior frontal gyrus left	69674	0.9	13.4	0.3	0.08 (±0.08)	0.20 (±0.16)	0.12
Frontal Lobe	Superior frontal gyrus right	73393	1.3	14.2	0.4	0.09 (±0.08)	0.22 (±0.19)	0.13
Central Structures	Caudate nucleus left	6042	68.6	96.4	68.6	0.32 (±0.10)	0.65 (±0.13)	0.33
Central Structures	Caudate nucleus right	6172	75.6	91.7	72.4	0.34 (±0.09)	0.55 (±0.15)	0.21
Central Structures	Nucleus accumbens left	580	29.7	92.2	29.7	0.24 (±0.10)	0.54 (±0.15)	0.31
Central Structures	Nucleus accumbens right	443	6.1	94.6	6.1	0.17 (±0.08)	0.53 (±0.09)	0.37
Central Structures	Pallidum (globus pallidus) left	1893	32.2	89.2	32.2	0.22 (±0.09)	0.50 (±0.11)	0.28
Central Structures	Pallidum (globus pallidus) right	1812	39.8	70.3	28.5	0.27 (±0.07)	0.37 (±0.11)	0.11
Central Structures	Putamen left	7819	32.6	96.2	32.6	0.24 (±0.07)	0.67 (±0.14)	0.43
Central Structures	Putamen right	7325	59.2	90.8	57.5	0.29 (±0.07)	0.48 (±0.14)	0.19
Central Structures	Thalamus left	11142	15.2	76.2	14.4	0.18 (±0.10)	0.40 (±0.11)	0.22
Central Structures	Thalamus right	10774	6.8	68.9	6.6	0.16 (±0.09)	0.39 (±0.13)	0.23

The average voxel-wise heritability in 72 brain structures from family-based and population-based estimates. N = total number of tested voxels, Significant Population (%) = percentage of N which was significant at False discovery Rate threshold q = 0.05 in the population-based estimates. Significant family (%) = percentage of N which was significant at False discovery Rate threshold q = 0.05 in family-based estimates. Significant both (%) = percentage of N which was significant at False discovery Rate threshold q = 0.05 in both family-based as population-based estimates. Family = mean family-based heritability estimates per brain structure, Population = mean population-based heritability estimates per brain structure. Δ h2 = difference in regional heritability (Family-Population).

Enhancement of association signal

We explored if applying our heritability map could enhance the statistical association signal of previously discovered genome-wide significant loci. As expected the T-allele of rs77956314 (*HRK*) associated with a smaller total volume of the hippocampus (p = 5.1 x 10^{-7}) and the C-allele of rs945270 (*KTN1*) significantly associated with larger total volume of the putamen (p = 4.3 x 10^{-3}). When excluding the less heritable voxels the average heritability in the remaining voxels increased **(Figure 3A and 3B)**. With rising average heritability we observed a gradual decrease in p-values (Figure 3C), and consequently a more significant association of HRK with the more heritable part of the hippocampus**.** The maximum enrichment of association was reached when the 10% most significantly heritable voxels when combining heritability information from family-based and population-based studies was used. This increase corresponds to a 95.9 times more significant association, as the p-value decreased from p = 5.1 x 10^{-7} to p = 5.4 x 10^{-9}. Using only the family-based estimates the association was 12.9 times more significant. A less substantial decrease in p-value was observed for the association of KTN1 with the more heritable part of the putamen **(Figure 3D).** The p-value decreased when restricting to voxels that belong to the 25% most heritable voxels from the only the family-based study. This corresponds to a 5.5 times more significant association (p-value decrease from p = 4.3 x 10^{-3} to p = 7.9 x 10^{-4}).

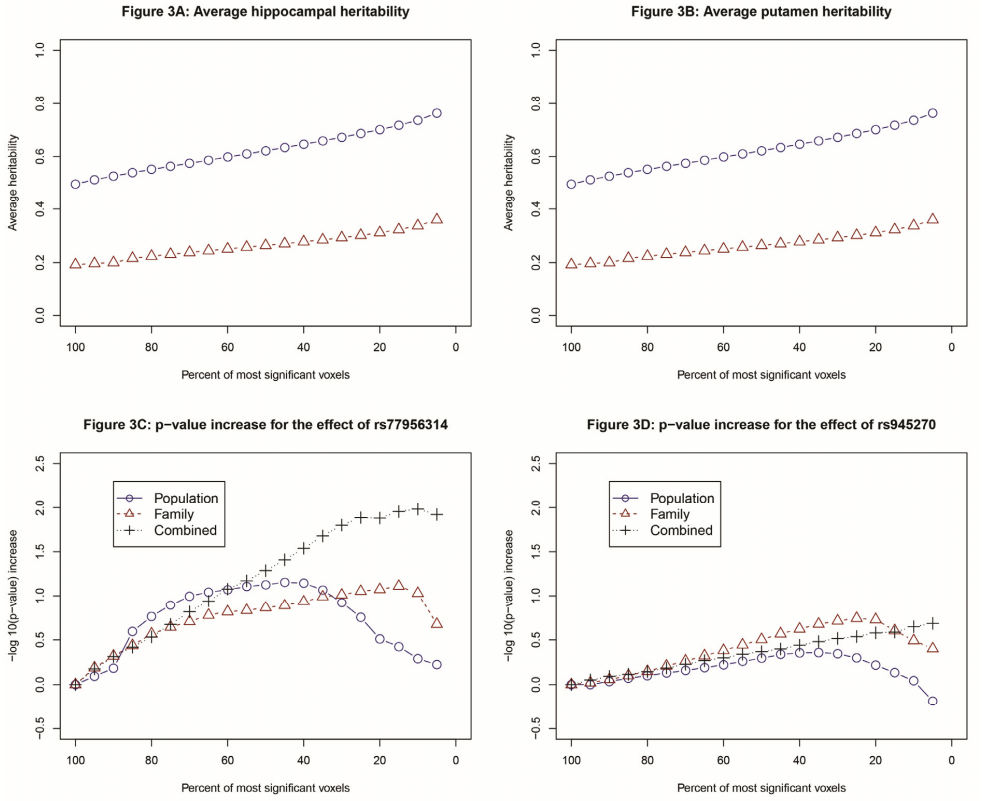


Figure 3. Enhancement of the association signal of variants with the most heritable voxels of the hippocampus and putamen. A,B: Average heritability (y-axis) of the voxels in hippocampus (A) and putamen (B) given a percentage of the most heritable voxels in that region (x-axis) in steps of 5%. **C,D:** The -log(p-value) increase comparing the p-value of association with subsets of the most heritable voxels and all voxels in the region. The –log(p-value) increase for association of hippocampal with rs77956314 (*HRK* gene) and putamen voxels with rs945270 (*KTN1* gene) is shown. Associations were corrected for age, sex, and the first three principal components.

Discussion

In this study we presented grey matter voxel heritability maps at resolution of 1×1×1 mm from population- and family-based studies. First we found that clusters of voxels that are significantly heritable in family-based heritability estimates as well as in an unrelated population-based study are predominantly located in subcortical regions. Second, when comparing the overall regional patterns of voxel-wise heritability the family-based estimates were always higher compared to population-based estimates

and predicted the population-based heritability estimates. Lastly, we showed that the heritability estimates from our studies could be used to enhance the association signal of two genetic variants with subcortical volumes.

Voxels with significant heritability formed clusters within mainly the subcortical structures. This is in line with the findings of previous studies that the volumes of subcortical structure are among the most heritable in the brain.[40] There are multiple explanations for this consistent finding. First, subcortical structures probably are under tight genetic control as they exert vital functions within the brain. The percentage of significantly heritable voxels was relatively low in the frontal and parietal lobes. Although intra-individual measurability was high throughout the brain (Supplementary Figure 1), intra-individual differences in cortical folding patterns could explain the lower heritability in frontal and parietal regions. These might give a reliable measurability of the voxels, while it makes comparisons of voxel values between individuals less meaningful, thus yielding a lower heritability compared with the subcortical structures. Finally, environmental effects could have a larger effect on cortical grey matter compared to subcortical structures. As the effects of non-genetic factors (e.g. lifestyle factors) accumulate over an individual's lifetime, the heritability of total brain volume and brain structures volume was found to reduce in adulthood up until old age[41] in line with the accumulation of environmental influences over age. Their reported maximum age was 70 years. We studied relatively old participants (~65 years), therefore study participants might have reduced estimated heritability because of their older age.

Apart from the subcortical structures, we found three cortical regions in the left hemisphere, the dominant hemisphere in over 95% of individuals, involved in speech production and word processing to have more than 5% significant voxels; the subcallosal area (also called Broca area), central part of the superior temporal gyrus (contains Wernicke's area) and the lingual gyrus. Moreover, their right counterparts contained less significant voxels compared to the left side. Language skills[42] and brain networks [43] are thought to be under tight genetic control and the left hemisphere language areas have been found more heritable than the right hemisphere before.[14] Regions with significant heritability could in theory be connected by white matter

connections, which in turn then also are under high genetic control, suggesting a common genetic architecture. In a recent report evidence for this theory was found.[44] Cortical thickness in some regions with high heritability, were connected by heritable white matter connections. These connections and the cortical regions were anatomically distant but showed significant genetically correlation.[44]

We found a relatively stable difference in the regional patterns of the total additive genetic heritability. The heritability calculated from familial relations was always higher than the total additive variance explained by all autosomal variants calculated in unrelated subjects. This known difference between family and population-based heritability estimates has been extensively described.[45,46] The difference can in part be explained by overestimation of heritability in families due to sharing of environmental factors within the family. These factors are interpreted as genetic effects and cause the overestimation of heritability in twin and nuclear family studies.[47] Subjects in multi-generational families share less environmental factors. Therefore multi-generational families, as ASPS-Fam and especially the ERF study, are more likely to yield an unbiased estimate of heritability. However, we assumed that all environmental factors affecting brain voxel volume are uncorrelated among family members (unique environmental effects) therefore some unassessed common environmental effects might be causing the higher heritability in our family-based estimates. At the same time an underestimation of the heritability calculated from genetic data in unrelated populations could occur because of an incomplete coverage of the causal variants and exclusion of rare variants. We used imputed data to increase coverage of the causal variants. Imputed data provide a much denser coverage of the genome than only genotyped variants, but we did exclude rare variants (MAF<0.01) which may in part be responsible for some missing heritability.

The overall regional patterns of heritability from families strongly predicted the population-based heritability. This suggests that the regional pattern of variance explained by additive genetic effects is similar across populations, despite different ways to measure heritability, study design and scanner types. On the website (http://www.imagene.nl/heritability) both the population-based estimates and the

family-based estimates can be viewed separately and can be downloaded. Combining current maps with results from other studies will further increase accuracy of the heritability estimates.

Heritability in genetic studies

Within the putamen and hippocampus we observed highly heritable clusters of grey matter voxels alternating with parts of the subcortical structures that were less heritable. Differences in heritability within structures might be due to technical limitations (e.g. voxels that are difficult to measure) or due to genetic or functional correlations. We hypothesized that studying the genetics of only highly heritable voxels could enhance signals in imaging genetics, either through reducing signal to noise ratio or through studying a more genetically homogeneous trait. We picked two genetic variants with a proven and strongly replicated biological effect, identified through genome-wide association studies, on the subcortical structure volume (hippocampus, putamen) to explore if enhancement was possible [5]. We show enhancement of the statistical signal of almost hundred-fold for the association of *HRK* (rs77956314) with hippocampal volume and a five-fold increase for the association of *KTN1* (rs945270) with putamen volume. Based on Figure 3 we can deduct that for future genetic studies in both examples a maximum power for association analyses was observed using voxels with a heritability over ~0.3 from the population-based heritability estimates and a heritability over ~0.7 from family-based heritability estimates. Despite these encouraging results there are limitations of our analysis. First, we only tested two genetic variants in two subcortical structures. While we expect that the increased signal of genetic variants with more heritable voxels will not be limited to the two variants tested in current study, future studies applying this method should be performed to determine whether this truly is the case. Second, we calculated heritability estimates and genetic association of *HRK* and *KTN1* variants with voxels in the same subjects of the Rotterdam Study. As voxels with a large (technical) measurement error have lower heritability and therefore were excluded first in our analysis, the decreased measurement error of the more heritable voxels could result in the more significant association of genetic variants. In other words, the enhancement of signal is a reflection of a higher signal to noise ratio. Also a higher test re-test reliability of the highly

heritable voxels, reduce signal to noise ratio. Third, we used the same data for the calculation of population-based heritability estimates and genetic testing, resulting in a possible inflation of the increase in signal due to non-independence.[48] However, when only the family-based heritability estimates were used to select the voxels for genetic associations (**Figure 3C,D**) the analyses were independent. In these analyses, we still observed an increase in the signal – and the enhancement was actually even stronger for the putamen – arguing against inflation due to non-independence. However, for the hippocampus the best enhancement was achieved using the combined sample when restricting to less than 55% the most significant voxels. While this could be due to non-independence, this is contradicted by the fact that the population-only results (i.e., fully dependent) are in fact worse at this and lower percentages. An explanation other than non-independence could be that the combined sample provides more accurate heritability estimates and therefore results in a better enhancement. Last, highly heritable voxels which are in close proximity of each other could share their genetic background. However finding a cluster of heritable voxels does not directly prove genetic correlation.

Strengths and limitations

Major strengths of this study are the large sample size of the population based study and unified imaging processing. Subjects from ERF and the Rotterdam Study subjects were scanned using the same 1.5T scanner, identical MRI protocols and images were processed with exactly the same software. The ASPS-Fam was scanned on a 3T scanner, but segmented using similar protocols and VBM processing was performed in the same way as ERF and the Rotterdam Study. Important to note is that softwares used for tissue segmentation are different, but both implement the same kNN algorithm.[26] The ERF and the Rotterdam Study both are both from the Netherlands, a genetically homogeneous country.[49] The ASPS-Fam study is from Austria, Austrians likely have slightly different genetic architecture than the Dutch. Maximum likelihood iterative optimization was used to estimation heritability. The iterations are prone to convergence failures when sample sizes are small. The percentage of voxels that did not converge was 9% in ASPS-Fam ($N_{participants} = 369$) and 36% in ERF ($N_{participants} = 122$). The methods used for population-based estimation of heritability always output an

estimate. It has been shown that not converging occurs frequently in small datasets in SOLAR producing conservative estimates.[47,50] We further note that using only VBM to assess heritability of brain morphology is a limitation of the current study. Cortical thickness, surface area and other MRI measures, including tensor-based (i.e. deformation) morphometry (TBM)[51,52] and shape analysis are all potentially interesting for future heritability and genetic studies. The differences between measures have been attributed both to biology[53,54] and methodology.[55,56] Most probably, these measures reflect a different genetic architecture[53] and should therefore be studied separately.

Future perspectives

Genetic association with several voxels within an anatomical structure is biologically relevant as it shows an important genetic contribution to a sub region of the structure. Apart from the biological relevance, this sub region of voxels could have clinical significance. For example, it was shown previously that subfields of the anatomically defined hippocampus contributed differently to schizophrenia[57] and β-Amyloid load.[58] If only highly heritability brain voxels are studied in future voxel-wise genome-wide association studies we do not expect statistical signals to be uniformly enhanced. However, for the tested genetic variant that was identified for putamen volume, we did find statistical enhancement. High heritability estimates capture a variety of sources that can affect power to detect associations, including lower signal to noise ratios and higher genetic homogeneity (i.e. genetic correlation). Using these benefits to increase statistical signal is desirable, irrespective of the underlying cause. Ideally we envision selecting groups of voxels for genetic studies based on high heritability and measured high genetic correlation. Genetic correlation can be calculated for any of the commonly used MRI-measures, but it would still require genetic testing of sufficiently powered (large) studies. A promising future direction would be to enable the calculation of genetic correlations, genetic association (millions of voxels times millions of genetic variants) and meta-analyses of these associations. Programs which make the calculation of genetic correlation and genetic association computationally possible in sufficiently powered studies (i.e. meta-analyses) are essential to the field. Currently these programs tailored to large scale genetic studies are developed and genetic studies started.[59] The

results of these studies will be able to prove to which extend clusters of heritable voxels have a common genetic architecture.

Conclusions

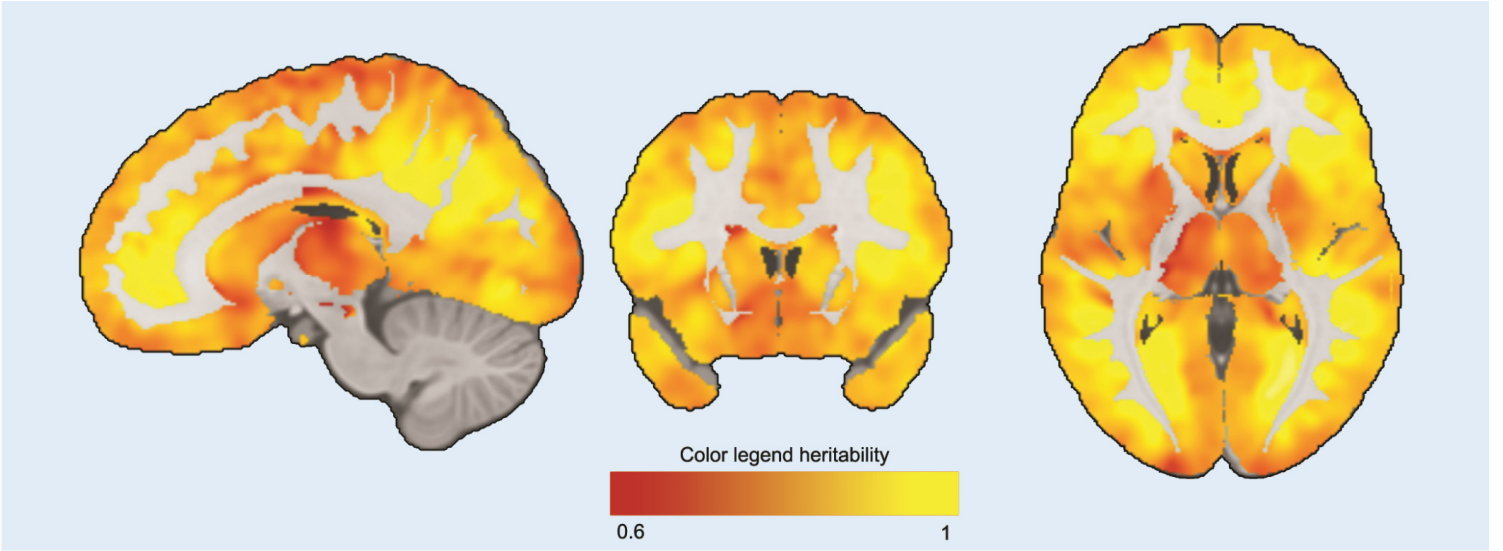
Heritability estimates can be reliably estimated using different methods and on different cohorts and combining heritability estimates from multiple studies leads to the construction of a reliable heritability map of grey matter. These maps can be used to prioritize highly heritable regions in future genetic imaging studies.

References:

1. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era--concepts and misconceptions. *Nat Rev Genet* 2008; **9**(4): 255-66.
2. Visscher PM, Medland SE, Ferreira MA, et al. Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet* 2006; **2**(3): e41.
3. Yang J, Benyamin B, McEvoy BP, et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 2010; **42**(7): 565-9.
4. Polderman TJ, Benyamin B, de Leeuw CA, et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet* 2015; **47**(7): 702-9.
5. Hibar DP, Stein JL, Renteria ME, et al. Common genetic variants influence human subcortical brain structures. *Nature* 2015; **520**(7546): 224-9.
6. Jahanshad N, Kochunov PV, Sprooten E, et al. Multi-site genetic analysis of diffusion images and voxelwise heritability analysis: a pilot project of the ENIGMA-DTI working group. *Neuroimage* 2013; **81**: 455-69.
7. Jahanshad N, Lee AD, Barysheva M, et al. Genetic influences on brain asymmetry: a DTI study of 374 twins and siblings. *Neuroimage* 2010; **52**(2): 455-69.
8. Kochunov P, Jahanshad N, Marcus D, et al. Heritability of fractional anisotropy in human white matter: a comparison of Human Connectome Project and ENIGMA-DTI data. *Neuroimage* 2015; **111**: 300-11.
9. Kochunov P, Glahn DC, Lancaster JL, et al. Genetics of microstructure of cerebral white matter using diffusion tensor imaging. *Neuroimage* 2010; **53**(3): 1109-16.
10. Ganjgahi H, Winkler AM, Glahn DC, Blangero J, Kochunov P, Nichols TE. Fast and powerful heritability inference for family-based neuroimaging studies. *Neuroimage* 2015; **115**: 256-68.
11. Kochunov P, Fu M, Nugent K, et al. Heritability of complex white matter diffusion traits assessed in a population isolate. *Hum Brain Mapp* 2016; **37**(2): 525-35.
12. Brouwer RM, Mandl RC, Peper JS, et al. Heritability of DTI and MTR in nine-year-old children. *Neuroimage* 2010; **53**(3): 1085-92.
13. Hulshoff Pol HE, Schnack HG, Posthuma D, et al. Genetic contributions to human brain morphology and intelligence. *J Neurosci* 2006; **26**(40): 10235-42.
14. Thompson PM, Cannon TD, Narr KL, et al. Genetic influences on brain structure. *Nat Neurosci* 2001; **4**(12): 1253-8.
15. Peper JS, Schnack HG, Brouwer RM, et al. Heritability of regional and global brain structure at the onset of puberty: a magnetic resonance imaging study in 9-year-old twin pairs. *Hum Brain Mapp* 2009; **30**(7): 2184-96.

16. Bis JC, DeCarli C, Smith AV, et al. Common variants at 12q14 and 12q24 are associated with hippocampal volume. *Nat Genet* 2012; **44**(5): 545-51.
17. Stein JL, Medland SE, Vasquez AA, et al. Identification of common variants associated with human hippocampal and intracranial volumes. *Nat Genet* 2012; **44**(5): 552-61.
18. Hofman A, Brusselle GG, Darwish Murad S, et al. The Rotterdam Study: 2016 objectives and design update. *Eur J Epidemiol* 2015; **30**(8): 661-708.
19. Willer CJ, Sanna S, Jackson AU, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* 2008; **40**(2): 161-9.
20. Ikram MA, van der Lugt A, Niessen WJ, et al. The Rotterdam Scan Study: design update 2016 and main findings. *Eur J Epidemiol* 2015; **30**(12): 1299-315.
21. Sayed-Tabatabaei FA, van Rijn MJ, Schut AF, et al. Heritability of the function and structure of the arterial wall: findings of the Erasmus Rucphen Family (ERF) study. *Stroke* 2005; **36**(11): 2351-6.
22. Ibrahim-Verbaas CA, Zorkoltseva IV, Amin N, et al. Linkage analysis for plasma amyloid beta levels in persons with hypertension implicates Abeta-40 levels to presenilin 2. *Hum Genet* 2012; **131**(12): 1869-76.
23. Schmidt R, Lechner H, Fazekas F, et al. Assessment of cerebrovascular risk profiles in healthy persons: definition of research goals and the Austrian Stroke Prevention Study (ASPS). *Neuroepidemiology* 1994; **13**(6): 308-13.
24. Schmidt R, Fazekas F, Kapeller P, Schmidt H, Hartung HP. MRI white matter hyperintensities: three-year follow-up of the Austrian Stroke Prevention Study. *Neurology* 1999; **53**(1): 132-9.
25. Seiler S, Pirpamer L, Hofer E, et al. Magnetization transfer ratio relates to cognitive impairment in normal elderly. *Front Aging Neurosci* 2014; **6**: 263.
26. Vrooman HA, Cocosco CA, van der Lijn F, et al. Multi-spectral brain tissue segmentation using automatically trained k-Nearest-Neighbor classification. *Neuroimage* 2007; **37**(1): 71-81.
27. Good CD, Johnsrude IS, Ashburner J, Henson RN, Friston KJ, Frackowiak RS. A voxel-based morphometric study of ageing in 465 normal adult human brains. *Neuroimage* 2001; **14**(1 Pt 1): 21-36.
28. Roshchupkin GV, Adams HH, van der Lee SJ, et al. Fine-mapping the effects of Alzheimer's disease risk loci on brain morphology. *Neurobiol Aging* 2016; **48**: 204-11.
29. Smith SM, Jenkinson M, Woolrich MW, et al. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 2004; **23 Suppl 1**: S208-19.
30. Fonov V, Evans AC, Botteron K, et al. Unbiased average age-appropriate atlases for pediatric studies. *Neuroimage* 2011; **54**(1): 313-27.
31. Hammers A, Allom R, Koepp MJ, et al. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Hum Brain Mapp* 2003; **19**(4): 224-47.
32. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; **86**(2): 420-8.
33. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011; **88**(1): 76-82.
34. Visscher PM, Yang J, Goddard ME. A commentary on 'common SNPs explain a large proportion of the heritability for human height' by Yang et al. (2010). *Twin Res Hum Genet* 2010; **13**(6): 517-24.
35. Adams HH, Verlinden VJ, Callisaya ML, et al. Heritability and Genome-Wide Association Analyses of Human Gait Suggest Contribution of Common Variants. *J Gerontol A Biol Sci Med Sci* 2016; **71**(6): 740-6.
36. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 1998; **62**(5): 1198-211.
37. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; **26**(17): 2190-1.

38. Blangero J, Williams JT, Almasy L. Variance component methods for detecting complex trait loci. *Adv Genet* 2001; **42**: 151-81.
39. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met* 1995; **57**(1): 289-300.
40. Blokland GA, de Zubicaray GI, McMahon KL, Wright MJ. Genetic and environmental influences on neuroimaging phenotypes: a meta-analytical perspective on twin imaging studies. *Twin Res Hum Genet* 2012; **15**(3): 351-71.
41. Batouli SA, Trollor JN, Wen W, Sachdev PS. The heritability of volumes of brain structures and its relationship to age: a review of twin and family studies. *Ageing Res Rev* 2014; **13**: 1-9.
42. Gayan J, Olson RK. Reading disability: evidence for a genetic etiology. *Eur Child Adolesc Psychiatry* 1999; **8 Suppl 3**: 52-5.
43. Budisavljevic S, Dell'Acqua F, Rijsdijk FV, et al. Age-Related Differences and Heritability of the Perisylvian Language Networks. *J Neurosci* 2015; **35**(37): 12625-34.
44. Shen KK, Dore V, Rose S, et al. Heritability and genetic correlation between the cerebral cortex and associated white matter connections. *Hum Brain Mapp* 2016; **37**(6): 2331-47.
45. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A* 2012; **109**(4): 1193-8.
46. Zuk O, Schaffner SF, Samocha K, et al. Searching for missing heritability: designing rare variant association studies. *Proc Natl Acad Sci U S A* 2014; **111**(4): E455-64.
47. Koran ME, Thornton-Wells TA, Jahanshad N, et al. Impact of family structure and common environment on heritability estimation for neuroimaging genetics studies using Sequential Oligogenic Linkage Analysis Routines. *J Med Imaging (Bellingham)* 2014; **1**(1): 014005.
48. Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci* 2009; **12**(5): 535-40.
49. Boomsma DI, Wijmenga C, Slagboom EP, et al. The Genome of the Netherlands: design, and project goals. *Eur J Hum Genet* 2014; **22**(2): 221-7.
50. Blangero J, Diego VP, Dyer TD, et al. A kernel of truth: statistical advances in polygenic variance component models for complex human pedigrees. *Adv Genet* 2013; **81**: 1-31.
51. Brun CC, Lepore N, Pennec X, et al. Mapping the regional influence of genetics on brain structure variability--a tensor-based morphometry study. *Neuroimage* 2009; **48**(1): 37-49.
52. Yoon U, Perusse D, Lee JM, Evans AC. Genetic and environmental influences on structural variability of the brain in pediatric twin: deformation based morphometry. *Neurosci Lett* 2011; **493**(1-2): 8-13.
53. Winkler AM, Kochunov P, Blangero J, et al. Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. *Neuroimage* 2010; **53**(3): 1135-46.
54. Voets NL, Hough MG, Douaud G, et al. Evidence for abnormalities of cortical development in adolescent-onset schizophrenia. *Neuroimage* 2008; **43**(4): 665-75.
55. Hutton C, Draganski B, Ashburner J, Weiskopf N. A comparison between voxel-based cortical thickness and voxel-based morphometry in normal aging. *Neuroimage* 2009; **48**(2): 371-80.
56. Blankstein U, Chen JYW, Mincic AM, McGrath PA, Davis KD. The complex minds of teenagers: Neuroanatomy of personality differs between sexes. *Neuropsychologia* 2009; **47**(2): 599-603.
57. Kuhn S, Musso F, Mobascher A, Warbrick T, Winterer G, Gallinat J. Hippocampal subfields predict positive symptoms in schizophrenia: first evidence from brain morphometry. *Transl Psychiatry* 2012; **2**: e127.
58. Schroeder C, Park MT, Germann J, et al. Hippocampal shape alterations are associated with regional Abeta load in cognitively normal elderly individuals. *Eur J Neurosci* 2016.
59. Roshchupkin GV, Adams HHH, Vernooij MW, et al. HASE: Framework for efficient high-dimensional association analyses. *Sci Rep* 2016; **6**(36076).



Supplementary Figure 1. Example of the intraclass correlation (ICC) in 83 individuals scanned twice within several weeks. In general voxels have a high ICC. All results can be interactive accessed (www.imagene.nl/heritability) and downloaded from the website.

Chapter 4.2

Fine-mapping the effects of Alzheimer's disease risk loci on brain morphology

Gennady V. Roshchupkin*, Hieab H.H. Adams*, Sven J. van der Lee*, Meike W. Vernooij, Cornelia M. van Duijn, André G. Uitterlinden, Aad van der Lugt, Albert Hofman, Wiro J. Niessen, M. Arfan Ikram.

*Authors contributed equally

Neurobiology of Aging. 2016 December; 48: 204-211.

Abstract

Background: The neural substrate of genetic risk variants for Alzheimer's disease (AD) remains unknown. We studied their effect on healthy brain morphology to provide insight into disease etiology in the pre-clinical phase.

Methods: We included 4071 non-demented, elderly participants of the population-based Rotterdam Study who underwent brain MRI and genotyping. We performed voxel-based morphometry (VBM) on all gray matter voxels for 19 previously identified, common AD risk variants. Whole-brain expression data from the Allen Human Brain Atlas was used to examine spatial overlap between VBM association results and expression of genes in AD risk loci regions.

Results: Brain regions most significantly associated with AD risk variants were the left postcentral gyrus with *ABCA7* (rs4147929, $p = 4.45\times10^{-6}$), right superior frontal gyrus by *ZCWPW1* (rs1476679, $p = 5.12\times10^{-6}$), and right postcentral gyrus by *APOE* ($p = 6.91\times10^{-6}$). Though no individual voxel passed multiple testing correction, we found significant spatial overlap between the effects of AD risk loci on VBM and the expression of genes (*MEF2C, CLU, SLC24A4)* in the Allen Brain Atlas. Results are available online on www.imagene.nl/ADSNPs/.

Conclusion: In this single largest imaging genetics dataset worldwide, we found that AD risk loci affect cortical gray matter in several brain regions known to be involved in AD, as well as regions that have not been implicated before.

Introduction

Alzheimer's disease (AD) is a complex neurodegenerative disease and the most common cause of dementia. It has a long preclinical phase, during which there are no symptoms but structural brain changes can already be detected, such cortical atrophy and localized atrophy of the hippocampus.[1,2]

In recent years, common genetic risk factors for AD have been discovered through large meta-analyses of genome-wide association studies (GWAS).[3] However, the underlying neurobiological substrate leading to AD for the genes assigned to these risk loci remains to be uncovered. Identifying the brain structures affected by these genes can increase our understanding of AD and aid future functional studies. Previous studies have investigated some of the AD risk loci in relation to neuroimaging measures.[4-7] However, they were generally focused on candidate regions that are known to play a role in AD, such as the hippocampus [4,5] or did not investigate all known risk loci.[6,7] Unbiased approaches for analyzing brain images have great potential to give novel insights that would not have been considered a priori. Voxel-based morphometry (VBM) is a hypothesis-free technique for analyzing brain imaging data that characterizes regional tissue concentration differences across the whole brain, without the need to predefine regions of interest.[8]. Using VBM, we studied the association of 19 AD genetic risk loci with gray matter morphology at the voxel level in 4071 non-demented elderly from the Rotterdam study. This study provides insight into non-diseased brain morphology. Such knowledge is complementary and intertwined with better understanding disease etiology in the pre-clinical phase. Subsequently, we co-localized our results with publicly available genetic expression data. We thus identified genetic associations with known as well as novel regions affected in AD.

Methods

Study Population

The Rotterdam Study is an ongoing population-based cohort study in the Netherlands investigating diseases in the elderly and currently consists of 14,926 residents of Rotterdam who were aged 45 years or more at baseline.[9,10] The initial cohort was started in 1990 and expanded in 2000 and 2005. The whole population is subject to a set of multidisciplinary examinations every four years. MRI was implemented in 2005 and 5,430 persons scanned until 2011 were eligible for this study. We excluded individuals with incomplete acquisitions, scans with artifacts hampering automated processing, participants with MRI-defined cortical infarcts, and subjects with dementia or stroke at the time of scanning. This resulted in a final study population of 4,071 non-demented persons with information available on both genome-wide genotyping and MRI data. The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC and by the Ministry of Health, Welfare and Sport of the Netherlands, implementing the Wet Bevolkingsonderzoek: ERGO (Population Studies Act: Rotterdam Study). All participants provided written informed consent to participate in the study and to obtain information from their treating physicians.

Imputation of genotypes

The Illumina 550K and 550K duo arrays were used for genotyping. Samples with low call rate (<97.5%), with excess autosomal heterozygosity (>0.336) or with sex-mismatch were excluded, as were outliers identified by the identity-by-state clustering analysis (outliers were defined as being >3 standard deviation (SD) from population mean or having identity-by-state probabilities >97%). A set of genotyped input SNPs with call rate >98%, MAF >0.001 and Hardy–Weinberg equilibrium (HWE) p-value > 10^{-6} was used for imputation. The Markov Chain Haplotyping (MACH) package version 1.0 software (Imputed to plus strand of NCBI build 37, 1000 Genomes phase I version 3) and minimac version 2012.8.6 were used for imputation. *APOE* status was genotyped separately, using a polymerase chain reaction, as described previously.[11] *APOEε4* was coded as the number of *APOEε4* alleles.

MRI data

From August 2005 onwards, a dedicated 1.5 Tesla MRI scanner (GE Healthcare, Milwaukee, Wisconsin, USA) is operational in the Rotterdam Study research center in Ommoord. This scanner is operated by trained research technicians and all imaging data are collected according to standardized image acquisition protocols.[10] Brain MRI scans included a high-resolution 3D T1-weighted fast RF spoiled gradient recalled acquisition in steady state with an inversion recovery pre-pulse (FASTSPGR-IR) sequence with thin slices (voxel size<1mm^3).[10]

Image processing

Voxel based morphometry (VBM) was performed according to an optimized VBM protocol.[12] First, all T1-weighted images were segmented into supratentorial gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) using a previously described k-nearest neighbor (kNN) algorithm, which was trained on six manually labeled atlases.[13] FSL software[14] was used for VBM data processing. Then, all GM density maps were non-linearly registered to the standard GM probability template. For this study we chose the ICBM MNI152 GM template (Montreal Neurological Institute) with a 1×1×1 mm^3 voxel resolution. The MNI152 standard-space T1-weighted average structural template is derived from 152 structural images, which have been warped and averaged into the common MNI152 co-ordinate system after high-dimensional nonlinear registration.

A spatial modulation procedure was used to avoid differences in absolute GM volume due to the registration. This involved multiplying voxel density values by the Jacobian determinants estimated during spatial normalization. All images were smoothed using a 3mm (FWHM 8mm) isotropic Gaussian kernel.

Statistical analysis

Linear regression models were fitted with voxel values of GM modulation density as the dependent variable and age, sex, and the number of reference alleles (risk alleles for Alzheimer's disease, **Supplementary Table 1**) as independent variables. In total 1,534,602 voxels were processed.

To perform a nonparametric permutation test, we randomly shuffled the genotype data between persons and performed the VBM association analysis with all 1,534,602 voxels in gray matter. This was repeated 10,000 times and for every permutation we saved the minimum p-value. Subsequently, we took the 5th percentile of this minimum p-value distribution to compute FWE p-value threshold, which was 3.0×10^{-7} .[15] This was then divided by 19 to account for the number of independent SNPs, resulting in the final threshold of 1.66×10^{-8}.

Genetic Risk Score

Genetic risk scores (GRS) were constructed by multiplying the number of risk alleles by their reported odds ratio (after natural logarithm transformation) for the disease, and summing this weighted allele score of each variant up into a disease risk score for AD.[16] We tested a GRS based on all 19 AD SNPs and second GRS excluding *APOEε4*.

APOE ε4 stratified analysis

To investigate whether it is possible to enrich association signal of AD variants on brain morphology we split our sample into groups with increased chance for AD pathology by stratifying it for *APOEε4* status. In total there were 1,168 carrier and 2,903 non-carrier in our data set.

The Allen Human Brain gene expression analysis

The Allen Human Brain Atlas (http://human.brain-map.org) includes RNA microarray data collected from the postmortem brains of six donors, with no known neuropsychiatric or neuropathological history. Around 500 samples per subject, per hemisphere were tested for expression profiles of 29,191 genes represented by 58,692 probes. The expression profiles were normalized across samples and across different brains as described previously.[17]

In our analysis we used the three Caucasian donors. For each of these donors we extracted expression profiles of 216 genes, which are located within ± 500kb from AD risk loci and used the MNI coordinates to map the location of the samples. For each probe we derived z-score statistics, which represent deviation of gene expression in that sample relative to background expression. Next, using the VBM association results from

all 19 tested AD SNPs, we formed clusters at the significance threshold of observed p-value<0.05 and identified all tissue samples localized inside these clusters or within 10 voxels from them.

We have performed 10,000 random VBM analyses to generate p-value maps of null associations. We formed clusters, based on a p-value threshold of < 0.05, and subsequently linked these to probes, exactly as described above. For three donors and all probes in the 216 genes (in total 667) we calculated the t-test statistic with a null hypothesis that expression of the gene within clusters is not significantly different from background expression. We saved the minimum p-values for every random VBM map. Subsequently, we took the 5th percentile of this minimum p-value distribution to compute the FWE p-value threshold. The obtained threshold was 1.7×10^{-5}. Then we performed the same t-test with the AD VBM maps. Thus, we compared expression of genes around AD risk loci in regions identified in the VBM analysis with their background expression in the brain.

Regional analysis

We used the Hammer atlas[18] to segment the gray matter into 36 regions for both hemispheres and compare effects on specific brain regions. We summed all voxels values inside segmented region to estimate gray matter volume. For every risk locus and brain region we run the same regression model as for the VBM analysis.

Visualization

To provide easy access to study results for the research community, we developed an online freely available interactive visualization tool (www.imagene.nl/ADSNPs/).

Results

Association of AD risk loci to voxel-based morphometry

The study population for VBM analysis consisted of 4,071 non-demented persons with information available on both genome-wide genotyping and MRI data from the population-based Rotterdam Study. The mean age was 64.7 (± 10.7) years and 2,251 (55%) subjects were women.

We studied the association of 19 AD risk loci with 1,534,602 voxels of gray matter. None of the associations reached the multiple-testing correction threshold 1.66×10^{-8}. **Table 1** shows all associations between AD risk loci and gray matter voxel density with suggestive evidence for association p-values < 1×10^{-5}. The strongest associations of gray matter voxel with AD risk loci were found in the left postcentral gyrus, right superior frontal gyrus, and right postcentral gyrus. In **Figure 1** we show the three-dimensional maps of the nominally significant ($p < 0.05$) associations for the *APOE* risk loci. The negative clusters of *APOE* are located close to the medial temporal lobe, in particular around the hippocampus, whereas positive clusters are mainly in the occipital lobe. The GRSs association also did not reach the correction threshold. The strongest signal for risk score with *APOE* was found in the postcentral gyrus right ($p = 8.02\times10^{-6}$) and for the risk score without *APOE* in the lateral remainder of the occipital lobe right ($p = 1.47\times10^{-5}$). On **Supplementary Figure 1** are shown maps for all risk loci from **Table 1**. **Supplementary Table 2** provides the full list of the top three associated clusters of voxels for each risk locus and more detailed statistical information. All study results are available and can interactively be explored on the ImaGene website: www.imagene.nl/ADSNPs/.

In *APOE*ε4 stratified analysis none of the signals passed the threshold, however variant in MEF2C loci showed much more significant association compare to full sample size analysis (**Supplementary Table 5**). Additionally, the association signal for non-carrier was in general less significant (**Supplementary Table 6**).

analysis.

Table 1. The most significant voxel-wise association signals with p-values < 10^{-5}. Brain region labeling based on the Hammer Atlas segmentation. Effect direction indicates beta sign, and demonstrates risk loci associated with increasing gray matter tissue (+) or decreasing gray matter tissue (-).

Risk variant	Gene*	Minimum p-value	Effect direction	Brain Region
rs4147929	*ABCA7*	4.46×10^{-6}	-	postcentral gyrus left
rs1476679	*ZCWPW1*	5.12×10^{-6}	+	superior frontal gyrus right
rs429358/rs7412	*APOEε4*	6.91×10^{-6}	+	postcentral gyrus right
rs11771145	*EPHA1*	8.91×10^{-6}	-	precentral gyrus right
rs190982	*MEF2C*	9.55×10^{-6}	+	lateral remainder of occipital lobe right
Genetic Risk Score	All	8.02×10^{-6}	+	postcentral gyrus right
Genetic Risk Score	Without APOE	1.47×10^{-5} **	+	lateral remainder of occipital lobe right

* Assigned risk gene according to Lambert et al.[3]

** *p*-value is not less than 10-5, shown to compare with GRS without exclusion APOE.

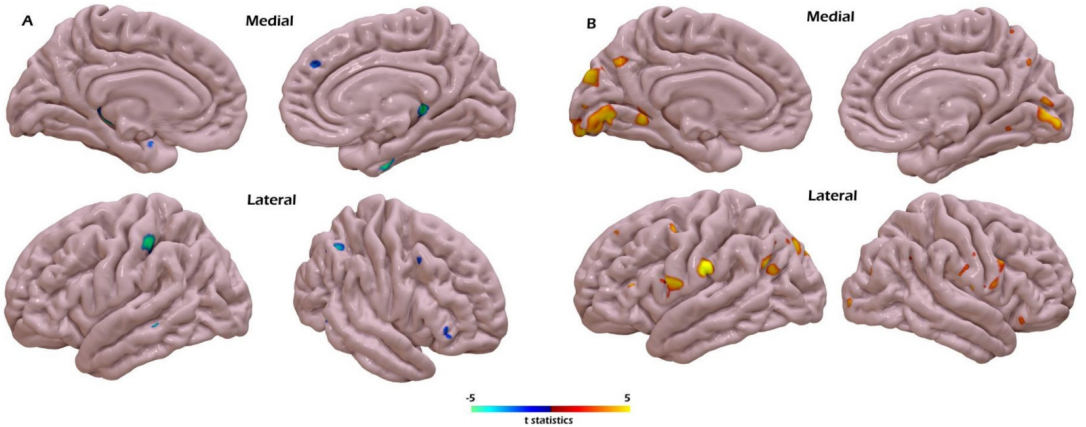


Figure 1. Projection of APOE risk loci association clusters from VBM to cortical surface. Colors reflect regression association: blue for negative (A), red for positive (B). Clusters formed based on nominal significant *p*-value threshold 0.05.

Spatial overlap between association maps and gene expression

To investigate whether the effect of AD risk loci on VBM overlaps with gene expression in the brain, we used the Allen Human Brain Atlas data. We overlapped brain regions identified through our VBM analysis with the maps of samples from three Allen Human Brain Atlas donors (**Figure 2**). We compared expression within the identified voxel clusters with background expression. In total we tested the expression profiles of 216 protein-coding genes, located ± 500kb from the AD variants (**Supplementary Table 1**). We found that *MEF2C, CLU, SLC24A4* were significantly expressed ($p < 1.7\times10^{-5}$) in the identified voxel clusters compared to other genes at that particular locus. Interestingly, these were the genes that were previously assigned as the risk genes at each respective locus based on a review of the available literature (Lambert et al., 2013) (**Table 2**). Additionally, we found genes showing significantly different expression, which are located in the risk loci but were previously not proposed as the causal gene for AD. These are: *NGEF* ($p = 7.57\times10^{-16}$) for the region around rs35349669 and *GSTK1* ($p = 1.01\times10^{-5}$) for the region around rs11771145. **Supplementary Table 2** provides the full list of genes and more detailed statistical information.

Table 2. Results of spatial overlap between VBM risk loci association and gene expression profiles of 3 Caucasian donors from the Allen Human Brain Atlas.

		Genes showing significant overlap				
Risk variant	**Putative causal gene [a]**	**Significant gene expression**	**Minimum p-value**	**Distance from risk loci, bp**	**Significant donors/ total number of donors**	**Significant probes/total number of probes**
rs10498633	*SLC24A4*	*SLC24A4*	1.50×10^{-5}	138.027	1/3	1/2
rs190982	*MEF2C*	*MEF2C*	1.41×10^{-5}	44.275	1/3	1/3
rs9331896	*CLU*	*CLU*	4.43×10^{-7}	13.252	1/3	1/1
rs35349669	*INPP5D*	*NGEF*	7.57×10^{-16}	325.080	3/3	2/2
rs11771145	*EPHA1*	*GSTK1*	1.01×10^{-5}	169.576	2/3	3/4

[a] Assigned causal gene according to Lambert et al.[3] The table shows genes in risk loci regions, for which expression differs significantly (at corrected threshold 1.7×10-5) from background expression in regions associated by VBM

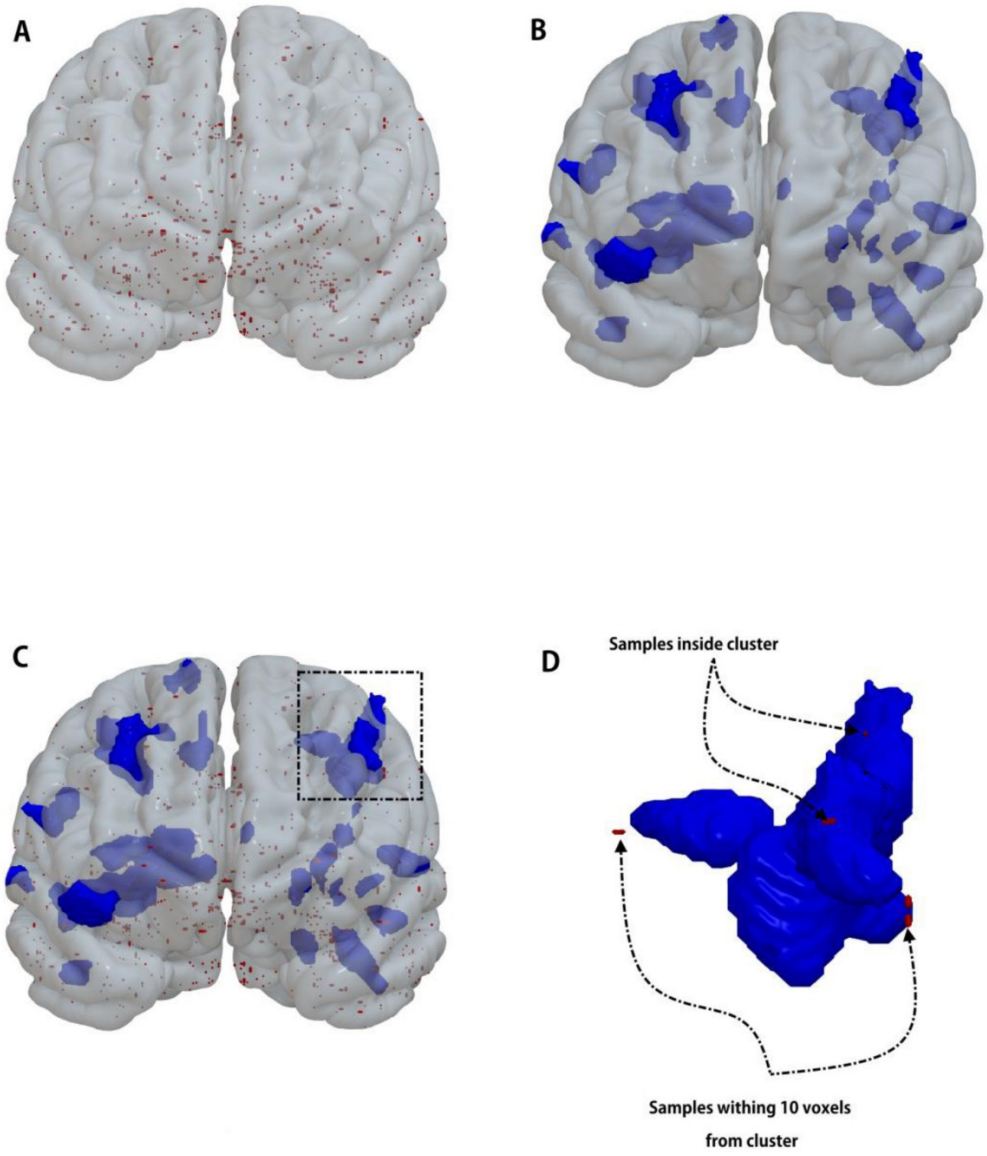


Figure 2. Example of spatial overlap between VBM association map for the MEF2C risk variant and MEF2C gene expression probes from Allen Human Brain Atlas. (A) –samples (red color) distribution from "donor9861" of Allen Human Brain Atlas; (B) – clusters of associated with MEF2C risk loci voxels (blue color) identified through VBM analysis formed using p-value threshold 0.05; (C) – Spatial overlap between Allen Brain probes and VBM clusters; (D) – example of VBM cluster and assigning sample location to them.

Regional analysis

Figure 3 provides a heat map showing all AD risk loci and their effect on different brain regions sorted by lobe. None of the association signals passed Bonferroni correction, however several loci showed nominal significant association ($p < 0.05$; cells with stars on **Figure 3**), among them variant in EPHA1 with less tissue in caudate and in insula, CELF1 with more tissue in accumbens and *APOE* with very strong positive effect in the occipital lobe. Variants in *APOE*, *FERMT2*, *PTK2B, CASS4* and *MS4A6A* showed the strongest effect on hippocampus and were associated with smaller gray matter volume. Risk variants in *EPHA1* and *SORL1* had the largest negative effect on deep gray matter structures: putamen, thalamus, and pallidum.

Discussion

This study presents the association of 19 genome-wide significant AD risk loci[3] with VBM of the gray matter, among 4,071 middle aged and elderly subjects from the population-based Rotterdam Study. The unprecedented sample size has enabled this unbiased whole grey matter investigation of established risk variants and their effect on brain morphology. We found nominally significant associations with the left postcentral gyrus, the right superior frontal gyrus and the right postcentral gyrus. Furthermore, through comparing our VBM results to the Allen Brain atlases of human gene expression, we found significant spatial overlap for genes previously assigned to be the causal gene in these loci (*CLU, SLC24A4* and *MEF2C*). Additionally, we identified two genes, not previously suggested to be the causal gene in AD (*GSTK1* and *NGEF*), of which the expression in the brain significantly overlaps with our VBM results.

There currently exists no consensus for voxel-wise genetics studies regarding the significance threshold for avoiding false positive findings while not to being too conservative.[19,20] A number of data processing and statistical analysis methods have been proposed in the literature to address this issue for neuroimaging analysis.[21-23] However, all these methods rely on a set of assumptions about the statistical structure of the data. Therefore, in our study we decided to use unbiased, but more conservative, non-parametric permutation methods to define the statistical threshold of significance.

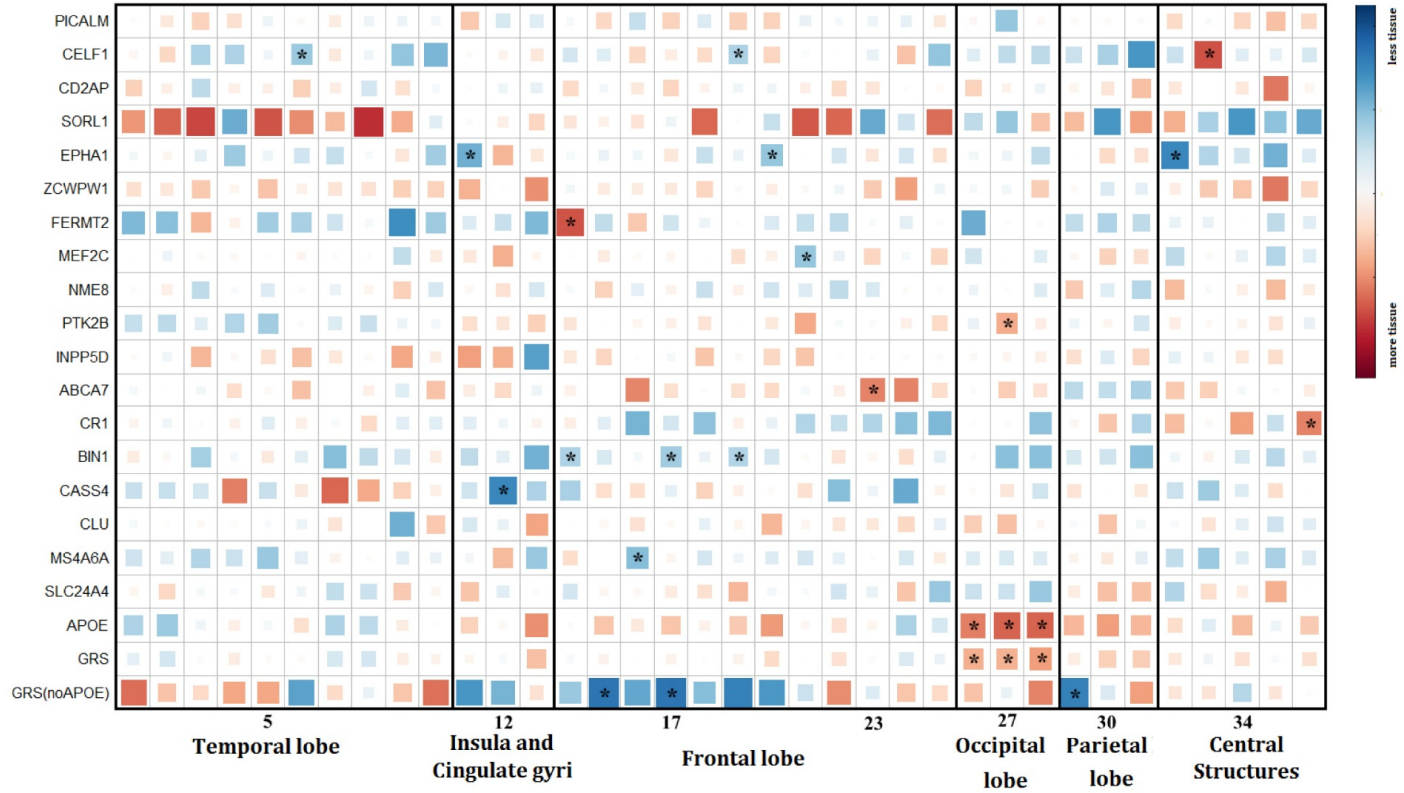


Figure 3. Heatmap of AD risk locil association effects from ROI analysis. The 19 AD risk loci are on the y-axis and brain regions, grouped by lobe, on the x-axis. Brain region labeling was based on the Hammer Atlas. Blue indicates that risk loci were associated with less gray matter tissue; red indicates association with more gray matter tissue. Supplementary Table 3 provides a coded structure list. Regions with nominal significant association (p-value < 0.05) marked with *. Assigned risk gene according to Lambert *et al*[3]

Although this is the largest genetic VBM study conducted to date, none of the voxels passed this conservative multiple testing correction. However, we have previously shown that AD risk loci are associated with cognitive functioning in the general population[11,16,24,25] as well as hippocampal volume in a larger sample (N = 9,232).[4] This showed that subclinical effects of AD risk loci exist and that effects on gray matter could be expected. Additionally, we constructed genetic risk scores, to explore the combined effect of all AD SNPs on brain morphology. The association signal of GRSs also did not pass correction threshold and the strongest signal for GRS with *APOE* was driven by *APOE* variant, while for GRS without *APOE* by *MEF2C* variant (**Supplementary Figure 2**).

Furthermore, it is reasonable to assume that the effects of the risk loci are not restricted to a single voxel, but rather to a cluster of voxels spanning a certain brain region. Therefore, we further explored the nominally significant associations we found by using the Allen Brain Human Atlas to analyze gene expression, and using Hammer brain atlas to estimate average effect on specific brain regions.

In Hammer regional analysis, we found that risk loci for Alzheimer's disease affect brain morphology in established regions such as the hippocampus (e.g. loci near *APOE, FERMT2, PTK2B*), putamen, thalamus (*SORL1, EPHA1*), as well as regions not often reported on including the insula (*EPHA1*) and occipital lobe (*APOE*). The heat map in **Figure 2** summarizes the association results over the whole brain.

Alzheimer's disease is a complex disorder with multiple variants from different pathways involved in its etiology.[26,27] Therefore, as previously shown,[5] the effect of these variants on brain morphology could also differ and have different directions. **Figure 2** provides a detailed map of such heterogeneous effects. For example, large brain structures, such as the temporal lobe and central regions, are affected differently. Also, some risk loci have a different direction of effects, e.g. *FERMT2* is associated with less tissue and *SORL1* with more tissue in the temporal lobe. Of particular interest is that we found the positive association of APOE with the occipital lobe, which could possibly be

explained by cerebral amyloid angiopathy (CAA). Indeed, CAA is linked to *APOE*ε4 carriership[28,29] and has a predilection for the occipital neocortex.[30] Moreover, CAA is involved in Alzheimer's disease[31] and is characterized by ß-amyloid deposition in the media and adventitia of small and medium sized arteries. In healthy subjects, this may be observed as an increase in gray matter tissue density because of the influx of cells to clear the deposits. More research on the effects of AD risk loci on brain morphology is needed to further unravel the biological substrates involved in disease etiology.

Previous case-control studies showed ambiguous differential expression of putative causal genes for AD in the brain[32] or reported that the regional expression of each of the risk loci did not match the pattern of brain regional distribution in Alzheimer pathology.[33] Most of AD variants are non-coding and for the follow up studies would be very important to explore the potential roles of these intronic and intergenic regions in the regulation of gene expression. Confirmed functional variants underlying validated GWAS hits are still sparse in the literature,[34,35] when considering all the diseases and traits studied, but each of these is extremely valuable to the respective research and clinical environments. In our study, we found significant spatial overlap between VBM results in the Allen Human Brain atlas with some of the previously identified genes (C*LU, SLC24A4* and *MEF2C*). This could mean that genetic variability in these genes could act on gray matter density through differences in expression. This is also in line with the fact that most trait-specific GWAS signals are non-coding and probably act through modulation of gene expression.[36] Our results also suggest that VBM analysis combined with expression data could provide evidence for new candidate genes in genetic loci, where the causal gene has not been strongly established by biological experiments.[37] In AD loci, examples are *NGEF* for rs35349669 and *GSTK1* for rs11771145. Although the index variant rs35349669 is located within *INPP5D*, this gene is expressed at low levels in the brain[3] and the linkage peak spans multiple genes with suggestive signals, including *NGEF*.[3] Neuronal Guanine Nucleotide Exchange Factor *(NGEF)*, among its related pathway is signaling by G protein–coupled receptors (GPCRs), which are involved at many stages of AD disease progression, and this class of receptors is a potential therapeutic target for AD.[38] Glutathione S-transferase Kappa 1 (*GSTK1*) is

member of the superfamily of enzymes that function in cellular detoxification. Interestingly, a significant decrease of glutathione transferase activity in different brain regions in patients with Alzheimer disease was previously reported,[39] suggesting a possible link to Alzheimer through diabetes.[40,41]

Our study also has several limitations. The 19 AD risk loci do not include all genetics variants associated with AD and the index variants used may not be the causal variants. Another consideration is that the cross-sectional nature of our analyses precludes us from inferring causality from the associations. Although reverse causality is unlikely for genetic variants, it remains unclear whether our findings represent developmental or degenerative effects. The absence of significant association, as we mentioned before, could be due to strict permutation threshold or lack of power of our study sample size compare to GWAS analysis where these risk loci were discovered.

Additionally, in the experiment to determine spatial overlap between gene expression and regions identified through the VBM analysis, a number of considerations need to be taken into account. First, the threshold to form the clusters is a manual parameter and could be set to a different statistical threshold. However, with decreasing p-value threshold the number and size of the clusters goes down and, due to quite sparse distribution of Allen Brain samples, there are not enough clusters linked to samples to perform such analysis. Second, gene expression depends on the time of measurement and could be different over the lifespan and even during the day.[42] Second, the association between a risk locus and tissue density does not necessarily require the causative gene to be expressed in the same brain region, but could also be through a downstream effect of a functional pathway. Third, given the difficulties in obtaining brain tissue samples, these analyses are all based on relatively small samples.

Conclusion

Using a voxel-based morphometry study in over 4,000 non-demented individuals, we provide a list of candidate brain regions that are potentially affected by AD risk loci and worthy of further study. Although detecting significant genetic effects on individual

voxels will require even larger sample sizes, we show that data can be exploited by incorporating additional information in the analysis, such as gene expression data. All results of our study are available online on www.imagene.nl/ADSNPs/.

References:

1. Thompson PM, Cannon TD, Narr KL, et al. Genetic influences on brain structure. *Nature neuroscience* 2001; **4**: 1253-8.
2. Weiner MW, Veitch DP, Aisen PS, et al. The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. *Alzheimer's & Dementia* 2012; **8**: S1-68.
3. Lambert J, Ibrahim-Verbaas C, Harold D. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature genetics* 2013; **45**: 1452-8.
4. Bis JC, DeCarli C, Smith AV, et al. Common variants at 12q14 and 12q24 are associated with hippocampal volume. *Nature Genetics* 2012; **44**: 545-51.
5. Chauhan G, Adams HHH, Bis JC, et al. Association of Alzheimer's disease GWAS loci with MRI markers of brain aging. *Neurobiology of Aging* 2015; **36**: 1765.e7-.e16.
6. Liu Y, Yu J-T, Wang H-F, et al. Association between NME8 Locus Polymorphism and Cognitive Decline, Cerebrospinal Fluid and Neuroimaging Biomarkers in Alzheimer's Disease. *PLoS ONE* 2014; **9**: e114777.
7. Morgen K, Ramirez A, Frölich L, et al. Genetic interaction of PICALM and APOE is associated with brain atrophy and cognitive impairment in Alzheimer's disease. *Alzheimer's & dementia : the journal of the Alzheimer's Association* 2014; **10**: 1-8.
8. Wright ICC, McGuire PK, Poline JB, et al. A voxel-based method for the statistical analysis of gray and white matter density applied to schizophrenia. *Neuroimage* 1995; **2**: 244-52.
9. Hofman A, van Duijn CM, Franco OH, et al. The Rotterdam Study: 2012 objectives and design update. *European journal of epidemiology* 2011; **26**: 657-86.
10. Ikram M, Lugt Avd. The Rotterdam Scan Study: design and update up to 2012. *European journal of epidemiology* 2011; **26**: 811-24.
11. Verhaaren BFJ, Vernooij MW, Koudstaal PJ, et al. Alzheimer's disease genes and cognition in the nondemented general population. *Biological Psychiatry* 2013; **73**: 429-34.
12. Good CD, Johnsrude IS, Ashburner J, Henson RN, Friston KJ, Frackowiak RS. A voxel-based morphometric study of ageing in 465 normal adult human brains. *NeuroImage* 2001; **14**: 21-36.
13. Vrooman HA, Cocosco CA, van der Lijn F, et al. Multi-spectral brain tissue segmentation using automatically trained k-Nearest-Neighbor classification. *NeuroImage* 2007; **37**: 71-81.
14. Smith SM, Jenkinson M, Woolrich MW, et al. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* 2004; **23 Suppl 1**: S208-19.
15. Churchill GA, Doerge RW. Empirical threshold values for quantitative trait mapping. *Genetics* 1994; **138**: 963-71.
16. Adams HHH, de Bruijn RFaG, Hofman A, et al. Genetic risk of neurodegenerative diseases is associated with mild cognitive impairment and conversion to dementia. *Alzheimer's & Dementia* 2015: 1-9.
17. ALLEN Human Brain Atlas Normalization, Microarray Data. 2013: 1-11.
18. Hammers A, Allom R, Koepp MJ, et al. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Human brain mapping* 2003; **19**: 224-47.
19. Fritsch V, Da Mota B, Loth E, et al. Robust regression for large-scale neuroimaging studies. *Neuroimage* 2015; **111**: 431-41.

20. Medland SE, Jahanshad N, Neale BM, Thompson PM. Whole-genome analyses of whole-brain data: working within an expanded search space. *Nat Neurosci* 2014; **17**(6): 791-800.
21. Bullmore ET, Suckling J, Overmeyer S, Rabe-Hesketh S, Taylor E, Brammer MJ. Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *IEEE Trans Med Imaging* 1999; **18**(1): 32-42.
22. Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 2002; **17**(2): 825-41.
23. Smith SM, Nichols TE. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage* 2009; **44**(1): 83-98.
24. Davies G, Armstrong N, Bis JC, et al. 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* 2015: 183-92.
25. de Bruijn RF, Bos MJ, Portegies ML, et al. The potential for prevention of dementia across two decades: the prospective, population-based Rotterdam Study. *BMC Med* 2015; **13**: 132.
26. Jones L. Convergent genetic and expression data implicate immunity in Alzheimer's disease. *Alzheimer's & Dementia* 2015; **11**: 658-71.
27. Mattson MP. Pathways towards and away from Alzheimer's disease. *Nature* 2004; **430**: 631-9.
28. Esiri M, Chance S, Joachim C, et al. Cerebral amyloid angiopathy, subcortical white matter disease and dementia: literature review and study in OPTIMA. *Brain pathology (Zurich, Switzerland)* 2015; **25**: 51-62.
29. Ringman JM, Sachs MC, Zhou Y, Monsell SE, Saver JL, Vinters HV. Clinical Predictors of Severe Cerebral Amyloid Angiopathy and Influence of APOE Genotype in Persons With Pathologically Verified Alzheimer Disease. *JAMA neurology* 2014; **71**: 878-83.
30. Nelson PT. APOE-ε2 and APOE-ε4 Correlate with Increased Amyloid Accumulation in Cerebral Vasculature. *Journal of Neuropathology & Experimental Neurology* 2013; **72**: 708-15.
31. Smith EE, Greenberg SM. Amyloid, blood vessels, and brain function. *Stroke* 2009; **40**: 2601-6.
32. Holton P, Ryten M, Nalls M, et al. Initial Assessment of the Pathogenic Mechanisms of the Recently Identified Alzheimer Risk Loci. *Annals of Human Genetics* 2013; **77**: 85-105.
33. Karch CM, Jeng AT, Nowotny P, Cady J, Cruchaga C, Goate AM. Expression of Novel Alzheimer's Disease Risk Genes in Control and Alzheimer's Disease Brains. *PLoS ONE* 2012; **7**.
34. Manolio TA. Genomewide association studies and assessment of the risk of disease. *N Engl J Med* 2010; **363**(2): 166-76.
35. Myers AJ, Gibbs JR, Webster JA, et al. A survey of genetic human cortical gene expression. *Nat Genet* 2007; **39**(12): 1494-9.
36. Ramasamy A, Trabzuni D, Guelfi S, et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nature Neuroscience* 2014; **17**: 1418-28.
37. Steinberg S, Stefansson H, Jonsson T, et al. Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. *Nature Genetics* 2015; **47**: 445-7.
38. Thathiah A, De Strooper B. The role of G protein-coupled receptors in the pathology of Alzheimer's disease. *Nature reviews Neuroscience* 2011; **12**: 73-87.
39. Lovell MA, Xie C, Markesbery WR. Decreased glutathione transferase activity in brain and ventricular fluid in Alzheimer's disease. *Neurology* 1998; **51**(6): 1562-6.
40. Shield AJ, Murray TP, Cappello JY, Coggan M, Board PG. Polymorphisms in the human glutathione transferase Kappa (GSTK1) promoter alter gene expression. *Genomics* 2010; **95**(5): 299-305.
41. Weyer C, Funahashi T, Tanaka S, et al. Hypoadiponectinemia in obesity and type 2 diabetes: close association with insulin resistance and hyperinsulinemia. *J Clin Endocrinol Metab* 2001; **86**(5): 1930-5.
42. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature genetics* 2003; **33 Suppl**: 245-54.

Supplementary material:

Supplementary Methods and Supplementary Tables can be accessed by scanning the following code or accessing the journals' website

Chapter 4.3

Novel genetic loci associated with brain lobar volumes

Sven J. van der Lee*, Hieab H.H. Adams*, Najaf Amin, Lisa R. Yanek, Rasika A. Mathias Albert Vernon Smith, Tamara B. Harris, Konstantinos Arfanakis, Lei Yu, Saima Hilal, Ching Yu Cheng, Tien Yin Wong, Wan Ting Zhao, Edith Hofer, Yasaman Saba, Derrek P. Hibar, Josh W. Cheung, Neda Jahanshad, Tomas Paus, Manon Bernard, Joshua C. Bis, Oscar L. Lopez, Kent D. Taylor, Ken Rice, WT Longstreth, Claudia Satizabal, Ganesh Chauhan, Joris Deelen, Erik van den Akker, Mariam Beekman, Paul A. Nyquist, Diane M. Becker, Lenore J. Launer, Vilmundur Gudnason, David A. Bennett, Philip L. De Jager, Christopher Chen, M. Kamran Ikram, Reinhold Schmidt, Helena Schmidt, Paul M. Thompson, Zdenka Pausova, Bruce M. Psaty, Sudha Seshadri, Stephanie Debette, Eline Slagboom Cornelia M. van Duijn, Meike W. Vernooij, M. Arfan Ikram, Charles S. Decarli

* Authors contributed equally

This chapter is in preparation

Abstract

Volumes of the four lobes of the human brain are heritable but specific genetic association studies are limited. We performed a genome-wide association study of the volume of the frontal, occipital, parietal and temporal lobes in 16,016 individuals from 20 cohorts in whom the lobar volumes were characterized using magnetic resonance imaging (MRI). Genome-wide we showed that up to 30-40% of the genetic background that determines lobar volume can be explained by common genetic variants. There was significant genetic correlation between the genetic determinants of parietal lobe and the temporal lobe (R_g=0.41, $p = 5.2\times10^{-5}$). In the single variant analysis we identified six genetic loci associated with specific lobar volumes. The locus 6q22.32 (rs1337736) associated with occipital lobe volume and was previously associated with intracranial volume (ICV). The locus 12q14.3 (rs61921502) associated with temporal lobe volume and was previously associated with hippocampal volume. Four loci were previously unknown to affect brain volume measurements, including locus 3q24 (rs2279829) associated with parietal lobe volume, and three loci associated with occipital lobe volume (1q22, rs124111216; 4p16.3, rs74921869; 14q23.1, rs147148763). These new associated variants were common (Minor allele frequency 0.13 to 0.46) and located in regions rich in epigenetic effects (*DAAM1* and *THBS3*) or located close to genes that cause brain related diseases with Mendelian inheritance patterns (*ZIC4* and *FGFRL1*). After correction for multiple testing we could not show a genetic overlap with neurological or psychiatric diseases. Our findings reveal part of the complex genetics underlying these brain volumes and suggest a role for regulatory regions in determining brain volumes.

Introduction

The four lobes of the brain are anatomically destict and several diseases are attributed to lobe specific structural changes. The brain functions of the frontal lobe include reasoning, movement, social behavior, planning, parts of speech and problem solving,[1] functions attributed to the parietal lobe include recognition and perception of stimuli,[2] functions attributed to the temporal lobe include memory and speech,[3] and lastly visual input is processed by the occipital lobe. Brain diseases with lobe specific abnormalities include frontotemporal lobar degeneration,[4] temporal lobe epilepsy,[5] PPA and cortical basilar ganglionic degeneration.

Environmental factors, such as smoking and hypertension, affect the volume of brain lobes, but previous studies show that also genetic differences across individuals contribute to the variability in brain volume measures.[6] To date no common variants influencing brain lobar volumes have been identified. Such studies may contribute to our understanding of brain lobe development and provide a biological link brain lobar volumes with brain related traits and diseases. The estimated heritability of brain lobar volumes is high.[7-13] Estimates ranging from 26% to 84% for the frontal lobe, from 32% to 74% for the occipital lobe, from 30% to 86% for the parietal lobe, and from 55% to 88% for the temporal lobe.[13] Bivariate genetic analysis, also shows that these regions appear to be under differential genetic influence.[13] The fact brain lobes are highly and differentially heritable makes them interesting targets for genetic studies. Recent large genome-wide association studies (GWAS) have shown that we can identify associations between genetic determinants and brain volumetric measures.[14,15]

To study the genetic influences of lobar brain volumes, we examined the genetics of brain lobar volumes in 16,016 individuals from 20 cohorts. Our primary aim was to find common genetic determinants of lobar brain volumes. With this study we shed light on the common genetic variants determining brain volume and allow for a deepened understanding of the genetic architecture of the brain lobes.

Methods

Populations

Study sample consisted of dementia and stroke-free individuals with quantitative brain MRI and genome-wide genotypes from 20 population-based cohort studies participating in the Cohorts of Heart and Aging Research in Genomic Epidemiology consortium (CHARGE) and Alzheimer's Disease Neuroimaging Initiative (ADNI). In total 16,016 participants were included, 15,269 participants of European ancestry, 405 African Americans, 211 Chinese and 131 Malay. Descriptive statistics of all populations are provided in **Supplementary Table 1**. DNA from whole blood was extracted and genome-wide genotyping was performed using a range of commercially available genotyping arrays. Genotype imputations were performed in each cohort using 1000 Genomes [16] (mainly Phase I version 3) as reference (**Supplementary Table 2**).

MRI Methods

High resolution brain magnetic resonance imaging (MRI) data was acquired by each participating cohort **(Supplementary data**). In each study, MRI scans were performed and processed with automatized protocols, without reference to clinical or genetic information. We studied the total volume (sum of white and grey matter, and the left and right hemisphere) of the frontal, parietal, temporal and occipital brain lobe. Although the methods of image segmentation varied across study cohorts, usual definitions of lobar boundaries were maintained. All participants (or their parents/guardians in the case of minors) provided written informed consent for study participation, use of brain MRI data and the use of their DNA for genetic research.

Estimation of heritability

Heritability of lobar brain volumes were estimated using family structure in the Framingham Heart Study (FHS) (n = 2,080) which constitutes a community based cohort of nondemented individuals without evidence of significant brain injury (e.g. stroke, multiple sclerosis). We calculated additive genetic heritability using variance-components analysis in SOLAR.[17]

GWAS of lobar volumes

Associations of lobar volumes with imputed genotype dosages were examined using linear regression analysis under an additive model. Associations were adjusted for age, age^2, sex, first four principal components to account for possible confounding due to population stratification, and study specific covariates. Linear mixed models with estimated kinships were used for association analysis in cohorts with related samples. Details on analysis methods used in each cohort are provided in the **Supplementary Tables 1 and 2.** Post-GWAS quality control was conducted using EasyQC [18] and filtering. Genetic variants with low imputation quality (R^2<0.5), minor allele count less than 10 and allelic or locational mismatching of SNPs with the reference panel were removed prior to the meta-analyses. Number of variants after filtering and genomic inflation per study are provided in **Supplementary Table 3**. After QC, summary statistics were adjusted by the genomic control method in each of the participating cohorts.[19] We then performed two inverse-variance weighted fixed-effect meta-analyses in METAL.[19] First we meta-analyzed all participants of European Ancestry (EA), then performed a multi-ethnic meta-analysis, including African Americans (N=405), Chinese (N=211) and Malay (N=131). After meta-analyses genetic variants with a total sample size of less than 5,000 were excluded. We performed conditional analysis on the index variants to determine if there were multiple independent genome-wide significant variants in a locus using the GCTA software (--cojo, --p-cojo), genotypes in the RSI study were used as reference for this analysis (n = 6,291). For each genome-wide significant locus we searched for candidate genes in the loci using publically available databases for differential expression of the SNPs (eQTL database in GTEx)[20] and HaploReg, an online tool that summarizes the ENCODE database for epigenetic markings and proteins binding to DNA.[21]

Variance explained by common genetic variation and genetic correlation

Variance explained by all single nucleotide polymorphisms (SNPs), or SNP-based heritability, was calculated from summary statistics using LDSCORE [22] implemented in GCTA.[23] Percentage variance explained for genome-wide significant single SNPs was determined based on meta-analysis results using the LD-hub.[24] We used the same

LDSCORE regression [22] to quantify the amount of genetic correlation between the four brain lobe meta-analyses summary statistics and to study complex relationships between the brain and diseases. By comparing genome-wide association and results of analysis, we can obtain genetic correlation estimates that are free of reverse causation and environmental confounding. These estimates approximate causal inferences that are difficult to obtain otherwise. LDSCORE regression was used to calculate genetic correlation between the lobar volumes and the meta-analysis of published studies of subcortical structures,[14] intracranial volume (ICV),[15] educational attainment, general cognitive ability,[25] neuroticism , schizophrenia, attention deficit disorder, autism, major depressive disorder, bipolar disorder, Parkinson's disease and Alzheimer's disease.[26]

Results

Heritability of lobar brain volumes

We found an age- and sex-adjusted heritability (h^2) for occipital lobe of 50%, for frontal lobe of 52%, for temporal lobe of 59% and parietal lobe of 59% (all: $p \leq 1.9\text{x}10^{-19}$) **(Supplementary Table 4).** In comparison the age- and sex-adjusted heritability estimate for the sum of the brain lobes was 34% ($p = 8.8\text{x}10^{-11}$).

Novel genome-wide significant variants associated with brain lobar volumes

Our multi-ethnic meta-analysis (n = 16,016 individuals of which 15,269 of European ancestry) revealed five independent, genome-wide significant loci associated with brain lobar volumes (**Figure 1, Table 1, Figure 2).** The quantile-quantile plots did not show genomic inflation ($\lambda_{GC} \leq 1.05$) (**Supplementary Figure 1**). In the European ancestry only meta-analysis one additional independent locus (1q22, rs12411216, $p_{EA\text{-}only} = 3.9\times10^{-8}$) was significantly associated with occipital lobe volume. In the multi-ethnic meta-analysis this locus was not significant ($p_{multi\text{-}ethnic} = 1.3\times10^{-7}$). Of these in total six loci, one locus associated with temporal lobe volume, one with parietal lobe volume and four loci with occipital lobe volume. The index variants were common (range minor allele frequency 0.13 to 0.46) associated with volume variations in the range of 0.48 to 0.95 cm^3 per allele, explaining up to 0.27% per allele of lobar volume variance (**Table 1**).

Table 1. Genetic variants at seven loci significantly associated with lobar brain volume

								European ancestry meta-analysis				Multi-ethnic meta-analysis					Variance explained
Lobe	Annotation	(nearby) Gene	Rs-ID	Chr	Postion	A1	A2	Frq	Effect	SE	P	Frq	Effect	SE	P	N	
PLV	UTR3	ZIC4	rs2279829	3	147106319	T	C	0.21	-0.921	0.158	$5.3x10^{-9}$	0.21	-0.954	0.153	$4.4x10^{-10}$	16015	0.24
TLV	intronic	MSRB3	rs146354218	12	65793942	A	G	0.37	0.687	0.113	$1.2x10^{-9}$	0.37	0.685	0.111	$6.4x10^{-10}$	15789	0.24
OLV	intergenic	DAAM1(dist=24302)	rs147148763	14	59631075	G	GTTGT	0.13	-0.831	0.143	$6.6x10^{-9}$	0.13	-0.847	0.143	$2.9x10^{-9}$	15220	0.23
OLV	intronic	FGFRL1	rs74921869	4	1013382	A	G	0.20	-0.840	0.144	$5.9x10^{-9}$	0.20	-0.818	0.141	$6.2x10^{-9}$	12424	0.27
OLV	intergenic	CENPW(dist=175626)	rs1337736	6	126845380	A	G	0.23	-0.639	0.119	$8.1x10^{-8}$	0.23	-0.637	0.116	$4.0x10^{-8}$	16016	0.19
OLV	upstream	MIR92B/THBS3	rs12411216	1	155164480	A	C	0.46	-0.5211	0.095	$3.9x10^{-8}$	0.46	-0.487	0.092	$1.4x10^{-7}$	16016	0.17

The allele frequency (frq) and effect size are given for A1. Effect sizes are given in units of cm^3 per effect allele. Results are provided for the discovery samples and the meta-analysis of all European ancestry and the multi-ethnic meta-analysis. The variance explained gives the percentage variance explained of a SNP.[14] Chr = Chromosome, A1=effect allele, A2 alternate allele, SE = standard error, P = p-value, N= number of participants.

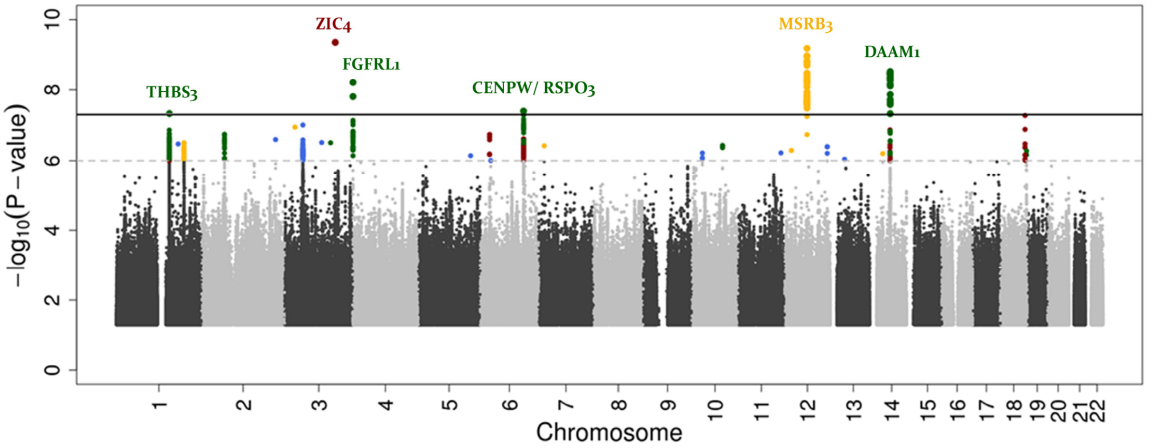


Figure 1: Common genetic variants associated with frontal-, parietal-, temporal- and occipital lobe volume. 'Manhattan' plot displaying the association p-value for each tested SNP (displayed as $-\log_{10}$ of the p-value). Genome-wide significance threshold is shown with a line at $p = 5\times10^{-8}$ (solid black line) and also the suggestive threshold at $p = 1\times10^{-5}$ (dashed line). Dots represent variants, results of the 4 lobes are shown in a single figure, the nearby gene is labeled. Above the suggestive threshold variants are colored by the associated lobe, yellow = temporal lobe, green = occipital lobe, red = parietal lobe, blue = frontal lobe.

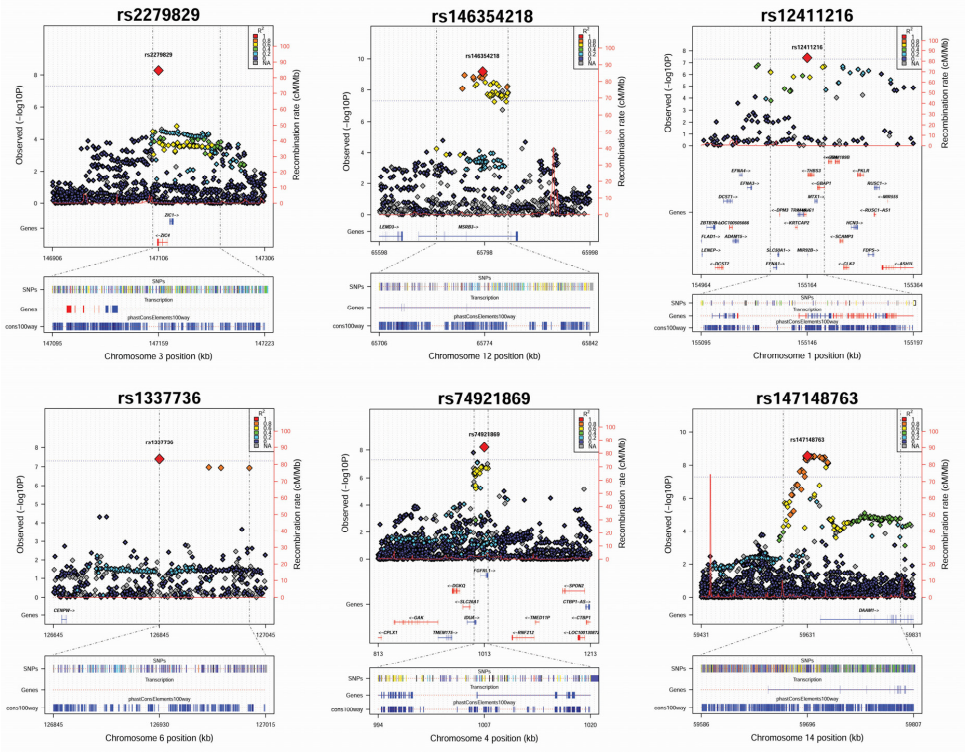


Figure 2: Regional view of the genome-wide significant loci. For each panel, zoomed in Manhattan plots (± kb from top SNP) are shown with gene models below (GENCODE version 19). Plots are zoomed in to highlight the genomic region that contains the index SNP and SNPs in LD with the index snp ($R^2 > 0.8$). Each plot was made using the LocusTrack software (http://gump.qimr.edu.au/general/gabrieC/LocusTrack/).

The variant rs146354218 (12q14.3, $p_{\text{multi-ethnic}} = 6.4\text{x}10^{-10}$) associated with temporal lobe volume and rs2279829 (3q24, $p_{\text{multi-ethnic}} = 4.4\text{x}10^{-10}$) associated with parietal lobe volume. Three loci associated with occipital lobe volume, index variants rs147148763 (small indel GTTGT→G, 14q23.1, $p_{\text{multi-ethnic}} = 2.9\times10^{-9}$), rs74921869 (4p16.3, $p_{\text{multi-ethnic}} = 6.2\times10^{-9}$) and rs1337736 (6q22.32, $p_{\text{multi-ethnic}} = 4.0\times10^{-8}$). No variants associated at genome wide significant level with frontal lobe volume. All variants showing genome-wide significant association with brain lobar volume are shown in **Supplementary Table 5** and annotated in **Supplementary Table 6.** Notably, the genome-wide significant variants identified here showed mostly effects on a single lobe rather than pleiotropic effects across multiple lobes (**Supplementary Table 7**). Only variant rs147148763 (14q23.1 locus) associated with decreased occipital lobe volume as well as a

suggestively associated with increased parietal lobe volume (effect = 0.91 cm^{-3}, p = 2.5×10^{-6}, variance explained = 0.15%).

Variance explained in lobar volumes by common variants

Common variants across the whole genome explained as much as 20.3% (SE 3.61%) of the variance in occipital lobe volume, 19.6% (SE 3.72%) of frontal lobe volume, 17.5% (SE 3.45%) of temporal lobe volume and 17.9% (SE 3.16%) of parietal lobe volume (**Supplementary Table 8**). Common genetic variants account for 30% to 41% of the total heritability lobar volumes (**Supplementary Table 8**).

Genetic overlap with other brain volumes, and brain related diseases

Although no overlap was observed among the top results, significant genetic correlations (R_g) were observed between the parietal and temporal lobe (R_g = 0.41, p = 5.2×10^{-5}) and nominally significant correlation between the parietal and occipital lobe (R_g = 0.28, p = 2.1×10^{-2}) (**Supplementary Table 9**). The frontal lobe showed no significant genetic correlation with the other lobes. Studying genetic correlation of brain lobar volumes with intracranial volume and other brain volume measurements showed suggestive association of the temporal lobar volume with intracranial volume (R_g=0.19, p = 0.04). This is even though our analyses were adjusted for intracranial volume. Temporal lobe volume also showed correlation with the thalamus volume (R_g = 0.33, $p = 3.7\times10^{-3}$), amygdala volume (R_g = 0.56, $p = 2.6\times10^{-3}$) and putamen volume (R_g = 0.20, p = 0.02). Parietal lobe volume showed genetic correlation with nucleus accumbens volume (R_g=0.24, p = 0.045) and occipital lobe volume with volume of the amygdala (R_g=0.33, p = 0.041). When studying brain diseases only frontal lobe volume showed suggestive genetic correlation with attention deficit disorder (ADD) (R_g=0.66, $p = 6.4\times10^{-3}$), after correction for multiple testing this association was not significant. There was no correlation with Parkinson's disease, Alzheimer's disease, depression, autism and other tested traits.

Discussion

The results of our genome-wide association study in up to 16,016 individuals of brain lobar volume we show that 30 to 40% of the genetic background that determines lobar brain volume can be explained by genome-wide common genetic variation. We found evidence for a joint genetic origin based on genetic correlation only for the parietal and temporal lobe. We identified six genome-wide significant loci associated with lobar brain volume. Four out or six loci were novel genetic loci associated with brain volume measures. Two loci were located in regions previously identified associated with brain volume measures. These new loci provide intriguing new insights in the genetics underlying brain lobar volumes.

We estimated that 17.5% to 20.3% of the variance in lobar volumes could be explained by common genetic variation. This forms 30% to 40% of the total heritability that we estimated suggesting a major contribution of common genetic variation in brain development. This is even though our analyses were adjusted for intracranial volume a highly heritable trait. More genetic variants associated with brain volume are to be discovered by increasing the sample sizes of genetic studies. Moreover, the majority of our sample was of European ancestry and therefore this does not necessarily reflect the variance explained in populations of other ancestry. Using the LDSCORE regression method we showed significant genetic correlation on a genome-wide scale between the parietal lobe and the adjacent temporal lobe as well as the occipital lobe. The absence of a strong genetic correlation between most of the lobes suggest that the genetic basis of the lobes is largely independent, underlining that the lobes and likely other brain structures should be studied independently.[13]

Concerning the identified genome-wide significant loci, two identified loci have been previously associated with brain volume measurements. The locus 12q14.3 associated with temporal lobe volume in our study and was previously associated with hippocampal volume.[15] Our index variant rs146354218 (p = 6.4x10^{-10}) is an intronic variant in the *MSRB3* gene and lies 39 kilobases (kb) from the previously published rs61921502 variant associated with hippocampal volume,[27] however the linkage

disequilibrium (LD) (R^2 =0.1, D'=1 , $p < 0.0001$)[28] is low. This previously published variant also showed some evidence of association with total temporal lobe volume (effect= 0.57 cm^3, $p = 6.5\times10^{-4}$). Thus the 12q14.3 locus not only influences hippocampal volume, but seems to have a more generalized effect on temporal lobe volume as would be expected by the significant genetic correlation between temporal lobe volume and hippocampal volume. The signal (rs1337736, $p = 4.0\times10^{-8}$) at 6q22.32 near to the gene *CENPW* (Centromere Protein W) is associated with occipital volume. This signal overlapped previously association signals with intracranial volume[14,15] and is further implicated in bone mineral density,[29] height,[30] waist-hip ratio[31] and infant length.[32] The index variant associated with ICV (rs11759026) and our top variant were in linkage equilibrium (R^2 = 0.07, D'=1, p = 0.0002)[28]. We also found suggestive associations between rs11759026 frontal (-1.0 cm^3, $p = 6.3\times10^{-5}$) and occipital lobar volume (-0.31cm^3, $p = 6.6\times10^{-3}$). Each novel locus was located in regions that are under epigenetic regulation in brain tissue (**Supplementary Table 6**), or close to genes or genomic loci associated with Mendelian brain related diseases. The variant rs2279829 (3q24) is located in the 3'-UTR of the *Zic Family Member 4 (ZIC4*) gene and close to the related *ZIC1* gene. This variant localizes within enhancer sites in predominantly neurological cell types, among which the brain germinal matrix (**Supplementary Table 6**) and both *ZIC4* and *ZIC1* are expressed in throughout the brain (**Supplementary Figure** 1). Heterozygous deletions of *ZIC1* and *ZIC4* cause Dandy-Walker malformation (DWM).[33] Children with this malformation have no vermis, the part connecting the two cerebellar hemispheres.[33] Gain of function mutations in *ZIC1* lead to coronal craniosynostosis and learning Disability.[34] Variant rs147148763 was located 24 kb from the dishevelled-associated activator of morphogenesis 1 (*DAAM1*). There is evidence for the most significant variants to localize within enhancer sites, as well as DNA-hypersensitivity sites in brain tissues and genome-wide significant SNPs in the locus are eQTLs of *DAAM1* in blood (**Supplementary Table 6**). Daam1 is a formin protein that has been linked to actin dynamics,[35] is regulated by RhoA [36] and is expressed in the shafts of dendrites.[37] Expression patterns in brain development of animals further suggest a role in neuronal cell differentiation and movement.[38] Variant rs4647940 is located in the 3'-UTR of fibroblast growth factor receptor (*FGFRL1*) and is in LD with a missense variant

in Alpha-L-Iduronidase (*IDUA*) (rs3755955, $R^2 = 0.87$) that was previously associated with bone mineral density ($p = 5.0 \times 10^{-15}$). Deletion of the 4p16.3 locus cause Wolf-Hirschhorn syndrome (WHS), a neurodevelopmental disorder characterized by mental retardation, craniofacial malformation, and defects in skeletal and heart development. Variant rs12411216 is located in an intron of *MIR92B* and *THBS3*, but the signal peak in this locus covers over 20 genes. Promotor histone marks overlap the variant and it is an eQTL for multiple genes, both in a multitude of different tissues among which brain tissues (**Supplementary Table 6**). In summary, these findings link genes that cause Mendelian syndromes that affect cranial skeletal malformations, brain malformations and intelligence with brain lobe volume in healthy individuals. One other interesting variant in tight LD with our index variant (rs4072037, $R^2 = 0.94$) is a missense SNP in the *MUC1* gene that decreased levels of blood magnesium concentrations.[39] It is not clear how decreased magnesium levels are involved in decreased occipital brain volume, but it is an interesting avenue to explore as magnesium is known to be important for neural transmissions [40] and magnesium infusions have anti-convulsive effects and is still used to prevent convulsions in pre-eclampsia.[41]

In our study we could not link the variant identified to disease outcomes. Using genetic correlation analysis the most significant genetic correlation we observed between brain lobar volumes and neurological and psychiatric traits was between frontal lobe volume and attention deficit disorders (ADD) ($R_g = 0.66$, $p = 6.4 \times 10^{-3}$). However, this finding is not significant after multiple testing correction leading us to report this finding with caution. Unfortunately we were not able to show other genetic correlations between other brain lobes and clinical diseases, this could be due to true absence of a genetic overlap but it could also mean our lobar volume GWASs and those for other diseases were still too underpowered to show significant genetic correlations.

In summary, brain lobar volumes are differentially heritable traits, which can be for a large part explained by common genetic variation. We identified 6 loci associated with specific brain lobes, four of which have not been implicated in brain morphology before. With this large effort we have just scratched the surface of the complex genetic factors

that determine brain morphology but pave the way for targeted functional research to identify the biology behind these genetic signals.

References:

1. Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* 2001; **24**: 167-202.
2. Culham JC, Valyear KF. Human parietal cortex in action. *Curr Opin Neurobiol* 2006; **16**(2): 205-12.
3. Levy DA, Bayley PJ, Squire LR. The anatomy of semantic knowledge: Medial vs. lateral temporal lobe. *P Natl Acad Sci USA* 2004; **101**(17): 6710-5.
4. Mendez MF, Selwood A, Mastri AR, Frey WH, 2nd. Pick's disease versus Alzheimer's disease: a comparison of clinical characteristics. *Neurology* 1993; **43**(2): 289-92.
5. de Lanerolle NC, Kim JH, Williamson A, et al. A retrospective analysis of hippocampal pathology in human temporal lobe epilepsy: evidence for distinctive patient subcategories. *Epilepsia* 2003; **44**(5): 677-87.
6. Hulshoff Pol HE, Schnack HG, Posthuma D, et al. Genetic contributions to human brain morphology and intelligence. *J Neurosci* 2006; **26**(40): 10235-42.
7. Batouli SA, Sachdev PS, Wen W, Wright MJ, Ames D, Trollor JN. Heritability of brain volumes in older adults: the Older Australian Twins Study. *Neurobiol Aging* 2014; **35**(4): 937 e5-18.
8. Baare WF, Hulshoff Pol HE, Boomsma DI, et al. Quantitative genetic modeling of variation in human brain morphology. *Cereb Cortex* 2001; **11**(9): 816-24.
9. Geschwind DH, Miller BL, DeCarli C, Carmelli D. Heritability of lobar brain volumes in twins supports genetic models of cerebral laterality and handedness. *Proc Natl Acad Sci U S A* 2002; **99**(5): 3176-81.
10. Gilmore JH, Schmitt JE, Knickmeyer RC, et al. Genetic and environmental contributions to neonatal brain structure: A twin study. *Hum Brain Mapp* 2010; **31**(8): 1174-82.
11. Yoon U, Fahim C, Perusse D, Evans AC. Lateralized genetic and environmental influences on human brain morphology of 8-year-old twins. *Neuroimage* 2010; **53**(3): 1117-25.
12. Wallace GL, Eric Schmitt J, Lenroot R, et al. A pediatric twin study of brain morphometry. *J Child Psychol Psychiatry* 2006; **47**(10): 987-93.
13. DeStefano AL, Seshadri S, Beiser A, et al. Bivariate heritability of total and regional brain volumes: the Framingham Study. *Alzheimer Dis Assoc Disord* 2009; **23**(3): 218-23.
14. Hibar DP, Stein JL, Renteria ME, et al. Common genetic variants influence human subcortical brain structures. *Nature* 2015; **520**(7546): 224-9.
15. Adams HH, Hibar DP, Chouraki V, et al. Novel genetic loci underlying human intracranial volume identified through genome-wide association. *Nat Neurosci* 2016.
16. Genomes Project C, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature* 2015; **526**(7571): 68-74.
17. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 1998; **62**(5): 1198-211.
18. Winkler TW, Day FR, Croteau-Chonka DC, et al. Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc* 2014; **9**(5): 1192-212.
19. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; **26**(17): 2190-1.

20. Consortium GT, Ardlie KG, Deluca DS, et al. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science (New York, N Y)* 2015; **348**(6235): 648-60.
21. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucl Acids Res* 2012; **40**(Database issue): D930-4.
22. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet* 2015; **47**(11): 1236-41.
23. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011; **88**(1): 76-82.
24. Zheng J, Erzurumluoglu AM, Elsworth BL, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* 2016.
25. Davies G, Armstrong N, Bis JC, et al. Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53 949). *Mol Psychiatr* 2015; **20**(2): 183-92.
26. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 2013; **45**(12): 1452-8.
27. Hibar DP, Adams HH, Jahanshad N, et al. Novel genetic loci associated with hippocampal volume. *Nat Commun* 2017; **8**: 13624.
28. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* 2015; **31**(21): 3555-7.
29. Estrada K, Styrkarsdottir U, Evangelou E, et al. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nature Genetics* 2012; **44**(5): 491-+.
30. Lango Allen H, Estrada K, Lettre G, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 2010; **467**(7317): 832-8.
31. Heid IM, Jackson AU, Randall JC, et al. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nature Genetics* 2011; **43**(11): 1164-.
32. van der Valk RJP, Kreiner-Moller E, Kooijman MN, et al. A novel common variant in DCST2 is associated with length in early life and height in adulthood. *Hum Mol Genet* 2015; **24**(4): 1155-68.
33. Grinberg I, Northrup H, Ardinger H, Prasad C, Dobyns WB, Millen KJ. Heterozygous deletion of the linked genes ZIC1 and ZIC4 is involved in Dandy-Walker malformation. *Nature Genetics* 2004; **36**(10): 1053-5.
34. Twigg SR, Forecki J, Goos JA, et al. Gain-of-Function Mutations in ZIC1 Are Associated with Coronal Craniosynostosis and Learning Disability. *Am J Hum Genet* 2015; **97**(3): 378-88.
35. Habas R, Kato Y, He X. Wnt/Frizzled activation of Rho regulates vertebrate gastrulation and requires a novel formin homology protein Daam1. *Cell* 2001; **107**(7): 843-54.
36. Matusek T, Djiane A, Jankovics F, Brunner D, Mlodzik M, Mihlaly J. The Drosophila formin DAAM regulates the tracheal cuticle pattern through organizing the actin cytoskeleton. *Development* 2006; **133**(5): 957-66.
37. Salomon SN, Haber M, Murai KK, Dunn RJ. Localization of the Diaphanous-related formin Daam1 to neuronal dendrites. *Neuroscience Letters* 2008; **447**(1): 62-7.
38. Kida Y, Shiraishi T, Ogura T. Identification of chick and mouse Daam1 and Daam2 genes and their expression patterns in the central nervous system. *Dev Brain Res* 2004; **153**(1): 143-50.

39. Meyer TE, Verwoert GC, Hwang SJ, et al. Genome-wide association studies of serum magnesium, potassium, and sodium concentrations identify six Loci influencing serum magnesium levels. *PLoS Genet* 2010; **6**(8).
40. Borges LF, Gucer G. Effect of magnesium on epileptic foci. *Epilepsia* 1978; **19**(1): 81-91.
41. Pritchard JA, Pritchard SA. Standardized treatment of 154 consecutive cases of eclampsia. *Am J Obstet Gynecol* 1975; **123**(5): 543-52.
42. Bulik-Sullivan BK, Loh PR, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 2015; **47**(3): 291-5.

Supplementary Table 1: Information on sampling and demographics of study populations.

Study	Study-type	Ancestry	Women(%)	Age (± sd)	MRI	Nmax
3C-Dijon	Population-based	European	63.1	72.2 ± 4.1	1.5 T Siemens	1397
ADNI	Case-Control (AD, MCI, healthy control)	European	45.6	76.1 ± 5.0	1.5 T GE, Philips, and Siemens MRI scanners	204(contols)
ADNI2GO	Case-Control (AD, MCI, healthy control)	European	55.2	72.6 ± 7.1	3.0 T GE, Philips, and Siemens MRI scanners	337(contols)
AGES	Population-based	European	60.8	76.0 ± 5.3	1.5 T Signa Twinspeed EXCITE system (GE Medical Systems, Waukesha, WI)	2478
ARIC	Population-based	European	61.3	72.7 ± 4.33	1.5 T MRI scanners (General Electric Medical Systems)	413
ARIC	Population-based	African-American	60.9	71.6 ± 4.4	1.5 T MRI scanners (General Electric Medical Systems)	389
ASPSFam	Population-based	European	60.8	65.2 ± 10.5	3.0 T Tim Trio (Siemens, Erlangen)	337
CHS	Population-based	European	61.4	78.9 ± 4.2	1.5 T MRI	648
EDIS-SCES	Population-based	Chinese	52.7	70.7 ± 6.2	3Tesla Siemens Magnetom Trio Tim scanner	211
EDIS-SiMES	Population-based	Malays	52.6	70.5 ± 6.7	3Tesla Siemens Magnetom Trio Tim scanner	131
ERF	Family-based study	European	51.7	64.3 ± 4.5	1.5 T MRI unit (GE Healthcare, Milwaukee, USA, software version 11×)	116
FHS	Population-based	European	51.5	58.5 ± 8.0	1 or 1.5 T Siemens Magnetom	2079
GeneSTAR	Family-based study	European	53.7	50.9 ± 10.6	3.0 T Philips Achieva 3.2.1	441
LLS	Family-based study	European	52.7	65.5 ± 6.65	3 Tesla (Philips Medical Systems)	355
ROS-MAP1	Population-based	European	72.5	84.0 ± 6.3	1.5 T MRI unit (GE Healthcare, Milwaukee, USA, software version 11×)	138
ROS-MAP2	Population-based	European	75.9	80.7 ± 6.7	1.5 T MRI unit (GE Healthcare, Milwaukee, USA, software version 11×)	86
RSI	Population-based	European	60.9	78.9 ± 4.9	1.5 T MRI unit (GE Healthcare, Milwaukee, USA, software version 11×)	894
RSII	Population-based	European	55.1	69.4 ± 6.0	1.5 T MRI unit (GE Healthcare, Milwaukee, USA, software version 11×)	1032
RSIII	Population-based	European	54.9	57.0 ± 6.3	1.5 T MRI unit (GE Healthcare, Milwaukee, USA, software version 11×)	2427
SYS	Family-based study	European	51.7	15.0 ± 1.84	1 T Philips	986

3C-Dijon = The Three-City Study, ADNI/ADNI2GO = Alzheimer's Disease Neuroimaging Initiative (ADNI), AGES = Age, Gene/Environment Susceptibility study, ARIC = The Atherosclerosis Risk in Communities Study, ASPSfam = Austrian Stroke Prevention Family study, CHS = Cardiovascular Health Study, EDIS-SCES = Epidemiology of Dementia in Singapore from the Singapore Chinese Eye Study, EDIS-SiMES = Epidemiology of Dementia in Singapore from the Singapore Malay Eye Study, ERF = Erasmus Rucphen Family study, FHS = Framingham Heart Study, GeneSTAR = genetic studies of Atherosclerosis Risk, LLS = Leiden Longevity study, ROS-MAP 1/2 = Religious Orders Study and Memory and Aging Project Study, RSI-II-III = Rotterdam Study I-II-III, SYS = the Saguenay Youth Study. MRI = Magnetic resonance imaging.

Supplementary Table 2: Information on genotyping and quality control.

Study Name	HWE	MAF	Call Rate	Association	Imputation	Reference panel	Genotype Platform
3C-Dijon	1x10-6	0.01	0.95	mach2qtl (1.1.2)	minimac (2012-11-16)	1000 GP (phase 1, version 3)	Illumina Human610-Quad BeadChip
ADNI	1x10-6	0.01	0.95	mach2qtl (1.1.2)	minimac (2012-05-29)	1000 GP (phase 1, version 3)	Illumina Human610-Quad BeadChip
ADNI2GO	1x10-6	0.01	0.95	mach2qtl (1.1.2)	minimac (2012-05-29)	1000 GP (phase 1, version 3)	Illumina Human Omni Express
AGES	1x10-6	0.01	0.97	probABEL	minimac (2012-11-16)	1000 GP (phase 1, version 3)	Illumina Hu370CNV
ARIC	1x10-6	0.01	0.95	probABEL	IMPUTE2 2.2.2	1000 GP (phase 1, version 3)	Affymetrix 6.0
ARIC	1x10-6	0.01	0.95	probABEL	IMPUTE2 2.2.2	1000 GP (phase 1, version 3)	Affymetrix 6.0
ASPS	1x10-6	0.01	0.98	plink v1.07	IMPUTE2 2.2.2	1000 GP (phase 1, version 3)	Illumina Human610-Quad BeadChip
CHS	1x10-5	-	0.97	R	minimac (2012-11-16)	1000 GP (phase 1, version 3)	Illumina Human370CNV BeadChip
EDIS	1x10-6	0.01	0.95	PLINK	minimac	1000 GP (phase 1, version 3)	Illumina Human 610 Quad BeadChips and Illumina Human OmniExpress BeadChips
EDIS	1x10-6	0.01	0.95	PLINK	minimac	1000 GP (phase 1, version 3)	Illumina Human 610 Quad BeadChips
ERF	1x10-6	0.01	0.98	probABEL	MaCH (Mach 1.0.18.c) and minimac (2012.8.15)	1000 GP (phase 1, version 3)	Illumina Human610- Quad BeadChip and/or 370k illumina Beadchip
FHS	1x10-6	0.01	0.97	Perl and R	MACH	1000 GP (phase 1, version 3)	Affymetrix 500K (250K Nsp & 250K Sty), MIPS 50K
GENESTAR	1x10-6	-	-	SAS proc genmod	IMPUTE2 2.2.2	1000 GP (phase 1, version 3)	Illumina 1M_v1c Human chip
LLS	1x10-6	0.01	0.95	SNPtest	IMPUTE2 2.2.2	1000 GP (phase 1, version 3)	Illumina Human660W-Quad
ROS-MAP1	1x10-6	0.01	0.95	plink v1.07	Beagle (3.3.2)	1000 GP (november 2010)	Perlegen-Affymetrix 5.0, Illumina 660K, Illumina 370K, Affymetrix 6.0 907K, Illumina Omni 1M
ROS-MAP2	1x10-6	0.01	0.95	plink v1.07	Beagle (3.3.2)	1000 GP (november 2010)	Illumina Human660W-Quad
RSI	1x10-6	0.01	0.98	probABEL	MaCH (Mach 1.0.18.c) and minimac (2012.8.15)	1000 GP (phase 1, version 3)	Affy 6.0
RSII	1x10-6	0.01	0.98	probABEL	MaCH (Mach 1.0.18.c) and minimac (2012.8.15)	1000 GP (phase 1, version 3)	Illumina OmniExpress
RSIII	1x10-6	0.01	0.98	probABEL	MaCH (Mach 1.0.18.c) and minimac (2012.8.15)	1000 GP (phase 1, version 3)	Illumina Human 550 (+duo) and 610-Quad BeadChip
SYS	1x10-6	0.01	0.95	probABEL palinear	IMPUTE v2.2.2	1000 GP (phase 1, version 3)	Illumina Human610-Quad BeadChip

3C-Dijon = The Three-City Study, ADNI/ADNI2GO = Alzheimer's Disease Neuroimaging Initiative (ADNI), AGES = Age, Gene/Environment Susceptibility study, ARIC = The Atherosclerosis Risk in Communities Study, ASPSfam = Austrian Stroke Prevention Family study, CHS = Cardiovascular Health Study, EDIS-SCES = Epidemiology of Dementia in Singapore from the Singapore Chinese Eye Study, EDIS-SiMES = Epidemiology of Dementia in Singapore from the Singapore Malay Eye Study, ERF = Erasmus Rucphen Family study, FHS = Framingham Heart Study, GeneSTAR = genetic studies of Atherosclerosis Risk, LLS = Leiden Longevity study, ROS-MAP 1/2 = Religious Orders Study and Memory and Aging Project Study, RSI-II-III = Rotterdam Study I-II-III, SYS = the Saguenay Youth Study.

Supplementary Table 3: Number of variants after QC and inflation factor per study.

	Frontal lobe		Temporal lobe		Parietal lobe		Occipital lobe	
Study Name	**N snps after filtering**	**inflation**	**N snps after filtering**	**inflation**	**N snps after filtering**	**inflation**	**N snps after filtering**	**inflation**
3C-Dijon	9078117	0.994	9078117	1.003	9078117	0.994	9078117	0.994
ADNI	8282479	0.993	8282479	0.982	8282479	1.004	8282479	0.998
ADNI2GO	8310693	1.001	8310693	1.000	8310693	0.995	8310693	1.010
AGES	9114618	1.085	9114618	1.050	9114618	1.052	9114618	1.070
ARIC	8775714	1.005	8775714	1.003	8775714	1.003	8775714	1.004
ARIC	14755666	1.020	14755666	1.010	14755666	1.002	14755666	1.004
ASPS	8230154	1.047	8230154	1.049	8230154	1.076	8230154	1.057
CHS	8069973	1.000	8069973	1.008	8069973	1.004	8069973	0.994
EDIS	6280387	1.012	6281593	1.007	6286244	1.019	6285575	1.009
EDIS	7182285	0.967	7186186	1.005	7175200	0.858	7185975	0.959
ERF	5337892	1.083	5337892	1.040	5337892	1.247	5337892	1.059
FHS	8451992	1.028	8451584	1.020	8452581	1.011	8451992	1.020
GENESTAR	7221424	1.027	7221424	1.039	7221424	1.032	7221424	1.050
LLS	8134125	1.070	8134136	1.021	8134128	1.044	8134140	1.023
ROS-MAP1	5687539	1.008	5648329	0.992	5708119	1.000	5706928	1.013
ROS-MAP2	5191901	1.020	5151042	1.013	5251192	1.005	5251192	1.016
RSI	8962350	1.003	8962350	1.002	8962350	1.002	8962350	1.007
RSII	9061092	1.002	9061092	1.012	9061092	1.007	9061092	1.009
RSIII	9757151	1.013	9757151	1.021	9757151	1.016	9757151	1.020
SYS	8511012	1.006	8511012	1.010	8511012	1.003	8511012	1.020

Supplementary Table 4: Heritability estimates of lobar brain volumes.

	Heritability	**p**
Frontal lobe:	0.52	1.9×10^{-19}
Occipital lobe:	0.50	4.3×10^{-21}
Parietal lobe:	0.59	3.1×10^{-29}
Temporal lobe:	0.59	4.6×10^{-28}
Total brain volume:	0.34	8.8×10^{-11}

Supplementary Table 5: Annotation of genome-wide significant variants associated with lobar volumes.

								Variance	EA-only meta-analysis				Multi-ethnic meta-analysis			
Lobe	RS-id	Annotation	(nearby)Gene	Chr	A1	A2	N-max	explained	Freq	Effect	SE	P.value	Frq	Effect	SE	P.value
OLV	**rs12411216**	**upstream; downstream**	**M1R92B;THBS3**	**1**	**A**	**C**	**16016**	**0.175**	**0.4625**	**-0.5211**	**0.0948**	**3.89×10-8**	**0.455**	**-0.487**	**0.092**	**1.37×10-7**
OLV	rs2974937	intronic	THBS3	1	T	C	16016	0.172	0.5384	0.5169	0.0948	4.94×10-8	0.546	0.483	0.092	1.71×10-7
OLV	rs13138250	intronic	IDUA	4	T	C	12287	0.259	0.6315	0.7471	0.1349	3.08×10-8	0.631	0.74	0.131	1.54×10-8
OLV	rs111588353	intergenic	IDUA(dist=2134),FGFRL1(dist=5159)	4	A	G	12200	0.236	0.3187	-0.6978	0.128	4.98×10-8	0.316	-0.671	0.125	7.30×10-8
OLV	**rs74921869**	**intronic**	**FGFRL1**	**4**	**A**	**G**	**12424**	**0.270**	**0.2019**	**-0.8401**	**0.1444**	**5.94×10-9**	**0.203**	**-0.818**	**0.141**	**6.17×10-9**
OLV	**rs1337736**	**intergenic**	**CENPW(dist=175626),RSPO3(dist=594668)**	**6**	**A**	**G**	**16016**	**0.188**	**0.2271**	**-0.6386**	**0.119**	**8.09×10-8**	**0.232**	**-0.637**	**0.116**	**3.98×10-8**
OLV	rs4901902	intergenic	DACT1(dist=502042),DAAM1(dist=38301)	14	A	C	15885	0.197	0.8734	0.7663	0.1399	4.31×10-8	0.874	0.778	0.139	2.37×10-8
OLV	rs8005394	intergenic	DACT1(dist=504526),DAAM1(dist=35817)	14	C	G	15885	0.165	0.8725	0.7668	0.1399	4.20×10-8	0.871	0.701	0.137	2.86×10-7
OLV	rs142965864	intergenic	DACT1(dist=504672),DAAM1(dist=35671)	14	T	TGGATAA	15220	0.207	0.8729	0.7955	0.1427	2.50×10-8	0.874	0.798	0.142	1.89×10-8
OLV	rs17833752	intergenic	DACT1(dist=506042),DAAM1(dist=34301)	14	A	G	15885	0.196	0.8737	0.7708	0.14	3.71×10-8	0.874	0.782	0.14	2.09×10-8
OLV	rs4901904	intergenic	DACT1(dist=509279),DAAM1(dist=31064)	14	T	C	15885	0.195	0.1259	-0.7672	0.1405	4.80×10-8	0.125	-0.779	0.14	2.60×10-8
OLV	rs73313052	intergenic	DACT1(dist=510959),DAAM1(dist=29384)	14	A	G	15885	0.187	0.1282	-0.8144	0.1404	6.58×10-9	0.128	-0.752	0.138	4.71×10-8
OLV	rs5809016	intergenic	DACT1(dist=512397),DAAM1(dist=27946)	14	A	AC	15220	0.231	0.1281	-0.8279	0.1431	7.26×10-9	0.127	-0.842	0.142	3.43×10-9
OLV	rs2164950	intergenic	DACT1(dist=512593),DAAM1(dist=27750)	14	A	G	15885	0.216	0.1275	-0.8057	0.1404	9.66×10-9	0.127	-0.82	0.14	4.63×10-9
OLV	rs74826997	intergenic	DACT1(dist=513571),DAAM1(dist=26772)	14	T	C	15885	0.215	0.8726	0.8052	0.1405	1.00×10-8	0.873	0.819	0.14	4.82×10-9
OLV	rs76341705	intergenic	DACT1(dist=513641),DAAM1(dist=26702)	14	A	G	15885	0.215	0.1274	-0.8054	0.1405	9.98×10-9	0.127	-0.819	0.14	4.79×10-9
OLV	rs201816193	intergenic	DACT1(dist=516037),DAAM1(dist=24304)	14	A	AGTT	15220	0.222	0.1277	-0.8171	0.1432	1.16×10-8	0.127	-0.832	0.143	5.46×10-9
OLV	**rs147148763**	**intergenic**	**DACT1(dist=516038),DAAM1(dist=24302)**	**14**	**G**	**GTTGT**	**15220**	**0.230**	**0.1287**	**-0.8311**	**0.1432**	**6.55×10-9**	**0.128**	**-0.847**	**0.143**	**2.89×10-9**
OLV	rs202194203	intergenic	DACT1(dist=516620),DAAM1(dist=23723)	14	A	AC	15220	0.223	0.8728	0.82	0.1435	1.10×10-8	0.874	0.834	0.143	5.25×10-9
OLV	rs79814107	intergenic	DACT1(dist=522465),DAAM1(dist=17878)	14	A	G	15885	0.220	0.8752	0.8173	0.1406	6.11×10-9	0.876	0.829	0.14	3.13×10-9
OLV	rs113269922	intergenic	DACT1(dist=523136),DAAM1(dist=17207)	14	A	G	15885	0.220	0.8752	0.817	0.1405	6.12×10-9	0.876	0.829	0.14	3.13×10-9
OLV	rs56115079	intergenic	DACT1(dist=527027),DAAM1(dist=13316)	14	T	G	15885	0.220	0.8753	0.8157	0.1404	6.18×10-9	0.876	0.828	0.14	3.19×10-9
OLV	rs112195026	intergenic	DACT1(dist=527964),DAAM1(dist=12379)	14	A	G	15885	0.219	0.8753	0.8152	0.1403	6.21×10-9	0.876	0.827	0.14	3.21×10-9
OLV	rs76850797	intergenic	DACT1(dist=528708),DAAM1(dist=11635)	14	T	C	15885	0.219	0.8753	0.8152	0.1402	6.14×10-9	0.876	0.826	0.14	3.32×10-9
OLV	rs78349529	intergenic	DACT1(dist=529248),DAAM1(dist=11095)	14	A	G	15885	0.220	0.8754	0.8161	0.1402	5.83×10-9	0.876	0.828	0.14	3.02×10-9
OLV	rs4901906	upstream	DAAM1	14	T	C	15885	0.221	0.1245	-0.8121	0.1399	6.43×10-9	0.124	-0.824	0.139	3.32×10-9
OLV	rs79360654	intronic	DAAM1	14	A	C	15885	0.217	0.1241	-0.81	0.1402	7.66×10-9	0.123	-0.822	0.14	3.94×10-9
OLV	rs77618329	intronic	DAAM1	14	T	C	15885	0.216	0.124	-0.8093	0.1404	8.16×10-9	0.123	-0.821	0.14	4.19×10-9
OLV	rs142169263	intronic	DAAM1	14	G	GT	15220	0.230	0.8756	0.8288	0.143	6.88×10-9	0.876	0.841	0.142	3.49×10-9
OLV	rs76256424	intronic	DAAM1	14	C	G	15885	0.216	0.8759	0.8089	0.1404	8.28×10-9	0.877	0.82	0.14	4.44×10-9

								Variance	EA-only meta-analysis				Multi-ethnic meta-analysis			
Lobe	**RS-id**	**Annotation**	**(nearby)Gene**	**Chr**	**A1**	**A2**	**N-max**	**explained**	**Freq**	**Effect**	**SE**	**P.value**	**Frq**	**Effect**	**SE**	**P.value**
OLV	rs76674533	intronic	DAAM1	14	T	C	15885	0.216	0.1241	-0.8087	0.1404	8.33×10-9	0.123	-0.82	0.14	4.46×10-9
OLV	rs143720575	intronic	DAAM1	14	A	G	15885	0.220	0.1155	-0.8475	0.1461	6.52×10-9	0.115	-0.858	0.145	3.57×10-9
OLV	rs78419480	intronic	DAAM1	14	C	G	15885	0.202	0.8795	0.8006	0.1434	2.34×10-8	0.88	0.81	0.143	1.35×10-8
OLV	rs75255901	intronic	DAAM1	14	T	C	15885	0.211	0.1292	-0.7847	0.138	1.31×10-8	0.128	-0.794	0.137	7.36×10-9
OLV	rs76787422	intronic	DAAM1	14	T	C	15885	0.212	0.1292	-0.7853	0.1381	1.30×10-8	0.128	-0.796	0.137	6.91×10-9
OLV	rs17255395	intronic	DAAM1	14	T	C	15883	0.212	0.1289	-0.7857	0.1381	1.27×10-8	0.128	-0.795	0.137	7.11×10-9
OLV	rs55649771	intronic	DAAM1	14	A	G	15885	0.208	0.1265	-0.7951	0.1402	1.41×10-8	0.126	-0.805	0.14	7.81×10-9
PLV	**rs2279829**	**3'-UTR**	**ZIC4**	**3**	**T**	**C**	**16015**	**0.242**	**0.2053**	**-0.9207**	**0.1577**	**5.27×10-9**	**0.21**	**-0.954**	**0.153**	**4.36×10-10**
TLV	rs2218069	intronic	MSRB3	12	A	G	16003	0.208	0.3238	0.6622	0.116	1.13×10-8	0.327	0.653	0.113	8.27×10-9
TLV	rs10878264	intronic	MSRB3	12	A	G	16003	0.233	0.6893	-0.6775	0.1164	5.80×10-9	0.692	-0.697	0.114	1.11×10-9
TLV	rs12831359	intronic	MSRB3	12	T	C	16003	0.228	0.705	-0.6987	0.119	4.35×10-9	0.708	-0.713	0.118	1.49×10-9
TLV	rs11175738	intronic	MSRB3	12	A	G	16003	0.225	0.3111	0.6763	0.1169	7.33×10-9	0.308	0.69	0.115	1.95×10-9
TLV	rs11175741	intronic	MSRB3	12	A	T	16003	0.212	0.6306	-0.6429	0.1121	9.68×10-9	0.629	-0.641	0.11	5.39×10-9
TLV	rs11175742	intronic	MSRB3	12	A	G	16003	0.214	0.3699	0.6451	0.1119	8.20×10-9	0.372	0.644	0.11	4.30×10-9
TLV	rs12816912	intronic	MSRB3	12	A	G	16003	0.228	0.2928	0.6993	0.1193	4.54×10-9	0.29	0.714	0.118	1.49×10-9
TLV	rs12825748	intronic	MSRB3	12	C	G	16003	0.225	0.3097	0.6806	0.1174	6.66×10-9	0.307	0.697	0.116	1.71×10-9
TLV	**rs146354218**	**intronic**	**MSRB3**	**12**	**A**	**G**	**15789**	**0.241**	**0.3678**	**0.6874**	**0.1129**	**1.15×10-9**	**0.37**	**0.685**	**0.111**	**6.43×10-10**
TLV	rs11519877	intronic	MSRB3	12	A	G	16003	0.234	0.3053	0.7144	0.1205	3.07×10-9	0.302	0.729	0.119	1.04×10-9
TLV	rs7299655	intronic	MSRB3	12	T	C	16003	0.221	0.3675	0.6539	0.1124	6.03×10-9	0.369	0.654	0.11	3.20×10-9
TLV	rs10784455	intronic	MSRB3	12	T	C	16003	0.197	0.371	0.6396	0.1116	9.91×10-9	0.381	0.613	0.109	1.96×10-8
TLV	rs4762095	intronic	MSRB3	12	T	C	16003	0.195	0.3698	0.6396	0.1116	9.81×10-9	0.379	0.609	0.109	2.39×10-8
TLV	rs7316273	intronic	MSRB3	12	C	G	16003	0.195	0.6291	-0.6406	0.1115	9.30×10-9	0.622	-0.61	0.109	2.07×10-8
TLV	rs11175755	intronic	MSRB3	12	A	G	16003	0.193	0.3711	0.6392	0.1115	9.91×10-9	0.383	0.606	0.109	2.71×10-8
TLV	rs11175756	intronic	MSRB3	12	A	G	16003	0.213	0.6299	-0.6432	0.1117	8.58×10-9	0.628	-0.642	0.11	4.71×10-9
TLV	rs1494505	intronic	MSRB3	12	T	C	16003	0.196	0.3699	0.6385	0.1115	1.02×10-8	0.379	0.611	0.109	2.05×10-8
TLV	rs4762081	intronic	MSRB3	12	A	G	16003	0.196	0.3711	0.6389	0.1115	1.00×10-8	0.383	0.611	0.109	2.13×10-8
TLV	rs2336716	intronic	MSRB3	12	T	C	16003	0.201	0.371	0.6404	0.1116	9.49×10-9	0.38	0.618	0.109	1.45×10-8
TLV	rs10878275	intronic	MSRB3	12	A	G	16003	0.196	0.3711	0.6388	0.1115	1.01×10-8	0.383	0.611	0.109	2.13×10-8
TLV	rs11175757	intronic	MSRB3	12	A	G	15991	0.197	0.6288	-0.6396	0.1117	1.02×10-8	0.617	-0.612	0.109	2.11×10-8
TLV	rs10878276	intronic	MSRB3	12	T	G	16003	0.196	0.3711	0.6387	0.1115	1.02×10-8	0.383	0.611	0.109	2.13×10-8
TLV	rs11175758	intronic	MSRB3	12	A	T	16003	0.195	0.6289	-0.6385	0.1115	1.03×10-8	0.617	-0.61	0.109	2.15×10-8
TLV	rs11175759	intronic	MSRB3	12	T	C	16003	0.197	0.3711	0.6379	0.1115	1.06×10-8	0.38	0.612	0.109	2.04×10-8

								Variance	EA-only meta-analysis				Multi-ethnic meta-analysis			
Lobe	**RS-id**	**Annotation**	**(nearby)Gene**	**Chr**	**A1**	**A2**	**N-max**	**explained**	**Freq**	**Effect**	**SE**	**P.value**	**Frq**	**Effect**	**SE**	**P.value**
TLV	rs10784456	intronic	MSRB3	12	T	C	16003	0.218	0.63	-0.6426	0.1118	8.93×10-9	0.628	-0.65	0.11	3.26×10-9
TLV	rs12229693	intronic	MSRB3	12	A	G	16003	0.185	0.6358	-0.6305	0.1129	2.35×10-8	0.625	-0.599	0.11	5.61×10-8
TLV	rs11561370	intronic	MSRB3	12	T	C	16003	0.201	0.6287	-0.6416	0.1115	8.79×10-9	0.619	-0.618	0.109	1.41×10-8
TLV	rs7314906	intronic	MSRB3	12	C	G	16003	0.215	0.6301	-0.644	0.1118	8.28×10-9	0.628	-0.646	0.11	3.87×10-9
TLV	rs1494512	intronic	MSRB3	12	A	C	16003	0.197	0.6296	-0.6422	0.1117	8.88×10-9	0.618	-0.613	0.109	1.97×10-8
TLV	rs1494502	intronic	MSRB3	12	A	G	16003	0.213	0.6372	-0.6389	0.1133	1.73×10-8	0.636	-0.648	0.111	5.77×10-9
TLV	rs11306160	intronic	MSRB3	12	G	GA	15348	0.208	0.6289	-0.6612	0.1139	6.43×10-9	0.617	-0.627	0.111	1.70×10-8
TLV	rs10784457	intronic	MSRB3	12	A	G	16003	0.195	0.629	-0.6392	0.1115	9.83×10-9	0.617	-0.609	0.109	2.31×10-8
TLV	rs201270036	intronic	MSRB3	12	T	TATAC	15348	0.201	0.391	0.67	0.1164	8.75×10-9	0.403	0.633	0.114	2.56×10-8
TLV	rs4762096	intronic	MSRB3	12	T	C	16003	0.194	0.3699	0.6388	0.1115	1.01×10-8	0.382	0.608	0.109	2.39×10-8
TLV	rs997188	intronic	MSRB3	12	A	C	16003	0.195	0.63	-0.6391	0.1115	1.00×10-8	0.618	-0.609	0.109	2.28×10-8
TLV	rs1587056	intronic	MSRB3	12	A	T	16003	0.189	0.3686	0.6375	0.1122	1.33×10-8	0.378	0.606	0.11	3.28×10-8
TLV	rs10533337	intronic	MSRB3	12	G	GTGAA	15348	0.213	0.6288	-0.6641	0.1141	5.81×10-9	0.619	-0.635	0.111	1.17×10-8
TLV	rs2336717	intronic	MSRB3	12	C	G	16003	0.194	0.6289	-0.6393	0.1116	1.01×10-8	0.617	-0.608	0.109	2.46×10-8
TLV	rs3858644	intronic	MSRB3	12	C	G	16003	0.184	0.3728	0.6284	0.1125	2.34×10-8	0.385	0.597	0.11	5.56×10-8
TLV	rs1494510	intronic	MSRB3	12	T	C	16003	0.195	0.3713	0.6396	0.1118	1.06×10-8	0.38	0.609	0.109	2.40×10-8
TLV	rs4762098	intronic	MSRB3	12	T	C	16003	0.195	0.6285	-0.6414	0.1118	9.59×10-9	0.617	-0.609	0.109	2.53×10-8
TLV	rs10878279	intronic	MSRB3	12	A	G	16003	0.209	0.2924	0.6805	0.1208	1.77×10-8	0.289	0.694	0.12	6.63×10-9
TLV	rs989271	intronic	MSRB3	12	A	G	16003	0.200	0.3719	0.6431	0.112	9.28×10-9	0.381	0.617	0.109	1.62×10-8
TLV	rs10878280	intronic	MSRB3	12	A	G	16003	0.211	0.6293	-0.6457	0.1122	8.68×10-9	0.628	-0.639	0.11	5.68×10-9
TLV	rs1494509	intronic	MSRB3	12	T	C	16003	0.196	0.372	0.6443	0.112	8.86×10-9	0.384	0.617	0.11	1.79×10-8
TLV	rs1494508	intronic	MSRB3	12	A	G	16001	0.198	0.3722	0.6447	0.112	8.60×10-9	0.384	0.62	0.11	1.50×10-8

OLV = Occipital lobar volume, TLV = Parietal lobar volume, TLV =Temporal lobar volume, Bold typed variants are the index variants in the locus.

Supplementary Table 6: Annotation of genome-wide significant variants associated with lobar volumes (continued) Epigenetic markers in brain (BRN) tissues are highlighted in bold.

RS-id	Chr	conserved-SiPhy	Promoter histone marks in:	Enhancer histone marks in:	DNAse hypersensitivity sites	Proteins bound:	eQTL hits	Motifs changed
rs12411216	1		ESC, ESDR, LNG, IPSC, FAT, STRM, BRST, BLD, MUS, **BRN**, SKIN, VAS, LIV, GI, ADRL, HRT, KID, PANC, PLCNT, THYM, OVRY, CRVX, BONE	PLCNT, GI, SPLN	ESC,ESDR,ESDR,ESDR,ESDR,ESC,LNG,IPSC,IPSC,BRST,BLD,BLD,BLD,BLD,BLD,BLD,SKIN,SKIN,SKIN,ADRL,**BRN**,**BRN**,HRT,GI,GI,KID,LNG,MUS,MUS,PLCNT,GI,THYM,GI,OVRY,PANC,LNG,BLD,CRVX,LIV,BRST,MUS,MUS,VAS,BLD,BLD,**BRN**,SKIN,LNG	POL2,POL24H8,TAF1,NFKB,JUND,TCF4,BRG1,INI1,STAT1,CCNT2,CMYC,ETS1,HEY1,HMGN3,IRF1,MAX,POL2B,TAF7,USF1,ZBTB7A	11 eQTL results	BAF155,ERalpha-a,Evi-1,HEY1,Osf2,PEBP,ZBTB33,p300
rs2974937	1			ESDR			12 eQTL results	ERalpha-a,Egr-1,Esr2,Rad21
rs13138250	4			ESDR, SKIN, **BRN**, SPLN			4 eQTL results	ZBTB33
rs111588353	4						4 eQTL results	
rs74921869	4			ESC, ESDR, IPSC, BRST, BLD, **BRN**, SKIN, ADRL, GI, MUS, PLCNT, OVRY, LNG			Lappalainen2013,Lymphoblastoid,FGFRL1	BCL,Egr-1,Ets,GATA,Irf,Maf,PU.1,Pax-5,SZF1-1
rs1337736	6			ESDR			Fairfax2014,Monocytes_LPS24,C6ORF173	Foxd1,Foxj1,Foxk1,Foxq1,Irf,XBP-1
rs4901902	14			BLD			6 eQTL results	CEBPA,CEBPB,HNF1,Rad21,p300
rs8005394	14							BDP1,Maf,STAT,Spz1,VDR
rs142965864	14							GATA,Hltf,Ik-2,SIX5
rs17833752	14			BLD, **BRN**			3 eQTL results	CEBPG,HMG-IY,Nkx2,Pou1f1
rs4901904	14							EBF
rs73313052	14			ESC, IPSC, KID	KID,LNG,MUS,GI	KAP1		
rs5809016	14	yes		ESDR, BLD, HRT, GI, PANC, LIV	ESDR,MUS,GI			Hand1,NF-kappaB,NRSF,Pou2f2,RP58,STAT,Sin3Ak-20
rs2164950	14	yes		ESDR, FAT, BLD, HRT, GI, PANC, LIV, **BRN**, LNG			3 eQTL results	
rs74826997	14		ESDR, FAT, STRM, BLD, MUS, **BRN**, SKIN, VAS, KID, GI, LIV, LNG, BONE	ESC, ESDR, IPSC, FAT, BLD, **BRN**, SKIN, GI, HRT, LNG, MUS, THYM, PANC, PLCNT, CRVX	ESDR,ESDR,IPSC,BLD,BLD,SKIN,HRT,GI,GI,KID,LNG,MUS,MUS,LIV,MUS,MUS,BLD,BLD,**BRN**,LNG	GR,POL2,TBP		NF-Y,Zbtb3
rs76341705	14		ESDR, FAT, STRM, BLD, MUS, **BRN**, SKIN, VAS, KID, GI, LIV, LNG, BONE	ESC, ESDR, IPSC, FAT, BLD, **BRN**, SKIN, GI, HRT, LNG, MUS, THYM, PANC, PLCNT, CRVX	ESDR,ESDR,LNG,IPSC,SKIN,SKIN,ADRL,HRT,GI,GI,KID,LNG,MUS,MUS,THYM,MUS,BLD,LIV,MUS,MUS,BLD,BLD,**BRN**,LNG	POL2		
rs201816193	14			ESC, IPSC				Evi-1,Foxa,Foxj2,GATA,HDAC2,Irf,p300
rs147148763	14			ESC, IPSC				Evi-1,Foxa,Foxj2,HDAC2,p300
rs202194203	14							Foxa,Foxo,Foxp1,HMG-IY,SIX5
rs79814107	14							AP-1,Dbx1,Lhx3,Prrx1
rs113269922	14							Cdx2,Evi-1,Hoxa10,Hoxb13,Hoxb9,Hoxc9,Hoxd10,Pax-4,TATA

RS-id	Chr	conserved-SiPhy	Promoter histone marks in:	Enhancer histone marks in:	DNAse hypersensitivity sites	Proteins bound:	eQTL hits	Motifs changed
rs56115079	14		BLD	BLD	BLD	POL2		
rs112195026	14							AFP1,Arid3a,FAC1,Foxp1,Mef2,Nanog,Pou5f1,SIX5
rs76850797	14			ESC, FAT, STRM, SKIN, VAS, HRT	SKIN,SKIN,OVRY,SKIN			
rs78349529	14			ESC, ESDR, FAT	ESC,ESDR,ESDR,IPSC,IPSC,SKIN,SKIN,SKIN,PLCNT,LNG,CRVX,MUS,**BRN**,SKIN,SKIN,LNG			NF-Y,NF-kappaB,STAT,UF1H3BETA
rs4901906	14	yes	ESC, ESDR, LNG, IPSC, FAT, STRM, BRST, BLD, MUS, **BRN**, SKIN, VAS, LIV, GI, ADRL, HRT, KID, PANC, PLCNT, THYM, OVRY, SPLN, CRVX, BONE		ESDR,ESDR,ESDR,IPSC,BRST,BLD,BLD,BLD,BLD,BLD,BLD,SKIN,SKIN,SKIN,ADRL,**BRN**,**BRN**,HRT,GI,GI,KID,LNG,MUS,MUS,PLCNT,GI,THYM,OVRY,PANC,GI,BLD,CRVX,LIV,MUS,VAS,BLD,BLD,SKIN,LNG	CMYC,E2F6,ETS1,GABP,GATA2,MAX,NRSF,POL2,POL24H8,POL2B,TAL1,YY1,GATA1,HAE2F1		Pax-5,Spz1
rs79360654	14		ESDR, **BRN**, GI, HRT, SKIN	ESDR, ESC, BRST, **BRN**, SKIN, FAT, LIV, GI, HRT, MUS, LNG, OVRY, PANC, PLCNT, THYM, VAS	BRST,SKIN,HRT,KID,BLD,BRST,SKIN			Foxa,Foxf2,TCF12,p300
rs77618329	14	yes	GI	ESDR, ESC, **BRN**, FAT, VAS, LIV, GI, HRT, KID, LNG, MUS, OVRY, PANC, BLD, SKIN	HRT,PANC,BLD,SKIN			BDP1,Egr-1,SETDB1,Znf143
rs142169263	14		VAS, HRT, GI	ESDR, **BRN**, FAT, LIV, GI, KID, LNG, MUS, HRT, OVRY, PANC, VAS, BLD, SKIN	KID,LNG,MUS,PANC			FXR,PRDM1
rs76256424	14		HRT, GI	SKIN, FAT, LIV, GI, **BRN**, LNG, MUS, HRT, VAS, BLD	HRT,LNG,MUS,OVRY,BLD			ERalpha-a,RXRA
rs76674533	14	yes	HRT	FAT, LIV, GI, LNG, MUS, VAS	HRT,LNG,MUS			AP-1
rs143720575	14			ADRL				Hoxa4,Pax-4,Pax-6
rs78419480	14			SKIN, FAT, GI, HRT, LNG, OVRY, BLD	IPSC,OVRY	CTCF		LUN-1
rs75255901	14			BLD, HRT	BLD,LNG,MUS,THYM,BLD			ERalpha-a,Esr2,RXRA,Znf143
rs76787422	14			HRT				NF-kappaB
rs17255395	14			HRT	OVRY		3 eQTL results	CEBPA,CEBPB,Foxj2,STAT,p300
rs55649771	14							YY1,ZID
rs2279829	3	yes	ESDR, **BRN**	ESC, ESDR, IPSC, **BRN**	ESDR,**BRN**	NANOG	4 eQTL results	
rs2218069	12						3 eQTL results	PLZF,XBP-1
rs10878264	12						3 eQTL results	Hic1
rs12831359	12						3 eQTL results	
rs111175738	12			SKIN				Foxl1,GATA,HMGN3,Nkx3,Nkx6-2,TATA
rs111175741	12							CEBPG,Hltf,Zbtb12
rs111175742	12			LNG, PANC				GR,Sin3Ak-20,Zbtb3
rs12816912	12						3 eQTL results	CEBPB,Fox,Foxa,Foxf2,Foxi1,Foxj1,Pou3f2,p300
rs12825748	12							
rs146354218	12							CEBPB,Mef2,Pax-4
rs11519877	12			ESDR, MUS				Irf,Pax-5,Sin3Ak-20,Zfp281
rs7299655	12							E2A,Pax-4,ZEB1
rs10784455	12						3 eQTL results	E2F,Mef2,Nkx3

RS-id	Chr	conserved-SiPhy	Promoter histone marks in:	Enhancer histone marks in:	DNAse hypersensitivity sites	Proteins bound:	eQTL hits	Motifs changed
rs4762095	12							Crx,Dbx1,Foxpl,Hoxa10,Ncx,Nkx6-1,Sox
rs7316273	12			LNG, BLD			3 eQTL results	Handl
rs11175755	12			FAT, BLD, ADRL, LNG, BRST	ESDR,BRST,SKIN		3 eQTL results	BATF,DMRT3,HMG-IY,Irf
rs11175756	12			BLD, LNG, BRST			3 eQTL results	Nkx3
rs1494505	12		LNG	SKIN, GI, LIV, BRST	MUS		4 eQTL results	Nanog,RXRA
rs4762081	12						3 eQTL results	Gfi1,Mef2,Pax-4,Pou2f2
rs2336716	12							HNF4
rs10878275	12							ERalpha-a,RORalpha1,RXRA,SF1
rs11175757	12						3 eQTL results	ERalpha-a
rs10878276	12							EWSR1-FLI1,NERF1a,Tel2
rs11175758	12							Pou5f1,Sox
rs11175759	12							HNF4,Hoxd10,Irf,STAT
rs10784456	12				GI,BLD			
rs12229693	12							HNF1
rs11561370	12				THYM			DMRT1,DMRT7,GR
rs7314906	12			ESC, IPSC				Fox,Foxa,Foxd3,Foxj2,GR,HDAC2,p300
rs1494512	12			VAS			3 eQTL results	CHOP::CEBPalpha,MAZ,Mef2,p300
rs1494502	12		STRM, LNG, VAS	ESDR, FAT, BRST, STRM, MUS, SKIN, GI, LNG, CRVX, **BRN**, BONE	SKIN,SKIN,SKIN,CRVX,MUS,VAS,**BRN**		3 eQTL results	p53
rs11306160	12			ESDR, FAT, STRM, BRST, MUS, SKIN, GI, VAS, **BRN**, LNG, BONE	**BRN**			CIZ,Evi-1,Gfi1,HDAC2,HMG-IY,HNF6,Hdx,Pou2f2,Sox
rs10784457	12				CRVX,LNG	CFOS		Maf,Mef2,Mrg1::Hoxa9,Pou5f1
rs201270036	12			ESDR, STRM, BRST, **BRN**, VAS		CFOS		Foxl1,Pou2f2,RBP-Jkappa
rs4762096	12			ESDR		TCF4	4 eQTL results	
rs997188	12			GI, VAS			3 eQTL results	Rad21
rs1587056	12							Foxc1,Nkx6-2,PLZF
rs10533337	12							AP-1,CDP,CEBPB,HNF1,Hbp1,Irf,Msx-1,Msx2,Pbx-1,Pbx3,Pou1f1,Pou2f2,Prrx2,RFX5,SP2,Sox
rs2336717	12						3 eQTL results	AP-1
rs3858644	12			GI, HRT	GI		3 eQTL results	Rad21
rs1494510	12			GI, HRT	GI,GI,GI		3 eQTL results	CHOP::CEBPalpha,RREB-1
rs4762098	12						3 eQTL results	
rs10878279	12						3 eQTL results	Foxp1,HDAC2
rs989271	12							GATA,HDAC2,STAT,Sox
rs10878280	12							SIX5,Znf143,p300
rs1494509	12							Arid5a,E2F,Nkx2,STAT
rs1494508	12						3 eQTL results	Nkx3

Supplementary Table 7: Pleiotropic effects of index genome-wide significant variants.

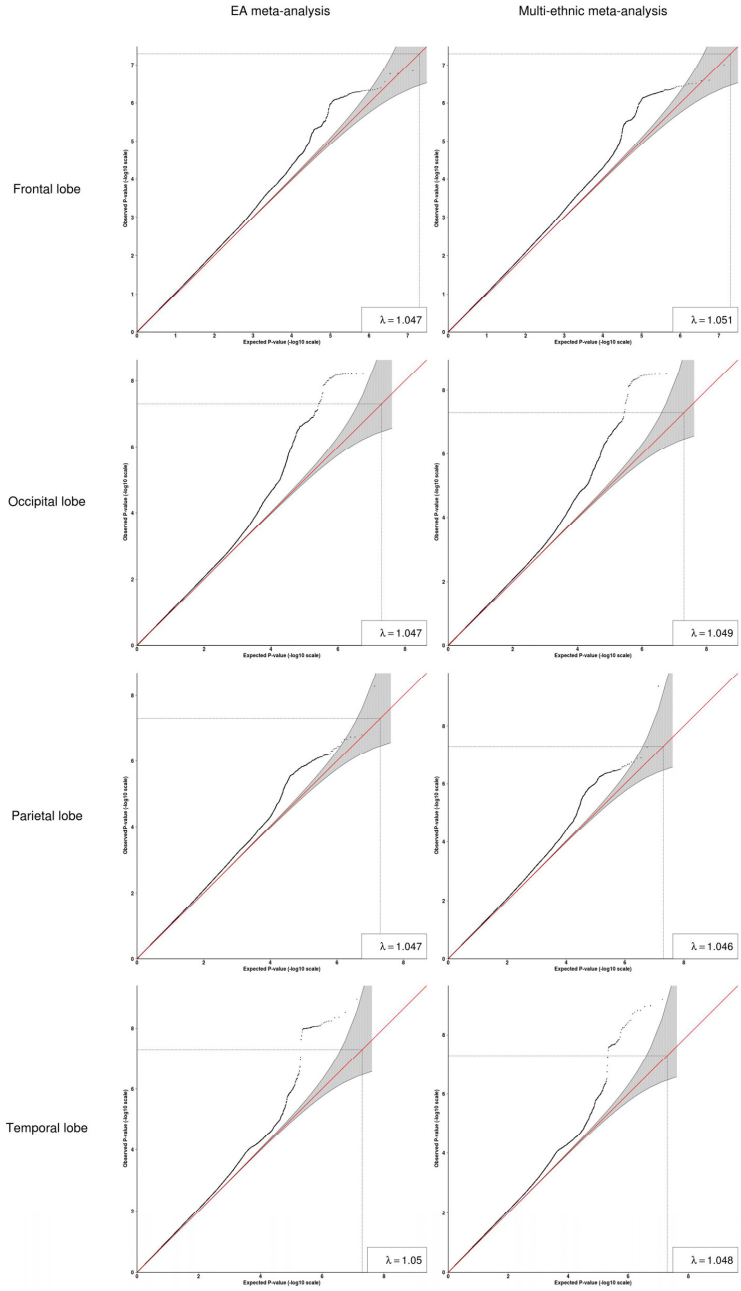
							EA-only meta-analysis				Multi-ethnic meta-analysis					Variance
Lobe	**(nearby)Gene**	**Marker**	**Chr**	**Postion**	**A1**	**A2**	**Freq**	**Effect**	**SE**	**P.value**	**Frq**	**Effect**	**SE**	**P.value**	**N**	**explained**
FLV	MIR92B;THBS3	rs12411216	1	155164480	A	C	0.4629	-0.4934	0.2027	1.5×10-2	0.457	-0.462	0.198	2.0×10-2	16015	0.034
OLV	**MIR92B;THBS3**	**rs12411216**	**1**	**155164480**	**A**	**C**	**0.4625**	**-0.5211**	**0.0948**	**3.9×10-8**	**0.455**	**-0.487**	**0.092**	**1.4×10-7**	**16016**	**0.175**
PLV	MIR92B;THBS3	rs12411216	1	155164480	A	C	0.4608	-0.558	0.1298	1.7×10-5	0.454	-0.537	0.127	2.4×10-5	16015	0.112
TLV	MIR92B;THBS3	rs12411216	1	155164480	A	C	0.4633	-0.042	0.1108	0.70	0.457	-0.009	0.108	0.93	16003	0.000
FLV	ZIC4	rs2279829	3	147106319	T	C	0.2057	0.052	0.2459	0.83	0.211	0.084	0.239	0.74	16015	0.001
OLV	ZIC4	rs2279829	3	147106319	T	C	0.203	0.1273	0.1164	0.27	0.21	0.064	0.112	0.57	16016	0.002
PLV	**ZIC4**	**rs2279829**	**3**	**147106319**	**T**	**C**	**0.2053**	**-0.9207**	**0.1577**	**5.3×10-9**	**0.21**	**-0.954**	**0.153**	**4.4×10-10**	**16015**	**0.242**
TLV	ZIC4	rs2279829	3	147106319	T	C	0.205	-0.2165	0.1349	0.11	0.21	-0.231	0.131	0.078	16003	0.019
FLV	FGFRL1	rs74921869	4	1013382	A	G	0.1999	-0.7564	0.3116	1.5×10-2	0.201	-0.72	0.305	1.8×10-2	12423	0.045
OLV	**FGFRL1**	**rs74921869**	**4**	**1013382**	**A**	**G**	**0.2019**	**-0.8401**	**0.1444**	**5.9×10-9**	**0.203**	**-0.818**	**0.141**	**6.2×10-9**	**12424**	**0.270**
PLV	FGFRL1	rs74921869	4	1013382	A	G	0.2005	-0.4555	0.1955	1.9×10-2	0.201	-0.382	0.192	4.6×10-2	12421	0.032
TLV	FGFRL1	rs74921869	4	1013382	A	G	0.2012	-0.1708	0.1671	0.306	0.202	-0.101	0.164	0.54	12413	0.003
FLV	CENPW(dist=175626),RSPO3(dist=594668)	rs1337736	6	126845380	A	G	0.2262	-0.0526	0.2541	0.84	0.23	-0.109	0.249	0.66	16015	0.001
OLV	**CENPW(dist=175626),RSPO3(dist=594668)**	**rs1337736**	**6**	**126845380**	**A**	**G**	**0.2271**	**-0.6386**	**0.119**	**8.1×10-8**	**0.232**	**-0.637**	**0.116**	**4.0×10-8**	**16016**	**0.188**
PLV	CENPW(dist=175626),RSPO3(dist=594668)	rs1337736	6	126845380	A	G	0.2255	-0.4532	0.1629	5.4×10-3	0.229	-0.51	0.159	1.4×10-3	16015	0.064
TLV	CENPW(dist=175626),RSPO3(dist=594668)	rs1337736	6	126845380	A	G	0.2259	0.1864	0.139	0.18	0.229	0.149	0.136	0.2741	16003	0.008
FLV	MSRB3	rs146354218	12	65793942	A	G	0.3699	0.29	0.2065	0.16	0.373	0.34	0.202	0.093	15791	0.018
OLV	MSRB3	rs146354218	12	65793942	A	G	0.3694	0.1037	0.0974	0.29	0.373	0.112	0.095	0.24	15792	0.009
PLV	MSRB3	rs146354218	12	65793942	A	G	0.3686	0.0878	0.1323	0.51	0.371	0.09	0.13	0.48	15791	0.003
TLV	**MSRB3**	**rs146354218**	**12**	**65793942**	**A**	**G**	**0.3678**	**0.6874**	**0.1129**	**1.2×10-9**	**0.37**	**0.685**	**0.111**	**6.4×10-10**	**15789**	**0.241**
FLV	DACT1(dist=516038),DAAM1(dist=24302)	rs147148763	14	59631075	G	GTTGT	0.1282	-0.6591	0.3044	3.0×10-2	0.128	-0.687	0.303	2.3×10-2	15219	0.034
OLV	**DACT1(dist=516038),DAAM1(dist=24302)**	**rs147148763**	**14**	**59631075**	**G**	**GTTGT**	**0.1287**	**-0.8311**	**0.1432**	**6.6×10-9**	**0.128**	**-0.847**	**0.143**	**2.9×10-9**	**15220**	**0.230**
PLV	DACT1(dist=516038),DAAM1(dist=24302)	rs147148763	14	59631075	G	GTTGT	0.1281	0.9436	0.1948	1.3×10-6	0.127	0.914	0.194	2.5×10-6	15219	0.146
TLV	DACT1(dist=516038),DAAM1(dist=24302)	rs147148763	14	59631075	G	GTTGT	0.1282	0.62	0.1676	2.2×10-4	0.127	0.593	0.167	3.8×10-4	15217	0.083

Supplementary Table 8: Variance explained by common SNPs.

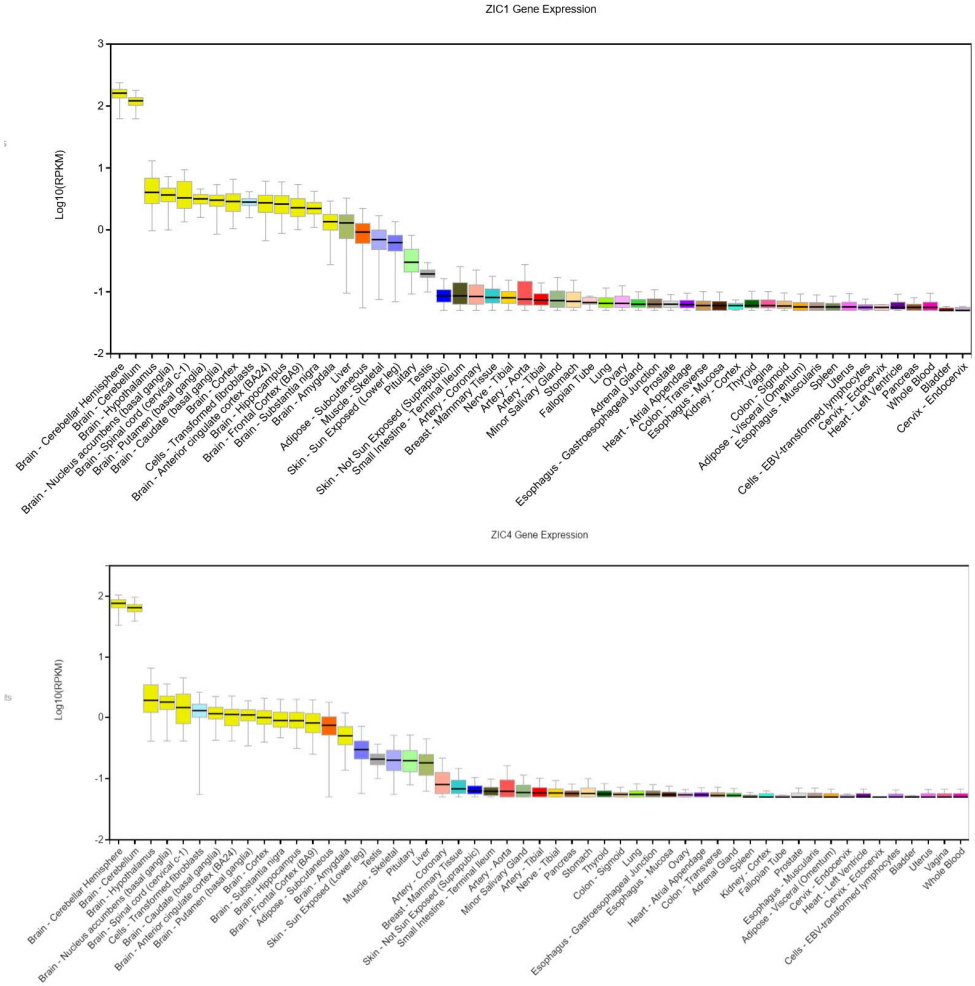
Lobe	Variance explained (se)	Heritability	% of heritability explained
Frontal lobe	0.196 (0.0372)	0.52	37.7%
Occipital lobe	0.203 (0.0361)	0.50	40.6%
Parietal lobe	0.179 (0.0316)	0.59	30.3%
Temporal lobe	0.175 (0.0345)	0.59	29.7%

Supplementary Table 9: Genetic correlations between brain lobar volumes, between brain lobar volumes and volumes of other brain structures, and between brain lobar volumes and neurological/psychiatric diseases using LD-SCORE regression.[24,42]

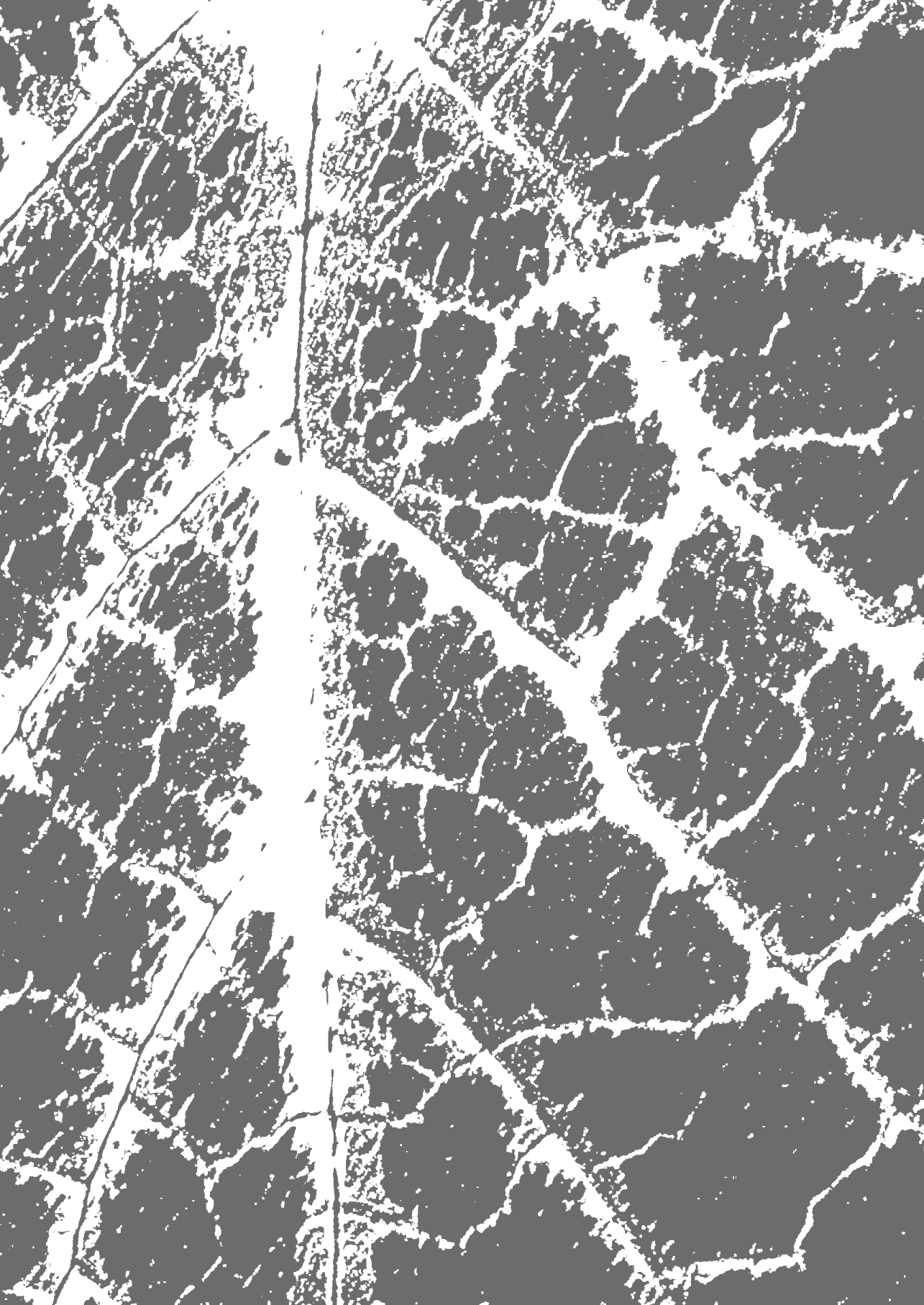
	Temporal Lobe		**Occipital lobe**		**Parietal lobe**		**Frontal lobe**	
Trait	**R_g**	**p**	**R_g**	**p**	**R_g**	**p**	**R_g**	**p**
Occipital lobe	0.22	0.063						
Parietal lobe	0.41	5.2×10^{-5}	0.28	2.1×10^{-2}				
Frontal lobe	0.02	0.905	0.09	0.436	0.12	0.358		
Amygdala volume	0.56	2.6×10^{-3}	0.33	0.041	0.13	0.498	0.16	0.377
Brainstem volume	0.05	0.599	0.04	0.666	0.10	0.283	0.14	0.121
Caudate volume	0.00	0.989	0.03	0.766	-0.06	0.548	-0.02	0.860
Extra ventricular volume	-0.12	0.372	0.04	0.767	0.11	0.369	-0.06	0.646
Hippocampal volume	0.21	6.0×10^{-2}	0.17	0.104	0.07	0.569	-0.09	0.460
Intracranial volume	0.19	4.1×10^{-2}	0.10	0.284	0.09	0.333	0.07	0.409
Nucleus accumbens volume	0.21	9.5×10^{-2}	0.19	0.116	0.24	4.5×10^{-2}	0.10	0.402
Pallidum volume	0.01	0.957	0.05	0.668	0.04	0.752	0.09	0.386
Putamen volume	0.20	2.2×10^{-2}	0.11	0.217	0.12	0.161	0.06	0.442
Thalamic volume	0.33	3.8×10^{-3}	0.11	0.308	0.09	0.400	0.17	0.084
White matter leasions	0.09	0.384	0.03	0.822	0.04	0.698	0.08	0.441
Alzheimer's Disease	-0.04	0.801	0.06	0.643	-0.02	0.899	0.07	0.624
attention deficit disorder	0.28	0.233	0.30	0.148	0.40	7.9×10^{-2}	0.66	6.5×10^{-3}
Autism	0.04	0.731	-0.11	0.337	-0.01	0.896	0.05	0.597
Bipolar disorder	-0.07	0.521	0.06	0.558	-0.08	0.421	0.11	0.340
General cognitive ability	0.02	0.846	0.02	0.828	-0.03	0.666	-0.03	0.626
Major depressive disorder	0.05	0.704	0.08	0.489	0.07	0.634	0.05	0.765
Neurotism	-0.13	0.433	-0.05	0.741	-0.06	0.687	0.15	0.299
Parkinsons disease	0.04	0.649	0.15	7.0×10^{-2}	0.03	0.736	0.12	0.173
Schizophrenia	0.00	0.966	0.06	0.459	-0.08	0.393	-0.07	0.495



Supplementary Figure 1: Quantile-quantile (QQ) plot of four lobar volume genome-wide association study meta-analysis. The plots are shown for the European Ancestry only and the multi-ethnic meta-analysis. All observed –log10 (p-values) are plotted against the expected –log10 (p-values) of a normal distribution. The grey lines depict the 95% confidence interval. The observed inflation was low lambda < 1.05 and was determined to be mainly a result of polygenicity using linkage disequilibrium score regression (intercept≈1).[42]



Supplementary Figure 2: GTEx expression of ZIC1 and ZIC



Chapter 5

Endophenotypes and risk factors of Alzheimer's disease in blood

Chapter 5.1

Genome-wide Association Study Links *APOE*ε4 and *BACE1* Variants with Plasma Amyloid β Levels

Vincent Chouraki,* Sven J. van der Lee,* Benjamin Grenier-Boley, Jeannette Simino, Hieab H. Adams, Giuseppe Tosto, Charles White, Natalie Terzikhan, Carlos Cruchaga, Shuo Li, Susanna Schraen, Megan L. Grove, Claudia Satizabal, Najaf Amin, Claudine Berr, Steven Younkin, Alzheimer's Disease Neuroimaging Initiative, Rebecca F. Gottesman, Luc Buée, Alexa Beiser, David S. Knopman, Andre Uitterlinden, Charles DeCarli, Jan Bressler, Anita DeStefano, Jean-François Dartigues, Qiong Yang, Eric Boerwinkle, Christophe Tzourio, Myriam Fornage, M. Arfan Ikram, Philippe Amouyel, Phil de Jager, Chritiane Reitz, Thomas H. Mosley, Jr. Jean-Charles Lambert,** Sudha Seshadri,** Cornelia M. van Duijn**

* authors contributed equally

** senior authors contributed equally

This chapter is submitted.

Abstract

Amyloid β (Aβ) peptides are the products of the catalytic processing of the Aβ precursor protein (APP) by the β-secretase, BACE1 and the γ-secretase complex. Impairment of the Aβ production/clearance balance is the major pathophysiological hypothesis in Alzheimer's disease (AD). Plasma Aβ levels are easy to measure in large numbers and therefore can be used to study the genetics of Aβ and its relevance to AD. Using a genome-wide association studies of plasma Aβ levels in 12,366 non-demented participants across 8 studies, we identified 21 variants reaching genome-wide significance across two loci spanning *APOE*, the most important genetic risk factor for AD, and *BACE1*. We also observed suggestive evidence of association around *APP* and *PSEN2*. Overall, this study was able to identify relevant and central actors of the APP metabolism in AD and strengthens the utility of plasma Aβ levels both as an endophenotype and a biomarker.

Introduction

Amyloid β (Aβ) peptides are the products of the catalytic processing of the Aβ precursor protein (APP) by the β-secretase, BACE1 and the γ-secretase complex. [1] Aβ peptides are mainly produced in the brain where *APP* and *BACE1* are both highly expressed,[1] but also in circulating blood platelets [2] and in the pancreas.[1] Aβ peptides are able to self-assemble in soluble Aβ oligomers but also in insoluble fibrils that can aggregate as plaques in the brain parenchyma or in the wall of pial blood vessels where they constitute defining hallmarks of Alzheimer's disease (AD)[3] and cerebral amyloid angiopathy (CAA),[4] respectively.

There is strong evidence pointing toward a central role of Aβ peptides in the pathophysiology of AD, although the exact role remains controversial.[5] Studies have shown that a large variety of individually rare mutations in genes involved in Aβ production, including *APP*, *PSEN1* and *PSEN2*, lead to autosomal dominant early-onset forms of AD (EOAD) and to lobar hemorrhage from cerebral amyloid angiopathy.[6] Also, Apolipoprotein E (*APOE*) ε4, the major genetic risk factor for late-onset AD (LOAD) in the general population,[7] has been implicated in Aβ aggregation, deposition and clearance, both in brain and in blood vessels.[8]

These results are the basis of drug discovery efforts targeting Aβ production and clearance that are currently under consideration in clinical trials, with negative results up to now.[9] Although these results appear to contradict the amyloid hypothesis, several explanations have been advanced.[10] One major concern is the ineffectiveness of intervention after the onset of clinical symptoms which are the result of irreversible neurodegeneration. Since the amyloid accumulation process likely precedes the clinical onset by decades,[11,12] early intervention in high risk groups such as Down syndrome patients and *APOE*ε4/ε4 carriers remains to be evaluated. Another concern is the lack of knowledge concerning the precise mechanisms involving Aβ in the pathophysiology of LOAD. Indeed, except for *APOE* and *SORL1*, only a small number of genes identified by genome-wide association studies (GWAS) of LOAD [13] have been linked to APP metabolism [14,15] and Aβ-related pathways have not yet emerged in formal enrichment

analyses.[16] Moreover, the genes involved in autosomal dominant forms of EOAD have not been detected in GWAS of LOAD.

To address these concerns, GWAS have turned to quantitative measures of Aβ peptides, either in the cerebrospinal fluid (CSF) or in the brain, through Pittsburgh Compound B (PiB) PET scan or autopsy. [17-20] Combining the effect of AD genetic loci resulted in statistically significant effects on CSF Aβ42, suggesting that amyloid metabolism is also involved in LOAD.[21] Nevertheless, these studies are limited in their sample size due to low acceptability of lumbar puncture and brain donation and high cost of PiB PET, and therefore may lack statistical power necessary in genetic association research. Aβ peptides produced in the brain can be degraded locally or transported into the CSF and the blood stream where they can be easily detected.[22] Although the brain-derived Aβ peptides in the circulation cannot be distinguished from Aβ derived from blood platelets or pancreas, plasma Aβ levels are modestly but significantly correlated with amyloid burden in the CSF and in the brain.[23] Our groups have also independently shown that plasma Aβ concentrations are prospectively associated with the future risk of developing AD,[24-27] suggesting that there is indeed a link between mechanisms controlling Aβ concentrations in plasma and AD pathophysiological processes in the brain and that circulating Aβ peptides can be used as a marker for brain amyloid metabolism. Alternatively, it has also been postulated that the relation of plasma Aβ and future AD is reflecting general physiological processes in the brain, platelets and kidney, thus giving information on Aβ peptides production/clearance in each of these tissues.[28]

Within this background, we set out to discover genetic determinants of circulating Aβ. We previously conducted a GWAS of plasma Aβ levels in 3,528 non-demented participants, but failed to find genome-wide significant associations due to lack of power.[28] The present study extends our previous work by performing a GWAS using a sample size (n=11,659) that is more than three times larger.

Methods

Study population

We included data from 12,366 European-descent participants from eight studies, the Framingham Heart Study (FHS; n=6,735), the Rotterdam study (n=1,958), the Three City Study (3C; n=1,954), the Atherosclerosis Risk in Communities Study (ARIC; n=830), the Washington Heights-Inwood Community Aging Project (WHICAP; n=190), the Epidemiological Prevention study Zoetermeer (EPOZ; n=397), the Alzheimer's Disease Neuroimaging Initiative (ADNI; n=173) and the Erasmus Rucphen Family Study (ERF; n=129). In each study, we excluded participants with prevalent dementia at the time of blood sampling used for plasma Aβ assessment (see **Supplementary Methods 1** for a detailed description of each study and **Supplementary Table 1** for baseline characteristics of the study populations).

Plasma Aβ assessment

Each study used different protocols for blood sampling, plasma extraction and storage and plasma Aβ assessment that have been detailed in previous publications. [24,25,27,29-31] In the FHS, Rotterdam and 3C study, plasma Aβ levels were measured at different times because of cost considerations. Various assays were used to quantify plasma Aβ1-40 and Aβ1-42 levels (see **Supplementary Methods 2** for a detailed description of the protocols used in each study).

Genotyping

Each study used different genotyping platforms as previously published.[13] After applying pre-imputation variant and sample filters, genotypes were imputed using the 1000 Genomes phase 1 version 3 (all ethnicities) imputation panel and various imputation pipelines (see **Supplementary Methods 3**). *APOE*ε genotyping was performed as part of protocols specific to each study (see **Supplementary Methods 4**).

Statistical analyses

Plasma Aβ levels

Plasma Aβ levels were expressed as pg/mL. In each study and for each Aβ dosage, we excluded values that were over or below 4 standard deviations around the mean. To study the variations of plasma Aβ levels in a consistent way across studies, we performed a ranked-based inverse normal transformation of plasma Aβ levels in each study. If they were significantly associated with plasma Aβ levels, this transformation was performed after adjusting for batch effect and other technical artifacts.

Genome-wide association studies

Each study performed genome-wide association studies of plasma Aβ1-40 and Aβ1-42 levels and Aβ1-42/Aβ1-40 ratio using 1000 Genomes imputed data. According to the imputation pipelines used, genetic information was available either as allele dosages or genotype probabilities. In each study, we excluded results from variants that had low imputation quality (r2 or info score < 0.3), variants with low frequency (minor allele frequency < 0.005 or minor allele count < 7) and variants that were available in small number of participant (n < 30). Association of genetic variations with plasma Aβ levels were assessed in linear regression models adjusted for sex and age at blood collection. If significantly associated with plasma Aβ levels, principal components were added in the models to account for population structure.

Genome-wide meta-analysis

Before meta-analysis, we applied a series of filters and quality check that were previously published (see Supplementary Figures 1 and 2).[32] We performed an inverse variance weighted genome-wide meta-analysis, accounting for genomic inflation factors using the METAL software.[33] Finally, we retained variants that had been meta-analyzed at least in the 3 largest available populations (FHS (n=6,735), Rotterdam Study (RS; n=1,958) and Three City Study (3C; n=1,954)). Statistical significance was defined as a p-value below 5×10-8. Signals with p-values between 1×10-5 and 5×10-8 were considered suggestive. Additional graphes and analyses were done using R v3.4.1.[34]

Confirmation of the APOEε4 signal

To confirm the APOE signal we obtained in our genome-wide meta-analysis, we reran our analysis using genotyped *APOE*ε4 and *APOE*ε2 status, adjusting for age and sex.

Annotation

Variant information were retrieved using the Feb 2009 (grch37) assembly of the human genome and dbSNP v147 in the UCSC Table Browser web tool (https://genome.ucsc.edu/cgi-bin/hgTables accessed on 2017-07-25)and the CADD database version 1.3 (http://cadd.gs.washington.edu/download accessed on 2017-07-26).[35] We considered that variants that were less than 250kb apart from one another belonged to the same locus. We then used the Ensembl Variant Effect Predictor (VEP, http://grch37.ensembl.org/info/docs/tools/vep/index.html accessed on 2017-07-25)[36] to relate those variants to nearby genes of potential interest. We also searched if those variants were also eQTL for nearby genes using data from the Genotype-Tissue Expression (GTEx) project (https://gtexportal.org/home/)[37] using the Ensembl REST API (http://rest.ensembl.org/documentation/info/species_variant, data retrieved on 2017-07-31). We corrected for multiple testing by computing False Discovery Rate (FDR), using an FDR threshold of 0.05. Finally, we cross-checked our results with previously published GWAS of AD[13] amyloid-related brain pathology,[20] and CSF Aβ42 levels.[21]

LD-score regression

To gain further insight, we performed an LD-score regression analysis of plasma Aβ1-42 and Aβ1-40 levels using previously published GWAS of AD, [13] hippocampal volume,[38] temporal lobe volume (unpublished data), parietal lobe volume (unpublished data), white matter lesions,[39] cognition[40] and CSF Aβ42 levels. .[21]

Pathway over-representation analysis

We used the ConsensusPathDB-human website (http://cpdb.molgen.mpg.de/ accessed on 2017-08-01)[41] and a curated list of genes related to Aβ[42] to check whether the genes we annotated using GTEx were over-represented in biochemical pathways (see Supplementary Table 2 for the complete list of genes). In order to detect new pathways, we performed a second pathway analysis after excluding genes from loci known for

their involvement in Aβ metabolism, namely *APOE* (including *APOE*, *PVRL2*, *NKPD1*, *CTB-129P6.4*, *APOC1 and VASP*), *APP* (including APP, *AP000230.1* and *AP001596.6*), *PSEN2* (including *PSEN2* and *ADCK3*) and *BACE1* (including *BACE1*, *SIDT2*, *TAGLN*, *RP11-109L13.1*, *RNF214* and *CEP164*). Statistical significance was assessed using hypergeometric tests and corrected for multiple testing using Q-value.

Results

Genome-wide significant variants associated with plasma Aβ levels

After applying quality control procedures in each study (**Supplementary Figures 1 and 2**), we performed a meta-analysis of GWAS and identified 21 variants reaching genome-wide significance across two loci (**Supplementary Figures 3 to 8**). The first locus was located on chromosome 19, in the APOE gene, with significant associations with plasma Aβ1-42 levels and plasma Aβ1-42/Aβ1-40 ratio (**Figures 1 and 2**). For both associations, the most significant variant was rs429358, with p-values of 9.01×10^{-13} and 6.46×10^{-20} for Aβ1-42 levels and Aβ1-42/Aβ1-40 ratio, respectively (**Table 1**).

Table 1. Associations of top variants associated within plasma Aβ and their association with amyloid-related traits.

	rs650585 (chr11:117110740, T/C, intron, *RNF214/BACE1*)					rs429358 (chr19:45411941, C/T, missense, *APOE*)				
	EAF	**Effect†**	**SE‡**	***p*-value**	**I2**	**EAF**	**Effect†**	**SE‡**	***p*-value**	**I2**
Plasma										
Aβ1-40	41.3%	-0.073	0.013	2.56×10^{-8}	3.1%	13.4%	0.023	0.023	0.311	23.7%
Aβ1-42	41.3%	-0.035	0.013	9.57×10^{-3}	27.8%	13.4%	-0.167	0.023	9.01×10-13	32.3%
Aβ1-42/Aβ1-40 ratio	41.4%	0.033	0.013	1.39×10^{-2}	0.0%	13.4%	-0.212	0.023	6.46×10^{-20}	52.6%
Cerebrospinal Fluid [a]										
Aβ42		0.003	0.005	0.543			-0.135	0.007	1.99×10^{-81}	
Brain Pathology [b]										
Neuritic Plaques		0.013		0.597			0.387		2.95×10^{-24}	
Diffuse Plaques		0.009		0.691			0.328		1.30×10^{-20}	
Neurofibrillary Tangles		-0.001		0.981			0.270		4.05×10^{-20}	
AD risk		1.04 [95%CI 1.01-1.08]		8.09×10^{-3}			3.86 [95%CI 3.66-4.07]		$<10^{-500}$	

Abbreviations. EAF: effect allele frequency; AD: Alzheimer's Disease

Notes. †: For AD, "Effect" is represented by Odds Ratio; ‡: For AD, "Standard Error" is represented by 95% confidence interval; For the other measures, "Effect" represents the mean variation of the standardized variable (i.e. transformed so that mean=0 and standard deviation=1). In each column block, the rs ID of the top SNP is followed by its GRCh37 position, Effect/Non-effect alleles, functional category, CADD score and closest genes. [a]: results obtained from Deming *et al.*[21]; [b]: results obtained from Shulman *et al.*[20] : results obtained from Lambert *et al.* [43]

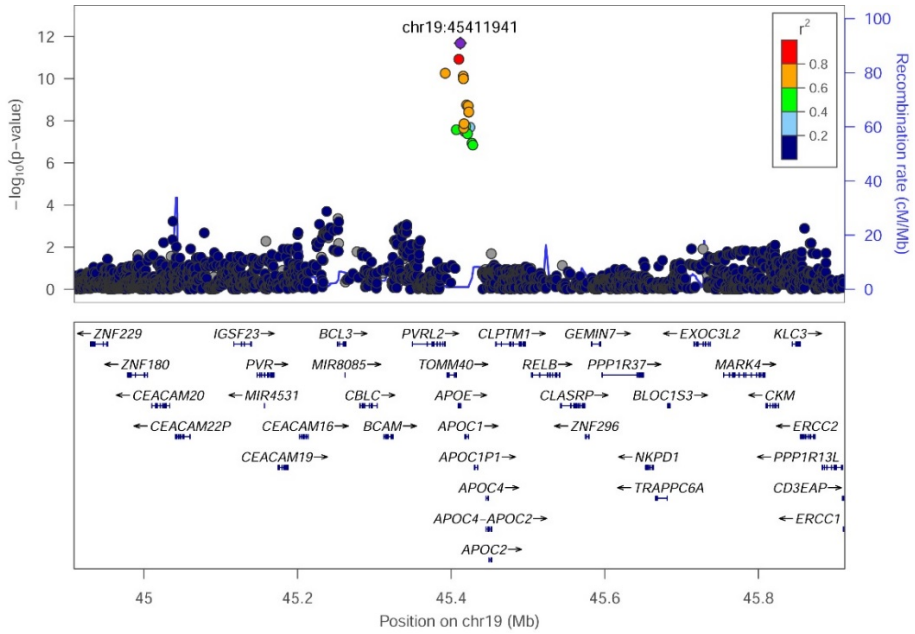


Figure 1. Association of frequent genetic variants with plasma Aβ1-42 at the *APOE* locus

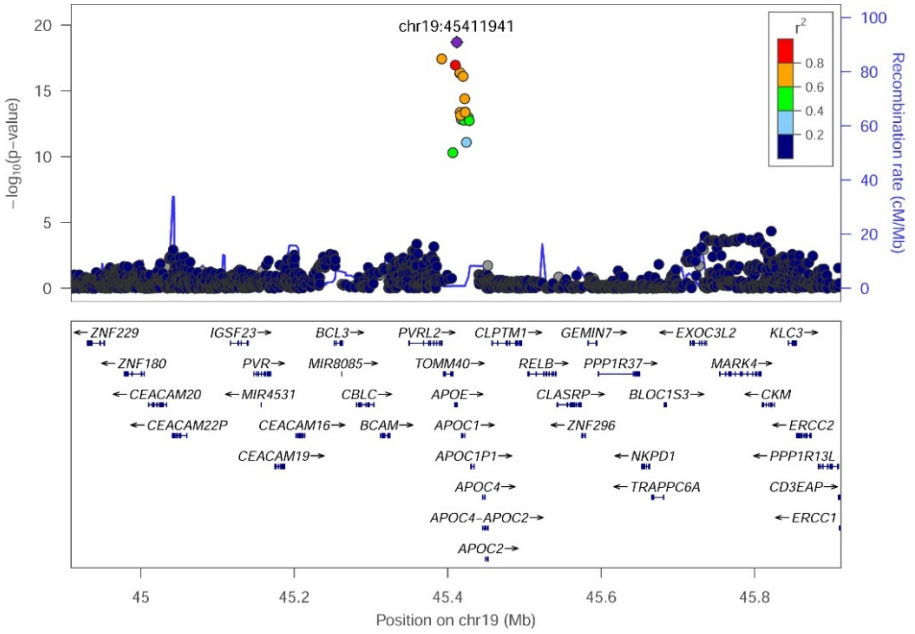


Figure 2. Association of frequent genetic variants with plasma Aβ1-42/Aβ1-40 ratio at the *APOE* locus.

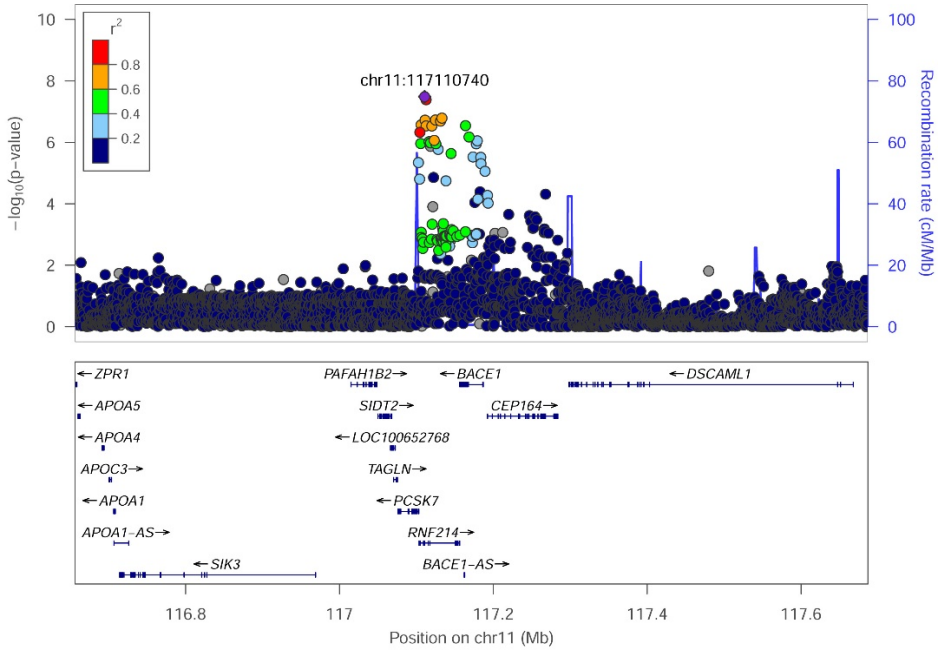


Figure 3. Association of frequent genetic variants with plasma Aβ1-40 at the *BACEI* locus.

5

GWAS suggestive hits

Besides genome-wide significant signals at the *APOE* and *BACEI* loci, signals reaching suggestive levels of association ($p<1 \times 10^{-5}$) for at least one of the three plasma Aβ measures were identified for 240 variants across 73 loci.

Interestingly, we found suggestive levels of association spread across three peaks within and nearby *APP*, one of the core genes of amyloid metabolism (**Figure 4**). Two independent variants located within APP were suggestively associated with plasma Aβ1-40 levels: rs150707803 (effect size=-0.184 SD; p-value=2.10×10^{-6}) and rs436011 (effect size=0.061 SD; p-value=3.92×10^{-6}) (**Figure 4A** and **Table 2**). SNPs within this second locus, including rs436011, were associated with APP expression in the "Esophagus – Muscularis" (FDR=6.29×10^{-7} for rs436011; **Supplementary Table 2**). The top variant of the third locus, rs199744263, was located upstream of *APP* (Figure 4B).

Table 2. Associations of top SNPs from the APP locus with plasma Aβ levels and amyloid-related traits

	rs150707803 (chr21:27272609, T/C, intron, CADD=8.85; *APP*)				
	EAF	**Effect†**	**Standard Error‡**	**P-value**	**I2**
Plasma					
Aβ1-40	3.2%	-0.184	0.039	$2.10x10^{-6}$	22.1%
Aβ1-42	3.2%	-0.078	0.040	$4.86x10^{-2}$	0.0%
Aβ1-42/Aβ1-40 ratio	3.2%	0.088	0.039	$2.48x10^{-2}$	0.0%
Cerebrospinal Fluid[a]					
Aβ42		0.020	0.015	0.193	
Brain Pathology[b]					
Neuritic Plaques		0.177		$4.79x10^{-2}$	
Diffuse Plaques		0.053		0.514	
Neurofibrillary Tangles		0.100		0.148	
AD risk[c]	3.7%	0.95	[0.87 ; 1.03]	0.217	26.4%
	rs436011 (chr21:27527690, A/C, intron, CADD=3.93; *APP*)				
	EAF	**Effect†**	**Standard Error‡**	**P-value**	**I2**
Plasma					
Aβ1-40	33.9%	0.061	0.013	$3.92x10^{-6}$	6.9%
Aβ1-42	33.9%	0.026	0.014	$5.54x10^{-2}$	30.4%
Aβ1-42/Aβ1-40 ratio	33.9%	-0.032	0.013	$1.88x10^{-2}$	0.0%
Cerebrospinal Fluid[a]					
Aβ42		0.010	0.005	$6.60x10^{-2}$	
Brain Pathology[b]					
Neuritic Plaques		-0.031		0.218	
Diffuse Plaques		-0.020		0.389	
Neurofibrillary Tangles		-0.030		0.119	
AD risk[c]	33.3%	0.97	[0.94 ; 1.01]	0.106	0.0%
	rs199744263 (chr21:27663374, D/I, unknown, CADD=1.08; AP000230.1/AP001595.1)				
	EAF	**Effect†**	**Standard Error‡**	**P-value**	**I2**
Plasma					
Aβ1-40	20.8%	0.034	0.016	$3.05x10^{-2}$	25.6%
Aβ1-42	20.8%	-0.037	0.016	$2.00x10^{-2}$	0.0%
Aβ1-42/Aβ1-40 ratio	20.8%	-0.075	0.016	$2.43x10^{-6}$	0.0%

Abbreviations. EAF: effect allele frequency; AD: Alzheimer's Disease

Notes. †: For AD, "Effect" is represented by Odds Ratio; ‡: For AD, "Standard Error" is represented by 95% confidence interval; For the other measures, "Effect" represents the mean variation of the standardized variable (i.e. transformed so that mean=0 and standard deviation=1). In each column block, the rs ID of the top SNP is followed by its GRCh37 position, Effect/Non-effect alleles, functional category, CADD score and closest genes. [a]: results obtained from Deming *et al.*[21]; [b]: results obtained from Shulman *et al.*[20] : results obtained from Lambert *et al.* [43]. For rs199744263 look-up results were not available.

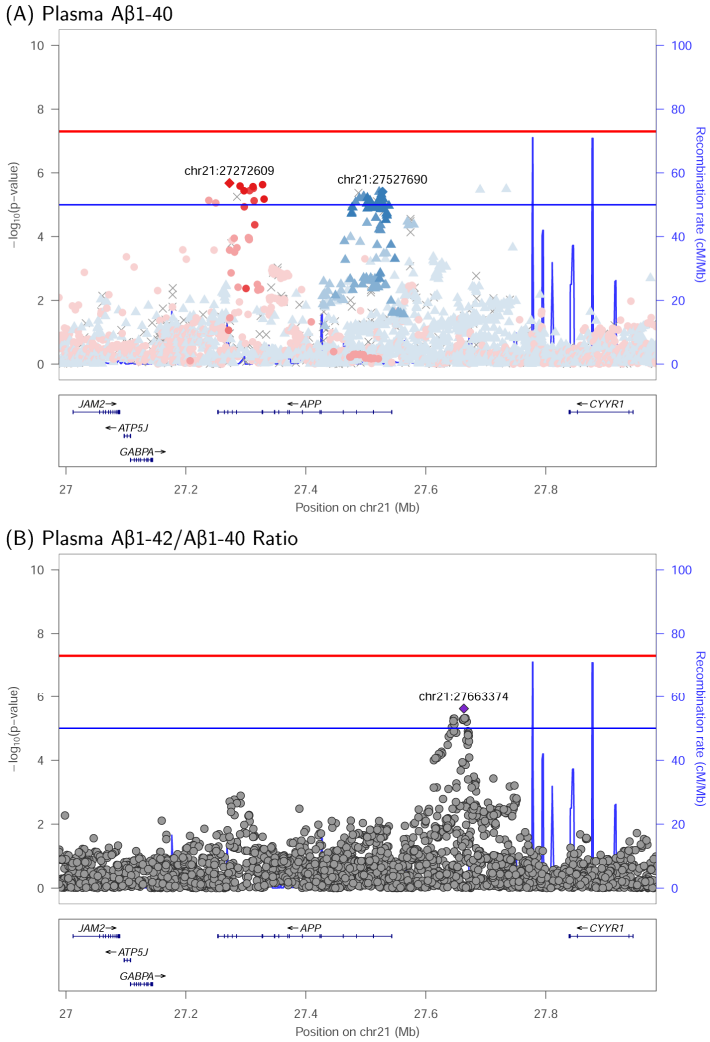


Figure 4. Association of frequent genetic variants with plasma Aβ1-40 (A) and Aβ1-42/Aβ1-40 ratio (B) in the *APP* locus

This variant was associated with lower Aβ1-42/Aβ1-40 ratio (effect size=-0.075 SD; p-value=2.43×10^{-6}; **Table 2**). SNPs from this locus were associated with brain expression of *AP000230.1*, a lincRNA located directly upstream of *APP* (e.g. for rs199744263, FDR=8.33×10^{-5} for cerebellar hemisphere expression; **Supplementary Table 2**).

In addition to *BACE1* and *APP*, we explored genetic associations around other genes closely involved with Aβ production, namely *PSEN1* and *PSEN2* which are part of the

γ-secretase complex and *ADAM10*, which encode for the α-secretase, and is involved in a competing, non-Aβ producing, processing of APP. There was a suggestive association with plasma Aβ1-40 at the *PSEN2* locus (**Table 3**, **Supplementary Figure 11**). In this locus, the top variant, rs2246221 (effect size= 0.057 SD; p-value=8.09×10^{-6}) was also associated with PSEN2 expression in spleen (FDR=2.35×10^{-6}), thyroid (FDR=1.77×10^{-5}), lung (FDR=8.08×10^{-5}), skin (FDR=1.76×10^{-2}) and transformed fibroblasts (FDR=6.25×10^{-23}; **Supplementary Table 2**). Finally, we found a suggestive association with plasma Aβ1-40 within *RGS6* which is located approximately 600kb upstream of *PSEN1* but did not find any evidence of eQTL linking the two loci (**Table 3**, **Supplementary Figure 12**).

To minimize risk of false positive signal among the remaining suggestive loci, we prioritized 12 loci containing exonic variants or variants with a CADD score higher than 10 (**Supplementary Table 3**). Among these variants, rs111123523 (CADD=12.2), located near the TMEM37 gene, showed the lowest p-value. This variant was associated with plasma Aβ1-40 levels (effect size=-0.074; p-value=2.80×10^{-7}) and was associated with expression of a nearby gene, *SCTR*, in thyroid (FDR=3.34×10^{-10}), testis (FDR=9.45×10^{-3}) and tibial nerve (FDR=1.63×10^{-2}). The only exonic variant was rs704, a missense variant of the VTN gene associated with plasma Aβ1-40 levels (effect size=-0.060; p-value=5.81×10^{-6}).

Genetic overlap with other Aβ-related traits and diseases

In order to put those genome-wide significant variants in the context of amyloid-related pathophysiology of AD, we compared them with results obtained from CSF Aβ42, AD brain pathology and AD risk in GWAS (**Table 1**). The *APOEε4* allele was also associated with lower CSF Aβ42 levels, higher brain levels of neuritic plaques, diffuse plaques and neurofibrillary tangles and higher risk of AD. For the *BACE1* locus, we did not find genome-wide significant or suggestive associations of rs650585 with any of the aforementioned traits. Among suggestive variants, no genome-wide significant or suggestive association was found for any of the other amyloid-related traits or for AD risk.

Table 3. Associations of top SNPs from the PSEN1 and PSEN2 locus with plasma Aβ levels and amyloid-related traits.

	rs2246221 (chr1:227075402, G/A, intron, CADD=1.67; *PSEN2*)				
	EAF	**Effect†**	**Standard Error‡**	**P-value**	**I2**
Plasma					
Aβ1-40	47.6%	0.057	0.013	8.09×10^{-6}	34.5%
Aβ1-42	47.5%	0.039	0.013	2.75×10^{-3}	0.0%
Aβ1-42/Aβ1-40 ratio	47.5%	-0.020	0.013	0.118	0.0%
Cerebrospinal Fluid[a]					
Aβ42		-0.003	0.007	0.623	
Brain Pathology[b]					
Neuritic Plaques		0.177		4.79×10^{-2}	
Diffuse Plaques		0.053		0.514	
Neurofibrillary Tangles		0.1		0.148	
AD risk[c]	48.1%	0.99	[0.96 ; 1.02]	0.358	32.5%

	rs12589291 (chr14:72840969, T/C, intron, CADD=2.12; *RGS6*)				
	EAF	**Effect†**	**Standard Error‡**	**P-value**	**I2**
Plasma					
Aβ1-40	1.7%	-0.227	0.051	8.69×10-6	0.0%
Aβ1-42	1.7%	-0.028	0.052	0.597	0.0%
Aβ1-42/Aβ1-40 ratio	1.7%	0.178	0.051	5.45×10^{-4}	0.0%
Cerebrospinal Fluid[a]					
Aβ42		-0.022	0.022	0.305	

Abbreviations. EAF: effect allele frequency; AD: Alzheimer's Disease

Notes. †: For AD, "Effect" is represented by Odds Ratio; ‡: For AD, "Standard Error" is represented by 95% confidence interval; For the other measures, "Effect" represents the mean variation of the standardized variable (i.e. transformed so that mean=0 and standard deviation=1). In each column block, the rs ID of the top SNP is followed by its GRCh37 position, Effect/Non-effect alleles, functional category, CADD score and closest genes. [a]: results obtained from Deming *et al.*[21]; [b]: results obtained from Shulman *et al.*[20] : results obtained from Lambert *et al.* [43]. For rs12589291 look-up results were not available.

To improve statistical power, we performed a genetic correlation analysis of GWAS results of multiple traits using LD regression. On a genome-wide scale (**Figure 5**), we observed a nominally significant negative genetic correlation between variants modulating plasma Aβ1-40 levels and hippocampal volume (r=-0.68, *p*=0.04), and a positive correlation between variants modulating plasma Aβ1-42 levels and white matter lesions (r=0.48, *p*=0.04).

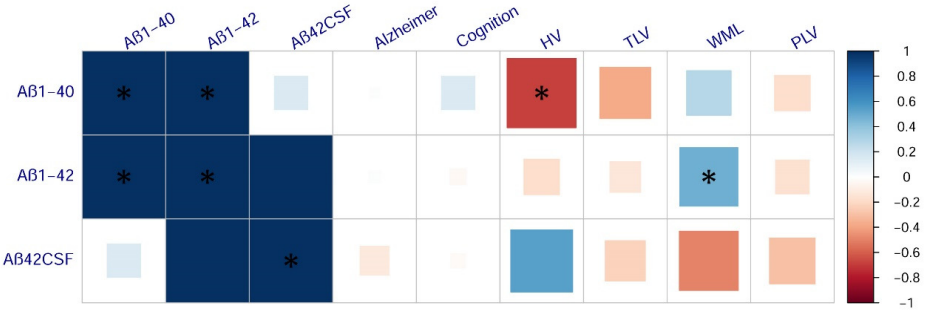


Figure 5. LD-score regression assessing genetic overlap between plasma Aβ levels, hippocampal, temporal and parietal brain volumes, white matter lesions, cognition and CSF Aβ42 levels

Pathway over-representation analysis

After annotation using GTEx, we obtained a list of 41 genes located nearby our suggestive and significant signals that we used to perform a pathway over-representation analysis. Using ConsensusPathDB-human and a curated list of genes related to Aβ as references, we found significant over-represention in several biochemical pathways, related to Alzheimer's disease (ConsensusPathDB Q-value=5.73×10^{-4}) and generation of Aβ (curated list p-value=4.18×10^{-3}; ConsensusPathDB Q-value=4.44×10^{-4}; Supplementary Table 4). These results were driven by genes previously known for their involvement in Aβ metabolism (*APOE*, *BACE1*, *APP*, *PSEN2*) as the AD and Aβ pathways were no longer significant after removing these genes from the analysis (**Supplementary Table 5**).

Discussion

Previous genome-wide association studies of plasma Aβ40 and Aβ42 levels have failed to uncover genome-wide significant findings. In this study, we identified two genome-wide significant loci associated with plasma Aβ levels in up to 11,969 non-demented subjects of European ancestry. The top variant in the first locus, rs429358, a well-known non-synonymous variant that encodes for the APOE4 isoform, was associated with lower circulating Aβ42 levels and Aβ42/40 ratio. In the second, located near *BACE1*, rs650585 was associated with lower plasma Aβ40 levels.

The *BACE1* region encompasses several genes (*PCSK7*, *RNF214*, *BACE1*, *CEP164*) and a *BACE1* anti-sense long non-coding RNA (*BACE1-AS*). Since the β-secretase activity of BACE1 is necessary for Aβ peptide production, it is likely that *BACE1* or a local regulation of *BACE1* expression are responsible for this signal. We also found suggestive associations with plasma Aβ40 levels near *APP* and *PSEN2*, two major actors of the Aβ metabolism. APP is obviously a central element of its own metabolism and PSEN2 is a key component of the γ-secretase which processes the APP C99 fragment into Aβ peptides.[1] We speculate that the effect of the variants on the expression/biological activation of these key elements of β-amyloid processing is strong enough to allow their detection at the plasma level, despite the influence of many other simultaneous biological processes, e.g. secretion, interaction with other proteins, degradation and/or clearance. Moreover, the top variants at the *PSEN2* and *BACE1* locus were also nominally associated with Aβ42 levels in the same direction as Aβ40 levels, which is in agreement with knowledge that *PSEN2* and *BACE1* activities indifferently produce Aβ40 and Aβ42 peptides.

Conversely, the *APOE* ε4 allele had the strongest association with Aβ42 levels but was not even nominally associated with Aβ40. In line with our previous comment, this suggests that the *APOE*4 isoform is not involved in the early process of Aβ peptide production but in more downstream events, such as Aβ aggregation or clearance. These results might also illustrate the greater ability to aggregate of Aβ42 peptides compared to Aβ40, and the influence of APOE isoforms in the regulation of this process.[8] Interestingly, associations of *APOE*ε2 with plasma Aβ levels were not significant and effect sizes were very small. Contrary to *APOE*ε4, the effect of *APOE* ε2 on amyloid markers has been much less studied and seems to be focused on specific brain regions, which could explain why we could not detect any association.[44] This could also suggest that other, Aβ-independent, mechanisms are involved in the lower risk of AD observed in *APOE*ε2 carriers.

Given the presence of suggestive associations around known APP/Aβ genes, it is likely that novel and relevant signals exist within this range of statistical significance, awaiting

definite validation. In order to minimize the risk of false positive result, we used exonic location and CADD score as additional filters to screen for SNPs of interest. Among those, we observed an association between a missense mutation of the VTN gene and lower plasma Aβ40 levels. *VTN* encodes for vitronectin, a glycoprotein present in abundance in the plasma and the extracellular matrix and involved in early regulation of thrombogenesis and tissue repair.[45] Vitronectin has been associated with amyloid deposits,[46] including Aβ, both at the level of amyloid plaques in the brain[47] and near the retinal pigment epithelium in the aging eye.[48] Modest correlation between plasma vitronectin and brain amyloid burden measured by PiB PET has been reported.[49] Vitronectin is involved in microglial activation,[50] which is relevant to AD pathophysiology.[51] Vitronectin might also be involved in small vessel disease. In a mouse model of CADASIL, an autosomal dominant disease responsible for stroke and cognitive impairment, reduction of vitronectin expression resulted in less white matter lesions.[52] Vitronectin could therefore represent a promising candidate to study mechanisms linking Aβ peptides with Aβ-related pathologies, such as AD, cerebral amyloid angiopathy and small vessel disease. Other potential genes of interest close to suggestive signals warrant further investigation. For example, *SCTR*, encoding for secretin receptor, is involved in a wide range of physiological functions, beyond the scope of this study. Interestingly though, a study reported that mice deficient for that gene display impaired hippocampal synaptic plasticity.[53]

Although research on variants associated with levels of circulating amyloid peptides is of general interest for Aβ physiology, we are interpreting the findings primarily from the perspective of brain disorders, especially AD. When cross-checking our results with other published GWAS of CSF and brain Aβ, and AD, we observed consistent results only with the *APOE*ε4 variant. As expected, this variant was associated with low plasma and CSF Aβ42, high brain Aβ and AD risk, consistent with a differential effect of APOE isoforms on Aβ aggregation and clearance from the brain. When interpreting the absence of significant association between variants in the other loci and those same traits, one should keep in mind that their effect on plasma Aβ levels was generally smaller compared to that of *APOE*ε4 so we might be underpowered to detect a

significant association. It is also of interest that the LD-score regression analysis suggested a positive correlation of variants modulating plasma Aβ1-42 levels and white matter lesions and a negative correlation between plasma Aβ40 and hippocampal volume. Nevertheless, these genetic correlations were only nominally significant and await replication.

Plasma Aβ is usually considered as a poor biomarker of AD in the literature. A previous meta-analysis reported that plasma Aβ levels were not useful to make a clinical diagnosis of AD.[54] Nevertheless many of the cohorts participating in the present study have previously reported that low plasma Aβ42 and Aβ42/40 ratio levels were associated with development of AD after several years of follow-up.[24-27] These results are consistent with an early, preclinical, involvement of Aβ in AD pathophysiology and are strengthened by our present observation that APOEε4, is both associated with low plasma Aβ42 and Aβ42/40 ratio and high AD risk. Some of those studies have also reported that this association remained significant after adjusting for *APOE*ε4,[27] and we might hypothesize that variations of plasma Aβ levels are not only a side-effect of *APOE*ε4, but are also involved in AD pathophysiology. As such, plasma Aβ levels would not be only useful as a biomarker of an active amyloid process in the brain but could also be considered as a therapeutic target. In favor of this hypothesis are reports that hemodialysis or peritoneal dialysis are able to lower Aβ in the brain.[55,56] The association we observed between variants near BACE1 and plasma Aβ40 is also of interest in the light of the ongoing trials testing BACE inhibitors, even though the lack of association of these variants with AD risk should be further investigated.[9]

Our study has several strengths. First, it is, to date, the largest study of circulating amyloid peptides. This enabled us to identify known actors of Aβ metabolism and, thus, to be optimistic about the relevance of some of our suggestive signals. Second, this study was conducted in non-demented participants and therefore is relevant for the study of early amyloid pathophysiological processes. Third, we carefully normalized the plasma Aβ data before running GWAS, thus taking into account some of the heterogeneity that has been described when using plasma Aβ levels.

Our study has also limitations worth mentioning. As stated before, the state of current knowledge makes it hard for us to extrapolate the role of these actors from the plasma compartments to the brain and further research in this area is needed. Second, most studies used assays that only measured free monomeric forms of Aβ. Therefore, our interpretation of the present results might differ from other studies in which assays used non-selectively measured both monomers and oligomers of Aβ.[57] Future studies should carefully choose assays that allow measurements of each form of Aβ as this will facilitate interpretation with regard to the balance between Aβ production, aggregation and clearance. Finally we tried to prioritize signals of interest using strict criteria, thus omitting to mention other interesting signals that might be real. Therefore we hope that the unfiltered results from this study, will be helpful as a resource to the scientific community to further decipher the physiology of Aβ peptides and its links to pathophysiology of AD and other Aβ-related diseases.

In summary, our results indicate that genetic determinants of plasma Aβ40 and Aβ42 levels are close to genes known to be central actors in APP metabolism in AD. Further increasing the statistical power of plasma Aβ analyses may potentially lead us to the identification of currently unknown players in Aβ metabolism, novel hypotheses and hopefully, new preventive or therapeutic targets against Alzheimer's disease.

References

1. Haass C, Kaether C, Thinakaran G, Sisodia S. Trafficking and Proteolytic Processing of APP. *Csh Perspect Med* 2012; **2**(5).
2. Chen M, Inestrosa NC, Ross GS, Fernandez HL. Platelets are the primary source of amyloid beta-peptide in human blood. *Biochem Biophys Res Commun* 1995; **213**(1): 96-103.
3. Hyman BT, Phelps CH, Beach TG, et al. National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. *Alzheimers Dement* 2012; **8**(1): 1-13.
4. Viswanathan A, Greenberg SM. Cerebral amyloid angiopathy in the elderly. *Ann Neurol* 2011; **70**(6): 871-80.
5. Selkoe DJ, Hardy J. The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO Mol Med* 2016; **8**(6): 595-608.
6. Cacace R, Sleegers K, Van Broeckhoven C. Molecular genetics of early-onset Alzheimer's disease revisited. *Alzheimers & Dementia* 2016; **12**(6): 733-48.
7. Genin E, Hannequin D, Wallon D, et al. APOE and Alzheimer disease: a major gene with semi-dominant inheritance. *Mol Psychiatry* 2011; **16**(9): 903-7.

8. Kanekiyo T, Xu H, Bu G. ApoE and Abeta in Alzheimer's disease: accidental encounters or partners? *Neuron* 2014; **81**(4): 740-54.
9. Mullard A. BACE inhibitor bust in Alzheimer trial. *Nat Rev Drug Discov* 2017; **16**(3): 155.
10. Karran E, Hardy J. A critique of the drug discovery and phase 3 clinical programs targeting the amyloid hypothesis for Alzheimer disease. *Ann Neurol* 2014; **76**(2): 185-205.
11. Bateman RJ, Xiong CJ, Benzinger TLS, et al. Clinical and Biomarker Changes in Dominantly Inherited Alzheimer's Disease. *New Engl J Med* 2012; **367**(9): 795-804.
12. Villemagne VL, Burnham S, Bourgeat P, et al. Amyloid beta deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. *Lancet Neurol* 2013; **12**(4): 357-67.
13. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 2013; **45**(12): 1452-8.
14. Bali J, Gheinani AH, Zurbriggen S, Rajendran L. Role of genes linked to sporadic Alzheimer's disease risk in the production of beta-amyloid peptides. *Proc Natl Acad Sci U S A* 2012; **109**(38): 15307-11.
15. Chapuis J, Flaig A, Grenier-Boley B, et al. Genome-wide, high-content siRNA screening identifies the Alzheimer's genetic risk factor FERMT2 as a major modulator of APP metabolism. *Acta Neuropathol* 2017; **133**(6): 955-66.
16. International Genomics of Alzheimer's Disease C. Convergent genetic and expression data implicate immunity in Alzheimer's disease. *Alzheimers Dement* 2015; **11**(6): 658-71.
17. Cruchaga C, Kauwe JS, Harari O, et al. GWAS of cerebrospinal fluid tau levels identifies risk variants for Alzheimer's disease. *Neuron* 2013; **78**(2): 256-68.
18. Ramanan VK, Risacher SL, Nho K, et al. APOE and BCHE as modulators of cerebral amyloid deposition: a florbetapir PET genome-wide association study. *Mol Psychiatry* 2014; **19**(3): 351-7.
19. Beecham GW, Hamilton K, Naj AC, et al. Genome-Wide Association Meta-analysis of Neuropathologic Features of Alzheimer's Disease and Related Dementias. *PLoS Genet* 2014; **10**(9).
20. Shulman JM, Chen K, Keenan BT, et al. Genetic susceptibility for Alzheimer disease neuritic plaque pathology. *JAMA Neurol* 2013; **70**(9): 1150-7.
21. Deming Y, Li Z, Kapoor M, et al. Genome-wide association study identifies four novel loci associated with Alzheimer's endophenotypes and disease modifiers. *Acta Neuropathol* 2017; **133**(5): 839-56.
22. Sagare AP, Bell RD, Zlokovic BV. Neurovascular Defects and Faulty Amyloid-beta Vascular Clearance in Alzheimer's Disease. *Journal of Alzheimers Disease* 2013; **33**: S87-S100.
23. Toledo JB, Vanderstichele H, Figurski M, et al. Factors affecting Abeta plasma levels and their utility as biomarkers in ADNI. *Acta Neuropathol* 2011; **122**(4): 401-13.
24. van Oijen M, Hofman A, Soares HD, Koudstaal PJ, Breteler MM. Plasma Abeta(1-40) and Abeta(1-42) and the risk of dementia: a prospective case-cohort study. *Lancet Neurol* 2006; **5**(8): 655-60.
25. Lambert JC, Schraen-Maschke S, Richard F, et al. Association of plasma amyloid beta with risk of dementia The prospective Three-City Study. *Neurology* 2009; **73**(11): 847-53.
26. Shah NS, Vidal JS, Masaki K, et al. Midlife blood pressure, plasma beta-amyloid, and the risk for Alzheimer disease: the Honolulu Asia Aging Study. *Hypertension* 2012; **59**(4): 780-6.
27. Chouraki V, Beiser A, Younkin L, et al. Plasma amyloid-beta and risk of Alzheimer's disease in the Framingham Heart Study. *Alzheimers Dement* 2015; **11**(3): 249-57 e1.
28. Chouraki V, De Bruijn RF, Chapuis J, et al. A genome-wide association meta-analysis of plasma Abeta peptides concentrations in the elderly. *Mol Psychiatry* 2014; **19**(12): 1326-35.
29. Figurski MJ, Waligorska T, Toledo J, et al. Improved protocol for measurement of plasma beta-amyloid in longitudinal evaluation of Alzheimer's Disease Neuroimaging Initiative study patients. *Alzheimers & Dementia* 2012; **8**(4): 250-60.
30. Ibrahim-Verbaas CA, Zorkoltseva IV, Amin N, et al. Linkage analysis for plasma amyloid beta levels in persons with hypertension implicates Abeta-40 levels to presenilin 2. *Hum Genet* 2012; **131**(12): 1869-76.

31. Reitz C, Cheng R, Schupf N, et al. Association between variants in IDE-KIF11-HHEX and plasma amyloid beta levels. *Neurobiol Aging* 2012; **33**(1): 199 e13-7.
32. Winkler TW, Day FR, Croteau-Chonka DC, et al. Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc* 2014; **9**(5): 1192-212.
33. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; **26**(17): 2190-1.
34. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; 2015.
35. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014; **46**(3): 310-5.
36. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biol* 2016; **17**(1): 122.
37. Consortium GT, Ardlie KG, Deluca DS, et al. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science (New York, N Y)* 2015; **348**(6235): 648-60.
38. Hibar DP, Adams HH, Jahanshad N, et al. Novel genetic loci associated with hippocampal volume. *Nat Commun* 2017; **8**: 13624.
39. Fornage M, Debette S, Bis JC, et al. Genome-wide association studies of cerebral white matter lesion burden: the CHARGE consortium. *Ann Neurol* 2011; **69**(6): 928-39.
40. Davies G, Armstrong N, Bis JC, et al. Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53949). *Mol Psychiatry* 2015; **20**(2): 183-92.
41. Herwig R, Hardt C, Lienhard M, Kamburov A. Analyzing and interpreting genome data at the network level with ConsensusPathDB. *Nat Protoc* 2016; **11**(10): 1889-907.
42. Campion D, Pottier C, Nicolas G, Le Guennec K, Rovelet-Lecrux A. Alzheimer disease: modeling an Abeta-centered biological network. *Mol Psychiatry* 2016; **21**(7): 861-71.
43. Lambert J, Ibrahim-Verbaas C, Harold D. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature genetics* 2013; **45**: 1452-8.
44. Grothe MJ, Villeneuve S, Dyrba M, Bartres-Faz D, Wirth M, Alzheimer's Disease Neuroimaging I. Multimodal characterization of older APOE2 carriers reveals selective reduction of amyloid load. *Neurology* 2017; **88**(6): 569-76.
45. Leavesley DI, Kashyap AS, Croll T, et al. Vitronectin--master controller or micromanager? *IUBMB Life* 2013; **65**(10): 807-18.
46. Winter M, Tholey A, Kruger S, Schmidt H, Rocken C. MALDI-mass spectrometry imaging identifies vitronectin as a common constituent of amyloid deposits. *J Histochem Cytochem* 2015; **63**(10): 772-9.
47. Akiyama H, Kawamata T, Dedhar S, McGeer PL. Immunohistochemical localization of vitronectin, its receptor and beta-3 integrin in Alzheimer brain tissue. *J Neuroimmunol* 1991; **32**(1): 19-28.
48. Thompson RB, Reffatto V, Bundy JG, et al. Identification of hydroxyapatite spherules provides new insight into subretinal pigment epithelial deposit formation in the aging eye. *Proc Natl Acad Sci U S A* 2015; **112**(5): 1565-70.
49. Kiddle SJ, Thambisetty M, Simmons A, et al. Plasma based markers of [11C] PiB-PET brain amyloid burden. *PLoS One* 2012; **7**(9): e44260.
50. Milner R, Campbell IL. The extracellular matrix and cytokines regulate microglial integrin expression and activation. *J Immunol* 2003; **170**(7): 3850-8.
51. Heppner FL, Ransohoff RM, Becher B. Immune attack: the role of inflammation in Alzheimer disease. *Nat Rev Neurosci* 2015; **16**(6): 358-72.
52. Capone C, Cognat E, Ghezali L, et al. Reducing Timp3 or vitronectin ameliorates disease manifestations in CADASIL mice. *Ann Neurol* 2016; **79**(3): 387-403.

53. Nishijima I, Yamagata T, Spencer CM, et al. Secretin receptor-deficient mice exhibit impaired synaptic plasticity and social behavior. *Hum Mol Genet* 2006; **15**(21): 3241-50.
54. Olsson B, Lautner R, Andreasson U, et al. CSF and blood biomarkers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis. *Lancet Neurol* 2016; **15**(7): 673-84.
55. Sakai K, Senda T, Hata R, et al. Patients that have Undergone Hemodialysis Exhibit Lower Amyloid Deposition in the Brain: Evidence Supporting a Therapeutic Strategy for Alzheimer's Disease by Removal of Blood Amyloid. *J Alzheimers Dis* 2016; **51**(4): 997-1002.
56. Jin WS, Shen LL, Bu XL, et al. Peritoneal dialysis reduces amyloid-beta plasma levels in humans and attenuates Alzheimer-associated phenotypes in an APP/PSI mouse model. *Acta Neuropathol* 2017.
57. Fullwood NJ, Hayashi Y, Allsop D. Plasma amyloid-beta concentrations in Alzheimer's disease: an alternative hypothesis. *Lancet Neurol* 2006; **5**(12): 1000-1; author reply 2-3.

Supplementary Tables

The supplementary methods and tables of this chapter can be accessed scanning the following code:

Supplementary Table 1. Characteristics of the study populations at baseline.

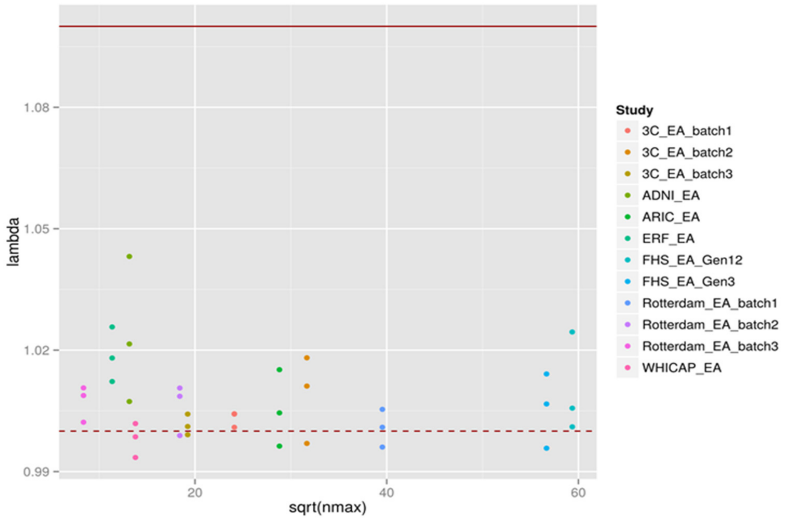
Supplementary Table 2. Association of gene expression with SNPs showing suggestive levels of association with plasma Aβ measures.

Supplementary Table 3. Associations of genome-wide suggestive coding or high-CADD score variants with plasma Aβ levels and amyloid-related traits.

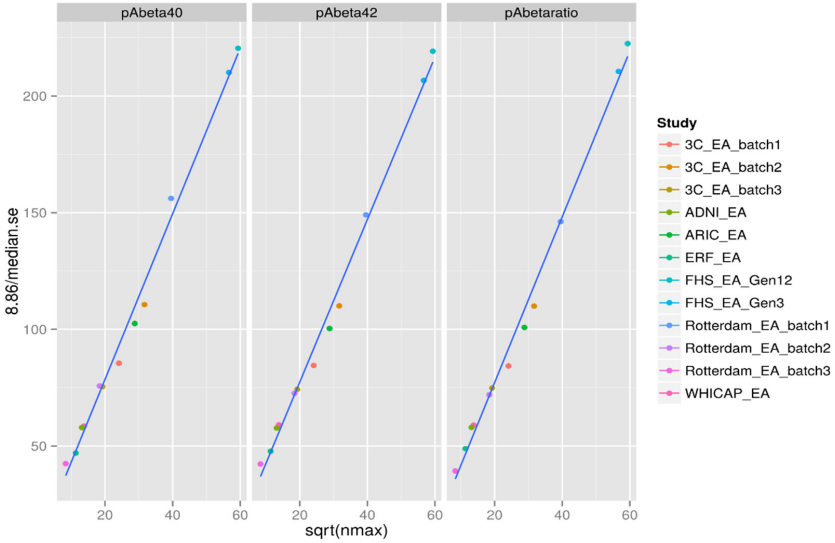
Supplementary Table 4. ConsensusPathDB pathway analysis (full list of genes).

Supplementary Table 5. ConsensusPathDB pathway analysis (after removing genes in the same locus as core Aβ genes from the list of genes).

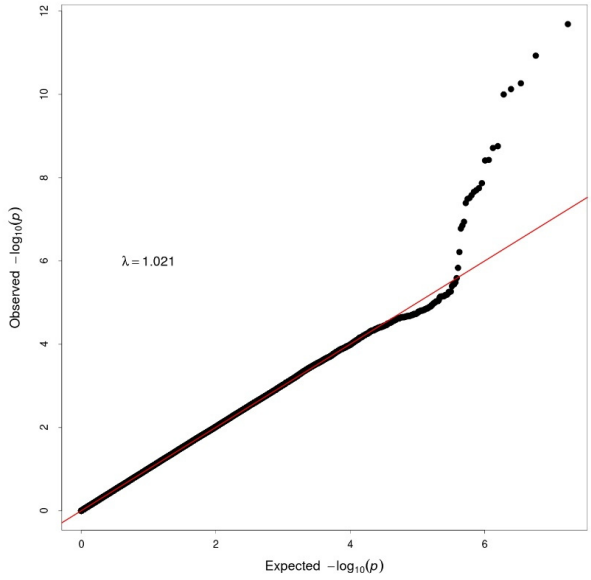
Supplementary Figures



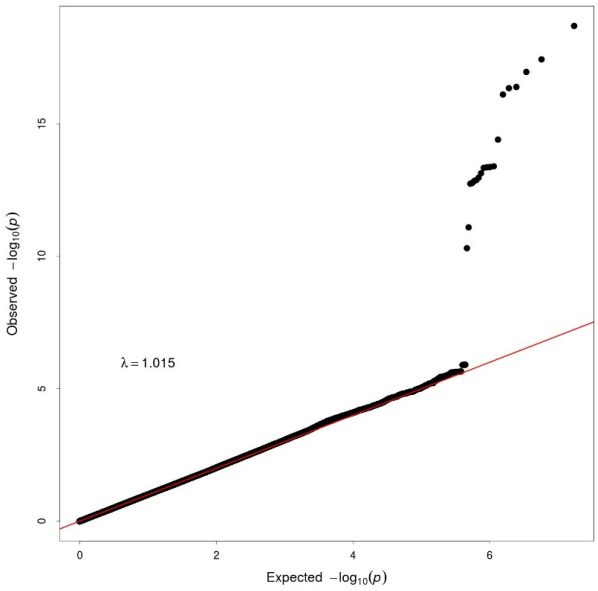
Supplementary Figure 1: Genomic inflation factors (λ) of individual GWAS of plasma Aβ1-40, Aβ1-42 and Aβ1-42/Aβ1-40 ratio according to sample size (adapted from Winkler *et al.*[33])



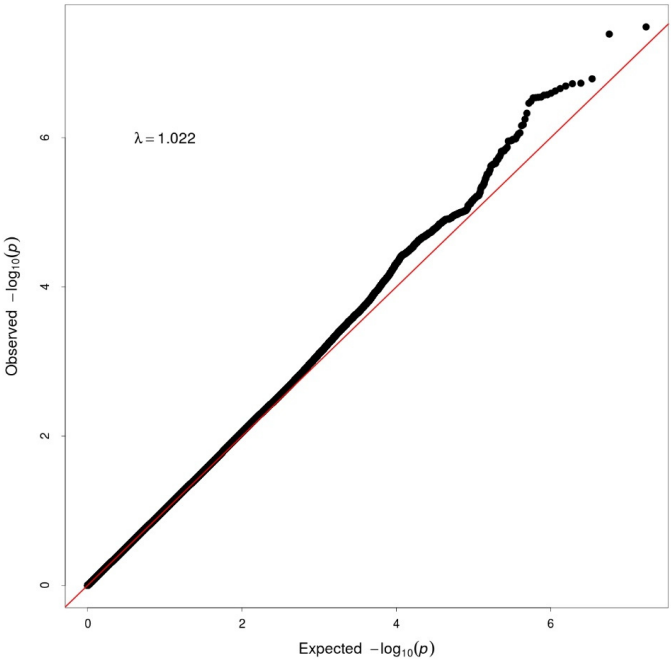
Supplementary Figure 2: Verification of the homogeneity of plasma Aβ measures across studies after inverse-normal transformation (adapted from Winkler *et al.*[33])



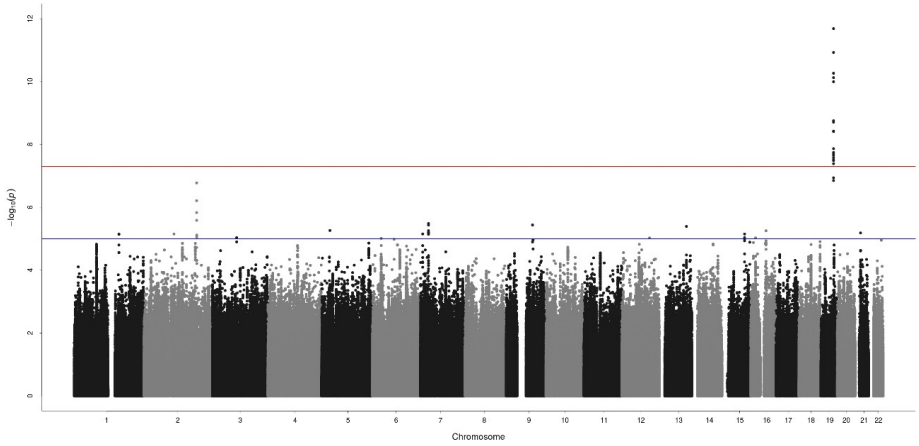
Supplementary Figure 3: Quantile-quantile plot of the genome-wide meta-analysis of plasma Aβ1-42 levels



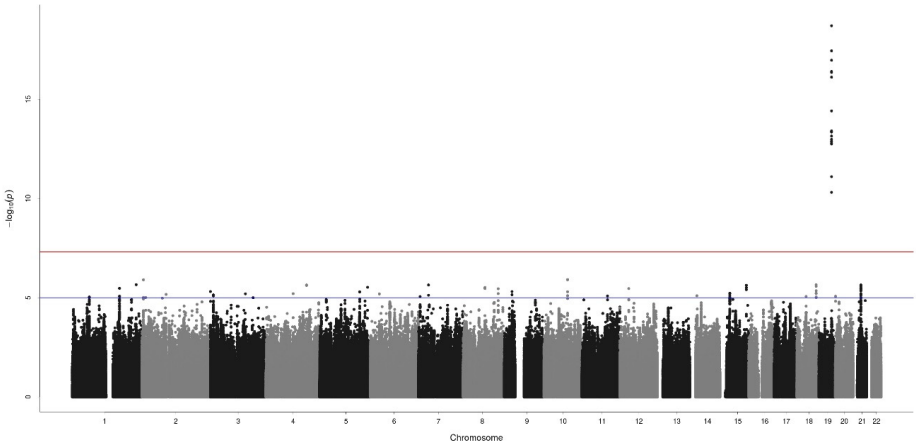
Supplementary Figure 4: Quantile-quantile plot of the genome-wide meta-analysis of plasma Aβ1-42/Aβ1-40 ratio



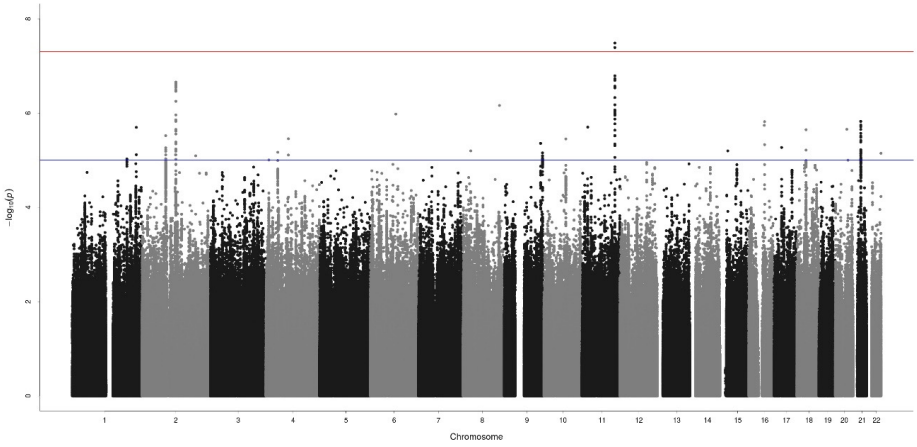
Supplementary Figure 5: Quantile-quantile plot of the genome-wide meta-analysis of plasma Aβ1-40 levels



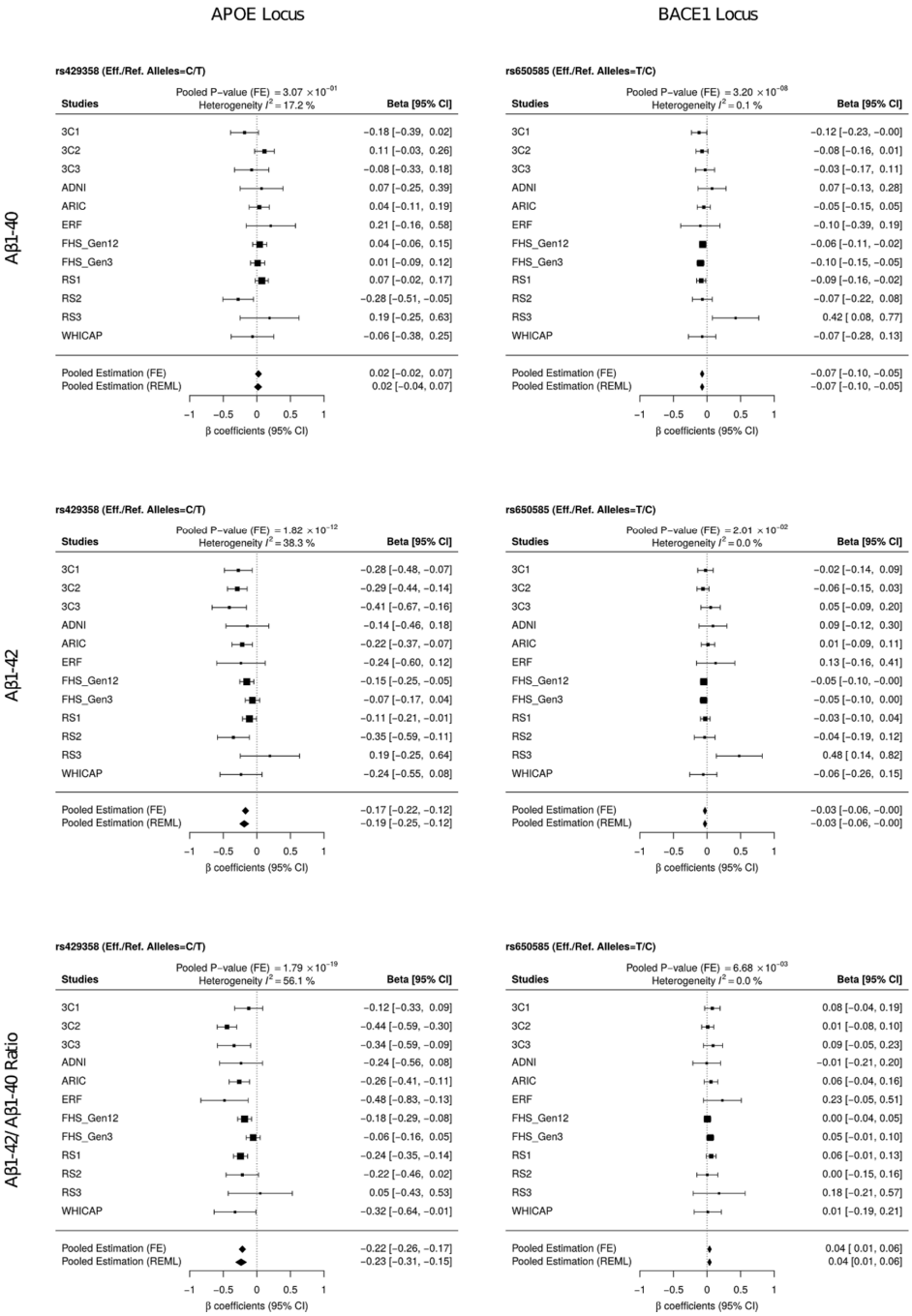
Supplementary Figure 6: Manhattan plot of the genome-wide meta-analysis of plasma Aβ1-42 levels



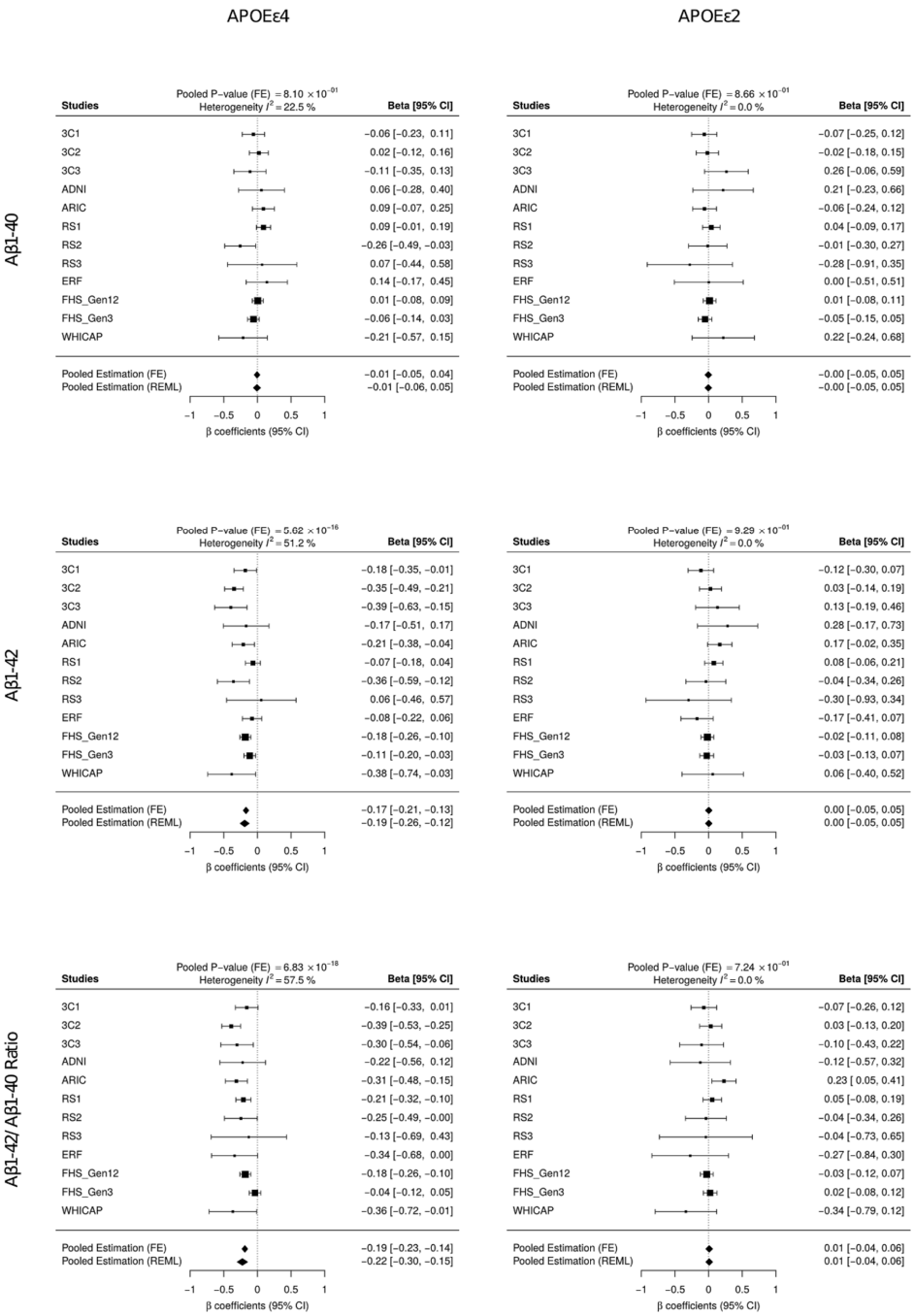
Supplementary Figure 7: Manhattan plot of the genome-wide meta-analysis of plasma Aβ1-42/Aβ1-40 ratio



Supplementary Figure 8: Manhattan plot of the genome-wide meta-analysis of plasma Aβ1-40 levels



Supplementary Figure 9: Association of top hits with plasma Aβ levels.



Supplementary Figure 10: Association of genotyped APOEε alleles with plasma Aβ levels.

Chapter 5.2

Association of branched chain amino acids and other circulating metabolites with risk of incident dementia and Alzheimer's disease: a prospective study in eight cohorts

Juho Tynkkynen*,Vincent Chouraki*, Sven J. van der Lee, Jussi Hernesniemi, Qiong Yang, Shuo Li, Alexa Beiser, Martin G Larson, Katri Sääksjärvi, Martin J Shipley, Archana Singh-Manoux, Robert E Gerszten, Thomas J. Wang, Aki S. Havulinna, Peter Wùrtz, Krista Fischer, Ayse Demirkan, M Arfan Ikram, Najaf Amin, Terho Lehtimäki, Mika Kähönen, Markus Perola, Andres Metspalu, Antti J. Kangas, Pasi Soininen, Mika Ala-Korpela, Ramachandran S Vasan, Mika Kivimäki, Cornelia M. van Duijn, Sudha Seshadri**, Veikko Salomaa**.

*Equal contribution as first authors

**Equal contribution as senior authors

This chapter is submitted

Abstract

Introduction: Metabolite, lipid and lipoprotein lipid profiling can provide novel insights into mechanisms underlying incident dementia and Alzheimer's disease (AD).

Methods: We studied eight prospective cohorts with 22,623 participants profiled by NMR or MS metabolomics. Four cohorts were used for discovery with replication undertaken in the other four to avoid false positives. For metabolites that survived replication, combined association results are presented.

Results: Over 246,698 person-years, 995 and 745 cases of incident dementia and AD were detected, respectively. Three branched-chain amino acids (BCAA) (isoleucine, leucine, valine), creatinine and two VLDL specific lipoprotein lipid subclasses were associated with lower dementia risk. One HDL (L-HDL-CE-%) and one VLDL (XL-VLDL-C-%) lipoprotein lipid subclass was associated with increased dementia risk. BCAAs were also associated with decreased and L-HDL-CE-% with increased AD risk.

Discussion: Further studies can clarify whether these molecules play a causal role in dementia pathogenesis or are merely markers of early pathology.

Introduction

Dementia, including Alzheimer's disease (AD) is a major public health problem with devastating physical, financial and social consequences for patients, their care-givers, families and society. Worldwide the cost of AD care in 2010 was $604 billion or 1% of the global gross domestic product.[1] However, despite over two decades of research on animal models and clinical trials, we still have no effective prevention or disease modifying therapy for late-onset clinical dementia and AD. Dementia is increasingly recognized as a heterogeneous syndrome that would be best addressed with a multi-pronged approach to prevention and treatment, analogous to the multi-pronged and individually tailored use of statins, anti-hypertensives, anti-platelet agents and vasodilators in persons with coronary artery disease. Identifying novel biology could suggest new circulating biomarkers for risk prediction and drug targets. Agnostic approaches such as genome wide genetic analyses have identified new biological pathways and molecules mediating microglial inflammation (TREM2) and endocytosis (BIN1, PICALM) as having a previously unsuspected key role in AD pathophysiology.[2-4]

Blood metabolomics is an attractive tool for agnostic exploration of disease pathways for several reasons. Metabolites are small molecules that reflect the interplay of genetic and environmental factors, readily cross the blood brain barrier and their levels are modifiable through dietary or pharmacological interventions. This recognition has spurred interest in using metabolomics as a tool to understand AD. For example longitudinal studies in mouse models of AD have implicated perturbed polyamine metabolism, disturbances in essential amino acids, branched-chain amino acids, and in the neurotransmitter serotonin along with imbalances in phospholipid and acylcarnitine homeostasis in both the brain and the blood.[5] Human studies in cerebrospinal fluid (CSF) and plasma have to date only compared AD cases to controls in cross-sectional settings or attempted to identify markers predicting conversion from mild cognitive impairment (MCI) to clinical dementia.[6-11] However, in persons with MCI or dementia it is not possible to determine if the observed metabolite changes are causal or secondary to disease related processes. There has only been one prior study of preclinical AD that failed to detect any consistently reproducible signal.[12]

We conducted a prospective study relating blood metabolites, lipid and lipoprotein lipids quantified by nuclear magnetic resonance (NMR) or mass spectrometry (MS) metabolomics to risk of incident dementia and AD in eight longitudinal studies with a total of 22,623 participants free of dementia at baseline: the FINRISK 1997 study, the DILGOM study, the Whitehall II Study, the Estonian biobank study (Estonian Genome Center, University of Tartu, EGCUT), the Health 2000, the Framingham Heart Study (FHS), the Rotterdam study (RS), and the Erasmus Ruchen Family (ERF) study. The first four cohorts were used in discovery analyses and the remaining four were used for replication. For metabolites taken to replication, we present overall association results combined across all eight samples by meta-analysis.

Methods

Cohorts

Eight prospective cohort studies were examined. Discovery cohorts were The National FINRISK Study 1997 (FINRISK 1997), Dietary, Lifestyle and Genetic determinants of Obesity and Metabolic Syndrome (DILGOM), Whitehall II, and the Estonian biobank study (Estonian Genome Center, University of Tartu, EGCUT). Replication cohorts were the Rotterdam study (RS), Erasmus Ruchen Family (ERF) study, Health 2000 study and the Framingham Heart study (FHS). More detailed descriptions of each study are provided in the supplemental material (**Supplements – Methods – Surveys**).

Altogether, 22,623 participants were included in this study. The sample size of each study is presented in **Supplementary Table 1**. No cognitive performance screening was conducted at baseline. Patients under 40 years of age were also excluded from all studies except ERF. All metabolite measurements were made from stored samples drawn at baseline of each cohort and no time-dependent covariates were used. Dementia identification in FINRISK 1997, DILGOM and Health 2000 cohorts was performed in the same manner, using country-wide, electronic health care registers: Causes of death Register (CDR), Hospital Discharge Register (HDR) and National Social Insurance Institution's Drug Reimbursement Register. In the EGCUT study cohort participants

were linked to the Estonian Health Insurance database containing detailed information on all contacts with healthcare services and prescriptions, and Estonian Causes of Death Registry, whereas prevalent disease information was additionally retrieved from recruitment questionnaires. In the Whitehall II study, participants were linked to electronic health records for dementia ascertainment using three databases: the national hospital episode statistics (HES) database, the Mental Health Services Data Set (MHSDS) and the mortality register. In the ERF survey we used register data from general practitioner's databases (9 to 14 years after baseline visit). In RS participants were screened for dementia at baseline and at follow-up examinations using a three-step protocol. Screen-positive participants subsequently underwent a more detailed examination and informant interview with the Cambridge Examination for Mental Disorders in the Elderly.[13] Additionally, the total cohort was continuously monitored for dementia through computerized linkage of the study database with digitized medical records from general practitioners.[13,14] In the FHS we screened participants at each examination, and between visits, for possible cognitive decline through a number of mechanisms, including an administration of the Folstein Mini-Mental Status Examination (MMSE), [15] participant and physician referrals, annual health status updates and review of medical records, and persons "flagged" as having possible cognitive decline underwent a more detailed neuropsychological and neurological evaluation. All cases were reviewed by a panel comprising at least one behavioural neurologist and one neuropsychologist. Details of endpoint detection are presented in the supplement (**Supplements – Methods - Surveys**).

Metabolomics analyses

A serum NMR metabolomics platform (Brainshake Ltd, Helsinki, Finland) was used to quantify 228 circulating metabolites, lipid or lipoprotein lipid measures in seven out of the eight cohorts.[16] All tested metabolites are listed in **Supplementary Table 2**. This high-throughput metabolomics platform provides simultaneous quantification of routine lipids, lipid concentrations of 14 lipoprotein subclasses and major subfractions, and further abundant fatty acids, amino acids, ketone bodies and gluconeogenesis related metabolites in absolute concentration units. The measured variables include 148 primary measures quantified in absolute concentrations as well as selected ratios,

primarily related to fatty acids and lipoprotein composition. The NMR platform has been applied extensively in epidemiological studies[17-19] and details of the experimentation have been described elsewhere.[16,20] NMR analysis was used in all cohorts except FHS. In the FHS cohort, liquid chromatography-tandem mass spectrometry (LC-MS) has been used. LC-MS data were acquired using either an AB SCIEX 4000 QTRAP triple quadrupole mass spectrometer (positively charged polar compounds and lipids) or an AB SCIEX 5500 QTRAP triple quadrupole mass spectrometer (negatively charged polar compounds). Detailed protocols for the quantification of metabolites in the FHS have been previously published[21-23] and are described in Supplementary material (**Supplement – Methods – Metabolite analysis**). A subset of 2,638 serum samples from the FINRISK 1997 study were additionally profiled with liquid-chromatography mass spectrometry (LC-MS) using the AbsoluteIDQ p180 Kit assay from Biocrates (Innsbruck, Austria). The correlations between circulating branched chain amino acids (BCAA) determined by NMR and LC-MS among the same individuals were fairly good as shown in **Supplement Figure 1,** the r^2 was 0.61 and 0.45 for valine and leucine respectively.

Statistical analysis

In the discovery stage we analysed the associations of all metabolites, lipid, and lipoprotein lipid measures quantified by NMR metabolomics (n=228) with incident dementia and Alzheimer's disease (AD). Two models were used: Model 1 with age, sex, education grade and number of *APOE* ε4 alleles as covariates, and Model 2 which additionally adjusted for systolic blood pressure, hypertension treatment, prevalent diabetes, current smoking, and any prevalent cardiovascular disease (atrial fibrillation, coronary heart disease, heart failure, stroke or peripheral artery disease). Because some of the NMR metabolites are highly correlated, we performed a principal component analysis (PCA) to estimate the number of independent tests and corrected the p-values for multiple testing accordingly. The PC-analysis was conducted in the FINRISK 1997 cohort and 95% of the variation of NMR metabolites, lipids and lipoprotein lipids was explained by 25 principal components, giving the corresponding p-value of 0.002 (0.05/25) as statistically significant (type I error correction). We used Cox proportional hazards regression model to test the metabolite associations. Time from the baseline

examination to incident dementia, AD, death or the end of the follow-up was used as the timescale and age was used as a covariate in all models. Hazard ratios (HRs) and their 95% confidence intervals (95%CI) are presented per 1-SD of the rank-inverse-normal transformed concentration; 1SD change in units is presented in **Supplementary Table 1**. Proportional hazard (PH) assumption was tested in two discovery cohorts (FINRISK 1997, DILGOM) and violations were observed for 14 metabolites (marked with asterisk in **Supplementary Table 4**), none of these metabolites or lipoprotein lipids were significant in the discovery analysis. All metabolites with p-value less than 0.002 in discovery meta-analyses were taken forward for testing in the replication cohorts. Since FHS did not use the NMR metabolomics platform there were only a limited number of metabolites which could be replicated in the FHS cohort: valine, leucine, isoleucine and creatinine. The statistical analyses were carried out with R, version 3.2.3 or 3.3.1 using 'survival' and 'meta' packages.[24,25]

Results

The study included 22,623 subjects with 246,698 person-years of follow-up. In discovery cohorts (FINRISK 1997, DILGOM, Whitehall II and EGCUT) we observed altogether 329, 181 and 1435 cases of incident dementia, AD and deaths from any cause, respectively. The EGCUT cohort did not record AD cases separately. The median follow-up times were 10.0, 7.9, 17.9 and 7.5 years (0.0, 0.1, 0.7 and 1.2 interquartile range) in FINRISK 1997, DILGOM, Whitehall II and EGCUT cohorts, respectively (**Supplementary Table 2**).

In replication cohorts we observed altogether 666, 466 and 1,405 cases of incident dementia, AD and deaths from any cause, respectively. (**Supplementary Table 3**) The baseline characteristics for all cohorts are presented in **Supplementary Tables 2-3**. The classical risk factors for dementia, used as covariates in our models, produced the expected results (**Supplementary Table 4**). Covariates are described in Supplement material with more details (**Supplement – Methods – Covariates**). As a positive control it is worth noting that the number of *APOE* ε4 alleles was strongly associated

with the risk of incident dementia (HR=2.51, 95% CI 2.00 – 3.16, p<0.001, **Supplementary Table 4**)

Discovery and replication findings

In the discovery analyses (n=15,161/329 study subjects/cases of incident dementia), altogether ten metabolites or lipoprotein lipids were associated with incident dementia (p < 0.002) (**Table 1**) and none with incident AD. All the results of discovery analyses are presented in Supplements (**Supplementary Tables 5-8**), for both models and both outcomes, dementia and AD. The ten metabolites and lipoprotein lipids were tested further in four replication cohorts (Health 2000, RS, ERF, FHS) and these results are shown in **Table 2** (Dementia, Model 2) and **Supplementary Tables 9-11** (Dementia Model 1 and AD Models 1-2).

Table 2. Replication results of the ten preliminary significant metabolites associated with incident dementia in four separate cohorts. HR and 95% CI are shown per one standard deviation (SD) of rank inverse normal transformed metabolite concentration. Adjusted for model 2*.

Metabolite	HEALTH2000 HR (95%CI)	ROTTERDAM HR (95%CI)	ERF HR (95%CI)	FHS HR (95%CI)
CREA	1.19 (0.87;1.63)	0.94 (0.84;1.06)	0.82 (0.57;1.18)	0.97 (0.78;1.2)
ILE	0.87 (0.61;1.25)	0.9 (0.8;1.01)	0.66 (0.42;1.03)	1.05 (0.84;1.3)
L-HDL-CE-PC	0.99 (0.71;1.38)	1.04 (0.94;1.17)	1.08 (0.61;1.91)	NA
L-HDL-PL-PC	1.09 (0.8;1.49)	1.04 (0.94;1.16)	1.25 (0.74;2.11)	NA
LEU	0.94 (0.69;1.3)	0.84 (0.74;0.94)	0.91 (0.6;1.37)	0.94 (0.75;1.19)
S-VLDL-C	0.89 (0.64;1.23)	0.91 (0.82;1.01)	1.08 (0.59;1.97)	NA
SFA-FA	1.08 (0.78;1.49)	0.94 (0.84;1.04)	1.15 (0.61;2.2)	NA
VAL	0.93 (0.68;1.26)	0.86 (0.76;0.96)	0.84 (0.55;1.29)	0.96 (0.76;1.19)
XL-VLDL-C-PC	0.87 (0.65;1.16)	1.07 (0.95;1.19)	0.67 (0.3;1.49)	NA
XL-VLDL-TG-PC	1.19 (0.87;1.63)	0.94 (0.84;1.06)	0.82 (0.57;1.18)	0.97 (0.78;1.2)

* Model 2 includes age, sex, education grade and number of APOE ε4 alleles plus systolic blood pressure, hypertension treatment, prevalent diabetes, current smoking, and any prevalent cardiovascular disease (atrial fibrillation, coronary heart disease, heart failure, stroke or peripheral artery disease) as covariates. Abbreviation: SFA-FA, Ratio of saturated fatty acids to total fatty acids; L-HDL-CE-%, Cholesterol esters to total lipids ratio in large HDL; S-VLDL-C, Total cholesterol in small VLDL; XL-VLDL-C-%, Total cholesterol to total lipids ratio in very large VLDL; XL-VLDL-TG-%, Triglycerides to total lipids ratio in very large VLDL

Table 1. Metabolites and lipoprotein lipids associating statistically significantly ($p<0.002$) with incident dementia in meta-analysis of discovery cohorts. Hazard ratios (HR) and 95% confidence intervals (CI) are shown per one standard deviation (SD) of rank inverse normal transformed metabolite concentration. Adjusted for model 2† and the additional discoveries row adjusted for model 1*.

Metabolite	FINRISK 1997 HR (95%CI)	DILGOM HR (95%CI)	WHITEHALL HR (95%CI)	EGCUT HR (95%CI)	FIXED EFFECT, HR (95%CI)	RANDOM EFFECT, HR (95%CI)	I2	p.fixed	p.random
Creatinine	0.92 (0.75;1.12)	0.78 (0.58;1.05)	0.64 (0.5;0.82)	0.85 (0.59;1.24)	0.8 (0.7;0.91)	0.79 (0.67;0.94)	0.383	<0.001	0.008
SFA-FA	1.13 (0.9;1.4)	1.13 (0.84;1.52)	1.31 (1.07;1.6)	1.6 (1.17;2.19)	1.26 (1.11;1.42)	1.26 (1.09;1.45)	0.227	<0.001	0.001
Iso-leucine	0.9 (0.7;1.16)	0.8 (0.59;1.08)	0.72 (0.58;0.9)	0.64 (0.42;0.96)	0.78 (0.68;0.89)	0.78 (0.68;0.89)	0	<0.001	<0.001
Leucine	0.86 (0.66;1.13)	0.72 (0.52;1)	0.75 (0.6;0.92)	0.54 (0.34;0.84)	0.75 (0.65;0.86)	0.74 (0.64;0.86)	0.047	<0.001	<0.001
Valine	0.81 (0.62;1.06)	0.64 (0.46;0.89)	0.87 (0.72;1.06)	0.59 (0.4;0.88)	0.78 (0.69;0.89)	0.76 (0.64;0.9)	0.354	<0.001	0.002
L-HDL-CE-%	1.06 (0.87;1.3)	1.25 (0.91;1.71)	1.4 (1.13;1.75)	1.39 (0.96;2.02)	1.23 (1.09;1.4)	1.24 (1.07;1.44)	0.211	0.001	0.004
S-VLDL-C	1.01 (0.78;1.3)	0.87 (0.62;1.23)	0.72 (0.59;0.87)	0.76 (0.53;1.08)	0.81 (0.71;0.92)	0.82 (0.69;0.97)	0.355	0.001	0.023
XL-VLDL-C-%	0.94 (0.68;1.29)	1.06 (0.72;1.56)	1.28 (1.07;1.53)	1.35 (1.07;1.71)	1.22 (1.08;1.38)	1.2 (1.03;1.4)	0.259	<0.002	0.017
XL-VLDL-TG-%	1.03 (0.75;1.42)	0.88 (0.59;1.32)	0.78 (0.67;0.91)	0.81 (0.62;1.07)	0.82 (0.73;0.93)	0.82 (0.73;0.93)	0	0.001	0.001
Additional discoveries adjusted for model 1									
L-HDL-PL-%	0.94 (0.75;1.17)	0.81 (0.58;1.13)	0.76 (0.62;0.93)	0.66 (0.44;0.97)	0.81 (0.71;0.92)	0.81 (0.71;0.93)	0.067	<0.002	0.002

*Model 1 includes age, sex, education grade and number of APOE ε4 alleles as covariates.

† Model 2 includes all of the above plus systolic blood pressure, hypertension treatment, prevalent diabetes, current smoking, and any prevalent cardiovascular disease (atrial fibrillation, coronary heart disease, heart failure, stroke or peripheral artery disease) as covariates. Abbreviation: SFA-FA, Ratio of saturated fatty acids to total fatty acids; L-HDL-CE-%, Cholesterol esters to total lipids ratio in large HDL; S-VLDL-C, Total cholesterol in small VLDL; XL-VLDL-C-%, Total cholesterol to total lipids ratio in very large VLDL; XL-VLDL-TG-%, Triglycerides to total lipids ratio in very large VLDL.

Meta-analysis of all cohorts

The results of the discovery and replication cohorts were combined in meta-analysis of all cohorts (n=22,623/995 study subjects/cases of incident dementia). All three BCAAs (isoleucine, leucine and valine) were inversely associated with incident dementia (Model 2), (**Figure 1, Supplementary Tables 12-15**). Also creatinine, total cholesterol in small VLDL (S-VLDL-C) and triglycerides to total lipids ratio in very large VLDL (XL-VLDL-TG-%) were inversely associated with incident dementia (**Figure 1**) (Supplementary Tables 12-13). The concentration of cholesterol esters relative to total lipids in large HDL (L-HDL-CE-%) and total cholesterol to total lipids ratio in very large VLDL (XL-VLDL-C-%) were directly associated with incident dementia (**Figure 1**) (**Supplementary Tables 12-13**). The HRs for AD were broadly similar but smaller numbers of incident AD cases compared to incident dementia diluted the statistical significance of the results (**Figure 1**) (**Supplementary Tables 14-15**). The correlations between the metabolites associated with incident dementia are presented in Supplementary Table 16. Of note are the moderately strong correlations between the BCAAs and S-VLDL-C (r=0.67 – 0.55).

Sensitivity analyses

We carried out two sensitivity analyses. First, to examine the influence of nutrition (that might be impacted at an early pre-clinical stage of dementia), we compared body mass index (BMI) at baseline between persons with and without subsequent incident dementia. Second, to examine the possible effects of selective mortality on our findings, we tested the associations of identified biomarkers with all-cause mortality in the FINRISK 1997 and DILGOM cohorts. These analyses are briefly described below. The BMI analysis was done in FINRISK 1997 and DILGOM studies separately. No statistically significant difference between the groups with and without incident dementia was seen in baseline levels of BMI using an ANOVA test adjusted for age, sex and *APOE* ε4 genotype. Next, we hypothesized that the inverse association of BCAA with incident dementia or AD might be caused by a competing risk of death since elevated levels of BCAA have been associated with metabolic syndrome, diabetes and cardiovascular events.

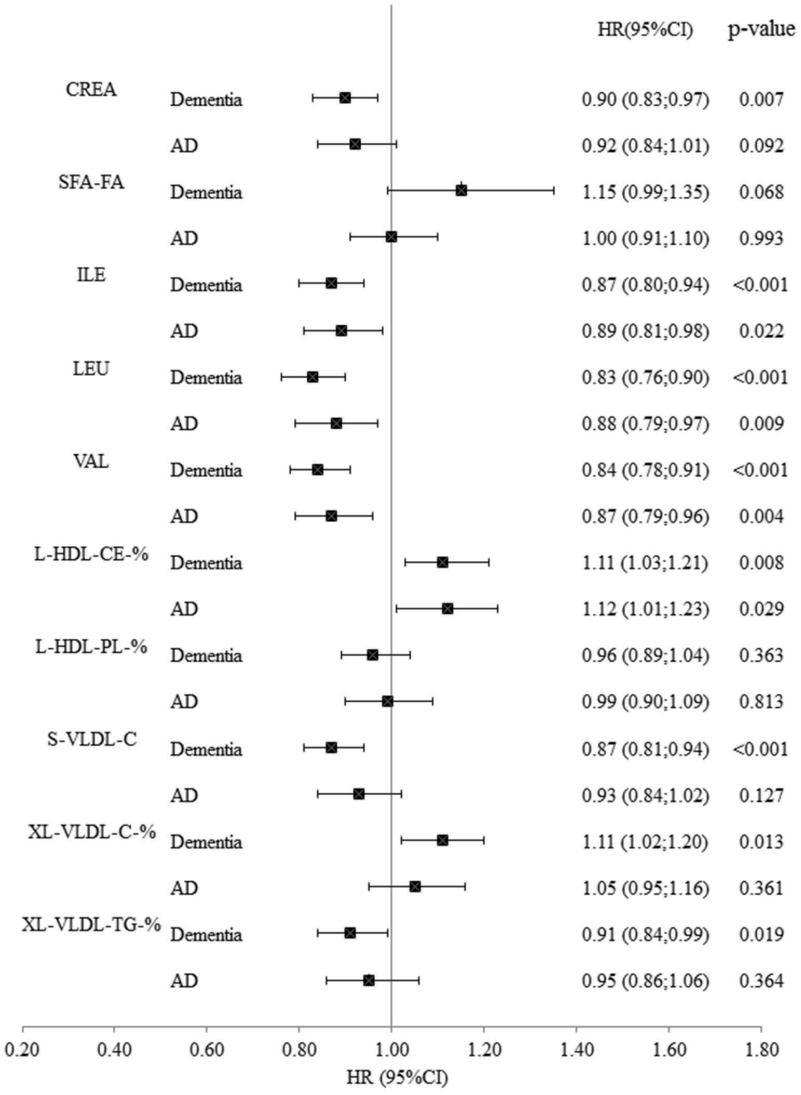


Figure 1: A forest plot describing the meta-analysis results of all eight cohorts combined (n=22,623/995, participants/ incident dementia cases). Results are adjusted for Model 2. Hazard ratios (HR), and 95% confidence intervals (CI) are shown per one standard deviation (SD) of rank inverse normal transformed metabolite concentration. Model 2 includes age, sex, education grade, number of *APOE* ε4 alleles, systolic blood pressure, hypertension treatment, prevalent diabetes, current smoking, and any prevalent cardiovascular disease (atrial fibrillation, coronary heart disease, heart failure, stroke or peripheral artery disease) as covariates.

No association was observed between BCAA or any other of the identified inversely associated lipid biomarkers and risk of death. We observed, however, saturated fatty acids to total fatty acids ratio (SFA-FA) and phospholipids to total lipids ratio in large HDL (L-HDL-PL-%) to be robustly associated with death from any cause in FINRISK 1997, DILGOM and Health 2000 cohorts (**Supplementary Table 17**). In a meta-analysis of FINRISK 1997, DILGOM and Health 2000 results we observed no interactions between metabolites, lipids, or lipoprotein lipids and *APOE* ε4 genotype or sex on the association ($p < 0.002$) with incident dementia or AD.

Discussion

Our study identified ten metabolites or lipoprotein lipids associated with the risk for clinically incident dementia across four discovery cohorts (n = 329 cases of incident dementia in 15,161 study subjects). Lower levels of branched chain amino acids such as valine were associated with an increased risk of both all dementia and of AD in this discovery cohort, and in a combined meta-analysis with a replication sample (n = 995 cases in 22,623 subjects). In addition, we observed inverse associations of creatinine, total cholesterol in small VLDL (S-VLDL-C) and triglycerides to total lipids ratio in very large VLDL (XL-VLDL-TG-%) with incident dementia, but not with AD in the discovery cohort alone. In meta-analysis, the concentration of cholesterol esters relative to total lipids in large HDL (L-HDL-CE-%) was associated with an increased risk of AD.

We tested 228 metabolic measures quantified by serum NMR metabolomics in the discovery sample. We chose a very conservative strategy of initial discovery followed by independent replication of a hard clinical end-point, to minimize the risk of reporting false-positive associations, a pitfall that has impacted earlier reports. This may have however reduced our ability to identify some true associations, for example with DHA which, in addition to earlier literature, was recently discovered to associate with higher general cognitive ability in a study based partly on the same cohorts as the present study (personal communication, Sven J. van der Lee, chapter 5.3).

To our knowledge the inverse association of BCAAs with clinical dementia has not reported previously. In line with our results, Toledo et al. observed higher valine level to be associated with slower cognitive decline and lesser cerebral atrophy change in the 'Alzheimer's Disease Neuroimaging Initiative' (ADNI) cohort.

Although there are biologically plausible explanations for such an association what we describe, subsequently we wished to explore whether these findings could represent reverse causality or selective survival. Valine, leucine and isoleucine are essential branched chain amino acids (BCAA), and circulating levels are largely determined by dietary intake. Thus, reduced levels of these essential amino acids might indicate subclinical nutritional deficiencies in persons with preclinical dementia and MCI.[26,27] Indeed, in later life weight loss is known to be associated with a higher risk of dementia and weight loss has been associated with declining BCAA levels. [28,29] However, plasma albumin, a robust measure of nutritional status was not related to dementia risk and baseline BMI did not differ between subjects with and without incident dementia. BCAA are also associated with muscle mass,[30] which is consistent with our observation of an inverse association of creatinine and dementia risk. Hence, it is possible that these metabolites are early markers of MCI, reduced physical activity and muscle mass. Competing risk of death in persons with elevated BCAA levels is not likely to explain the observed inverse associations with incident dementia or AD based on the absence of an association with all-cause mortality in our discovery cohorts and in the study of Fischer et al. where no association between BCAAs and all- cause mortality was observed in the EGCUT study.[18] In a recent study, Pedersen et al demonstrated that gut microbiota are an independent source of BCAAs, but these have not been examined in relation to dementia risk even though changes in gut microbiota have been associated with other neuropsychiatric diseases.[31]

Lower CSF valine has been recorded in persons with AD dementia compared to controls,[32] although some studies have shown low valine in MCI and higher values in AD.[33] In a small cross sectional study no differences were seen in BCAA blood levels among healthy controls, patients with mild cognitive impairment (MCI) and those with

AD,[34] but an independent cross-sectional study identified valine, among other metabolites, as an indicator for disease progression from MCI to AD.[33] It is hypothesized that circulating BCAAs could have an important role in glutamate synthesis and could also buffer toxic levels of glutamate.[35] Glutamate is the most abundant excitatory neurotransmitter and binds to cell surface receptors like AMPA- and NMDA-receptors.[36] Since NMDA-receptor hypofunction seems to be related to calcium ion dysregulation and impaired synaptic plasticity,[37] it is possible that the association of reduced levels of BCAA to dementia and AD is mediated through this pathway.[33] Higher valine has also been associated with increased apoptosis in maple-syrup urine disease and lower valine in preclinical AD could be a compensatory phenomenon in response to activation of apoptotic pathways.[38]

We did not observe any robust associations of cholesterols, triglycerides, phospholipids, apolipoproteins, fatty acids, glycolysis related metabolites, ketone bodies, fluid balance metabolites or inflammation markers with incident dementia or AD. In previous studies several specific phospholipids have been shown to associate with the risk of conversion of mild cognitive impairment (MCI) to AD[7,11] but the metabolomics platform employed here does not share the same phospholipids reported on in those earlier studies.

Our results also differ from prior studies which suggested an association of sphingomyelin and DHA with dementia and AD[7,39-41]; these differences could be due to chance or due to differences in dietary patterns or genetic risk between the cohorts studied here and in early reports. Docosahexaenoic acid (DHA) and eicosapentaenoic acid (EHA) dietary supplements have been associated with better cognitive function[42,43] and lower odds of dementia and AD (personal communication, Sven J. van der Lee, chapter 5.3) in some studies but two large meta-analyses did not show an association suggesting that this area deserves further scrutiny including possible gene-environment interactions.[44] Our conservative statistical strategy and lower sensitivity to detect some associations may partly explain these differences between our study and previous studies as described above.[44,45]

We did not observe any effect modification of the metabolites or lipoprotein lipids by *APOE* genotype or sex in their effect on dementia or AD risk. We did not observe any robust associations with absolute concentrations of lipoproteins and incident dementia or AD. This is in line with previous genetic studies where no causal effect of circulating levels of HDL-cholesterol, serum total cholesterol, LDL-cholesterol or triglycerides on incident AD was seen despite the genetic associations with the *APOE* and *ABCA7* loci.[46,47] Age, education level, APOE genotype and prevalent diabetes were the only covariates associated with incident dementia in this study. We did not observe statistically significant association between incident dementia and smoking, systolic blood pressure, heart failure or atrial fibrillation. This is not surprising, since the analyses were not designed to investigate these associations and survival bias likely affects these results.

The strengths of our study include the large samples and prospective population-based design with separate discovery and replication cohorts. Furthermore, all NMR metabolomics measurements were carried out in the same laboratory following the same protocol and only one cohort used a different methodology. Limitations of the study include some differences between cohorts in the methods used to identify cases of incident dementia. Most cohorts relied on electronic health registers which leads to virtually complete follow-up and high specificity but may have limited sensitivity.[48] The lower sensitivity potentially reduces the power to detect statistically significant associations but should not result in spurious associations.[49] Moreover, the findings were replicated in prospective cohorts such as the Rotterdam study and FHS that used more sensitive, surveillance-based outcome ascertainment. Some cohorts were unable to distinguish AD from other dementias, which led to a reduced statistical power for the AD analysis. Ethnic homogeneity of our largely Caucasian sample limits the generalizability of these results to other populations with different ethnic backgrounds. Also, the Framingham Heart Study (FHS) used a mass-spectrometry (MS) platform and therefore only a limited set of metabolites were in common between the FHS and other cohorts. The main association we report here, of BCAA, were however directionally

consistent in FHS and the other cohorts and the correlations between the BCAAs measured using MS and NMR were reasonably strong.

Several next steps can be considered to clarify our findings further. The earlier studies that have reported an association between BCAA and metabolic syndrome and diabetes risk, should be examined for availability of cognitive end-points and possible confirmation of the present findings. Furthermore, a Mendelian randomization study on BCAA and other metabolites and incident dementia or AD should help with causal inferences. If further evidence supporting causality is obtained, a clinical trial supplementing BCAA in diet could be considered but such a trial should be carefully designed and monitored also for diabetes and other metabolic outcomes. A wider angle is the rapidly increasing availability of metabolomics profiling. Researchers should consider establishing an international metabolomics consortium aiming at harmonizing measures and analysis protocols across metabolomics platforms and cohorts, so that data already collected can be meta-analyzed. This would also encourage metabolomics studies in other geographical areas and other ethnic groups which are clearly needed. Another line of research that is worth pursuing in parallel to confirming the discovery findings is the prediction of dementia or AD risk. Even if not causal, these biomarkers may improve the prediction of incident dementia or AD over and above the currently used risk scores. This in turn could enable earlier starting and better targeting of medical and other treatments which has the potential to slow down the cognitive decline.

In conclusion, our large prospective study identified lower branched-chain amino acids levels to be associated with an increased risk of incident dementia, independent of other conventional risk factors. Further, creatinine, one HDL and three VLDL lipoprotein subclasses were also associated with dementia risk. Further studies are needed to explore whether these metabolites play a role in the aetiology and pathogenesis or dementia, or reflect reverse causation, i.e., are biomarkers for systemic or lifestyle changes in the preclinical stages of dementia. In either case, if corroborated in other

studies, these biomarkers may help in early identification of persons at risk of dementia and initiation of preventive and treatment measures.

References:

1. Wimo A, Jonsson L, Bond J, Prince M, Winblad B, Alzheimer Disease I. The worldwide economic impact of dementia 2010. *Alzheimers Dement* 2013; 9(1): 1-11 e3.
2. Jonsson T, Stefansson H, Steinberg S, et al. Variant of TREM2 associated with the risk of Alzheimer's disease. *N Engl J Med* 2013; 368(2): 107-16.
3. Chouraki V, Seshadri S. Genetics of Alzheimer's disease. *Adv Genet* 2014; 87: 245-94.
4. Seshadri S, Fitzpatrick AL, Ikram MA, et al. Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA* 2010; 303(18): 1832-40.
5. Pan X, Nasaruddin MB, Elliott CT, et al. Alzheimer's disease-like pathology has transient effects on the brain and blood metabolome. *Neurobiol Aging* 2016; 38: 151-63.
6. Proitsi P, Kim M, Whiley L, et al. Association of blood lipids with Alzheimer's disease: A comprehensive lipidomics analysis. *Alzheimers Dement* 2017; 13(2): 140-51.
7. Li D, Misialek JR, Boerwinkle E, et al. Plasma phospholipids and prevalence of mild cognitive impairment and/or dementia in the ARIC Neurocognitive Study (ARIC-NCS). *Alzheimers Dement (Amst)* 2016; 3: 73-82.
8. Wood PL, Locke VA, Herling P, et al. Targeted lipidomics distinguishes patient subgroups in mild cognitive impairment (MCI) and late onset Alzheimer's disease (LOAD). *BBA Clin* 2016; 5: 25-8.
9. Ellis B, Hye A, Snowden SG. Metabolic Modifications in Human Biofluids Suggest the Involvement of Sphingolipid, Antioxidant, and Glutamate Metabolism in Alzheimer's Disease Pathogenesis. *J Alzheimers Dis* 2015; 46(2): 313-27.
10. Graham SF, Chevallier OP, Elliott CT, et al. Untargeted metabolomic analysis of human plasma indicates differentially affected polyamine and L-arginine metabolism in mild cognitive impairment subjects converting to Alzheimer's disease. *PLoS One* 2015; 10(3): e0119452.
11. Mapstone M, Cheema AK, Fiandaca MS, et al. Plasma phospholipids identify antecedent memory impairment in older adults. *Nat Med* 2014; 20(4): 415-8.
12. Casanova R, Varma S, Simpson B, et al. Blood metabolite markers of preclinical Alzheimer's disease in two longitudinally followed cohorts of older individuals. *Alzheimers Dement* 2016; 12(7): 815-22.
13. Schrijvers EM, Verhaaren BF, Koudstaal PJ, Hofman A, Ikram MA, Breteler MM. Is dementia incidence declining?: Trends in dementia incidence since 1990 in the Rotterdam Study. *Neurology* 2012; 78(19): 1456-63.
14. de Bruijn RF, Bos MJ, Portegies ML, et al. The potential for prevention of dementia across two decades: the prospective, population-based Rotterdam Study. *BMC Med* 2015; 13: 132.
15. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975; 12(3): 189-98.
16. Soininen P, Kangas AJ, Wurtz P, Suna T, Ala-Korpela M. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circ Cardiovasc Genet* 2015; 8(1): 192-206.
17. Wurtz P, Havulinna AS, Soininen P, et al. Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation* 2015; 131(9): 774-85.
18. Fischer K, Kettunen J, Wurtz P, et al. Biomarker profiling by nuclear magnetic resonance spectroscopy for the prediction of all-cause mortality: an observational study of 17,345 persons. *PLoS Med* 2014; 11(2): e1001606.
19. Wurtz P, Wang Q, Soininen P, et al. Metabolomic Profiling of Statin Use and Genetic Inhibition of HMG-CoA Reductase. *J Am Coll Cardiol* 2016; 67(10): 1200-10.

20. Soininen P, Kangas AJ, Wurtz P, et al. High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism. *Analyst* 2009; 134(9): 1781-5.
21. Rhee EP, Cheng S, Larson MG, et al. Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. *J Clin Invest* 2011; 121(4): 1402-11.
22. Wang TJ, Ngo D, Psychogios N, et al. 2-Aminoadipic acid is a biomarker for diabetes risk. *J Clin Invest* 2013; 123(10): 4309-17.
23. Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes. *Nat Med* 2011; 17(4): 448-53.
24. (2008) RDCT. R: A language and environment for statistical computing. 2013.
25. (2015). TT. A Package for Survival Analysis in S. version 2.38, https://CRAN.R-project.org/package=survival.
26. Orsitto G, Fulvio F, Tria D, Turi V, Venezia A, Manca C. Nutritional status in hospitalized elderly patients with mild cognitive impairment. *Clin Nutr* 2009; 28(1): 100-2.
27. Alhurani RE, Vassilaki M, Aakre JA, et al. Decline in Weight and Incident Mild Cognitive Impairment: Mayo Clinic Study of Aging. *JAMA Neurol* 2016; 73(4): 439-46.
28. Zhao X, Han Q, Liu Y, Sun C, Gang X, Wang G. The Relationship between Branched-Chain Amino Acid Related Metabolomic Signature and Insulin Resistance: A Systematic Review. *J Diabetes Res* 2016; 2016: 2794591.
29. Knopman DS, Edland SD, Cha RH, Petersen RC, Rocca WA. Incident dementia in women is preceded by weight loss by at least a decade. *Neurology* 2007; 69(8): 739-46.
30. Lustgarten MS, Price LL, Chale A, Phillips EM, Fielding RA. Branched chain amino acids are associated with muscle mass in functionally limited older adults. *J Gerontol A Biol Sci Med Sci* 2014; 69(6): 717-24.
31. Moos WH, Faller DV, Harpp DN, et al. Microbiota and Neurological Disorders: A Gut Feeling. *Biores Open Access* 2016; 5(1): 137-45.
32. Basun H, Forssell LG, Almkvist O, et al. Amino acid concentrations in cerebrospinal fluid and plasma in Alzheimer's disease and healthy control subjects. *J Neural Transm Park Dis Dement Sect* 1990; 2(4): 295-304.
33. Ibanez C, Simo C, Martin-Alvarez PJ, et al. Toward a predictive model of Alzheimer's disease progression using capillary electrophoresis-mass spectrometry metabolomics. *Anal Chem* 2012; 84(20): 8532-40.
34. Oresic M, Hyotylainen T, Herukka SK, et al. Metabolome in progression to Alzheimer's disease. *Transl Psychiatry* 2011; 1: e57.
35. Yudkoff M. Interactions in the Metabolism of Glutamate and the Branched-Chain Amino Acids and Ketoacids in the CNS. *Neurochem Res* 2017; 42(1): 10-8.
36. Chung C. NMDA receptor as a newly identified member of the metabotropic glutamate receptor family: clinical implications for neurodegenerative diseases. *Mol Cells* 2013; 36(2): 99-104.
37. Foster TC, Kyritsopoulos C, Kumar A. Central role for NMDA receptors in redox mediated impairment of synaptic function during aging and Alzheimer's disease. *Behav Brain Res* 2017; 322(Pt B): 223-32.
38. Vilela TC, Scaini G, Furlanetto CB, et al. Apoptotic signaling pathways induced by acute administration of branched-chain amino acids in an animal model of maple syrup urine disease. *Metab Brain Dis* 2017; 32(1): 115-22.
39. Mielke MM, Haughey NJ, Bandaru VV, et al. Plasma sphingomyelins are associated with cognitive progression in Alzheimer's disease. *J Alzheimers Dis* 2011; 27(2): 259-69.
40. Tan ZS, Harris WS, Beiser AS, et al. Red blood cell omega-3 fatty acid levels and markers of accelerated brain aging. *Neurology* 2012; 78(9): 658-64.

41. Schaefer EJ, Bongard V, Beiser AS, et al. Plasma phosphatidylcholine docosahexaenoic acid content and risk of dementia and Alzheimer disease: the Framingham Heart Study. *Arch Neurol* 2006; 63(11): 1545-50.
42. Janssen CI, Kiliaan AJ. Long-chain polyunsaturated fatty acids (LCPUFA) from genesis to senescence: the influence of LCPUFA on neural development, aging, and neurodegeneration. *Prog Lipid Res* 2014; 53: 1-17.
43. Zhang Y, Chen J, Qiu J, Li Y, Wang J, Jiao J. Intakes of fish and polyunsaturated fatty acids and mild-to-severe cognitive impairment risks: a dose-response meta-analysis of 21 cohort studies. *Am J Clin Nutr* 2016; 103(2): 330-40.
44. Burckhardt M, Herke M, Wustmann T, Watzke S, Langer G, Fink A. Omega-3 fatty acids for the treatment of dementia. *Cochrane Database Syst Rev* 2016; 4: CD009002.
45. Wu S, Ding Y, Wu F, Li R, Hou J, Mao P. Omega-3 fatty acids intake and risks of dementia and Alzheimer's disease: a meta-analysis. *Neurosci Biobehav Rev* 2015; 48: 1-9.
46. Ostergaard SD, Mukherjee S, Sharp SJ, et al. Associations between Potentially Modifiable Risk Factors and Alzheimer Disease: A Mendelian Randomization Study. *PLoS Med* 2015; 12(6): e1001841; discussion e.
47. Proitsi P, Lupton MK, Velayudhan L, 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): e1001713.
48. Solomon A, Ngandu T, Soininen H, Hallikainen MM, Kivipelto M, Laatikainen T. Validity of dementia and Alzheimer's disease diagnoses in Finnish national registers. *Alzheimers Dement* 2014; 10(3): 303-9.
49. Brenner H, Savitz DA, Jockel KH, Greenland S. Effects of nondifferential exposure misclassification in ecologic studies. *Am J Epidemiol* 1992; 135(1): 85-95.

Supplements – Methods - Surveys

FINRISK 1997 and DILGOM

The FINRISK Studies 1997 and 2007 a{population-based health examination surveys which were carried out in 6 areas in Finland.[1] DILGOM (Dietary, Lifestyle and Genetic determinants of Obesity and Metabolic Syndrome) was an extension of the FINRISK 2007 survey, focused on obesity. The original random samples were stratified by area, sex, and 10-year age group according to the World Health Organization (WHO) MONICA (MONItoring trends and determinants of CArdiovascular disease) protocol.[2] The samples comprised of men and women aged 25 to 74. Participation rates were 72% and 60.9% in years 1997 and 2007, respectively.[1] In the FINRISK 1997 sample we excluded 11 persons with prevalent dementia or AD at baseline, 8 persons with prevalent MS, 121 with prevalent epilepsy, 105 with prevalent stroke and 76 women who were pregnant. In the DILGOM sample we excluded 6 persons with prevalent dementia or AD, 11 persons with prevalent MS, 81 persons with prevalent epilepsy, 85 with prevalent stroke and 24 women with pregnancy. In addition, all participants aged less than 40 years at baseline examination were excluded from the analyses. The final analyzed sample size was 4,581 in FINRISK 1997 and 3,399 in DILGOM. These surveys were conducted in accordance with the Declaration of Helsinki. For the FINRISK 1997 study, ethics approval was obtained from the Ethics Committee of the National Public Health Institute and for DILGOM from the Coordinating Ethics Committee of the Helsinki and Uusimaa Hospital District. The participants gave a written, informed consent.

FINRISK 1997 participants were advised to fast for four hours before the scheduled examination and avoid heavy meals earlier during the day. The length of fasting and the time of blood drawing were recorded. DILGOM participants were examined in the morning after an overnight fast. A self-administered questionnaire included questions on medical history and socioeconomic factors.[1] In the health examination, specially trained research nurses measured participants' height and weight following the standardized WHO MONICA protocol.[2] BMI was calculated as weight(kg)/height(m2). Blood pressure was measured in each survey from the right arm after 5 minutes of sitting using a mercury sphygmomanometer. After blood pressure measurement, a venous blood specimen was taken. In both surveys blood samples were analyzed for traditional

cardiovascular risk factors in the same central laboratory (National Public Health Institute, Helsinki, Finland), standardized with the Center for Disease Control (CDC), Atlanta, Georgia. Blood sampling, sample handling, laboratory methods, accuracy and precision for total- and HDL cholesterol (HDL-C) measurements have been described previously.[3,4] The samples were genotyped on a number of different genome-wide genotyping arrays and subsequently imputed using ShapeIt2 and IMPUTE2 standard protocols. The apo E genotype was obtained from the imputed GWAS data as described in more detail elsewhere. (Tynkkynen, et al. in Press DOI: 10.1007/s11357-016-9950-x). Description of covariates (prevalent diseases and antihypertensive medication) in FINRISK 1997 and DILGOM studies covariate has been presented previously.[5,6] Education was graded into tertiles by years of education adjusted for the birth year. Similar methods were used in the Health 2000 cohort except for antihypertensive medication data were obtained from the National Social Insurance Institution's Drug Reimbursement Register. Identification of dementia cases was carried out by annual record linkage of the FINRISK and DILGOM data on the basis of personal ID code to the Finnish Hospital Discharge Register (HDR), Causes of Death Register (CDR) and national Social Insurance Institution's Drug Reimbursement Register. These registers cover all deaths and hospitalizations in Finland, purchases of all drugs prescribed by a doctor and all drug reimbursements. With these country-wide electronic registers the coverage of the follow-up is 100% as long as the participant continues to live in Finland. Incident dementia was identified if a participant was hospitalized or died with the ICD-10 codes of F00, F01, F02, F03 or G30 or he/she made drug purchases of acetylcholinesterase inhibitors (AChRis) or NMDA-receptor antagonist, memantine, more than three times during the follow-up. A participant entitled to reimbursements for these therapeutic agents with the indication of dementia, was considered as having incident dementia. To be granted the right to such reimbursements, a person must present a written statement by a neurologist or a geriatrician documenting the findings which have led to the diagnosis of dementia. Incident AD was identified only if the drug reimbursement was granted with the ICD-10 code of G30. For these analyses the follow-up was available until Dec. 31st, 2013.

Health 2000

The Health 2000 Survey (BRIF8901) was a population-based health examination survey carried out in 80 areas throughout Finland in 2000-2001.[7,8] The sample collection was performed in two stages with stratified cluster sampling. The study population consisted of 8,028 participants (85% of the sample) and was representative of the Finnish population aged ≥30 years. The Health 2000 Survey included interviews, self-administered questionnaires and a comprehensive health examination. In-depth cardiovascular examinations were performed in the cardiovascular disease and diabetes subcohort (SVT-D, sample size 1,867 and participation rate 82%), the participants of which were aged 45 to 74 years and living in the area of Helsinki, Kuopio, Oulu, Tampere, Turku or Joensuu. Data on education, smoking, and previous diseases (e.g. type 2 diabetes and cardiovascular diseases), and antihypertensive medication were self-reported in a health interview or a self-administered questionnaire at baseline. Height, weight and waist circumference were measured at a health examination, and body mass index (BMI) was calculated. Blood pressure was measured three times from the right arm after 10 minutes of sitting with automatic OMRON M4 device. Fasting blood samples were taken and stored at – 70°C. More detailed description of the study protocol is given elsewhere.[7] The linkage to the Finnish national registers is done in the same way as described for FINRISK 1997. *APOE* genotyping was performed using the MassARRAY System (Sequenom, San Diego, CA, USA) with a modified protocol, which has been described elsewhere.[9]

Rotterdam - RS

This study included participants from the Rotterdam Study (RS), which is a prospective population-based cohort study. In 1990, all residents aged 55 and older residing in Ommoord, a district of Rotterdam, the Netherlands, were invited to participate in the study. Of the 10,215 invited inhabitants, 7,983 agreed to participate in the baseline examinations. In 2000, 3,011 participants (out of 4,472 invitees) who had become 55 years of age or moved into the study district since the start of the study were added to the cohort. Follow-up examinations take place every 3 to 4 years.[10] The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC and by

the Ministry of Health, Welfare and Sport of the Netherlands, implementing the Population Studies Act: Rotterdam Study (Erasmus Rotterdam Gezondheid Onderzoek). All participants provided written informed consent to participate in the study and to obtain information from their treating physicians. Measurements of the co-variates are described in detail elsewhere.[11] Participants were screened for dementia at baseline and at follow-up examinations using a three-step protocol. [12] Screening was done using the Mini-Mental State Examination (MMSE) and the Geriatric Mental Schedule (GMS) organic level. Screen-positives (MMSE <26 or GMS organic level >0) subsequently underwent an examination and informant interview with the Cambridge Examination for Mental Disorders in the Elderly.[12] Additionally, the total cohort was continuously monitored for dementia through computerized linkage of the study database with digitized medical records from general practitioners and the Regional Institute for Outpatient Mental Health Care. When information on neuroimaging was required and available, it was used for decision making on the diagnosis. Ultimately, a consensus panel, led by a neurologist, decided on the fil diagnosis in accordance with standard criteria for dementia (DSM-III-R) and Alzheimer's disease (NINCDS-ADRDA).[11,12] Of the 2938 individuals with metabolite data 507 participants developed dementia of which 13 were classified as Parkinson's' disease (excluded), 13 were rare specific causes of dementia and 346 were classified as AD.

Erasmus Ruchen Family Study - ERF

The Erasmus Ruchen Family study (ERF) is a prospective family-based study from a genetically isolated population in Southwest of the Netherlands. The ERF study is described in detail previously.[13,14] The baseline demographic data and measurements of the ERF participants were collected around 2002 to 2006. Blood samples were collected through venipuncture in the morning after at least 8 hours fasting at baseline. From 2015 until May 2016 records of 1368 participants in the original cohort had complete covariate data, were scanned for incident dementia in general practitioner's databases (9 to 14 years after baseline visit) and contributed to the analysis. Three individuals developed Parkinson's' disease dementia and were excluded, 145 died without dementia, 35 participants developed dementia and 25 were classified as AD.

Whitehall II

The Whitehall II study is a prospective cohort study of 10,308 participants (70% men), aged 35–55 years and recruited between 1985 and 1989 from 20 London-based Civil service departments (https://www.ucl.ac.uk/whitehallII). Clinical examinations have been performed in 1991-1994, 1997-1999, 2002-2004, 2007-2009, 2012-2013, 2015-2016 with the data from circulating metabolomic traits and cognitive testing for the present study obtained from the 1997-1999 clinic phase. The participants have been linked to electronic medical records. The study was approved by the UCL Research Ethics Committee. All participants gave written informed consent to each aspect of the study.

We used comprehensive tracing of electronic health records for dementia ascertainment using three databases: the national hospital episode statistics (HES) database, the Mental Health Services Data Set (MHSDS) and the mortality register. Record linkage until 31st of March 2015, using International Classification of Diseases Tenth Edition (ICD-10) codes F00, F01, F02, F03, F05.1, G30, G31.0, G31.1 and G31.8 identified cases of dementia, following the National Health Service (NHS) guidelines. In the UK (England, Scotland, Wales), the NHS provides most of the health care, including out- and in-patient care. Private medical insurance, held by around 12% of the UK population (1997 figures), is mainly used for elective surgery rather than chronic conditions such as dementia. MHSDS is a national database which contains information for persons in contact with mental health services in hospitals, outpatient clinics, and the community.

Framingham Heart Study - FHS

The Framingham Heart Study (FHS) is an ongoing community-based prospective cohort study initiated in 1948 with the enrollment of 5,209 women and men aged 28 to 74 years (Original Cohort). [15] In 1971, offspring of the Original Cohort and the spouses of these offspring (n=5,124; age, 5–70 years; 3,548 biological offspring and 1,576 offspring spouses) were enrolled in the Framingham Offspring Cohort. [16] They have been examined every 4 to 8 years since, 9 times to date, for a core examination.[17] In addition, the Offspring cohort has been under ongoing surveillance for cognitive decline and

dementia since the fifth examination (1991-1995, n=3,799). A total of 2,526 participants among those who attended this fifth examination had their plasma metabolome measured. In FHS cohort education was graded to three levels: college graduate, high school degree and no high school degree. We screened participants at each examination for possible cognitive decline through a number of mechanisms, including an administration of the Folstein Mini-Mental Status Examination (MMSE).[18] Persons "flagged" as having possible cognitive decline or otherwise being at risk for developing dementia underwent a more detailed neuropsychological and neurological evaluation and when required a structured family interview was administered to one or more family members and caregivers over the telephone. All persons were assigned a Clinical Dementia Rating [19] scale score. We then determined whether each person fulfilled criteria for a diagnosis of dementia, the probable date of onset, and type of dementia at a consensus review conducted by a panel comprising at least one behavioral neurologist and one neuropsychologist. Participants with dementia met criteria outlined in the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders criteria.[20] Participants with AD met national Institute of Neurological and Communicative Diseases and Stroke/Alzheimer's Disease and Related Disorders Association criteria for definite, probable, or possible AD.[21] For the present analyses, data for incident dementia obtained till December 2013 were used.

Briefly, the metabolome was assessed in FHS Offspring on plasma samples collected following an overnight fast and stored at -80°C. Plasma metabolites were analyzed using liquid chromatography-tandem mass spectrometry (LC-MS). LC-MS data were acquired using either an AB SCIEX 4000 QTRAP triple quadrupole mass spectrometer (positively charged polar compounds and lipids) or an AB SCIEX 5500 QTRAP triple quadrupole mass spectrometer (negatively charged polar compounds). Polar, positively charged metabolites were separated using hydrophobic interaction liquid chromatography (HILIC) and analyzed using multiple reaction monitoring (MRM) in the positive ion mode. Polar, negatively charged compounds, including central and polar phosphorylated metabolites, were separated using a Luna NH2 column (150×2 mm, Luna NH2, Phenomenex) and analyzed using MRM in the negative ion mode. Lipids

were separated on a Prosphere C4 HPLC column and underwent full scan MS analysis in the positive ion mode. MultiQuant software (Version 1.2, AB SCIEX) was used for automated peak integration and manual review of data quality prior to statistical analysis.

Estonia Biobank (Estonian Genome Center, University of Tartu, EGCUT)

Estonian Biobank (Estonian Genome Center, University of Tartu) is a population-based biobank that recruited a cohort of 51,830 participants, including adults from all counties in Estonia and accounting for approximately 5% of the Estonian adult population during the recruitment period (2002-2011). At baseline, an extensive phenotype questionnaire was conducted together with a measurement panel. In ECCUT cohort education was graded to four levels: university, secondary, basic and primary or less. Follow-up data is available from linkage with national health-related registries and the Estonian Health Insurance database. The NMR-biomarkers are available for a random subset of 11000 individuals from the cohort.

References – Supplement:

1. Vartiainen E, Laatikainen T, Peltonen M, Juolevi A, Mannisto S, Sundvall J, Jousilahti P, Salomaa V, Valsta L, Puska P (2010) Thirty-five-year trends in cardiovascular risk factors in Finland. Int J Epidemiol 39, 504-518.
2. Pajak A, Kuulasmaa K, Tuomilehto J, Ruokokoski E. (1988) Geographical variation in the major risk factors of coronary heart disease in men and women aged 35-64 years. The WHO MONICA Project. World Health Stat Q 41, 115-140.
3. Sundvall J, Leiviska J, Alfthan G, Vartiainen E (2007) Serum cholesterol during 27 years: assessment of systematic error and affecting factors and their role in interpreting population trends. Clin Chim Acta 378, 93-98.
4. Leiviska J, Sundvall J, Alfthan G, Tahtela R, Salomaa V, Jauhiainen M, Vartiainen E (2013) What have we learnt about high-density lipoprotein cholesterol measurements during 32 years? Experiences in Finland 1980-2012. Clin Chim Acta 415, 118-123.
5. Kastarinen M, Antikainen R, Peltonen M, Laatikainen T, Barengo NC, Jula A, Salomaa V, Jousilahti P, Nissinen A, Vartiainen E, Tuomilehto J (2009) Prevalence, awareness and treatment of hypertension in Finland during 1982-2007. J Hypertens 27, 1552-1559.
6. Blankenberg S, Zeller T, Saarela O, Havulinna AS, Kee F, Tunstall-Pedoe H, Kuulasmaa K, Yarnell J, Schnabel RB, Wild PS, Munzel TF, Lackner KJ, Tiret L, Evans A, Salomaa V, MORGAM Project (2010) Contribution of 30 biomarkers to 10-year cardiovascular risk estimation in 2 population cohorts: the MONICA, risk, genetics, archiving, and monograph (MORGAM) biomarker project. Circulation 121, 2388-2397.
7. Heistaro S http://www.terveys2000.fi/doc/methodologyrep.pdf.
8. Heistaro S (2005) Menetelmäraportti. Terveys 2000 -tutkimuksen toteutus, aineisto ja menetelmät. Helsinki: National Public Health Institute.

9. Janis MT, Siggins S, Tahvanainen E, Vikstedt R, Silander K, Metso J, Aromaa A, Taskinen MR, Olkkonen VM, Jauhiainen M, Ehnholm C (2004) Active and low-active forms of serum phospholipid transfer protein in a normal Finnish population sample. J Lipid Res 45, 2303-2309.
10. Hofman A, Brusselle GG, Darwish Murad S, van Duijn CM, Franco OH, Goedegebure A, Ikram MA, Klaver CC, Nijsten TE, Peeters RP, Stricker BH, Tiemeier HW, Uitterlinden AG, Vernooij MW (2015) The Rotterdam Study: 2016 objectives and design update. Eur J Epidemiol 30, 661-708.
11. de Bruijn RF, Bos MJ, Portegies ML, Hofman A, Franco OH, Koudstaal PJ, Ikram MA (2015) The potential for prevention of dementia across two decades: the prospective, population-based Rotterdam Study. BMC Med 13, 132-015-0377-5.
12. Schrijvers EM, Verhaaren BF, Koudstaal PJ, Hofman A, Ikram MA, Breteler MM (2012) Is dementia incidence declining?: Trends in dementia incidence since 1990 in the Rotterdam Study. Neurology 78, 1456-1463.
13. Sayed-Tabatabaei FA, van Rijn MJ, Schut AF, Aulchenko YS, Croes EA, Zillikens MC, Pols HA, Witteman JC, Oostra BA, van Duijn CM (2005) Heritability of the function and structure of the arterial wall: findings of the Erasmus Rucphen Family (ERF) study. Stroke 36, 2351-2356.
14. Aulchenko YS, Heutink P, Mackay I, Bertoli-Avella AM, Pullen J, Vaessen N, Rademaker TA, Sandkuijl LA, Cardon L, Oostra B, van Duijn CM (2004) Linkage disequilibrium in young genetically isolated Dutch population. Eur J Hum Genet 12, 527-534.
15. DAWBER TR, MEADORS GF, MOORE FE,Jr (1951) Epidemiological approaches to heart disease: the Framingham Study. Am J Public Health Nations Health 41, 279-281.
16. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP (1975) The Framingham Offspring Study. Design and preliminary data. Prev Med 4, 518-525.
17. Au R, Seshadri S, Wolf PA, Elias M, Elias P, Sullivan L, Beiser A, D'Agostino RB (2004) New norms for a new generation: cognitive performance in the framingham offspring cohort. Exp Aging Res 30, 333-358.
18. Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 12, 189-198.
19. Berg L (1988) Clinical Dementia Rating (CDR). Psychopharmacol Bull 24, 637-639.
20. American Psychiatric Association and American Psychiatric Association. Task Force on DSM-IV. Diagnostic and statistical manual of mental disorders: DSM-IV. 1994.
21. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM (1984) Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 34, 939-944.

Supplementary Table 1. Descriptive statistics of discovery cohorts

	FINRISK 1997	DILGOM	Whitehall II	EGCUT
Participants/Incident dementia/Incident AD/Death	4580/105/78 /471	3399/69/65/176	4612/114/35/379	2570/41/NA/409
Follow-up time, y Max/Median (IQR)/mean	10.0/10.0 (0.0)/9.6	7.9/7.9 (0.1)/7.7	17.9/17.0 (0.7)/16.7	12.8/7.5 (1.2)/7.03
Age,y Mean (SD)	55.1(9.19)	57.54(9.49)	55.9 (6.0)	59.6 (12.4)
Sex, M %	50.4	47.0	73.5	41.1
SBP(mmHg), Mean (SD)	140.4 (20.2)	140.2 (20.5)	121.0 (15.9)	133.7 (17.2)
BP medication %	18.1	26.8	10.7	42.2
*Education % high/intermediate/low	32.0/34.0/34.0	29.9/35.4/34.7	34.9/28.1/37.0	22.1/60.5/16.8 (4.2 very low)
Current Smoking, yes %	21.7	15.9	8.7	23.6
Prevalent diseases				
Diabetes%	7.1	10.7	3.3	6.6
Atrial fibrillation%	2.3	3.1	0.3	4.2
CHD%	3.5	3.6	5.1	11.4
Heart failure%	2.8	2.8	NA	9.5
PAD%	0.9	0.9	NA	3.1
APOE ε4 (none/1/2) %	66.1/30.4/3.5	64.4/31.6/4.0	72.1/25.5/2.4	76.1/22.5/1.4

* see education description Supplement – Methods – Covariates
NA-not recorded separately

Supplementary Table 2. List of all metabolites quantified by NMR in the present study. This table can be viewed by scanning the code below:

Supplementary Table 3 Descriptive statistics of replication cohorts

	RS	ERF	Health 2000	FHS
Participants/Incident dementia/Incident AD/Death	2759/481/325/620	1366/35/25/126	1166/44/38/111	2171/106/78/548
Follow-up time, y Max/Median (IQR)/mean	11.7/9.4(4.3)/7.63	14.1/11.6(0.7)/11.31	10.0/10.0(0)/9.5	22.6/17.9(6.8)/ 15.6
Age,y Mean (SD)	75(6.6)	49.2(14.1)	56.7(8.0)	56.7 (9.0)
Sex, M %	42.4	42.5	42.9	47.4
SBP(mmHg), Mean (SD)	152.1(21.3)	139.4(20.2)	138.5(21.7)	126.7 (18.7)
BP medication %	49.8	34.3	27.3	20.0
*Education % high/intermediate/low	14.3/42.7/30.8 (11.6 very low)	34.5/43.9/18.2 (very low 3.5)	38.8/30.4/30.8	6.0/59.5/34.5
Current Smoking, yes %	40.3	42.3	18.1	18.0
Prevalent diseases				
Diabetes%	16.6	7.8	4.8	6.8
Atrial fibrillation %	11.8	NA	2.8	2.1
Coronary heart disease%	NA	NA	3.6	9.1 (CVD)
Heart failure%	NA	NA	3.6	NA
Peripheral artery disease %	NA	NA	1.3	NA
APOE ε4 (none/1/2) %	72.6/25.4/2	64.4/31.3/4.2	66.0/30.2/3.8	77.2/21.2/1.7

* see education description Supplement – Methods – Covariates
NA-not recorded separately, CVD-cardiovascular disease

Supplementary Table 4. Associations of covariates with incident dementia. HR and 95% CI is presented

	FINRISK 97 HR (95%CI)	DILGOM HR (95%CI)	WHITEHALL HR (95%CI)	HEALTH 2000 HR (95%CI)	RS HR (95%CI)	ERF HR (95%CI)	FSH HR (95%CI)	FIXED EFFECTHR (95%CI)	RANDOM EFFECT HR (95%CI)	I2	p.fixed	p.random
Age, years	1.21 (1.17;1.24)	1.25 (1.19;1.31)	1.23 (1.18;1.28)	1.24 (1.17;1.32)	1.12 (1.1;1.14)	1.19 (1.12;1.25)	1.08 (1.07;1.1)	1.12 (1.11;1.13)	1.18 (1.13;1.23)	0.939	<0.001	<0.001
Sex, HR for women	1.6 (1.07;2.38)	1.01 (0.63;1.63)	0.66 (0.44;0.99)	0.71 (0.38;1.33)	1.15 (0.91;1.45)	0.58 (0.25;1.34)	0.98 (0.82;1.17)	1.01 (0.9;1.14)	0.98 (0.79;1.21)	0.555	0.850	0.853
Education level, in three levels compared to the lowest	0.96 (0.76;1.21)	0.6 (0.44;0.82)	1.02 (0.81;1.29)	0.59 (0.37;0.93)	0.87 (0.77;0.99)	0.66 (0.32;1.4)	0.79 (0.58;1.07) college*	0.86 (0.79;0.95)	0.83 (0.7;0.98)	0.564	0.002	0.026
apo E (number of ε4 alleles)	2.84 (2.14;3.77)	2.78 (1.96;3.95)	2.83 (2.11;3.78)	3.77 (2.31;6.15)	2.41 (2.01;2.9)	2.45 (1.35;4.44)	1.58 (1.33;1.89)	2.25 (2.04;2.49)	2.51 (2;3.16)	0.776	<0.001	<0.001
Systolic blood pressure, mmHg	1 (0.99;1.01)	0.99 (0.98;1.01)	1 (0.99;1.01)	1 (0.99;1.02)	1 (1;1.01)	1 (0.98;1.01)	1 (1;1.01)	1 (1;1)	1 (1;1)	0	0.776	0.776
Antihypertensive medication	1.4 (0.91;2.15)	0.88 (0.53;1.46)	0.69 (0.38;1.25)	1.01 (0.52;1.95)	0.9 (0.72;1.12)	NA	1.14 (0.95;1.37)	1.03 (0.91;1.17)	1.02 (0.87;1.2)	0.251	0.620	0.818
Current smoking	1.12 (0.62;2.01)	0.57 (0.21;1.59)	0.77 (0.34;1.78)	1.26 (0.5;3.18)	1.01 (0.9;1.13)	0.49 (0.17;1.36)	0.92 (0.67;1.27)	0.99 (0.89;1.1)	0.99 (0.89;1.1)	0	0.814	0.814
Prevalent												
Diabetes	0.97 (0.52;1.81)	1.64 (0.82;3.28)	1.96 (0.94;4.05)	0.74 (0.17;3.24)	1.13 0.86;1.49)	2.61 (1.04;6.56)	1.1 (0.84;1.43)	1.19 (1.01;1.41)	1.21 (1;1.46)	0.102	0.044	0.053
Atrial fibrillation	1.34 (0.55;3.24)	0.86 (0.33;2.26)	NA	2.35 (0.58;9.59)	NA	NA	1.11 (0.66;1.86)	1.17 (0.79;1.73)	1.17 (0.79;1.73)	0	0.426	0.426
Coronary heart disease	0.88 (0.42;1.86)	0.69 (0.26;1.85)	NA	2.61 (0.99;6.9)	1.15 (0.81;1.63)	NA	NA	1.13 (0.85;1.51)	1.13 (0.76;1.7)	0.293	0.392	0.541
Heart failure	1.01 (0.46;2.2)	2.4 (1.06;5.43)	NA	1.13 (0.3;4.22)	NA	NA	NA	1.45 (0.87;2.44)	1.45 (0.81;2.6)	0.18	0.158	0.213
Peripheral artery disease	0.83 (0.24;2.93)	1.43 (0.33;6.17)	NA	NA	NA	NA	NA	1.05 (0.4;2.73)	1.05 (0.4;2.73)	0	0.921	0.921
Cardiovascular disease	NA	NA	1.24 (0.64;2.42)	NA	NA	NA	1.05 (0.83;1.34)	1.07 (0.86;1.34)	1.07 (0.86;1.34)	0	0.545	0.545

* HR and 95% CI for a college education. Not included in the meta-analysis result. NA - not detected separately.

Due to their size Supplementary Table 5,6,7,8 can be viewed by scanning the code below:

Supplementary Table 5. HRs and 95% CIs of metabolites, lipids and lipoprotein lipids for incident dementia. Adjusted for Model 1. *Metabolite which did not fulfill the proportional hazard assumption (see Methods)

Supplementary Table 6. HRs and 95% CIs of metabolites, lipids and lipoprotein lipids for incident dementia. Adjusted for Model 2.

Supplementary Table 7. HRs and 95% CIs of metabolites, lipid and lipoprotein lipids for incident AD. Adjusted for Model 1

Supplementary Table 8. HRs and 95% CIs of metabolites, lipid and lipoprotein lipids for incident AD. Adjusted for Model 2.

Supplementary Table 9. Replication results of the ten preliminary significant metabolites associated with incident dementia in four separate cohorts. HRs and 95% CIs for incident dementia. Adjusted for Model 1. (Model 2 results are presented in Table 2)

Metabolite, lipoprotein lipid	HEALT2000 HR (95%CI)	ROTTERDAM HR (95%CI)	ERF HR (95%CI)	FHS HR (95%CI)
CREA	1.19 (0.88;1.63)	0.97 (0.88;1.08)	0.88 (0.62;1.24)	0.98 (0.79;1.2)
SFA-FA	1.12 (0.82;1.54)	0.97 (0.88;1.07)	1.2 (0.66;2.16)	NA
ILE	0.92 (0.66;1.3)	0.95 (0.86;1.05)	0.84 (0.55;1.31)	1.09 (0.88;1.35)
LEU	0.99 (0.72;1.35)	0.88 (0.79;0.97)	1.07 (0.71;1.61)	0.98 (0.78;1.23)
VAL	0.95 (0.7;1.29)	0.9 (0.81;0.99)	1.08 (0.71;1.64)	0.9 (0.79;1.22)
L-HDL-CE-%	1.01 (0.73;1.39)	1.02 (0.92;1.12)	0.99 (0.55;1.78)	NA
L-HDL-PL-%	1.06 (0.78;1.44)	1.02 (0.93;1.13)	1.23 (0.71;2.12)	NA
S-VLDL-C	0.86 (0.63;1.19)	0.94 (0.86;1.04)	1.09 (0.6;2)	NA
XL-VLDL-C-%	0.82 (0.62;1.08)	1.02 (0.93;1.12)	0.57 (0.24;1.34)	NA
XL-VLDL-TG-%	1.15 (0.86;1.54)	0.99 (0.89;1.08)	0.76 (0.37;1.54)	NA

Supplementary Table 10. Replication results of the ten preliminary significant metabolites associated with incident dementia in four separate cohorts. HRs and 95% CIs for incident AD. Adjusted for Model 1

Metabolite, lipoprotein lipid	HEALT2000 HR (95%CI)	ROTTERDAM HR (95%CI)	ERF HR (95%CI)	FHS HR (95%CI)
CREA	1.21 (0.87;1.69)	0.95 (0.85;1.08)	0.89 (0.59;1.35)	0.99 (0.77;1.25)
SFA-FA	1.07 (0.76;1.5)	0.98 (0.87;1.09)	1.32 (0.68;2.58)	NA
ILE	1.1 (0.76;1.58)	0.94 (0.84;1.06)	0.89 (0.52;1.51)	1.07 (0.83;1.36)
LEU	1.13 (0.81;1.57)	0.9 (0.8;1.02)	1.05 (0.64;1.72)	0.95 (0.73;1.24)
VAL	1.08 (0.78;1.48)	0.92 (0.82;1.03)	1.08 (0.66;1.78)	0.93 (0.72;1.2)
L-HDL-CE-%	0.96 (0.68;1.37)	1.07 (0.96;1.2)	0.66 (0.33;1.32)	NA
L-HDL-PL-%	1.12 (0.8;1.56)	0.98 (0.88;1.1)	1.41 (0.72;2.77)	NA
S-VLDL-C	0.91 (0.64;1.3)	0.97 (0.87;1.08)	1.04 (0.51;2.13)	NA
XL-VLDL-C-%	0.75 (0.56;1.01)	1.05 (0.94;1.18)	0.67 (0.26;1.73)	NA
XL-VLDL-TG-%	1.23 (0.9;1.69)	0.99 (0.88;1.11)	0.56 (0.25;1.29)	NA

Supplementary Table 11. Replication results of the ten preliminary significant metabolites associated with incident dementia in four separate cohorts. HRs and 95% CIs for incident AD. Adjusted for Model 2.

Metabolite, lipoprotein lipid	HEALT2000 HR (95%CI)	ROTTERDAM HR (95%CI)	ERF HR (95%CI)	FHS HR (95%CI)
CREA	1.21 (0.86;1.7)	0.94 (0.81;1.08)	0.83 (0.54;1.26)	0.97 (0.75;1.25)
SFA-FA	1.05 (0.74;1.49)	0.94 (0.83;1.07)	1.3 (0.62;2.76)	NA
ILE	1.03 (0.7;1.52)	0.87 (0.76;1)	0.71 (0.41;1.21)	1.04 (0.8;1.34)
LEU	1.08 (0.76;1.52)	0.85 (0.74;0.98)	0.92 (0.56;1.51)	0.93 (0.71;1.22)
VAL	1.05 (0.76;1.47)	0.86 (0.75;0.98)	0.85 (0.51;1.41)	0.93 (0.72;1.2)
L-HDL-CE-%	0.94 (0.66;1.35)	1.1 (0.97;1.25)	0.75 (0.38;1.48)	NA
L-HDL-PL-%	1.15 (0.82;1.62)	1.02 (0.89;1.15)	1.41 (0.76;2.65)	NA
S-VLDL-C	0.95 (0.67;1.36)	0.94 (0.83;1.07)	1.08 (0.54;2.16)	NA
XL-VLDL-C-%	0.77 (0.56;1.06)	1.08 (0.95;1.23)	0.78 (0.34;1.83)	NA
XL-VLDL-TG-%	1.17 (0.84;1.64)	0.98 (0.86;1.12)	0.63 (0.3;1.35)	NA

Supplementary Table 12. Results of the ten preliminary significant metabolites associated with incident dementia. Discovery and replication cohorts combined. HRs and 95% CIs for incident dementia. Adjusted for Model 1.

Metabolite, Lipoprotein lipid	FINRISK 97 HR (95%CI)	DILGOM HR (95%CI)	WHITEHALL HR (95%CI)	EGCUT HR (95%CI)	HEALT 2000 HR (95%CI)	ROTTERDAM HR (95%CI)	ERF HR (95%CI)	FSH HR (95%CI)	FIXED EFFECT, HR 95%CI	RANDOM EFFECT, HR 95%CI	I2	P fixed	P random
CREA	0.92 (0.75;1.13)	0.81 (0.6;1.09)	0.64 (0.5;0.81)	0.86 (0.59;1.25)	1.19 (0.88;1.63)	0.97 (0.88;1.08)	0.88 (0.62;1.24)	0.98 (0.79;1.2)	0.92 (0.86;0.99)	0.9 (0.8;1.01)	0.501	0.028	0.077
SFA-FA	1.14 (0.92;1.42)	1.12 (0.84;1.49)	1.25 (1.03;1.52)	1.62 (1.18;2.21)	1.12 (0.82;1.54)	0.97 (0.88;1.07)	1.2 (0.66;2.16)	NA	1.07 (1;1.15)	1.16 (1.01;1.33)	0.575	0.064	0.037
ILE	0.92 (0.71;1.18)	0.84 (0.63;1.13)	0.71 (0.58;0.88)	0.64 (0.43;0.97)	0.92 (0.66;1.3)	0.95 (0.86;1.05)	0.84 (0.55;1.31)	1.09 (0.88;1.35)	0.91 (0.85;0.98)	0.88 (0.79;0.99)	0.412	0.010	0.031
LEU	0.89 (0.68;1.16)	0.75 (0.55;1.04)	0.73 (0.59;0.89)	0.54 (0.34;0.84)	0.99 (0.72;1.35)	0.88 (0.79;0.97)	1.07 (0.71;1.61)	0.98 (0.78;1.23)	0.86 (0.8;0.93)	0.85 (0.77;0.95)	0.346	<0.001	0.004
VAL	0.83 (0.64;1.07)	0.7 (0.51;0.96)	0.86 (0.72;1.04)	0.59 (0.41;0.86)	0.95 (0.7;1.29)	0.9 (0.81;0.99)	1.08 (0.71;1.64)	0.98 (0.79;1.22)	0.88 (0.82;0.94)	0.87 (0.8;0.95)	0.191	<0.001	0.002
L-HDL-CE-%	1.05 (0.85;1.29)	1.29 (0.95;1.76)	1.43 (1.15;1.78)	1.35 (0.93;1.96)	1.01 (0.73;1.39)	1.02 (0.92;1.12)	0.99 (0.55;1.78)	NA	1.09 (1.01;1.17)	1.14 (1.01;1.29)	0.452	0.022	0.039
L-HDL-PL-%	0.94 (0.75;1.17)	0.81 (0.58;1.13)	0.76 (0.62;0.93)	0.66 (0.44;0.97)	1.06 (0.78;1.44)	1.02 (0.93;1.13)	1.23 (0.71;2.12)	NA	0.95 (0.89;1.03)	0.91 (0.79;1.04)	0.524	0.208	0.164
S-VLDL-C	1.02 (0.79;1.32)	0.84 (0.6;1.17)	0.72 (0.6;0.86)	0.77 (0.54;1.1)	0.86 (0.63;1.19)	0.94 (0.86;1.04)	1.09 (0.6;2)	NA	0.9 (0.83;0.96)	0.87 (0.78;0.98)	0.339	0.003	0.019
XL-VLDL-C-%	0.92 (0.67;1.26)	1.07 (0.73;1.57)	1.29 (1.08;1.53)	1.34 (1.06;1.69)	0.82 (0.62;1.08)	1.02 (0.93;1.12)	0.57 (0.24;1.34)	NA	1.07 (0.99;1.14)	1.06 (0.91;1.23)	0.606	0.090	0.458
XL-VLDL-TG-%	1.06 (0.77;1.46)	0.89 (0.6;1.32)	0.79 (0.68;0.91)	0.83 (0.63;1.09)	1.15 (0.86;1.54)	0.99 (0.89;1.08)	0.76 (0.37;1.54)	NA	0.93 (0.87;1)	0.92 (0.82;1.04)	0.41	0.048	0.175

Supplementary Table 13. Results of the ten preliminary significant metabolites associated with incident dementia. Discovery and replication cohorts combined. HRs and 95% CIs for incident dementia. Adjusted for Model 2.

Metabolite, lipoprotein lipid	FINRISK 97 HR (95%CI)	DILGOM HR (95%CI)	WHITEHALL HR (95%CI)	EGCUT HR (95%CI)	HEALT 2000 HR (95%CI)	ROTTERDAM HR (95%CI)	ERF HR (95%CI)	FSH HR (95%CI)	FIXED EFFECT, HR 95%CI	RANDOM EFFECT, HR 95%CI	I2	P fixed	P random
CREA	0.92 (0.75;1.12)	0.78 (0.58;1.05)	0.64 (0.5;0.82)	0.85 (0.59;1.24)	1.19 (0.87;1.63)	0.94 (0.84;1.06)	0.82 (0.57;1.18)	0.97 (0.78;1.2)	0.9 (0.83;0.97)	0.88 (0.79;0.99)	0.434	0.007	0.034
SFA-FA	1.13 (0.9;1.4)	1.13 (0.84;1.52)	1.31 (1.07;1.6)	1.6 (1.17;2.19)	1.08 (0.78;1.49)	0.94 (0.84;1.04)	1.15 (0.61;2.2)	NA	1.07 (0.99;1.15)	1.15 (0.99;1.35)	0.639	0.106	0.068
ILE	0.9 (0.7;1.16)	0.8 (0.59;1.08)	0.72 (0.58;0.9)	0.64 (0.42;0.96)	0.87 (0.61;1.25)	0.9 (0.8;1.01)	0.66 (0.42;1.03)	1.05 (0.84;1.3)	0.87 (0.8;0.94)	0.85 (0.76;0.94)	0.293	<0.001	0.002
LEU	0.86 (0.66;1.13)	0.72 (0.52;1)	0.75 (0.6;0.92)	0.54 (0.34;0.84)	0.94 (0.69;1.3)	0.84 (0.74;0.94)	0.91 (0.6;1.37)	0.94 (0.75;1.19)	0.83 (0.76;0.9)	0.83 (0.76;0.9)	0.042	<0.001	<0.001
VAL	0.81 (0.62;1.06)	0.64 (0.46;0.89)	0.87 (0.72;1.06)	0.59 (0.4;0.88)	0.93 (0.68;1.26)	0.86 (0.76;0.96)	0.84 (0.55;1.29)	0.96 (0.76;1.19)	0.84 (0.78;0.91)	0.84 (0.77;0.92)	0.073	<0.001	<0.001
L-HDL-CE-%	1.06 (0.87;1.3)	1.25 (0.91;1.71)	1.4 (1.13;1.75)	1.39 (0.96;2.02)	0.99 (0.71;1.38)	1.04 (0.94;1.17)	1.08 (0.61;1.91)	NA	1.11 (1.03;1.21)	1.14 (1.02;1.27)	0.264	0.008	0.017
L-HDL-PL-%	0.94 (0.75;1.17)	0.85 (0.6;1.19)	0.78 (0.63;0.96)	0.65 (0.43;0.98)	1.09 (0.8;1.49)	1.04 (0.94;1.16)	1.25 (0.74;2.11)	NA	0.96 (0.89;1.04)	0.93 (0.81;1.06)	0.486	0.363	0.266
S-VLDL-C	1.01 (0.78;1.3)	0.87 (0.62;1.23)	0.72 (0.59;0.87)	0.76 (0.53;1.08)	0.89 (0.64;1.23)	0.91 (0.82;1.01)	1.08 (0.59;1.97)	NA	0.87 (0.81;0.94)	0.87 (0.79;0.95)	0.147	<0.001	0.003
XL-VLDL-C-%	0.94 (0.68;1.29)	1.06 (0.72;1.56)	1.28 (1.07;1.53)	1.35 (1.07;1.71)	0.87 (0.65;1.16)	1.07 (0.95;1.19)	0.67 (0.3;1.49)	NA	1.11 (1.02;1.2)	1.1 (0.96;1.24)	0.452	0.013	0.165
XL-VLDL-TG-%	1.03 (0.75;1.42)	0.88 (0.59;1.32)	0.78 (0.67;0.91)	0.81 (0.62;1.07)	1.06 (0.78;1.44)	0.98 (0.88;1.1)	0.78 (0.4;1.52)	NA	0.91 (0.84;0.99)	0.9 (0.81;1.01)	0.274	0.019	0.064

Supplementary Table 14. Results of the ten preliminary significant metabolites associated with incident dementia. Discovery and replication cohorts combined. HRs and 95%CIs for incident AD. Adjusted for Model 1. EGCUT did not record AD cases separately.

Metabolite, lipoprotein lipid	FINRISK 97 HR (95%CI)	DILGOM HR (95%CI)	WHITEHALL HR (95%CI)	HEALT 2000 HR (95%CI)	ROTTERDAM HR (95%CI)	ERF HR (95%CI)	FSH HR (95%CI)	FIXED EFFECT, HR 95%CI	RANDOM EFFECT, HR 95%CI	I2	P fixed	P random
CREA	0.9 (0.7;1.14)	0.82 (0.6;1.12)	0.66 (0.41;1.05)	1.21 (0.86;1.7)	0.94 (0.81;1.08)	0.83 (0.54;1.26)	0.97 (0.75;1.25)	0.94 (0.86;1.03)	0.94 (0.86;1.03)	0	0.167	0.167
SFA-FA	0.9 (0.71;1.15)	0.85 (0.62;1.15)	0.64 (0.41;1)	1.21 (0.87;1.69)	0.95 (0.85;1.08)	0.89 (0.59;1.35)	0.99 (0.77;1.25)	1.01 (0.93;1.1)	1.01 (0.93;1.1)	0	0.808	0.808
ILE	1.06 (0.83;1.37)	1.09 (0.81;1.47)	1.02 (0.73;1.41)	1.07 (0.76;1.5)	0.98 (0.87;1.09)	1.32 (0.68;2.58)	NA	0.95 (0.87;1.03)	0.95 (0.87;1.03)	0	0.203	0.203
LEU	1.04 (0.78;1.4)	0.83 (0.61;1.12)	0.67 (0.46;0.98)	1.1 (0.76;1.58)	0.94 (0.84;1.06)	0.89 (0.52;1.51)	1.07 (0.83;1.36)	0.91 (0.83;1)	0.91 (0.83;1)	0	0.043	0.043
VAL	1 (0.73;1.35)	0.75 (0.54;1.05)	0.74 (0.52;1.05)	1.13 (0.81;1.57)	0.9 (0.8;1.02)	1.05 (0.64;1.72)	0.95 (0.73;1.24)	0.91 (0.83;0.99)	0.91 (0.83;0.99)	0	0.035	0.035
L-HDL-CE-%	0.93 (0.69;1.26)	0.7 (0.5;0.97)	0.8 (0.59;1.1)	1.08 (0.78;1.48)	0.92 (0.82;1.03)	1.08 (0.66;1.78)	0.93 (0.72;1.2)	1.1 (1;1.2)	1.11 (0.98;1.25)	0.186	0.048	0.096
L-HDL-PL-%	1.13 (0.89;1.44)	1.23 (0.9;1.7)	1.53 (1.03;2.25)	0.96 (0.68;1.37)	1.07 (0.96;1.2)	0.66 (0.33;1.32)	NA	0.96 (0.87;1.05)	0.94 (0.81;1.09)	0.387	0.360	0.409
S-VLDL-C	0.96 (0.74;1.24)	0.84 (0.59;1.2)	0.61 (0.41;0.89)	1.12 (0.8;1.56)	0.98 (0.88;1.1)	1.41 (0.72;2.77)	NA	0.94 (0.86;1.03)	0.94 (0.86;1.03)	0	0.212	0.212
XL-VLDL-C-%	0.98 (0.73;1.32)	0.86 (0.61;1.21)	0.76 (0.54;1.07)	0.91 (0.64;1.3)	0.97 (0.87;1.08)	1.04 (0.51;2.13)	NA	1.03 (0.94;1.14)	1.02 (0.86;1.2)	0.437	0.472	0.861
XL-VLDL-TG-%	0.94 (0.65;1.37)	1.12 (0.76;1.65)	1.36 (1;1.83)	0.75 (0.56;1.01)	1.05 (0.94;1.18)	0.67 (0.26;1.73)	NA	0.96 (0.88;1.06)	0.95 (0.82;1.1)	0.352	0.441	0.477

Supplementary Table 15. Results of the ten preliminary significant metabolites associated with incident dementia. Discovery and replication cohorts combined. HRs and 95% CIs for incident AD. Adjusted for Model 2. EGCUT did not record AD cases separately.

Metabolite, lipoprotein lipid	FINRISK 97 HR (95%CI)	DILGOM HR (95%CI)	WHITEHALL HR (95%CI)	HEALT 2000 HR (95%CI)	ROTTERDAM HR (95%CI)	ERF HR (95%CI)	FSH HR (95%CI)	FIXED EFFECT, HR 95%CI	RANDOM EFFECT, HR 95%CI	I2	p.fixed	p.random
CREA	0.9 (0.7;1.14)	0.82 (0.6;1.12)	0.66 (0.41;1.05)	1.21 (0.86;1.7)	0.94 (0.81;1.08)	0.83 (0.54;1.26)	0.97 (0.75;1.25)	0.92 (0.84;1.01)	0.92 (0.84;1.01)	0	0.092	0.092
SFA-FA	1.07 (0.83;1.38)	1.1 (0.81;1.49)	1.11 (0.78;1.58)	1.05 (0.74;1.49)	0.94 (0.83;1.07)	1.3 (0.62;2.76)	NA	1 (0.91;1.1)	1 (0.91;1.1)	0	0.993	0.993
ILE	1.03 (0.77;1.39)	0.79 (0.58;1.08)	0.72 (0.48;1.08)	1.03 (0.7;1.52)	0.87 (0.76;1)	0.71 (0.41;1.21)	1.04 (0.8;1.34)	0.89 (0.81;0.98)	0.89 (0.81;0.98)	0	0.022	0.022
LEU	0.99 (0.72;1.36)	0.72 (0.51;1.01)	0.81 (0.55;1.19)	1.08 (0.76;1.52)	0.85 (0.74;0.98)	0.92 (0.56;1.51)	0.93 (0.71;1.22)	0.88 (0.79;0.97)	0.88 (0.79;0.97)	0	0.009	0.009
VAL	0.93 (0.68;1.27)	0.64 (0.46;0.9)	0.87 (0.62;1.23)	1.05 (0.76;1.47)	0.86 (0.75;0.98)	0.85 (0.51;1.41)	0.93 (0.72;1.2)	0.87 (0.79;0.96)	0.87 (0.79;0.96)	0	0.004	0.004
L-HDL-CE-%	1.15 (0.9;1.47)	1.2 (0.87;1.65)	1.45 (0.97;2.16)	0.94 (0.66;1.35)	1.1 (0.97;1.25)	0.75 (0.38;1.48)	NA	1.12 (1.01;1.23)	1.12 (1.01;1.23)	0	0.029	0.029
L-HDL-PL-%	0.96 (0.74;1.24)	0.89 (0.62;1.26)	0.66 (0.44;0.98)	1.15 (0.82;1.62)	1.02 (0.89;1.15)	1.41 (0.76;2.65)	NA	0.99 (0.9;1.09)	0.98 (0.85;1.12)	0.256	0.813	0.717
S-VLDL-C	0.97 (0.72;1.31)	0.89 (0.63;1.27)	0.73 (0.52;1.04)	0.95 (0.67;1.36)	0.94 (0.83;1.07)	1.08 (0.54;2.16)	NA	0.93 (0.84;1.02)	0.93 (0.84;1.02)	0	0.127	0.127
XL-VLDL-C-%	0.95 (0.66;1.39)	1.09 (0.73;1.61)	1.32 (0.97;1.81)	0.77 (0.56;1.06)	1.08 (0.95;1.23)	0.78 (0.34;1.83)	NA	1.05 (0.95;1.16)	1.03 (0.89;1.19)	0.254	0.361	0.658
XL-VLDL-TG-%	1.03 (0.71;1.5)	0.86 (0.57;1.3)	0.77 (0.59;1.02)	1.17 (0.84;1.64)	0.98 (0.86;1.12)	0.63 (0.3;1.35)	NA	0.95 (0.86;1.06)	0.95 (0.85;1.07)	0.086	0.364	0.373

Supplementary Table 16. Inter-correlation table of eight preliminary significant metabolites associated with incident dementia. All correlations were statistically significant (p<0.001, Pearson, pairwise deletion).

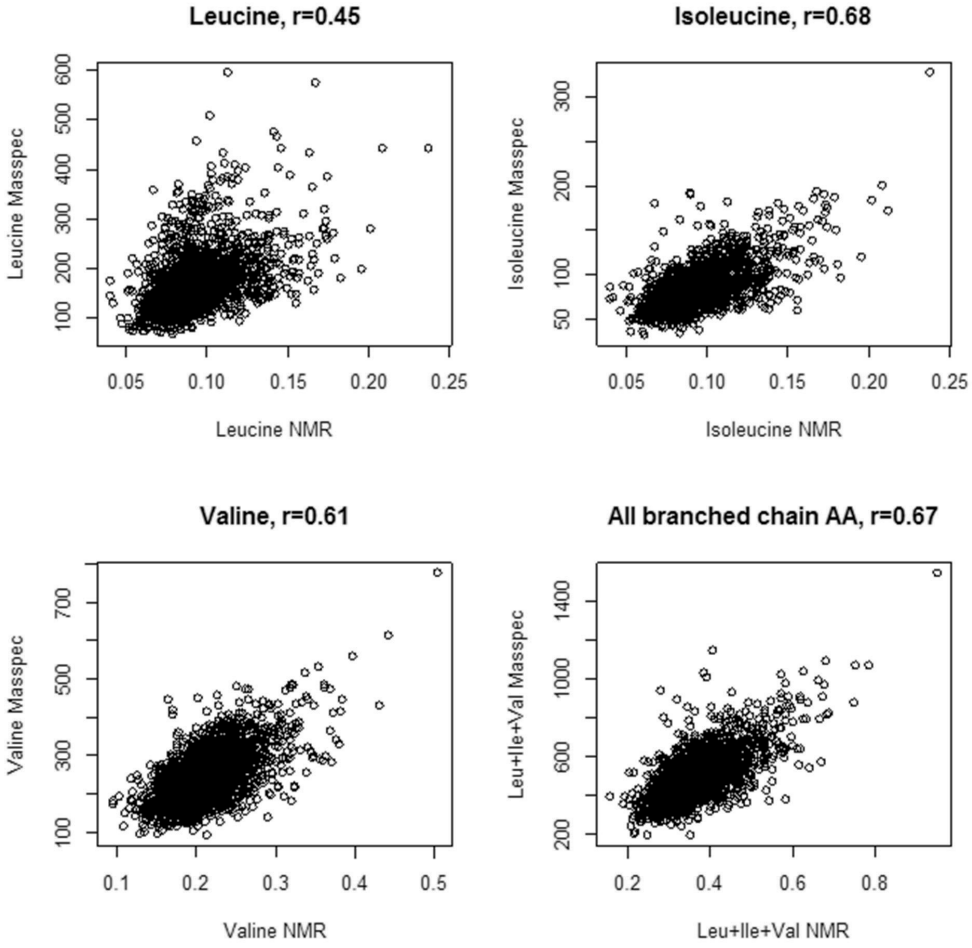
	CREA	**SFA-FA**	**ILE**	**LEU**	**VAL**	**L-HDL-CE-%**	**L-HDL-PL-%**	**S-VLDL-C**	**XL-VLDL-C-%**	**XL-VLDL-TG-%**
CREA	1.00									
SFA-FA	-0.05	1.00								
ILE	0.35	0.34	1.00							
LEU	0.30	0.40	0.92	1.00						
VAL	0.26	0.36	0.82	0.89	1.00					
L-HDL-CE-%	-0.20	0.15	-0.17	-0.09	-0.04	1.00				
L-HDL-PL-%	0.13	-0.29	-0.04	-0.10	-0.15	-0.84	1.00			
S-VLDL-C	0.19	0.32	0.62	0.67	0.55	-0.24	-0.10	1.00		
XL-VLDL-C-%	-0.11	0.31	0.07	0.22	0.30	0.39	-0.48	0.39	1.00	
XL-VLDL-TG-%	0.12	-0.31	-0.07	-0.22	-0.29	-0.35	0.44	-0.41	-0.97	1.00

Supplementary Table 17. HRs and 95% CIs of the discovered metabolites and lipoprotein lipids for death from any cause. Adjusted for Model 1.

Metabolite, lipoprotein lipid	FINRISK 1997 HR (95%CI)	DILGOM HR (95%CI)	Health 2000 HR (95%CI)	FIXED EFFECT, HR (95%CI)	RANDOM EFFECT, HR (95%CI)	I2	P fixed	P random
CREA	0.94 (0.85;1.05)	0.86 (0.71;1.03)	1.05 (0.85;1.28)	0.94 (0.87;1.02)	0.94 (0.87;1.02)	0.041	0.148	0.163
SFA-FA	1.13 (1.02;1.25)	1.26 (1.05;1.5)	1.02 (0.84;1.24)	1.14 (1.05;1.23)	1.14 (1.03;1.25)	0.193	0.002	0.009
ILE	1.08 (0.96;1.21)	0.95 (0.79;1.14)	0.9 (0.73;1.11)	1.01 (0.92;1.11)	1 (0.9;1.11)	0.242	0.804	0.990
LEU	1.07 (0.94;1.21)	0.89 (0.72;1.08)	0.83 (0.68;1.02)	0.97 (0.88;1.06)	0.94 (0.8;1.1)	0.619	0.509	0.435
VAL	1 (0.88;1.13)	0.79 (0.64;0.97)	0.86 (0.7;1.05)	0.92 (0.84;1.01)	0.89 (0.77;1.04)	0.553	0.079	0.137
L-HDL-CE-%	0.98 (0.89;1.08)	1.16 (0.95;1.41)	1.11 (0.91;1.35)	1.03 (0.95;1.11)	1.05 (0.94;1.17)	0.339	0.505	0.396
L-HDL-PL-%	0.94 (0.85;1.04)	0.78 (0.64;0.96)	0.89 (0.74;1.08)	0.9 (0.83;0.98)	0.89 (0.81;0.99)	0.198	0.017	0.030
S-VLDL-C	1.04 (0.92;1.18)	0.91 (0.73;1.12)	0.79 (0.65;0.96)	0.95 (0.87;1.05)	0.92 (0.78;1.09)	0.645	0.304	0.338
XL-VLDL-C-%	0.88 (0.75;1.03)	0.99 (0.76;1.28)	1.03 (0.86;1.24)	0.95 (0.85;1.06)	0.95 (0.85;1.06)	0	0.341	0.341
XL-VLDL-TG-%	1.14 (0.97;1.34)	1.08 (0.83;1.4)	0.99 (0.82;1.19)	1.07 (0.96;1.2)	1.07 (0.96;1.2)	0	0.201	0.201

Supplementary Table 18. HRs and 95% CIs of the discovered metabolites and lipoprotein lipids for death from any cause. Adjusted for Model 2.

Metabolite, lipoprotein lipid	FINRISK 1997 HR (95%CI)	DILGOM HR (95%CI)	Health 2000 HR (95%CI)	FIXED EFFECT HR (95%CI)	RANDOM EFFECT HR (95%CI)	I2	P Fixed	P random
CREA	0.92 (0.83;1.02)	0.87 (0.73;1.04)	1.02 (0.83;1.26)	0.92 (0.85;1)	0.92 (0.85;1)	0	0.050	0.050
SFA-FA	1.09 (0.98;1.21)	1.22 (1.01;1.46)	1.01 (0.83;1.22)	1.1 (1.01;1.19)	1.1 (1.01;1.19)	0	0.027	0.027
ILE	1.04 (0.93;1.17)	0.89 (0.74;1.08)	0.86 (0.68;1.07)	0.98 (0.89;1.07)	0.95 (0.83;1.08)	0.429	0.589	0.457
LEU	1.03 (0.91;1.17)	0.84 (0.69;1.03)	0.79 (0.64;0.98)	0.94 (0.85;1.03)	0.9 (0.75;1.07)	0.671	0.181	0.234
VAL	0.97 (0.86;1.1)	0.74 (0.6;0.91)	0.81 (0.65;1)	0.89 (0.81;0.97)	0.85 (0.71;1.01)	0.661	0.012	0.071
L-HDL-CE-%	0.97 (0.89;1.07)	1.22 (1;1.49)	1.14 (0.93;1.39)	1.03 (0.96;1.11)	1.08 (0.93;1.26)	0.616	0.425	0.299
L-HDL-PL-%	0.96 (0.87;1.07)	0.74 (0.6;0.91)	0.87 (0.72;1.06)	0.91 (0.83;0.98)	0.87 (0.74;1.01)	0.624	0.019	0.075
S-VLDL-C	1.03 (0.91;1.17)	0.89 (0.72;1.1)	0.83 (0.68;1.02)	0.95 (0.87;1.05)	0.93 (0.81;1.07)	0.439	0.343	0.327
XL-VLDL-C-%	0.94 (0.81;1.1)	0.99 (0.75;1.32)	1.12 (0.92;1.35)	1.01 (0.9;1.13)	1.01 (0.9;1.13)	0	0.904	0.904
XL-VLDL-TG-%	1.05 (0.89;1.23)	1.09 (0.83;1.44)	0.9 (0.74;1.09)	1 (0.89;1.12)	1 (0.89;1.12)	0	0.975	0.975



Supplementary Figure 1. Correlation plots of BCAAs determined by NMR (Brainshake Ltd) and LC-MS (Biocrates p180). A subset of 2,638 serum samples from FINRISK 1997.

Chapter 5.3

Circulating metabolites and general cognitive ability and dementia: Evidence from 11 cohort studies

Sven J. van der Lee, Charlotte E. Teunissen, René Pool, Martin J. Shipley, Alexander Teumer, Vincent Chouraki, Debora Melo van Lent, Juho Tynkkynen, Krista Fischer, Jussi Hernesniemi, Andres Metspalu, Archana Singh-Manoux, Aswin Verhoeven, Gonneke Willemsen, Francien A. de Leeuw, Holger Wagner, Jenny van Dongen, Johannes Hertel, Kathrin Budde, Ko Willems van Dijk, Leonie Weinhold, M. Arfan Ikram, Maik Pietzner, Markus Perola, Michael Wagner, Nele Friedrich, P.E. Slagboom, Philip Scheltens, Qiong Yang, Robert E. Gertzen, Sarah Egert, Shuo Li, Thomas Hankemeier, Catharine E.M. van Beijsterveldt, Vasan Ramachandran, Wolfgang Maier, Carel F.W. Peeters, Hans Jörgen Grabe, Alfredo Ramirez, Sudha Seshadri, Toomas Haller, Mika Kivimäki, Veikko Salomaa, Ayşe Demirkan, Dorret Boomsma, Wiesje M. van der Flier, Najaf Amin, Cornelia M. van Duijn

This chapter is submitted

Abstract

Introduction: Identifying circulating metabolites that are associated with cognition and dementia may improve our understanding of the pathogenesis of dementia and provide crucial read-outs for preventive and therapeutic interventions.

Methods: We studied 299 metabolites in relation to cognition (general cognitive ability) in two discovery cohorts (N-total=5,658). Metabolites significantly associated with cognition after adjusting for multiple testing were replicated in four independent cohorts (N-total=6,652) and the associations with dementia and Alzheimer's disease (N=25,872) and lifestyle factors (N=5,168) were examined. Next, we examined whether the 15 metabolites associated with general cognitive ability were associated with dementia. We compared (maximum) 1990 dementia patients, of whom 1,356 were AD cases, with 23,882 controls.

Results: We discovered and replicated 15 metabolites associated with cognition including sub-fractions of high-density lipoprotein, docosahexaenoic acid, ornithine, glutamine and glycoprotein acetyls. These associations were independent of classical risk factors including HDL-cholesterol, LDL-cholesterol, triglycerides, glucose and *APOE* genotypes. Six of the cognition-associated metabolites were related to the risk of dementia and lifestyle factors.

Discussion: Circulating metabolites were consistently associated with cognition, dementia and lifestyle factors, opening new avenues for prevention of cognitive decline and dementia.

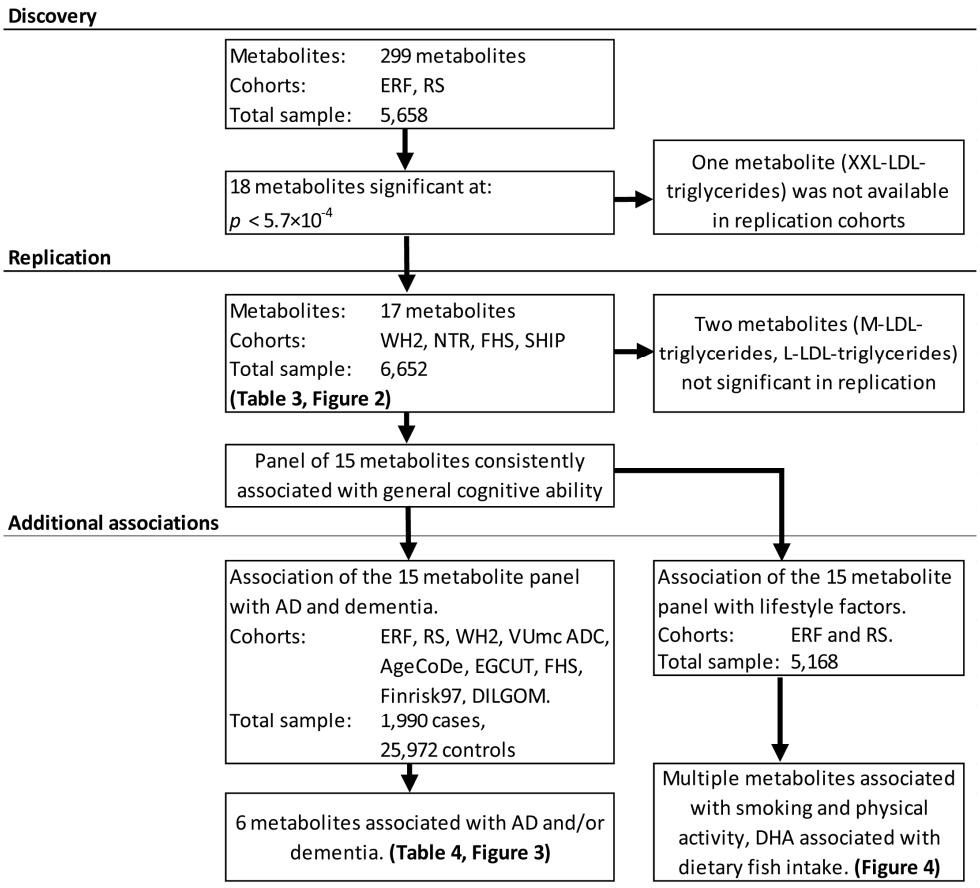
Introduction

Cognitive function is an important determinant of health and well-being and a key component of the dementia spectrum, including Alzheimer disease (AD), the most common cause of dementia.[1] Vascular dysfunction and metabolic dysregulation contribute to impairment in cognitive performance.[2] Clinical and population-based studies suggest a relationship of cognitive function with midlife hypertension, high blood levels of total cholesterol (TC), low density lipoprotein-cholesterol (LDL-C), triglycerides and glucose and low levels high density lipoprotein-cholesterol (HDL-C).[3-5] The recent decrease in incidence of dementia in longitudinal studies has been attributed to improved control of vascular and metabolic factors.[6-9] These findings have fueled speculation that discovery of other circulating metabolites influencing cognition and future dementia may not only improve our understanding of the determinants of cognition, but may also facilitate prevention through interventions on lifestyle factors and dedicated medication.[10] Previous studies have shown circulating metabolites in blood (e.g. lipoproteins, amino acids, fatty acids, and other small molecules) to be associated with cognitive function and conversion from normal cognition to dementia or AD.[11-17] However, these studies were relatively small and findings have not been replicated,[15,18] emphasizing the need for studies in large well-characterized populations where findings are replicated.[10,19]

We performed a comprehensive metabolic analysis to study the role of circulating metabolites in cognitive function. Discovery of novel measures associated with cognitive function was performed in two large population-based studies in the Netherlands, the Rotterdam Study and the Erasmus Rucphen Family study (ERF). We determined whether the associations were independent of known vascular and metabolic risk factors. Metabolites independently associated with cognition were replicated in independent cohort studies and their relation to the risk of dementia and Alzheimer disease was validated in 8 cohort studies. Finally, we assessed whether lifestyle factors, including dietary fish intake, smoking and physical activity, were associated with the identified metabolites.

Methods

For a schematic overview of the analysis setup see **Figure 1**.



Study names: Erasmus Rucphen Family study (ERF); Rotterdam Study (RS); Whitehall II (WHII); Netherlands Twin Registry (NTR); Study of Health in Pomerania - Trend (SHIP-Trend); Framingham Heart Study (FHS); VUmc Amsterdam Dementia Cohort (VUmc ADC); The National FINRISK study(Finrisk97); Dietary, Lifestyle and Genetic determinants of Obesity and Metabolic Syndrome (DILGOM); Estionian biobank (EGCUT); German Study on Ageing, Cognition, and Dementia (AgeCoDe).

Discovery and replication populations for research of cognitive function

Metabolomics profiling in multiple cohorts from the Netherlands was done as part of the 4th Rainbow Project of the BioBanking for Medical Research Infrastructure of the Netherlands (BBMRI-NL). These include the two discovery cohorts (ERF and the Rotterdam Study). A short description of the cohort studies included in this paper can

be found in **Supplementary Table 1.** ERF is a prospective family-based study (ERF, N=2,683) from the Southwest of the Netherlands, and the Rotterdam Study is a prospective population-based cohort study that started in 1990 in Ommoord, a district of Rotterdam. In this analysis we used the fourth wave of the baseline cohort (N=2,975). Replication cohorts included the Netherlands Twin Register (NTR, N=338; also part of the 4th Rainbow Project of BBMRI-NL), the Whitehall II study (WHII, N=4,612)[20], the Framingham Heart Study (FHS, N=2,356) and the Study of Health in Pomerania - Trend (SHIP-Trend, N=944).

Cohorts for extrapolation to dementia and Alzheimer's disease

Dementia and Alzheimer's disease (AD) was assessed in eight cohorts; the ERF study, Rotterdam study, a series of dementia patients and controls from the VUMC Amsterdam Dementia Cohort metabolically characterized as part of the 4th Rainbow Project of BBMRI-NL (N= 1,306)[21], two cohorts used in the replication of cognitive findings (WHII=4,612, FHS=2,356), FINRISK97 (the National FINRISK Studies 1997; N=7,517), DILGOM (Dietary, Lifestyle and Genetic determinants of Obesity and Metabolic Syndrome; N=4,788), the Estonian biobank (EGCUT; N=2,572) and AgeCoDe (The German Study on Ageing, Cognition, and Dementia, N=310).

Assessment of cognitive function and dementia

Participants underwent cognitive tests using a highly variable battery of assessments which varied across studies; details on the cognitive tests used can be found in **Supplementary Table 2.** Cognitive function tests were assessed at the same time point in all studies. To enable meta-analyses of results from the heterogeneous set of tests efficiently we constructed a general cognitive ability score to capture information from a wide variety of cognitive tests reliably into a single cognitive measure.[22,23] A person's general cognitive ability was calculated by principal component analysis of the different cognitive tests, with the first principal component being the measure representing general cognitive ability.[22,24,25] General cognitive ability can be reliably estimated over the life course [26] and shows a decline at older age.[26,27] Additionally it has been shown that general cognitive ability derived from different cognitive test batteries are very similar.[22,23] To ensure comparability for the general cognitive ability in our study only

studies were included that had cognitive tests covering at least three different cognitive domains. The domains covered are shown in **Supplementary Table 3**. General cognitive ability accounted for between 35-58% of variance in cognitive tests in various studies (**Supplementary Table 3).** Correlations between the individual test measures and the derived principal component (loadings) by study are shown in **Supplementary Table 3.** In all studies the general cognitive ability had a high correlation with multiple single cognitive measures, showing the factor was not driven by a single measure.

Details on the ascertainment of dementia and AD for the various cohorts can be found in **Supplementary Table 2.** The diagnosis of dementia is based on continuous follow-up health records in ERF, WHII, EGCUT, Finrisk97, and DILGOM. Studies that additionally used periodic visits to a research center were RS, FHS and AgeCoDe. The ascertainment of dementia and AD in the VUMC Amsterdam Dementia Cohort was done by clinical visits of participants.

Assessment of genetic and environmental factors

In the two discovery cohorts, ERF and Rotterdam Study, *APOE* ε4 genotypes were determined by direct genotyping.[28,29] In both studies, lifestyle factors, including smoking (current vs. past and never smokers), physically activity (yes/no) and dietary fish (oil) intake,[30] were ascertained using questionnaires as described previously.[31,32] Glucose, TC, HDL-C, LDL-C and triglycerides were measured in mainly fasting blood samples by standard procedures;[33,34] further details in **Supplementary Table 1.** Multiple metabolites associated with smoking and physical activity, DHA associated with fish intake. **(Figure 4)**

Assessment of blood metabolites

In the ERF and Rotterdam Study the metabolic biomarkers were quantified from fasted EDTA plasma samples using high-throughput proton nuclear Nuclear Magnetic Resonance (NMR) metabolomics (Nightingale Ltd, Helsinki, Finland). This method provides simultaneous quantification of metabolites, i.e. routine lipids, lipoprotein subclass profiling with lipid concentrations within 14 subclasses, fatty acid composition, and various low-molecular weight metabolites including amino acids, ketone bodies

and gluconeogenesis-related metabolites in molar concentration units. Details of the experimentation and applications of this NMR metabolomics platform have been described previously.[35,36] Metabolomics measurements of the ERF study further included two NMR experiments.[36,37] If a metabolite was measured in a study by multiple experiments, the experiment measuring the largest number of samples was used. In total, 299 unique metabolite concentrations were measured in ERF and of these 242 metabolites were also available in the Rotterdam study. The summary statistics of metabolomics measurements in the discovery cohorts are shown in **Supplementary Table 4**. The cohorts used for replication of the cognitive findings and extrapolation to dementia used NMR-based platforms or mass spectrometry (MS) techniques (**Supplementary Table 1**). The Nightingale NMR-platform was also used in: NTR, VUMC, EGCUT, WHII, FINRISK97 and DILGOM. In NTR additional NMR experiments [36,37] were measured and again the experiment with the largest number of observations was used. Measurements of cognitive function and blood drawn for metabolite measurements (same day/visit to the study center) were concurrent in all metabolite measurements from our discovery and 73.6% of the samples in the replication cohorts **(Supplementary Table 3)**. Samples used in replication and extrapolation were collected after overnight fasting. Except for the samples at the VUMC Alzheimer center and FINRISK97, which were non-fasting or 'semi-fasting' (participants were instructed to fast for 4 hours before the scheduled examination). The summary statistics of metabolomics measurements in the cohorts used for the replication of the cognitive findings and the cohorts used for extrapolation to dementia are shown in **Supplementary Table 5**.

Statistical analyses

Histograms of classical blood measurements and metabolites in the discovery cohorts were visually inspected for non-normality, if necessary natural logarithmic or rank-transformations were applied (**Supplementary Table 4**). Individuals who had suffered from a stroke or who were diagnosed with dementia at time of cognitive assessment were excluded. Linear regression analyses were used to assess the relation of standardized measures of TC, HDL-C, LDL-C, triglycerides and glucose with general cognitive ability, adjusting for age, sex, lipid-lowering medication (yes/no) and BMI as

covariates. The effect of *APOE ε*4 on general cognitive ability was assessed using an additive model coding non-carriers as 0, heterozygotes as 1 and homozygotes as 2. The association of 299 standardized metabolites with general cognitive ability was assessed using linear regression with age, sex, BMI and lipid lowering medication as covariates (model 1). To test if the identified associations were independent of the classically measured and frequently studied circulating markers, we ran a second model (model 2) where we additionally included TC, HDL-C, LDL-C, triglycerides and glucose as covariates to model. Finally, we tested if the identified metabolites-general cognitive ability association were confounded by *APOE ε*4 (model 3).

Since metabolites are highly correlated we used the method of Li and Ji [38] to correct for multiple testing. The method calculates the number of independent tests in correlated measures. In this study testing 299 metabolites corresponded to 87 independent tests (p for significance=0.05/87 =5.7×10^{-4}). To assess the relation of metabolites found to be associated with cognitive function to incident dementia and AD, we used Cox-proportional hazard models when data came from prospective studies. Again, we standardized the metabolite levels and adjusted for age (at entry), sex, BMI and lipid lowering medication. For VUMC Alzheimer center, we used logistic regression adjusted for age and sex. The relations with incident dementia and AD were evaluated in a second model additionally adjusting for *APOE ε*4 genotypes.

In the discovery cohorts we used linear regression analysis to study associations of lifestyle factors (smoking, physical activity and fish (oil) consumption) with metabolites and cognitive function, adjusting for age, sex, BMI and lipid-lowering medication. All analyses were performed in R (version 3.2.1, 2015-06-18). Summary statistics by cohort were combined with inverse variance-weighted fixed-effects meta-analysis using the '*rmeta'* package (version 2.16). The association magnitudes are reported in units of standard deviation (SD) or odds ratio (OR) change per 1-SD increase in each metabolite easing comparison of effects.[39,40]

Results

Clinical characteristics of all cohorts analyzed in this study are provided in **Table 1**. Compared to the Rotterdam Study participants, ERF participants (N=2,683, 56.1% women, mean age 48.9 ± 14.2 years) were younger, resulting in a difference in age-related clinical characteristics (N=2,505, 58.2% women, mean age 74.6 ± 6.2 years). Results of the association of general cognitive ability with baseline clinical characteristics in the discovery cohorts are shown in **Table 2.** As expected, general cognitive ability was higher in participants with higher education and was inversely associated with increasing age and the presence of *APOE ε*4 allele (**Table 2**). Increased HDL-C was associated with higher general cognitive ability (0.034 SD higher general cognitive ability per one SD higher HDL-C concentration; p=6.4×10^{-3}), while fasting glucose levels were associated with lower cognitive ability (0.039 SD; p=2.2×10^{-3}).

The metabolic profile of general cognitive ability

We identified 18 metabolites that were significantly associated (p<5.9×10^{-4} – Model 1) with general cognitive ability (listed as top 18 associations in **Supplementary Table 6).** Of the 18 metabolites XXL-LDL-triglycerides was not measured in the replication cohorts, therefore 17 metabolites were tested for replication in independent cohorts (**Supplementary Table 7** – model 1, Nmax=6,652). Out of these 17, we found 15 to be associated with general cognitive ability in the replication cohorts (preplication<0.05, **Table 3, Figure 2**). Thirteen metabolites surpassed the more stringent Bonferroni corrected threshold for significance in the replication ($p_{replication}$<2.9×10^{-3}). Combining discovery and replication data (**Table 3**), 12 metabolites were associated with higher general cognitive ability and 3 were associated with lower general cognitive ability. The metabolites associated with increased higher cognitive ability include 11 HDL sub-fractions, the most significant being free cholesterol in HDL (0.078 SD; p=2.3×10^{-15}) and docosahexaenoic acid (DHA or 22:6(n-3)) an omega-3-fatty acid (0.060 SD; p=9.8×10^{-11}). The three metabolites that were associated with lower general cognitive ability include glycoprotein acetyls (-0.075 SD; p=5.4×10^{-13}), glutamine (-0.042 SD; p=2.8×10^{-7}) and ornithine (-0.057 SD; p=8.5×10^{-7}).

Table 1: Baseline characteristics of all studied 11 cohorts.

	ERF	RS	WHII	NTR	SHIP-Trend	FHS	VUmc ADC	Finrisk97	DILGOM	EGCUT	AgeCoDe
Number of samples in cognitive analysis	2683	2505	4235	338	944	1508	-	-	-	-	-
Age (years)	48.9±14.2	74.2±6.2	55.8±6.0	40.7±12.4	50.1±13.6	55.7±9.8	64.1±9.0	48.8±13.5	52.3±13.5	59.1±12.4	84.1±3.1
N-Women (%)	56.1	58.2	26.2	62.4	56.4	52.5	45.1	54.7	55.8	58.9	69.4
Education (1-4 scale)	2.1±0.9	2.4±0.9	2.0±0.8	3.2±0.8	2.4±0.9	2.3±0.6 $	5.0±1.0 #	2.0±0.8 $	2.1±0.8$	3.0±0.8	-
Body-mass index (kg/m2)	27±4.7	27.4±4.1	25.9±3.8	24.7±4	27.4±4.6	27.5±4.9	25.3±3.8	26.7±4.6	27.2±4.9	28±5.1	25.6±3.7
Lipid lowering medication (%)	12.7	22.8	3.0	7.0	7.4	7.6	19.5	3.49	15.6	24	21.6
APOE ε4 carriers (%)	37.7	27.6	27.7	26.6	22.5	22.5	51.7	35.1	24	23.6	20.3
Diastolic blood pressure (mmHg)	80±10	79±11	77±10.3	76±9.8	76.7±10	75±10	86±11	83±11.24	79±11	82±10	78±8.0
Systolic blood pressure (mmHg)	140±20	152±21	122±15	124±12	124±16	126±18	141±19	136±20	137±20	134±17	136±15
Established blood measures											
TC (mmol/l)	5.6±1.1	5.6±1.0	5.9±1.0	5.1±1.1	5.5±1.1	5.3±1.0	4.9±1.0	5.5±1.1	5.3±1.0	5.8±1.1	5.8±1.1
LDL-Cholesterol (mmol/l)	3.7±1.0	3.5±0.9*	3.8±0.9	3.1±1.0	3.4±0.9	3.3±0.9	1.7±0.5	3.5+0.9*	3.2±0.8*	2.2±0.7	3.5±1.0
HDL-Cholesterol (mmol/l)	1.3±0.4	1.5±0.4	1.5±0.4	1.4±0.3	1.5±0.4	1.3±0.4	1.5±0.4	1.39±0.4	1.44±0.4	1.7±0.4	1.6±0.4
Triglycerides (mmol/l)	1.3±0.8	1.49±0.7	1.3±0.8	1.4±0.7	1.4±0.9	1.7±1.4	1.4±0.7	1.51±1.1	1.43±0.9	1.8±1.1	1.4±0.6
Glucose (mmol/l)	4.7±1.1	5.9±1.5	5.2±1	5.4±0.7	5.4±0.6	5.6±1.5	5.7±1.7	4.61±1.1	4.12±0.8	4.6±1.7	-
Dementia analysis											
Number of samples	1532	2010	4612	-	-	2356	1303	7517	4788	2572	310
Follow-up time (years)	11.3±1.7	7.6±3.6	16.7±1.6	-	-	15.7±5	-	9.67±1.35	7.68±0.9	7.03±2.22	4.5±1.8
Maximum follow-up	13.6	11.7	17.9	-	-	22.6	-	10	7.9	12.9	6.4
Number of AD cases	28	346	35	-	-	81	665	100	75	-	75
Number of dementia cases	39	506	114	-	-	110	917	141	81	41	82

$ education in 1-3 scale , # education in 1-7 scale , * LDL-C estimated using he Friedewald estimation

Study names: Erasmus Rucphen Family study (ERF), Rotterdam Study (RS), Whitehall II (WHII), Netherlands Twin Registry (NTR), Study of Health in Pomerania - Trend (SHIP-Trend), Framingham Heart Study (FHS), VUmc Amsterdam Dementia Cohort (VUmc ADC), The National FINRISK study(Finrisk97), Dietary, Lifestyle and Genetic determinants of Obesity and Metabolic Syndrome (DILGOM), Estionian biobank (EGCUT), German Study on Ageing, Cognition, and Dementia (AgeCoDe). TC= total Cholesterol, LDL-Cholesterol = cholesterol in low density lipoprotein, HDL- Cholesterol = cholesterol in high density lipoprotein, *APOE* = Apolipoprotein E (genotype).

Table 2: Association of characteristics with general cognitive ability in discovery cohorts.

	ERF study			Rotterdam study			Meta-analysis		
Phenotype	**Effect (±SE)**	**p**	**N**	**Effect (±SE)**	**p**	**N**	**Effect (±SE)**	**p**	**N**
Age	-0.042 (±0.001)	2.1×10-280	2699	-0.078 (±0.003)	1.1×10-154	2483	-0.046 (±0.001)	<1×10-500	5182
Sex (male vs female)	-0.020 (±0.029)	0.49	2699	0.142 (±0.035)	3.8×10-5	2483	0.048 (±0.022)	3.1×10-2	5182
BMI	0.005 (±0.003)	0.15	2694	-0.015 (±0.004)	3.1×10-4	2483	-0.003 (±0.003)	0.30	5177
Education	0.410 (±0.017)	1.4×10-117	2699	0.277 (±0.020)	1.5×10-41	2483	0.355 (±0.013)	<1×10-500	5182
APOE ε4	-0.046 (±0.028)	0.09	2342	-0.157 (±0.035)	9.0×10-6	2378	-0.088 (±0.022)	5.2×10-5	4720
Lipid lowering medication	-0.131 (±0.047)	5.8×10-3	2690	0.013 (±0.042)	0.76	2397	-0.050 (±0.031)	0.11	5087
Classical blood measures									
TC	-0.011 (±0.016)	0.50	2635	0.022 (±0.019)	0.24	2481	0.003 (±0.012)	0.8	5116
LDL-Cholesterol*	-0.026 (±0.016)	0.10	2621	0.016 (±0.019)	0.41	2265	-0.009 (±0.012)	0.45	4886
HDL-Cholesterol	0.037 (±0.016)	2.3×10-2	2635	0.029 (±0.019)	0.13	2401	0.034 (±0.012)	6.4×10-3	5036
Triglycerides	-0.014 (±0.016)	0.36	2637	-0.017 (±0.018)	0.35	2345	-0.015 (±0.012)	0.19	4982
Glucose	-0.047 (±0.017)	6.2×10-3	2623	-0.029 (±0.019)	0.12	2401	-0.039 (±0.013)	2.2×10-3	5024

TC= total Cholesterol, LDL-C = cholesterol in low density lipoprotein, HDL-C = cholesterol in high density lipoprotein, *APOE* = Apolipoprotein E (genotype). *LDL-C in the Rotterdam study was estimated using the Friedewald estimation. Multivariate analysis of general cognitive ability with age was adjusted for sex, with sex it was adjusted for age. Association of general cognitive ability with BMI, educational level, APOE and lipid lowering medication use for age and sex. Associations of general cognitive ability with blood measures (TC, LDL-C, HDL-C, Triglycerides and glucose) were adjusted for age, sex and lipid lowering medication use. The association magnitudes are reported in units of standard deviation (SD) change (±standard error) per 1-SD increase in each metabolite.[39,40]

Table 3: Association of metabolites with general cognitive ability

	Discovery			Replication			Meta-analysis				
Metabolite	**Effect (±SE)**	***p***	**N**	**Effect (±SE)**	***p***	**N**	**Effect (±SE)**	***p***	**N**	**I^2**	***p*-I^2**
HDL-free cholesterol	0.067 (±0.013)	1.5×10^{-7}	4791	0.094 (±0.015)	1.2×10^{-9}	4542	0.078 (±0.010)	2.3×10^{-15}	9333	63	6.7×10^{-2}
L-HDL-free cholesterol	0.059 (±0.013)	3.4×10^{-6}	4793	0.106 (±0.016)	6.2×10^{-11}	4542	0.077 (±0.010)	1.5×10^{-14}	9335	80	7.3×10^{-3}
L-HDL-total cholesterol	0.052 (±0.013)	5.0×10^{-5}	4792	0.108 (±0.016)	3.5×10^{-11}	4542	0.073 (±0.010)	3.5×10^{-13}	9334	79	8.2×10^{-3}
Glycoprotein acetyls	-0.064 (±0.014)	4.5×10^{-6}	3778	-0.087 (±0.015)	1.4×10^{-8}	4542	-0.075 (±0.010)	5.4×10^{-13}	8320	0	0.60
L-HDL-cholesterol esters	0.048 (±0.013)	1.5×10^{-4}	4792	0.108 (±0.016)	3.4×10^{-11}	4542	0.071 (±0.010)	1.6×10^{-12}	9334	80	7.6×10^{-3}
L-HDL-phospholipids	0.045 (±0.013)	4.2×10^{-4}	4791	0.100 (±0.016)	3.3×10^{-10}	4542	0.067 (±0.010)	2.2×10^{-11}	9333	78	9.7×10^{-3}
HDL-phospholipids	0.050 (±0.013)	9.5×10^{-5}	4790	0.089 (±0.015)	9.8×10^{-9}	4542	0.066 (±0.010)	2.5×10^{-11}	9332	68	4.3×10^{-2}
HDL-total cholesterol	0.048 (±0.013)	2.1×10^{-4}	4796	0.088 (±0.016)	1.5×10^{-8}	4542	0.064 (±0.010)	9.8×10^{-11}	9338	66	5.4×10^{-2}
22:6, docosahexaenoic acid	0.047 (±0.014)	5.4×10^{-4}	3772	0.070 (±0.012)	2.1×10^{-8}	5480	0.060 (±0.009)	9.8×10^{-11}	9252	67	4.6×10^{-2}
M-HDL-phospholipids	0.063 (±0.013)	8.2×10^{-7}	4799	0.057 (±0.014)	7.2×10^{-5}	4542	0.060 (±0.010)	2.5×10^{-10}	9341	0	0.48
M-HDL-total cholesterol	0.046 (±0.013)	3.0×10^{-4}	4799	0.054 (±0.014)	1.5×10^{-4}	4542	0.050 (±0.010)	1.8×10^{-7}	9341	0	0.62
M-HDL-cholesterol esters	0.046 (±0.013)	2.5×10^{-4}	4799	0.052 (±0.014)	2.6×10^{-4}	4542	0.049 (±0.009)	2.4×10^{-7}	9341	0	0.63
Glutamine	-0.052 (±0.012)	2.5×10^{-5}	4715	-0.034 (±0.011)	1.8×10^{-3}	6652	-0.042 (±0.008)	2.8×10^{-7}	11367	74	9.8×10^{-3}
S-HDL-free cholesterol	0.059 (±0.012)	8.4×10^{-7}	4796	0.029 (±0.014)	4.0×10^{-2}	4542	0.047 (±0.009)	3.5×10^{-7}	9338	0	0.55
Ornithine	-0.083 (±0.018)	4.5×10^{-6}	2228	-0.039 (±0.015)	1.0×10^{-2}	2750	-0.057 (±0.012)	8.5×10^{-7}	4978	43	0.18
L-LDL-triglycerides	-0.055 (±0.012)	3.7×10^{-6}	4797	0.018 (±0.015)	2.3×10^{-1}	4542	-0.027 (±0.009)	4.2×10^{-3}	9339	90	3.6×10^{-5}
M-LDL-triglycerides	-0.042 (±0.012)	5.1×10^{-4}	4800	0.019 (±0.015)	1.9×10^{-1}	4542	-0.018 (±0.009)	6.3×10^{-2}	9342	85	1.3×10^{-3}

LDL = low density lipoprotein, HDL = high density lipoprotein, *APOE* = Apolipoprotein E (genotypes), L = large particles, M = medium particles, S = small particles, I^2 = measure for heterogeneity in the meta-analysis in percent, *p*-I^2 = p-value for heterogeneity. Glycoprotein acetyls are mainly alpha-1-acid glycoprotein. The association magnitudes are reported in units of standard deviation (SD) change (±standard error) per 1-SD increase in each metabolite.[39,40] Shown associations of the metabolites with general cognitive ability are adjusted for age, sex, body mass index and lipid-lowering medication.

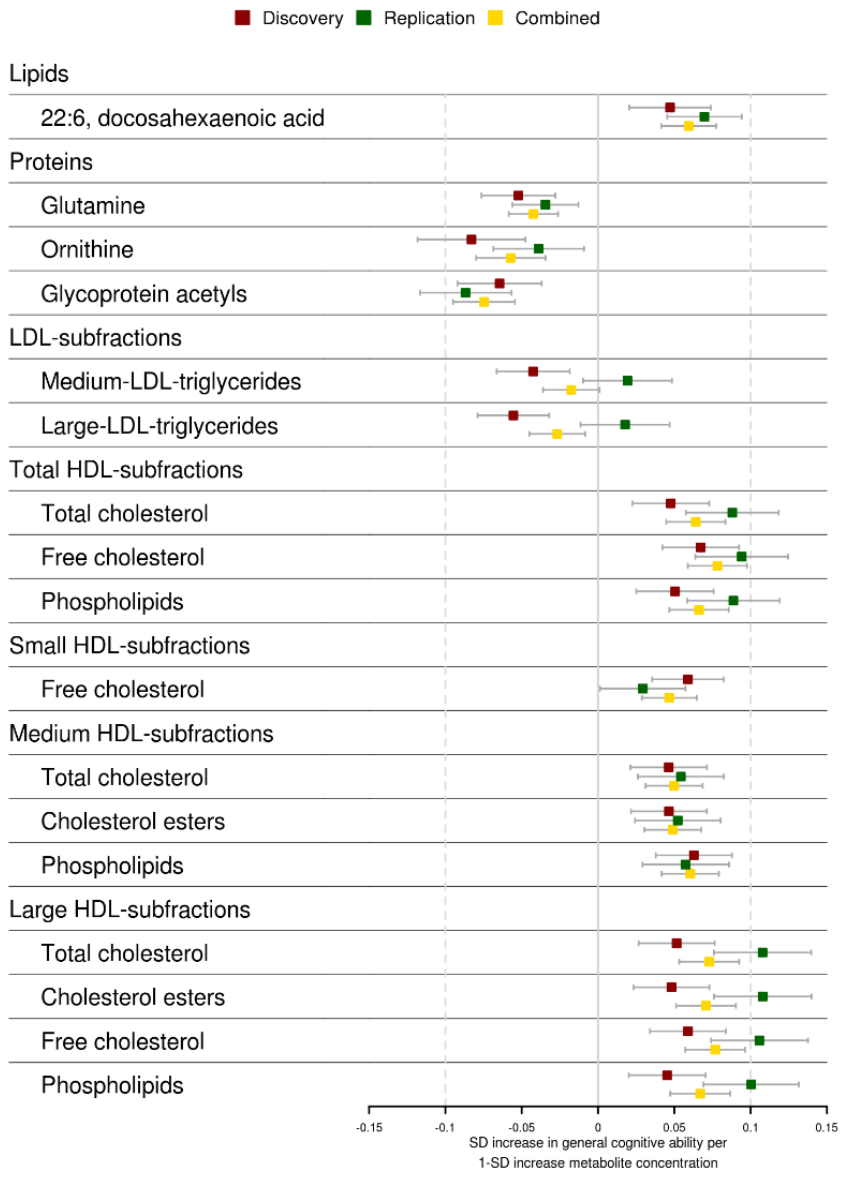


Figure 2: Associations of metabolites with general cognitive ability. The standardized effect estimates on general cognitive ability of metabolites adjusted for age, sex, body-mass index (BMI) and lipid lowering medication use are shown. The estimates are shown for the discovery (red), replication (green) and the combined (yellow) analysis. Point estimates are shown as boxes with whiskers denoting the 95% confidence interval of the effect estimates.

Of the 15 metabolites significantly associated with general cognition, only two metabolites, HDL cholesterol esters ($p_{model2}=9.9\times10^{-3}$) and medium HDL total cholesterol ($p_{model2}= 7.9\times10^{-3}$), lost their significance in the combined analysis when additionally adjusting for Glucose, TC, HDL-C, LDL-C and triglycerides (**Supplementary Table 7** – Model 2)**.** However, **(Supplementary Figure 1** – Model 2**)** adjustment did not result in a major change in effect estimates for these two metabolites, suggesting an independent effect. Adjusting for *APOE ε*4 did not explain any of the 15 associations (**Supplementary Table 7** – Model 3 and **Supplementary Figure 1** – Model 3). In **Box 1** we summarize the functions of the metabolites we found associated with general cognitive ability.

Association of the metabolic profile with dementia and Alzheimer's disease

Next, we examined whether the 15 metabolites associated with general cognitive ability were associated with dementia. We compared (maximum) 1990 dementia patients, of whom 1,356 were AD cases, with 23,882 controls. Six metabolites were associated with dementia and three of these were also associated with AD ($p<0.05$; **Table 4**, for all association results, see **Supplementary Table 8** and **Figure 3**). Free cholesterol in small HDL associated most significantly with a lower risk of dementia (OR=0.85 per 1-SD increase in metabolite concentration; 95%CI 0.80-0.91; $p=6.3\times10^{-7}$) and AD (OR=0.87; 95%CI 0.81-0.94; $p=2.3\times10^{-4}$). Other metabolites associated with a lower dementia risk were DHA (OR=0.92; 95%CI 0.86-0.97; $p=3.4\times10^{-3}$; AD, $p=1.5\times10^{-3}$) and sub-fractions of medium size HDL particles (phospholipids $p=2.5\times10^{-3}$, total cholesterol $p=0.025$ and cholesterol esters $p=0.025$). Higher glutamine levels were associated with an increased risk of dementia (OR=1.08; 95%CI 1.02-1.15; $p=0.011$) and AD (OR=1.11; 95%CI 1.04-1.20; $p=3.0\times10^{-3}$)**.** The association of free cholesterol in small HDL and DHA surpassed the more stringent Bonferroni corrected threshold for significance ($p<1.5\times10^{-3}$). After additionally adjusting for the number of *APOE ε4* alleles the associations of dementia and AD with sub-fractions of medium size HDL particles were no longer significant ($p>0.05$) **(Supplementary Table 8** and **Supplementary Figure 2**).

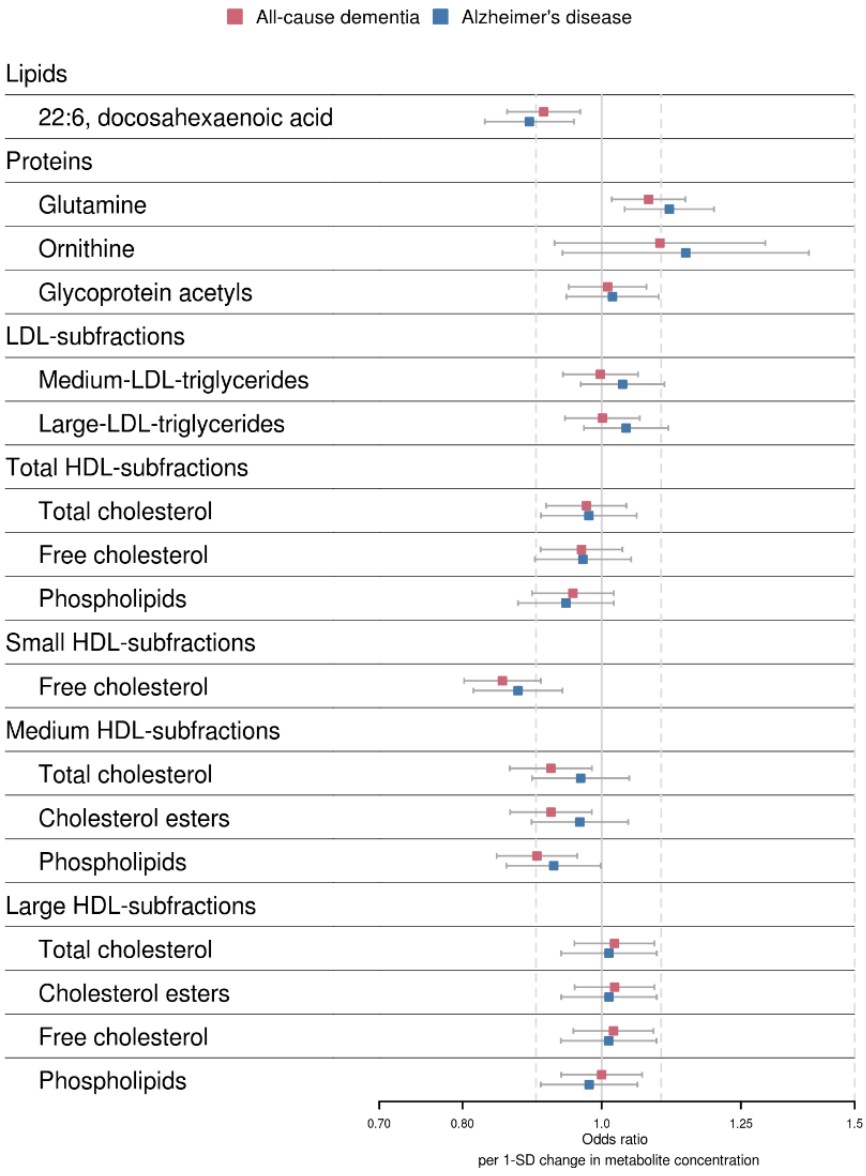


Figure 3: Association of metabolites with all-cause dementia and Alzheimer's disease. The standardized odds ratio of metabolites with all-cause dementia (red) and Alzheimer's Disease (blue) shown as point estimates with whiskers denoting the 95% confidence interval of the OR. Associations shown are adjusted for age (at entry), sex, and if available body mass index and lipid lowering medication.

Table 4: Metabolite concentrations associated with dementia and AD

	AD				Dementia			
Metabolite	**OR**	**P**	**N-cases**	**N-total**	**OR**	**P**	**N-cases**	**N-total**
S-HDL-free cholesterol	0.87 [0.81-0.81]	2.3×10^{-4}	1276	22880	0.85 [0.80-0.80]	4.1×10^{-7}	1881	25868
M-HDL-phospholipids	0.93 [0.86-0.86]	4.7×10^{-2}	1276	22884	0.90 [0.85-0.85]	1.8×10^{-3}	1881	25872
22:6, docosahexaenoic acid	0.89 [0.83-0.83]	1.5×10^{-3}	1334	22466	0.91 [0.86-0.86]	1.9×10^{-3}	1938	25417
Glutamine	1.11 [1.04-1.04]	3.1×10^{-3}	1356	25181	1.08 [1.02-1.02]	1.3×10^{-2}	1990	25640
M-HDL-cholesterol esters	0.97 [0.89-0.89]	0.38	1276	22884	0.92 [0.86-0.86]	1.6×10^{-2}	1881	25872
M-HDL-total cholesterol	0.97 [0.89-0.89]	0.40	1276	22884	0.92 [0.86-0.86]	1.6×10^{-2}	1881	25872

HDL = high density lipoprotein, L = large, M = medium, AD= Alzheimer's disease. OR = Odds ratio for the increase or decrease in AD or dementia risk per one SD increase of metabolite concentration. Sorted by the P for dementia. Combined results from Cox-proportional hazard models and logistic regression models are presented as OR. Associations shown are adjusted for age (at entry), sex, and if available BMI and lipid lowering medication. N-total is the sum of cases and controls.

Association of the metabolic profile with lifestyle factors

The analysis of the association of lifestyle factors with metabolites and general cognitive ability are shown in **Supplementary Table 9** and summarized in **Figure 4**. Fish (oil) intake was strongly associated with DHA blood concentrations (p=9.9×10^{-53}). Physical activity was associated with increased (p<0.05) levels of metabolites that were associated with higher cognitive function (medium and large HDL sub-fractions) and decreased levels of metabolites that were associated with lower cognitive function (glycoprotein acetyls, ornithine, glutamine). Smokers had decreased concentrations of all HDL sub-fractions associated with higher cognitive function and increased concentrations of metabolites that associated with decreased cognitive function (**Figure 4**).

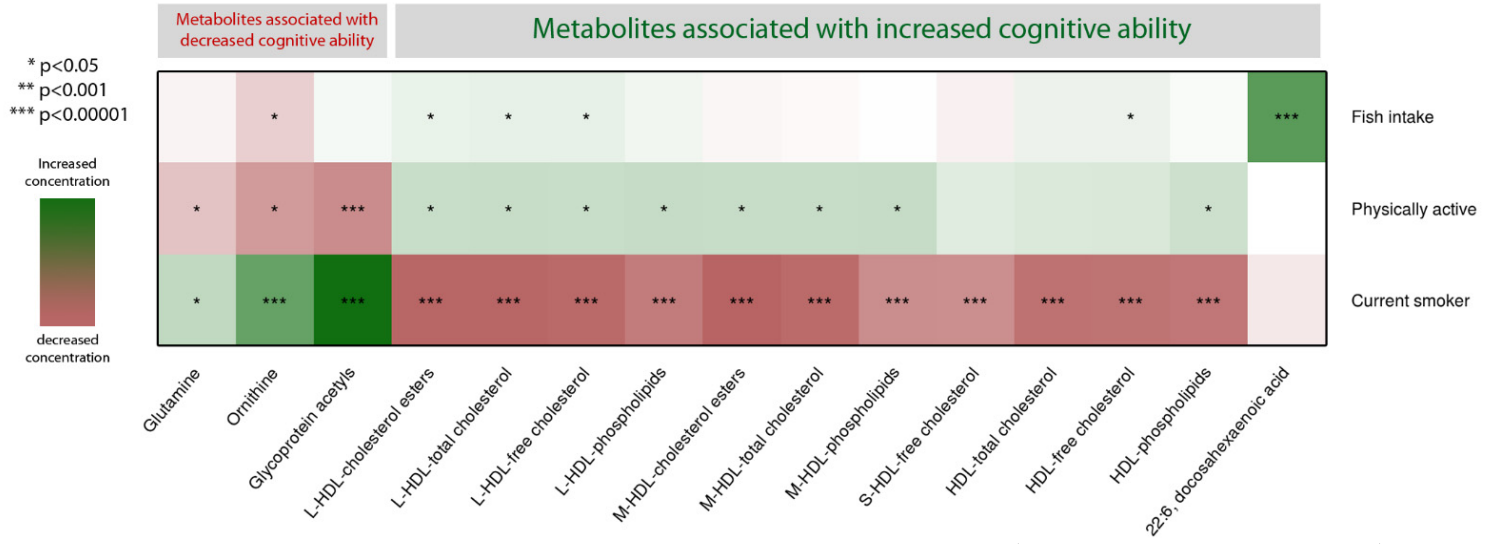


Figure 4: Association of metabolites with lifestyle factors. Lifestyle factors including smoking (current vs. past and never smokers), physically activity (yes/no) and dietary fish (oil) intake. Metabolites are grouped based on their association with general cognitive ability and subfraction of HDL. Colors represent standardized effects estimates. Green shows the lifestyle factor associated with an increase in the metabolite concentration and red shows the lifestyle factor associated with a decrease in the metabolite concentration. Significance of the associations are shown as stars: * P<0.05, **P<0.001, ***P<1×10-5. HDL=high-density lipoprotein

Discussion

In this study, we discovered and replicated 15 metabolites associated with general cognitive ability. This metabolic profile includes sub-fractions of HDL, DHA, ornithine, glutamine and glycoprotein acetyls. We show that metabolites in the profile are independent of classical cardiometabolic blood correlates of cognitive function. Of the 15 replicated metabolites, 6 were associated with dementia and 3 of these also with Alzheimer's disease. Further, we show that lifestyle factors such as diet, smoking and physical activity, have strong effects on metabolites in the profile.

The most interesting metabolite in the profile is DHA, a long-chain omega-3 polyunsaturated fatty acid. As the largest cross-sectional study to date studying DHA in relation to cognitive function, the present study showed compelling evidence that DHA levels in blood were associated with higher cognitive function (p=9.8×10^{-11}). This finding is in line with a many previous studies, summarized by Cederholm and colleagues,[41] suggesting a relation between nutritive DHA intake, or fish (oil) intake as its proxy, and better cognition. Blood levels of DHA are raised by eating fat fish, as also in our study. DHA from diet is most likely actively transported over the blood brain barrier by Msfd2a,[42,43] where it is abundant in grey matter [44] and found in lower concentrations in brains of individuals with AD.[45] We showed for the first time that DHA in blood was associated with a lower risk of AD and dementia, using blood measures of DHA in up to 22,887 individuals. Taken together, our study implies that high levels of DHA could be beneficial for cognitive function, potentially also reducing the risk of dementia and AD.

Beyond the association of high HDL cholesterol with better cognitive function,[3-5] the present study points towards a role of cholesterol, free cholesterols and phospholipids in small, medium, large subclasses of HDL. Free cholesterols and phospholipids in HDL showed the strongest association with cognitive function, dementia and AD. We showed that these associations were largely independent of HDL-cholesterol. However, current knowledge of the functions of HDL subclasses is limited, thus we can only

speculate on the pathways through which the metabolites that we observed exert their effect on cognitive function.[46] We speculate that the phospholipids have a direct effect as they are the main constituents of neuronal membrane structures, like the presynaptic and postsynaptic membranes of synapses. Breakdown of membrane phospholipids leading to degeneration neuronal membranes has been linked to synapse loss in AD.[47] Possibly circulating phospholipids and free cholesterols in HDL form a buffer to repair damaged membranes. A recent study supports this by observing that both AD patients and patients with mild cognitive impairments have lower circulating levels of nutrients involved in phospholipid synthesis in blood and cerebrospinal fluid.[48] Alternatively the free cholesterols in the phospholipid layer of HDL tag the presence of other important proteins that are transported to or are disposed from the brain. Previously it has been shown that HDL contains up to 95 proteins and lipids that may segregate into distinct subclasses of HDL and lead to subclass specific effects.[49,50] Regions in membranes of both neurons and astrocytes [51] where HDL related free cholesterols, sphingomyelins and free fatty acids (such as DHA) concentrate are called lipid rafts. Changes in lipid raft composition may be an early marker of neurodegenerative diseases.[52] A hypothesis that requires further study is that increased free cholesterols in (small) HDL and DHA in blood affects lipid raft quantity, composition or cell-signaling leading to beneficial effects on the brain.

In our study, levels of glycoprotein acetyls, mainly alpha-1-acid glycoprotein (also known as orosomucoid), were associated with lower cognitive function. Levels of this protein are strongly associated with smoking and physical activity and glycoprotein acetyl concentration has been shown to be a strong predictor of 10-year mortality.[53,54] The protein is a marker of acute phase reactions and may be implicated in this way in depression,[55] diabetes,[56] cardiovascular disease [57] and cancer.[58] A major genetic determinant of glycoprotein acetyl levels in blood is located close to the gene coding for haptoglobin (*HP*).[36] This protein may link our findings of HDL sub-fractions to that of glycoprotein acetyls, as the haptoglobin protein has been found in specific HDL sub-fractions.[50] Furthermore, the *HP* gene was previously associated with the risk of cognitive impairment in type 2 diabetes with poor glycemic control.[59]

Two non-essential amino-acids that were associated with lower cognitive function were ornithine, which is part of the urea cycle, and glutamine. Ornithine accumulation causes hyperornithinemia–hyperammonemia–homocitrullinuria (HHH) syndrome,[60] a disease with a currently poorly known pathogenesis which is clinically characterized by mental retardation.[61] Glutamine and its closely related neurotransmitter glutamate have been found to be differentially expressed in brains of AD patients [62]. In the brain, glutamate is considered harmful [62] and our population-based studies suggest that in the circulation glutamine is associated with lower cognition. Both ornithine and glutamine are interesting targets for further studies.

A major strength of the current study is the large sample size both in the discovery and replication. To our knowledge, this is the largest study exploring the association of a large array of blood based metabolites with general cognitive ability to date. Other strengths are the similar methods across studies used to determine metabolites and the use of general cognitive ability to harmonize the studied cognitive outcome.[23] We chose to analyze the associations of metabolites with cognitive ability in the largest sample size available, accepting that subtle differences cognitive testing, metabolite measuring, study design and populations would introduce heterogeneity of effects. Then followed by a replication in independent samples. This approach is modelled to the standard approach followed in genome-wide association studies.[63] A potential limitation of our cross-sectional study of cognition is that we cannot determine causality of the association with circulating metabolites. However, in our extrapolation to dementia and AD we mostly studied incident cases with metabolites measured before the disease onset, suggesting that at least the 6 dementia associated metabolites are most likely in a causal pathway. We note that the associations of the HDL sub fractions associated with dementia were attenuated by the *APOE* ε4 genotype, suggesting they could be in the causal pathway of *APOE* ε4 to dementia or the associations found are pleiotropic effects of *APOE* ε4. Last but not least we did not adjust for education because they are highly correlated [64] if measured at the same time. In fact, there is still debate on whether education determines cognitive ability or vice-versa.[65] A recent study of 70,000 children

in the UK, suggested that general cognitive ability is a determinant of education. In this study the first principal component from the Cognitive Abilities Test (CAT) battery taken at age 11 years correlated about 0.8 with grades on the General Certificate of Secondary Education (GCSE) examinations taken at age 16 years.[26] Several other studies have reported correlations between 0.6 to 0.96,[66,67] On the other hand, several studies claimed that education influences development of intelligence.[68,69] In fact, there is a very high genetic correlation between educational attainment and cognitive ability (R^2 = 0.55 based on LD-SCORE regression).[22,70,71] This shared genetic background is probably the primary reason for the high correlation between education and cognitive ability.[64,72] The strong correlation of education and cognitive ability has led researchers to use educational attainment as a proxy-phenotype of cognitive ability.[73] Given the high genetic correlation we decided that adjusting for education as a covariate in the model would lead to over-adjustment and ultimately false-negative findings in the study.

In conclusion, we discovered and replicated the relation of 15 metabolites in blood to cognitive function in cognitively healthy individuals. We found that six metabolites were associated with dementia and three with Alzheimer disease. The association of lifestyle factors to the metabolites associated with cognitive ability and dementia opens new avenues for targeted prevention.

References:

1. Alzheimer's A. 2015 Alzheimer's disease facts and figures. *Alzheimers Dement* 2015; **11**(3): 332-84.
2. Snyder HM, Corriveau RA, Craft S, et al. Vascular contributions to cognitive impairment and dementia including Alzheimer's disease. *Alzheimers Dement* 2015; **11**(6): 710-7.
3. Solomon A, Kareholt I, Ngandu T, et al. Serum cholesterol changes after midlife and late-life cognition: twenty-one-year follow-up study. *Neurology* 2007; **68**(10): 751-6.
4. Crichton GE, Elias MF, Davey A, Sullivan KJ, Robbins MA. Higher HDL cholesterol is associated with better cognitive function: the Maine-Syracuse study. *J Int Neuropsychol Soc* 2014; **20**(10): 961-70.
5. Corley J, Starr JM, Deary IJ. Serum cholesterol and cognitive functions: the Lothian Birth Cohort 1936. *Int Psychogeriatr* 2015; **27**(3): 439-53.
6. Satizabal CL, Beiser AS, Chouraki V, Chene G, Dufouil C, Seshadri S. Incidence of Dementia over Three Decades in the Framingham Heart Study. *New Engl J Med* 2016; **374**(6): 523-32.
7. Rocca WA, Petersen RC, Knopman DS, et al. Trends in the incidence and prevalence of Alzheimer's disease, dementia, and cognitive impairment in the United States. *Alzheimers & Dementia* 2011; **7**(1): 80-93.

8. Schrijvers EM, Verhaaren BF, Koudstaal PJ, Hofman A, Ikram MA, Breteler MM. Is dementia incidence declining?: Trends in dementia incidence since 1990 in the Rotterdam Study. *Neurology* 2012; **78**(19): 1456-63.
9. Langa KM, Larson EB, Crimmins EM, et al. A Comparison of the Prevalence of Dementia in the United States in 2000 and 2012. *JAMA Intern Med* 2016.
10. Henriksen K, O'Bryant SE, Hampel H, et al. The future of blood-based biomarkers for Alzheimer's disease. *Alzheimers Dement* 2014; **10**(1): 115-31.
11. Hye A, Riddoch-Contreras J, Baird AL, et al. Plasma proteins predict conversion to dementia from prodromal disease. *Alzheimers Dement* 2014; **10**(6): 799-807 e2.
12. Muenchhoff J, Poljak A, Song F, et al. Plasma protein profiling of mild cognitive impairment and Alzheimer's disease across two independent cohorts. *J Alzheimers Dis* 2015; **43**(4): 1355-73.
13. Song F, Poljak A, Crawford J, et al. Plasma apolipoprotein levels are associated with cognitive status and decline in a community cohort of older individuals. *PLoS One* 2012; **7**(6): e34078.
14. Proitsi P, Kim M, Whiley L, et al. Association of blood lipids with Alzheimer's disease: A comprehensive lipidomics analysis. *Alzheimers Dement* 2016.
15. Li D, Misialek JR, Boerwinkle E, et al. Plasma phospholipids and prevalence of mild cognitive impairment and/or dementia in the ARIC Neurocognitive Study (ARIC-NCS). *Alzheimers Dement (Amst)* 2016; **3**: 73-82.
16. Mapstone M, Cheema AK, Fiandaca MS, et al. Plasma phospholipids identify antecedent memory impairment in older adults. *Nat Med* 2014; **20**(4): 415-8.
17. Fiandaca MS, Zhong X, Cheema AK, et al. Plasma 24-metabolite Panel Predicts Preclinical Transition to Clinical Stages of Alzheimer's Disease. *Front Neurol* 2015; **6**: 237.
18. Casanova R, Varma S, Simpson B, et al. Blood metabolite markers of preclinical Alzheimer's disease in two longitudinally followed cohorts of older individuals. *Alzheimers Dement* 2016; **12**(7): 815-22.
19. Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. *Nature* 2014; **505**(7485): 612-3.
20. Marmot MG, Smith GD, Stansfeld S, et al. Health Inequalities among British Civil-Servants - the Whitehall-Ii Study. *Lancet* 1991; **337**(8754): 1387-93.
21. van der Flier WM, Pijnenburg YA, Prins N, et al. Optimizing patient care and research: the Amsterdam Dementia Cohort. *J Alzheimers Dis* 2014; **41**(1): 313-27.
22. Davies G, Armstrong N, Bis JC, et al. Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53949). *Mol Psychiatry* 2015; **20**(2): 183-92.
23. Johnson W, Nijenhuis J, Bouchard TJJ. Still just 1 g: Consistent results from five test batteries. *Intelligence* 2008; **36**(1): 81–95.
24. Sabia S, Gueguen A, Marmot MG, Shipley MJ, Ankri J, Singh-Manoux A. Does cognition predict mortality in midlife? Results from the Whitehall II cohort study. *Neurobiol Aging* 2010; **31**(4): 688-95.
25. Spearman C. General intelligence, objectively determined and measured. *American Journal of Psychology* 1904; **15**: 201-93.
26. Deary IJ, Pattie A, Starr JM. The stability of intelligence from age 11 to age 90 years: the Lothian birth cohort of 1921. *Psychol Sci* 2013; **24**(12): 2361-8.
27. Hoogendam YY, Hofman A, van der Geest JN, van der Lugt A, Ikram MA. Patterns of cognitive function in aging: the Rotterdam Study. *Eur J Epidemiol* 2014; **29**(2): 133-40.
28. Isaacs A, Sayed-Tabatabaei FA, Aulchenko YS, et al. Heritabilities, apolipoprotein E, and effects of inbreeding on plasma lipids in a genetically isolated population: the Erasmus Rucphen Family Study. *Eur J Epidemiol* 2007; **22**(2): 99-105.
29. de Bruijn RFAG, Bos MJ, Portegies MLP, et al. The potential for prevention of dementia across two decades: the prospective, population-based Rotterdam Study. *BMC Med* 2015; **13**: 132.

30. Kalmijn S, Launer LJ, Ott A, Witteman JC, Hofman A, Breteler MM. Dietary fat intake and the risk of incident dementia in the Rotterdam Study. *Ann Neurol* 1997; **42**(5): 776-82.
31. Hofman A, Brusselle GG, Darwish Murad S, et al. The Rotterdam Study: 2016 objectives and design update. *Eur J Epidemiol* 2015; **30**(8): 661-708.
32. Sayed-Tabatabaei FA, van Rijn MJ, Schut AF, et al. Heritability of the function and structure of the arterial wall: findings of the Erasmus Rucphen Family (ERF) study. *Stroke* 2005; **36**(11): 2351-6.
33. van Gent CM, van der Voort HA, de Bruyn AM, Klein F. Cholesterol determinations. A comparative study of methods with special reference to enzymatic procedures. *Clin Chim Acta* 1977; **75**(2): 243-51.
34. Neeley WE. Simple automated determination of serum or plasma glucose by a hexokinase-glucose-6 -phosphate dehydrogenase method. *Clin Chem* 1972; **18**(6): 509-15.
35. Soininen P, Kangas AJ, Wurtz P, et al. High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism. *Analyst* 2009; **134**(9): 1781-5.
36. Kettunen J, Demirkan A, Wurtz P, et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat Commun* 2016; **7**: 11122.
37. Demirkan A, Henneman P, Verhoeven A, et al. Insight in genome-wide association of metabolite quantitative traits by exome sequence analyses. *PLoS Genet* 2015; **11**(1): e1004835.
38. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* 2005; **95**(3): 221-7.
39. Wurtz P, Makinen VP, Soininen P, et al. Metabolic signatures of insulin resistance in 7,098 young adults. *Diabetes* 2012; **61**(6): 1372-80.
40. Wurtz P, Havulinna AS, Soininen P, et al. Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation* 2015; **131**(9): 774-85.
41. Cederholm T, Salem N, Palmblad J. omega-3 Fatty Acids in the Prevention of Cognitive Decline in Humans. *Adv Nutr* 2013; **4**(6): 672-6.
42. Nguyen LN, Ma DL, Shui GH, et al. Mfsd2a is a transporter for the essential omega-3 fatty acid docosahexaenoic acid. *Nature* 2014; **509**(7501): 503-+.
43. Ben-Zvi A, Lacoste B, Kur E, et al. Mfsd2a is critical for the formation and function of the blood-brain barrier. *Nature* 2014; **509**(7501): 507-11.
44. Svennerholm L. Distribution and fatty acid composition of phosphoglycerides in normal human brain. *J Lipid Res* 1968; **9**(5): 570-9.
45. Soderberg M, Edlund C, Kristensson K, Dallner G. Fatty acid composition of brain phospholipids in aging and in Alzheimer's disease. *Lipids* 1991; **26**(6): 421-5.
46. Vitali C, Wellington CL, Calabresi L. HDL and cholesterol handling in the brain. *Cardiovasc Res* 2014; **103**(3): 405-13.
47. Prasad MR, Lovell MA, Yatin M, Dhillon H, Markesbery WR. Regional membrane phospholipid alterations in Alzheimer's disease. *Neurochem Res* 1998; **23**(1): 81-8.
48. van Wijk N, Slot RER, Duits FH, et al. Nutrients required for phospholipid synthesis are lower in blood and cerebrospinal fluid in mild cognitive impairment and Alzheimer's disease dementia. *Alzheimers Dement (Amst)* 2017; **8**: 139-46.
49. Gordon SM, Deng J, Tomann AB, Shah AS, Lu LJ, Davidson WS. Multi-dimensional co-separation analysis reveals protein-protein interactions defining plasma lipoprotein subspecies. *Mol Cell Proteomics* 2013; **12**(11): 3123-34.
50. Li H, Gordon SM, Zhu X, et al. Network-Based Analysis on Orthogonal Separation of Human Plasma Uncovers Distinct High Density Lipoprotein Complexes. *J Proteome Res* 2015; **14**(8): 3082-94.
51. Sebastiao AM, Colino-Oliveira M, Assaife-Lopes N, Dias RB, Ribeiro JA. Lipid rafts, synaptic transmission and plasticity: impact in age-related neurodegenerative diseases. *Neuropharmacology* 2013; **64**: 97-107.

52. Sonnino S, Aureli M, Grassi S, Mauri L, Prioni S, Prinetti A. Lipid rafts in neurodegeneration and neuroprotection. *Mol Neurobiol* 2014; **50**(1): 130-48.
53. Fischer K, Kettunen J, Wurtz P, et al. Biomarker profiling by nuclear magnetic resonance spectroscopy for the prediction of all-cause mortality: an observational study of 17,345 persons. *PLoS Med* 2014; **11**(2): e1001606.
54. Singh-Manoux A, Shipley MJ, Bell JA, Canonico M, Elbaz A, Kivimaki M. Association between inflammatory biomarkers and all-cause, cardiovascular and cancer-related mortality. *CMAJ* 2016.
55. Harley J, Roberts R, Joyce P, et al. Orosomucoid influences the response to antidepressants in major depressive disorder. *J Psychopharmacol* 2010; **24**(4): 531-5.
56. El-Beblawy NMS, Andrawes NG, Ismail EAR, Enany BE, Abou El-Seoud HS, Erfan MA. Serum and Urinary Orosomucoid in Young Patients With Type 1 Diabetes: A Link Between Inflammation, Microvascular Complications, and Subclinical Atherosclerosis. *Clin Appl Thromb-Hem* 2016; **22**(8): 718-26.
57. Carriere I, Dupuy AM, Lacroux A, Cristol JP, Delcourt C, St POLA. Biomarkers of inflammation and malnutrition associated with early death in healthy elderly people. *Journal of the American Geriatrics Society* 2008; **56**(5): 840-6.
58. Bruno R, Olivares R, Berille J, et al. alpha-1-acid glycoprotein as an independent predictor for treatment effects and a prognostic factor of survival in patients with non-small cell lung cancer treated with docetaxel. *Clinical Cancer Research* 2003; **9**(3): 1077-82.
59. Guerrero-Berroa E, Ravona-Springer R, Heymann A, et al. Haptoglobin genotype modulates the relationships of glycaemic control with cognitive function in elderly individuals with type 2 diabetes. *Diabetologia* 2015; **58**(4): 736-44.
60. Filosto M, Alberici A, Tessa A, Padovani A, Santorelli FM. Hyperornithinemia-hyperammonemia-homocitrullinuria (HHH) syndrome in adulthood: a rare recognizable condition. *Neurol Sci* 2013; **34**(9): 1699-701.
61. Fonteh AN, Harrington RJ, Tsai A, Liao P, Harrington MG. Free amino acid and dipeptide changes in the body fluids from Alzheimer's disease subjects. *Amino Acids* 2007; **32**(2): 213-24.
62. Zhou Y, Danbolt NC. Glutamate as a neurotransmitter in the healthy brain. *J Neural Transm (Vienna)* 2014; **121**(8): 799-817.
63. Conneely KN, Boehnke M. Meta-analysis of genetic association studies and adjustment for multiple testing of correlated SNPs and traits. *Genet Epidemiol* 2010; **34**(7): 739-46.
64. Johnson WM, M.; Lacono, W.G. Disruptive behavior and school grades: genetic and environmental relations in 11-year-olds. *J Edu Psychol* 2005; **97**(3): 391-405.
65. Deary IJ, Johnson W. Intelligence and education: causal perceptions drive analytic processes and therefore conclusions. *Int J Epidemiol* 2010; **39**(5): 1362-9.
66. S W. Environmental and innate factors and educational attainment: chapter in Genetic and Environmental Factors in Human Ability 1966.
67. S.F. BTJF. Twins reared together: what they tell about human diversity, individuality and determinism; 1984.
68. G. BPR. Cohort effects in cognitive development in children as revealed by cross-sectional sequences. *Dev Psychol* 1969; **1**: 169-77.
69. Schmidt WHO. Socio-economic status, schooling, intelligence, and scholastic progress in a community in which education is not yet compulsory. *Paedogogica Europa* 1967; **2**: 275-86.
70. Zheng J, Erzurumluoglu AM, Elsworth BL, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* 2016.
71. Okbay A, Beauchamp JP, Fontana MA, et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* 2016; **533**(7604): 539-42.

72. Johnson W, McGue M, Iacono WG. Genetic and environmental influences on academic achievement trajectories during adolescence. *Dev Psychol* 2006; **42**(3): 514-32.
73. Rietveld CA, Esko T, Davies G, et al. Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proc Natl Acad Sci U S A* 2014; **111**(38): 13790-4.
74. Shah AS, Tan L, Long JL, Davidson WS. Proteomic diversity of high density lipoproteins: our emerging understanding of its importance in lipid transport and beyond. *J Lipid Res* 2013; **54**(10): 2575-85.
75. Schwendeman A, Sviridov DO, Yuan WM, et al. The effect of phospholipid composition of reconstituted HDL on its cholesterol efflux and anti-inflammatory properties. *Journal of Lipid Research* 2015; **56**(9): 1727-37.
76. Fellows K, Uher T, Browne RW, et al. Protective associations of HDL with blood-brain barrier injury in multiple sclerosis patients. *J Lipid Res* 2015; **56**(10): 2010-8.
77. Guemez-Gamboa A, Nguyen LN, Yang HB, et al. Inactivating mutations in MFSD2A, required for omega-3 fatty acid transport in brain, cause a lethal microcephaly syndrome. *Nature Genetics* 2015; **47**(7): 809-+.
78. Morris MC, Brockman J, Schneider JA, et al. Association of Seafood Consumption, Brain Mercury Level, and APOE epsilon4 Status With Brain Neuropathology in Older Adults. *JAMA* 2016; **315**(5): 489-97.
79. Huang TL. Omega-3 fatty acids, cognitive decline, and Alzheimer's disease: a critical review and evaluation of the literature. *J Alzheimers Dis* 2010; **21**(3): 673-90.
80. Nakanishi S. Molecular diversity of glutamate receptors and implications for brain function. *Science* 1992; **258**(5082): 597-603.
81. Israili ZH, Dayton PG. Human alpha-1-glycoprotein and its interactions with drugs. *Drug Metab Rev* 2001; **33**(2): 161-235.
82. Huang ZQ, Ung T. Effect of Alpha-1-Acid Glycoprotein Binding on Pharmacokinetics and Pharmacodynamics. *Current Drug Metabolism* 2013; **14**(2): 226-38.

Box 1: Description of pathways of metabolites in the context of cognitive function

HDL lipoprotein sub-fractions

The specific lipoprotein sub-fractions could point to the specific functions of lipoprotein sub-fractions. HDL fractions are well known to be individually tasked for different functions across lipid metabolism, inflammation, anti-oxidation, and host defence [74,75]. Additionally specific protein pairs on specific HDL subspecies exist that maintain stable compositions [49]. Previous research reported links between HDL cholesterol profiles and changes in vascular health with plaque accumulation in arteries of the brain, damage to the blood brain barrier [76] and occurrence of thrombosis. All possibly leading to progressive vascular brain damage resulting in loss of white matter microstructural organization.

Docosahexaenoic acid (DHA)

DHA levels in blood are highly associated with omega-3 fatty acid intake through diet [40] and it cannot be de novo synthesized in the brain and is therefore actively transported over the blood brain barrier through the MSFD2A [42,77]. DHA is essential for normal brain development in early life and is frequently associated with cognition [41]. High intake might also be beneficial in late life as DHA and fish oil intake associated with less Alzheimer disease pathology [78]. The evidence of the attributed beneficial effects of DHA on the brain in literature is inconsistent [79].

Glutamine

In the brain, glutamine is not only used for energy production and protein synthesis, as in other cells, but is also an essential precursor for biosynthesis of amino acid neurotransmitters. It is involved in the glutamine-glutamate/GABA cycle, a well-studied concept in excitatory signalling in the brain [62]. The cycle involves transfer of glutamine from astrocytes to neurons and neurotransmitter glutamate or GABA from neurons to astrocytes. The leading opinion in the field is that in the brain an excess of glutamate, excitotoxicity, is seen as detrimental and glutamine in the brain as beneficial [80].

Glycoprotein Acetyls

The measured glycoprotein acetyls is mainly alpha-1-acid glycoprotein (AGP) [35], also called orosomucoid, is an acute phase plasma alpha-globulin glycoprotein. The protein is widely studied and has previously been found to predict 10-year mortality [53]. Increased plasma levels of Glycoprotein acetyls as reaction to various diseases (cancer and inflammatory diseases) or following trauma (surgery) might explain the association with increased mortality, and could partially explain the association with general cognitive ability as chronic diseases decrease cognitive abilities. Another function of AGP is to carry mainly neutrally charged medications in blood, for example anti-depressants [81]. The plasma concentration of AGP is relatively low and there is only one drug-binding site in each AGP molecule [82], leading to lower antidepressant response in higher AGP concentrations [55].

Ornithine

Ornithine as a non-proteinogenic amino acid is an important intermediate product in arginine degradation and urea cycle. Hyperornithinemia is also the biochemical hallmark of an inherited metabolic disease, hyperornithinemia–hyperammonemia–homocitrullinuria (HHH) syndrome [60]. This disease is clinically characterized by mental retardation whose pathogenesis is still poorly known.

Supplementary information

The supplementary tables of this chapter can be accessed scanning the following code.

Supplementary Table 1

A short description of all studies included in this study, which metabolites were measured, and a short description of the methods used to measure these metabolites.

Supplementary Table 2

Description of the cognitive tests used to calculate general cognitive ability and the methods for the methods for the diagnosis of dementia and Alzheimer's disease.

Supplementary Table 3

The table shows by study; the time between cognitive measurements and the blood draw used for metabolite measurements in years (mean ± SD) (1); the total variance of all tests accounted for by the first principal component (general cognitive ability) by study (2); cognitive tests included (alphabetical order) and the correlations (or loadings) of cognitive test with the first principal component by study (general cognitive ability) (3); the cognitive domains covered by the tests by study (4).

Supplementary Table 4

Descriptive statistics of the metabolites in the discovery cohorts. Some metabolites were measured on two platforms, and therefore listed twice, in the Erasmus Rucphen Family study, the platform that was determined in the largest sample was used in the meta-analysis.

Supplementary Table 5

Descriptive statistics of the metabolites in the replication cohorts and cohorts associating the metabolites with dementia and AD.

Supplementary Table 6

Association results of 299 metabolites tested in the discovery. Metabolites significantly associated with general cognitive ability are marked in grey.

Supplementary Table 7

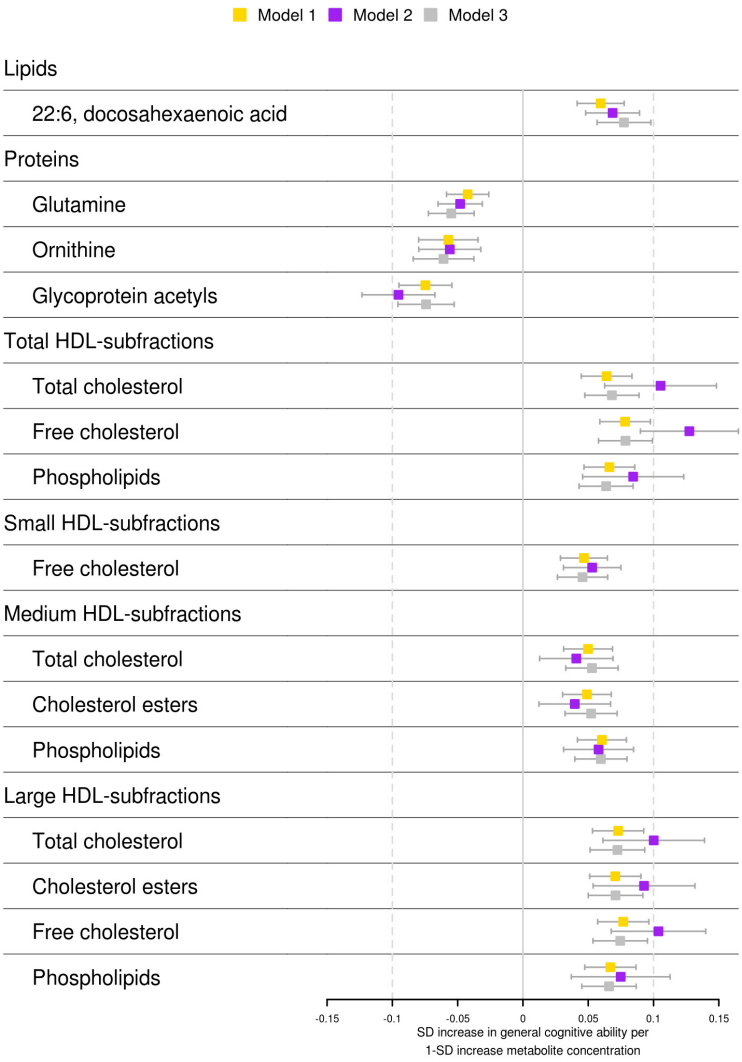
The betas, standard errors, and p-values per metabolite, per study of the metabolites significantly associated with general cognitive ability in the meta-analysis of the discovery and replication are shown. The results are shown for model 1 (adjusted for age, sex, BMI, lipid lowering medication), model 2 (Model 1 + Glucose, total cholesterol, HDL-cholesterol, LDL-cholesterol, triglycerides) and model 3 (Model 1 + *APOE* ε4). The table corresponds to supplementary figure 1.

Supplementary Table 8

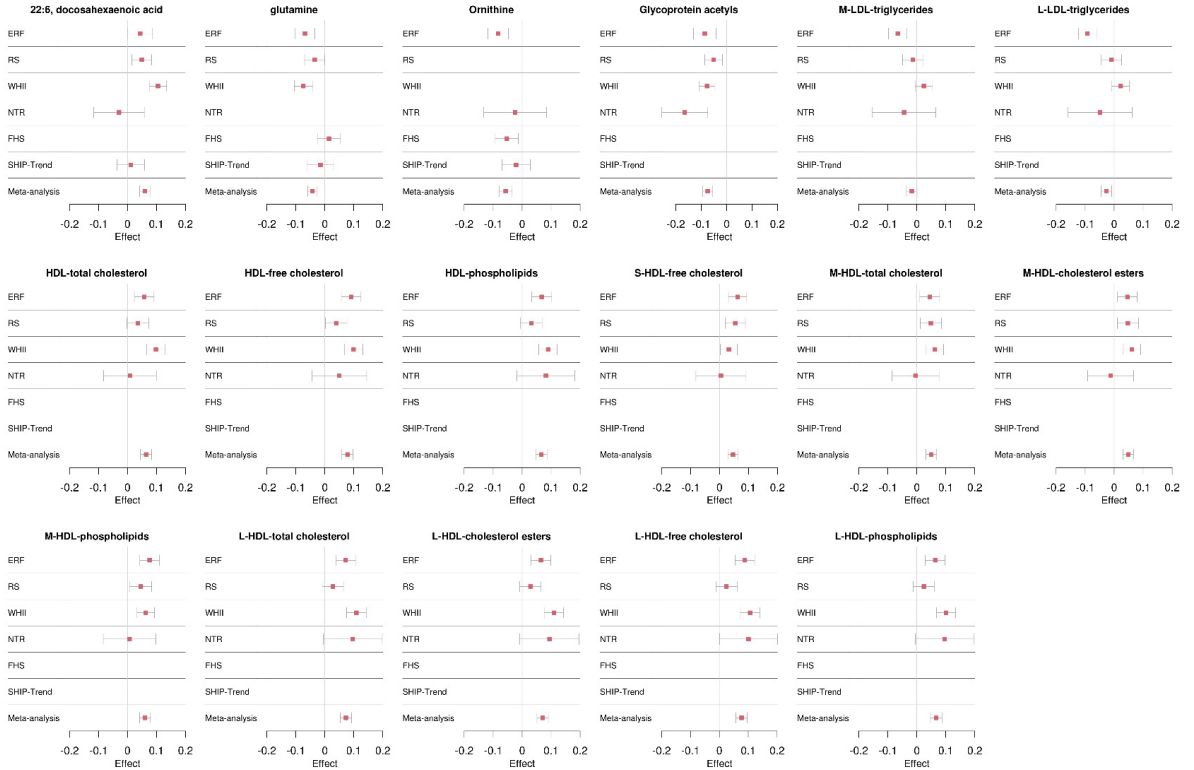
Association results of metabolites with AD and dementia. Associations were adjusted for age, sex, BMI, lipid lowering medication) and additionally adjusted for *APOE* ε4. The table corresponds to figure 3 and supplementary figure 2.

Supplementary Table 9

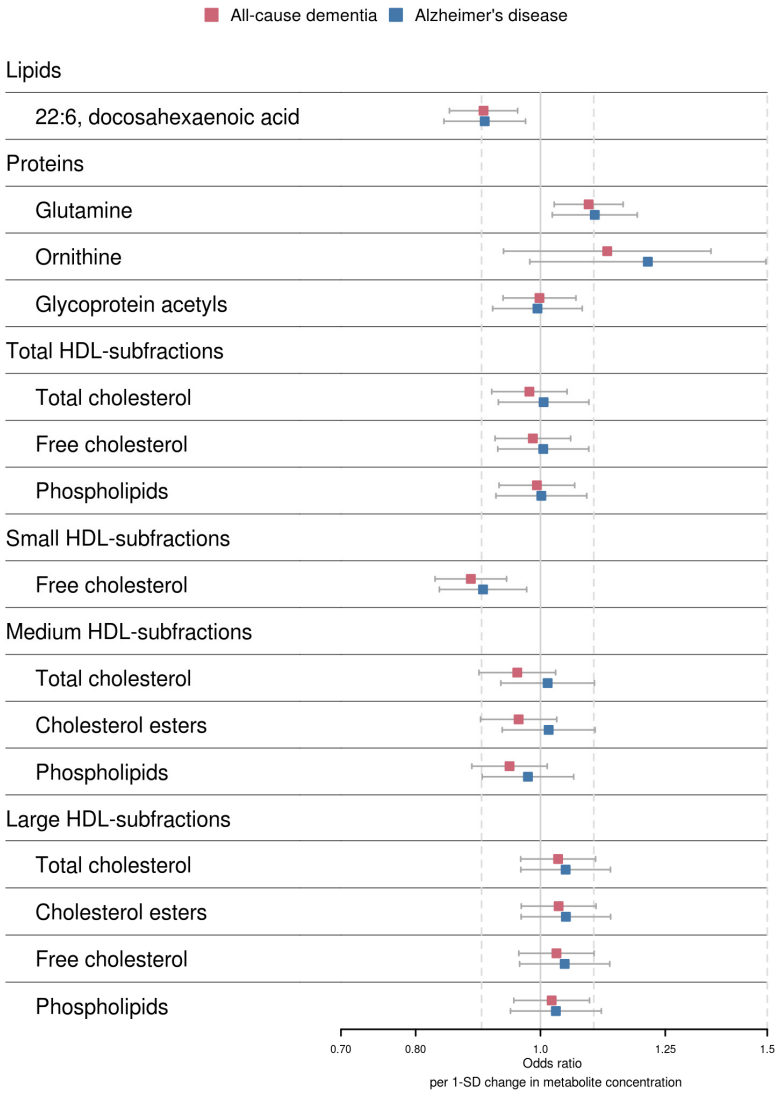
Association results of metabolites with lifestyle factors, including smoking (current vs. past and never smokers), physically activity (yes/no) and dietary fish (oil) intake. The table corresponds to figure 4.



Supplementary Figure 1: Comparing the effect estimates of the associations of metabolites with general cognitive ability over the three statistical models. Only replicated metabolites are shown. The standardized effect estimates are shown for model 1 (covariates: age, sex, BMI and lipid lowering medication), model 2 (covariates in model 1 and TC, HDL-C, LDL-C, triglycerides and glucose) and, model 3 (covariates in model 1 and *APOE* ε4). The standardized effect estimates of metabolites with general cognitive ability shown as point estimates with 95% confidence interval whiskers.



Supplementary Figure 2: Metabolite associations with general cognitive ability by study. The standardized effect estimates adjusted for age, sex, body-mass index (BMI) and lipid lowering medication of metabolites with general cognitive ability by study are shown. Discovery includes the Erasmus Rucphen Family study (ERF) and the Rotterdam study (RS). The replication studies include the Whitehall II (WHII) study, Netherlands Twin Registry (NTR), the Framingham Heart Study (FHS) and the Study of Health in Pomerania - Trend (SHIP-Trend). Point estimates are shown as boxes with whiskers denoting the 95% confidence interval of the effect estimates.



Supplementary Figure 3: Association with all-cause dementia and Alzheimer's disease adjusted for number of *APOE* ε4 alleles. The standardized odds ratio of metabolites with all-cause dementia (red) and Alzheimer's Disease (blue) shown as point estimates with whiskers denoting the 95% confidence interval of the OR. Associations shown are adjusted for age (at entry), sex, body mass index (if available), lipid lowering medication (if available) and number of *APOE* ε4 alleles.

Chapter 5.4

Metabolic profiling of intracranial arteriosclerosis

Sven J. van der Lee*, Dina Vojinovic*, Nicolien A. van Vliet, Cornelia M. van Duijn, Meike W. Vernooij, Diana van Heemst, Jeroen van der Grond, Maryam Kavousi, Marian Beekman, P. Eline Slagboom, Najaf Amin, Thomas Hankemeier, Ayşe Demirkan, M. Arfan Ikram, Aad van der Lugt, Daniel Bos

* These authors contributed equally

This chapter is in preparation

Abstract

Etiologic underpinnings of intracranial arteriosclerosis remain poorly studied. Additionally, evidence accumulates that intracranial arteriosclerosis develops under the influence of a differential risk factor profile than arteriosclerosis in other parts of the arterial system. Given the metabolic nature of risk factors for intracranial arteriosclerosis, such as diabetes mellitus and hypercholesterolemia, studying circulating metabolites represents a source with potential to provide novel insights. To study metabolites underlying intracranial arteriosclerosis, we investigated associations of a wide range of metabolites measured by proton nuclear magnetic resonance (NMR) with intracranial carotid artery calcification (ICAC), as a proxy of intracranial arteriosclerosis, in a sample of 1,111 participants from the population-based Rotterdam Study. We compared the metabolic association pattern of ICAC with that of extracranial carotid artery calcification (ECAC), aortic arch calcification (AAC) and coronary artery calcification (CAC). CAC results were from a combined sample of 1,501 participants from the Rotterdam Study and the Leiden Longevity Study. One standard deviation (SD) increase in concentration of 3-hydroxybutyrate, a ketone body, was significantly associated with a 0.11 SD increase in (p-value = 1.8×10^{-4}) ICAC volume. Furthermore, ICAC was nominally associated with metabolites related with glucose metabolism, whereas the other vessel beds showed association with lipid and lipoprotein measures. Interestingly, glycoprotein acetyls were associated with calcification in all studied vessel beds. These associations were strongest in men. To conclude, we found that higher circulating level of 3-hydroxybutyrate was associated with increase in ICAC. Furthermore, we found differences in metabolic association patterns of ICAC and calcifications in the other vessel beds providing further evidence for location-specific differences of arterial calcification.

Introduction

Intracranial arteriosclerosis is established as one of the leading causes of stroke worldwide,[1-4] but its etiologic underpinnings remain poorly studied. Importantly, increasing evidence is accumulating that intracranial arteriosclerosis develops under the influence of a differential risk factor profile than arteriosclerosis in other parts of the arterial system.[5-7] Especially, the contribution of specific metabolic risk factors, namely disrupted glucose metabolism (diabetes mellitus) and cholesterol metabolism (hypercholesterolemia), seem to be much more prominent for intracranial arteriosclerosis.[6,8,9] Yet, it is important to acknowledge that, apart from glucose and cholesterol, there is a large spectrum of circulating metabolites that may contribute to intracranial arteriosclerosis. Further in-depth investigation of the metabolic underpinnings of intracranial arteriosclerosis may expose important knowledge which may ultimately contribute to the development of therapeutic and preventive strategies. It has only recently become possible to study large spectra of active metabolites in relation to disease,[10,11] due to rapid technological improvements in the field of nuclear magnetic resonance (NMR) of circulating metabolites. Metabolites can now be inexpensively and reproducibly quantified on a large-scale, enabling metabolomics studies in large population-based cohorts. Such studies have already led to successful metabolic profiling of type 2 diabetes,[12,13] and cardiovascular events.[14-17]

Against this background, we investigated associations of a broad range of metabolites with intracranial carotid artery calcification (ICAC), as a proxy of intracranial arteriosclerosis. To study the putative differential risk factor profiles of arteriosclerosis in major vessel beds we compared the metabolic association profile of ICAC with that of extracranial carotid artery calcification (ECAC), aortic arch calcification (AAC) and coronary artery calcification (CAC), in participants from the population-based Rotterdam Study and the family-based Leiden Longevity Study.

Methods

Study population

Our study population consisted of participants from the Rotterdam Study and the Leiden Longevity Study.

The Rotterdam study is a prospective population-based cohort study among individuals aged 45 years and over, who are living in the well-defined Ommoord district in Rotterdam, the Netherlands.[18] The study started in 1990, with 7,983 participants (first Rotterdam Study cohort, RS-I), and was extended in 2000/2001 (RS-II, 3,011 participants) [18] All participants were invited for extensive re-examinations every 3-4 years. At each visit blood was drawn after overnight fasting. The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC and by the Ministry of Health, Welfare and Sport of the Netherlands, implementing the Wet Bevolkingsonderzoek: ERGO (Population Studies Act: Rotterdam Study).[18] All participants provided written informed consent to participate in the study and to obtain information from their treating physicians.

The Leiden Longevity Study is a cohort study consisting of 421 Dutch Caucasian families. Recruitment took place between 2002 and 2006 based on the following inclusion criteria: (1) there were at least two living siblings per family, who fulfilled the age criteria and were willing to participate; (2) men had to be aged ≥89 years and women had to be aged ≥91 years; and (3) the sib pairs had to have the same parents.[19] Additionally, offspring of these long-lived siblings and their partners as controls were enrolled to enable future association studies.[20] Non-fasting blood samples were drawn at baseline. The Leiden Longevity Study has been approved by the Medical Ethical Committee of the Leiden University Medical Centre, and written informed consent was obtained from all participants.

Population for analysis

Metabolites were available for two independent datasets of the Rotterdam Study. The first encompassed all individuals from the first cohort RS-I that were included in the

fourth visit to the study center. This subset included 2,975 individuals of which 730 also underwent a computed tomography (CT) scan. The second was the subset of samples that was previously included in the Biobank-based Integrative Omics Studies Consortium (BIOS Consortium).[21,22] This subset includes 768 participants of which 381 underwent a CT scan. The summary statistics of metabolites in the two Rotterdam Study datasets are shown in Supplementary Table 1. Additionally, metabolites were available for 2,343 offspring and controls from the Leiden Longevity Study of whom 390 participants (206 offspring and 184 controls) also underwent a CT-scan. Metabolite and CT measures have not been performed at the same time point, in Rotterdam study dataset 1 the CT scan was performed a median 4 months (IQR 2-4 months) after metabolite measuring, in the second dataset this was 6 years before scanning (IQR 5.9 -6.2) and in the Leiden longevity study the 6.17 (5.67 - 6.67)

The characteristics of the study population are shown in Table 1. Compared to the first group of Rotterdam Study participants (N=730, 51.2% women, mean age 73.8 ± 5.5 years) the second group of Rotterdam Study participants (N=381, 53.0% women, mean age 64.9 ± 3.2 years) were significantly younger, resulting in differences in age-related clinical characteristics and average volume of calcifications (Table 1). However, the prevalence of ICAC was comparable being 83.0% and 80.6 % in the first and second group respectively. Leiden Longevity Study participants (N=390, 52.3% women, mean age 65.9 ± 6.5) were comparable in age and CAC volume in comparison to the second group of the Rotterdam study (Table 1).

Metabolite quantification

The metabolites were quantified from EDTA plasma samples using high-throughput proton Nuclear Magnetic Resonance (NMR) metabolomics (Nightingale Health, Helsinki, Finland). This method provides simultaneous quantification of metabolic measures, i.e. routine lipids, lipoprotein subclass profiling with lipid concentrations within 14 subclasses, fatty acid composition, and various low-molecular weight metabolites including amino acids, ketone bodies and gluconeogenesis-related metabolites in molar concentration units. The lipoprotein subclasses include very low-density lipoprotein (VLDL), intermediate-density lipoprotein (IDL), low-density

lipoprotein (LDL) and high-density lipoprotein (HDL). In these subclasses, the concentration is measured as well as the subfraction of lipids, triglycerides, cholesterol esters, free cholesterol, and phospholipids. Details of the experimentation and applications of this NMR metabolomics platform have been described previously.[23,24] For this study we analyzed in total 166 non-derived metabolites that were measured across all cohorts.

Table 1. Descriptive characteristic of study population

	Rotterdam Study dataset 1	**Rotterdam Study dataset 2**	**Leiden Longevity Study**
Number of Participants	730	381	390
Age at CT scan, years	73.8 ± 5.5	64.9 ± 3.2	65.9 ± 6.5
Women	374 (51.2%)	202 (53.0%)	204 (52.3%)
BMI, kg/m2	27.3 ± 4	27.8 ± 3.8	25.4 ± 3.5
Smoking (never/past/current) (%)	206/403/91 (28.2/55.2/12.5)	110/200/59 (28.9/52.5/15.5)	122/223/43 (31.3/57.2/11.0)
Total cholesterol, mmol/l	5.6 ± 1.0	5.8 ± 1.0	5.7 ± 1.2
HDL-Cholesterol, mmol/l	1.4 ± 0.4	1.5 ± 0.4	1.4 ± 0.4
Hypercholesterolemia	351 (48.5%)	186 (49.7%)	146 (37.4%)
Diastolic blood pressure, mmHg	79.7 ± 11.4	81.3 ± 10.5	81.8 ± 10.2
Systolic blood pressure, mmHg	151.2 ± 21.2	142.3 ± 18.0	144 .0± 20.0
Hypertension	406 (56.2%)	162 (43.7%)	200 (51.3%)
Glucose, mmol/l	5.8 ± 1.4	5.6 ± 1.2	5.9 ± 1.2
Participants at CT with diabetes	105 (14.4%)	29 (7.6%)	55 (14.1%)
Participants at CT with cardiovascular disease	95 (13.1%)	27 (7.1%)	11 (2.8%)
Participants at CT with coronary heart disease	72 (9.9%)	17 (4.5%)	4 (1.0%)
Participants at CT with stroke	32 (4.4%)	11 (2.9%)	8 (2.1%)
ICAC volume, median (IQR), cm3	64.8 (13.0-205.6)	22.1 (3.8-75.4)	-
ECAC volume, median (IQR), cm3	48 (3.1-176.7)	10.4 (0-60.4)	-
AAC volume, median (IQR), cm3	437.1 (123.7-1324)	88.8 (9.8-391.4)	-
CAC volume, median (IQR), cm3	96.8 (10.2-404.3)	20.6 (0.1-130.1)	17 (0-141)

Abbreviations: BMI- body mass index; HDL- high-density lipoprotein; ICAC – intracranial carotid artery calcification; ECAC- extracranial carotid artery calcification; AAC – aortic arch calcification; CAC – coronary artery calcification; IQR – interquartile range. Values are means ± standard deviation for continuous variables and number (percentages) for dichotomous variables.

Assessment of Arteriosclerosis

In the Rotterdam study a 16-slice (n = 785) or 64-slice (n = 1,739) multidetector CT scanner (Somatom Sensation 16 or 64; Siemens, Forchheim, Germany) was used to perform non-enhanced scanning of the intracranial carotid arteries. Using a cardiac scan and a scan that reached from the aortic arch to the intracranial vasculature (1 cm above the sella turcica), we scanned the following vessel beds: coronary arteries, aortic arch, extracranial carotid arteries, and intracranial carotid arteries. Detailed information regarding the imaging settings of both scans is provided elsewhere.[4] As proxy for intracranial arteriosclerosis, we measured ICAC bilaterally in the intracranial internal carotid artery.[8] For quantification of ICAC, we used a semi-automated scoring method that is described in detail elsewhere.[4] Briefly, we manually drew regions of interest around calcification in the course of the intracranial internal carotid arteries in consecutive CT sections. Next, we calculated calcification volumes by multiplying the number of pixels in excess of 130 Hounsfield units by the pixel size and the increment.[8] Calcification volumes in the coronary arteries, aortic arch, and extracranial internal carotid arteries were quantified using dedicated commercially available software (Syngo Calcium Scoring; Siemens).[4] All calcification volumes are expressed in cubic millimeters. Correlations between calcification across the 4 vessel beds ranged from 0.5 to 0.6.[4,25]

In the Leiden Longevity study, coronary atherosclerosis was evaluated by measurement of the Agatston total coronary artery calcium score using a 320-multidetector row CT scanner (Aquilion ONE, Toshiba, Otawara, Japan).[26] The scan range was planned between the carina and cardiac apex. Detailed information regarding imaging settings is provided elsewhere.[26] Calcification volumes in the coronary arteries were quantified using dedicated CT calcium score analysis software (VScore, Vital Images). Pixels exceeding the threshold value of 130 HU were automatically recognized by the post processing tool. These areas were manually encircled in the course of the coronary arteries. The amount of coronary artery calcification was automatically calculated according to Agatston et al.[27] All imaging data were analyzed blinded, both for group (offspring or partner) and clinical information.[26]

Other measurements

Information on cardiovascular risk factors included age, gender, hypertension, diabetes mellitus, hypercholesterolemia, smoking, body mass index (BMI) and history of cardiovascular disease. Hypertension was defined as a systolic blood pressure ≥140 mmHg, diastolic blood pressure ≤90 mmHg, or use of medication for the treatment of hypertension.[28] Diabetes was defined as fasting plasma glucose levels above 7 mmol/L or use of medication indicated for the treatment of diabetes.[28] Hypercholesterolemia was defined as a total cholesterol ≥6.2 mmol/L or use of lipid-lowering medication.[28] BMI was calculated as weight in kilograms divided by square of height in meters. A history of cardiovascular disease was defined as previous myocardial infarction, percutaneous transluminal coronary angioplasty, coronary artery bypass graft or stroke.[8,28]

Statistical analysis

Distribution of metabolic measures was visually inspected for non-normality, and if necessary natural logarithmic or rank-transformations were applied to obtain approximately normal distribution (Supplementary Table 1). The metabolites were scaled to standard deviation (SD) units. ICAC measures were natural logarithmic transformed and subsequently scaled to SD units. To assess the relation of metabolites with ICAC, we performed linear regression analysis in R while adjusting for age, gender, and lipid-lowering medication (Model 1). The associations were further adjusted for hypertension, diabetes, hypercholesterolemia, smoking, and BMI (Model 2). Finally, the associations were tested with additional adjustment for history of cardiovascular disease (Model 3). The summary statistic results were meta-analyzed using inverse variance-weighted fixed-effect meta-analysis. Meta-analysis of two independent Rotterdam study datasets was performed for ICAC, ECAC, AAC and of two Rotterdam study datasets and the Leiden Longevity Study for CAC. Additionally, all analyses were performed in males and females separately.

As metabolic measures are highly correlated (median absolute correlation coefficient = 0.24 Inter Quartile Range (IQR) = 0.11-0.50), we used the method of Li and Ji[29] to

correct for multiple testing. The method calculates the number of independent variables (and thus tests) in correlated measures. The 166 metabolites corresponded to 33 independent variables. Bonferroni correction was applied for the number of independent variables tested (*p*-value threshold for significance: $0.05 / 33 = 1.5 \times 10^{-3}$).

Results

We associated each metabolite with ICAC volume and identified significant positive association of 3-hydroxybutyrate (*p*-value=1.8×10^{-4}) with ICAC (1-standard deviation (SD) increase in concentration of 3-hydroxybutyrate related to a 0.11 SD increase in ICAC volume) (Table 2). Adjustments for traditional cardiovascular risk factors or history of cardiovascular disease did not influence this association (Table 2).

Table 2. Association of 3-hydroxybutyrate with ICAC volume.

Models	Effect (±SE)*	P	N
Model 1			
Age, sex and, lipid lowering medication	0.107(±0.029)	1.76×10-4	1095
Model 2			
Model 1 + hypertension, diabetes, hypercholesterolemia, smoking, and BMI	0.092(±0.030)	2.10×10-3	1059
Model 3			
Model 2 + history of cardiovascular disease	0.092(±0.030)	2.02×10-3	1054

* Effect estimates are SD change in ICAC per 1-SD 3-hydroxybutyrate concentration.

Next, we compared the metabolic association pattern of all measured metabolites for ICAC with the association patterns of the metabolites for ECAC, AAC, and CAC. The effect estimates and statistical significance of the associations are highlighted in heatmaps (Figure 1, Supplementary Table 2). We found differences in the metabolic association pattern of ICAC compared to calcification in the other vessel beds. Among the glycolysis-related metabolic measures, 3-hydroxybutyrate which was significantly associated with ICAC also showed nominally significant association with ECAC (effect=0.07, *p*-value=0.015), but not with CAC and AAC, and glucose was nominally significant associated with ICAC (effect=0.07 , *p*-value=0.012), and ECAC (effect=0.06, *p*-value=0.026) and significantly associated with CAC (effect=0.09, *p*-value=$1.76\text{x}10^{-4}$) (Supplementary Table 2). Importantly, lipoprotein subfractions did not associate with

ICAC, except nominally significant with triglycerides in medium sized LDL (effect=0.06, *p*-value=0.03, Figure 1A, Supplementary Table 2). Furthermore, ICAC was also not associated with the lipid-related metabolites (total triglycerides, total cholesterols, lipoproteins serum lipid extracts and derived lipid measures, Figure 1B, Supplementary Table 2), whereas multiple subfractions of lipoproteins and lipid measures were associated with CAC, AAC and ECAC **(Figure 1, Supplementary Table 2)**. Interestingly, glycoprotein acetyls were associated with calcification in all studied vessel beds.

When we stratified the analysis by sex the association of 3-hydroxybutyrate was nominally significant in both men (effect=0.12, *p*-value=2.8×10^{-3}) and women (effect=0.08, *p*-value=0.036) (Supplementary Figure 1, Supplementary Table 3 and Supplementary Table 4). Interestingly, the association of glycoprotein acetyls with calcification volume was mainly driven by men and was only observed with AAC in women (Supplementary Figure 1). However, in model 2 and model 3 glycoprotein acetyls in men were not associated with ICAC and ECAC (*p*-value>0.05), but were nominally associated with CAC (model 3: effect=0.08, *p*-value=0.029) and AAC (effect=0.12, *p*-value=9.7×10^{-3}). Other metabolites that were significantly associated with ECAC in men were: the ratio of 18:2 linoleic acid to total fatty acids (effect=-0.17, *p*-value=4.6×10^{-5}) and the ratio of omega-6 fatty acids to total fatty acids (effect=-0.15, *p*-value=3.4×10^{-4}). These associations were not modified by additional adjustments in model 2 and 3 (Supplementary Table 3 and 4). Acetate was significantly associated with CAC in women (effect=-0.10, *p*-value=$2.8\text{x}10^{-4}$).

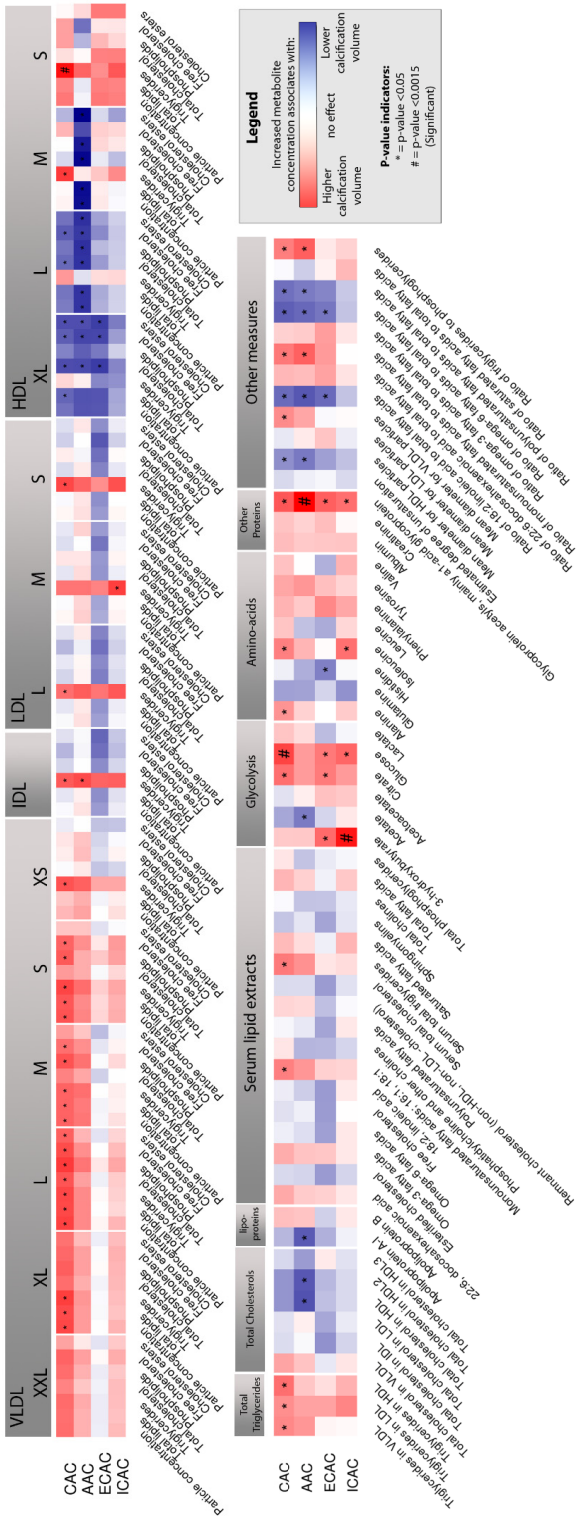


Figure 1. Metabolic association profiles of ICAC, ECAC, AAC and CAC. $^{*}p < 0.05$, $\#p < 1.5\times10^{-3}$. The colors represent standardized effect estimates of the metabolites with the calcification volume. The effect estimates of lipoprotein subfractions with calcification in all studied vessel beds are shown in **Figure 1A** grouped by classes of lipoproteins. The effect estimates of other metabolite measures are shown in **Figure 1B** grouped by type of metabolic measure.

Discussion

In the current population-based study we examined the contribution of a broad range of metabolites to ICAC. We found that higher level of 3-hydroxybutyrate was associated with a larger volume of ICAC. When comparing the metabolic association profile of ICAC with that of calcification in other vessel beds, we found differences. ICAC was associated with metabolites related with glucose metabolism, whereas the other vessel beds were associated with lipid and lipoprotein measures.

In this first study that associated ICAC with metabolomics, the most intriguing finding was the association of 3-hydroxybutyrate (also called beta-hydroxybutyric acid) with ICAC. The ketone 3-hydroxybutyrate is the most abundant of the three ketones bodies (acetoacetate, 3-hydroxybutyrate and acetone) and represents an alternative energy source that is produced by the liver during fasting and prolonged exercise especially for the brain.[30] In addition, fasting-induced 3-hydroxybutyrate was found to enhance expression of the glucose transporter Glut1 in brain endothelial cells, which play an important role in glucose transport across the blood brain barrier.[31] In general, ketone bodies are considered to exert beneficial effects on brain functioning.[32] In this light, our finding that higher concentrations of 3-hydroxybutyrate relate to larger volumes of ICAC seems to contrast these beneficial effects, especially because ICAC is a risk factor for (subclinical) stroke, cognitive decline and dementia.[33,34] Yet, a potential mechanism underlying this association may be found in the property of 3-hydroxybutyrate to form polymers known as Poly-(R)-3-hydroxybutyrates (PHB)s. These short-chain PHBs reside in the lipid core of lipoprotein(a) (Lp(a)), a lipoprotein with profound atherogenic effects and also causally related to coronary heart disease risk.[35-38] Another explanation for the relation of 3-hydroxybutyrate with ICAC may be impaired glucose tolerance. Impaired glucose tolerance is the (pre-) clinical state of diabetes mellitus type 2 (DM2) and is associated with an elevated risk of and a poor prognosis after cardiovascular events.[39,40] 3-hydroxybutyrate levels were increased in individuals with impaired glucose tolerance and in patients with DM2, in whom it predicted worsening of hyperglycemia and incident DM2 in the next 5 years.[41] These data could hypothetically place 3-hydroxybutyrate in the pathway that leads from an impaired

glucose tolerance to increased ICAC and eventually cardiovascular events. Another explanation may be that higher levels of 3-hydroxybutyrate compensate for defective transport of 3-hydroxybutyrate across blood brain barrier due to intracranial arteriosclerosis. Finally, a partial common genetic background might explain the relation between 3-hydroxybutyrate and ICAC.

The most widespread metabolic determinant we observed was glycoprotein acetyls with aortic calcifications in the entire sample, and glycoprotein acetyls associated with all calcifications in men, suggesting etiological differences in men and women.[42,43] Attenuation of these associations in model 2, suggest glycoprotein acetyls might in part reflect pathology related with smoking. Levels of this protein are strongly associated with smoking and physical activity and glycoprotein acetyl concentration has been shown to be a strong predictor of 10-year mortality.[44,45] The protein is a marker of acute-phase reactions and may be implicated in this way in depression,[46] diabetes,[47] cardiovascular disease[48] and cancer.[49] In the both cohorts from our study participants, these risk factors possibly are more prominent in men.

We also compared the metabolic association patterns of ICAC with that of calcification in three other major vessel beds, namely the coronary arteries, aortic arch and the extracranial carotid arteries. We observed a different association pattern for ICAC compared to calcification in the other vessel beds, with the most prominent differences with CAC. These findings further confirm the previously suggested hypothesis of markers for location-specific arterial calcification.[5,25,50] Most strikingly, our findings suggest a prominent role for glycolysis-related metabolic measures in the formation of ICAC and a less prominent role for the metabolic measures that are classically considered to be important in atherogenesis, especially cholesterol and triglycerides in lipoproteins. In contrast, we did observe associations of high concentrations of VLDL subfractions and lower concentrations of HDL subfractions as well as a high percentage of mono-saturated fatty acids with CAC.

The strength of the study includes the large sample with standardized assessments of metabolic measures and arterial calcification in multiple vessel beds, enabling

comparisons of the metabolic association patterns of ICAC with calcification in the other vessel beds. The metabolomics platform that we used contains a large proportion of lipoprotein or other lipid measures which provides an excellent opportunity to study arteriosclerosis.[15,23,24,51] However, it should be acknowledged that many other metabolites can be measured,[52] which may be of importance to arteriosclerosis. There are also certain limitations of our study that should be noted. First, even though calcification is a validated marker of arteriosclerosis, its most common subtype, atherosclerosis, contains more than solely calcification. Especially, these non-calcified components of the atherosclerotic plaque may also be influenced by the studied metabolites.[53] Another limitation of the current study is that metabolites and CT scan have not been taken at the same time. Future studies should address this point too. Lastly, lack of other studies that measure ICAC that could have been used to increase the study population and the statistical power can be considered as a possible limitation even though we studied ICAC in the largest population-based dataset. Therefore, we urge future replication efforts of our findings in independent datasets.

Conclusions

We found that 3-hydroxybutyrate was specifically associated with ICAC. When comparing the metabolic association profile of ICAC with that of calcification in other vessel beds, we found differences. Larger studies could further study the metabolic differences and investigation of the underlying biological mechanisms for the identified association should be the subject of future biological studies.

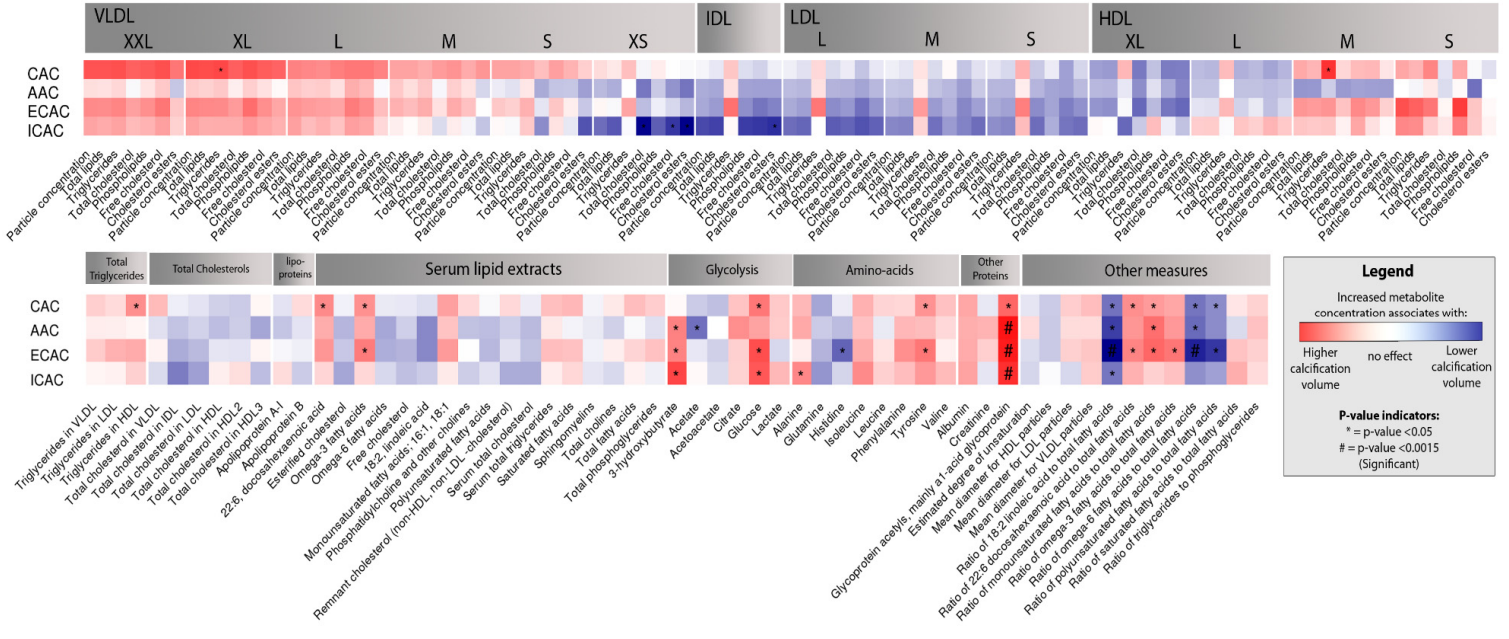
References:

1. Gorelick PB, Wong KS, Bae HJ, Pandey DK. Large artery intracranial occlusive disease: a large worldwide burden but a relatively neglected frontier. *Stroke* 2008; **39**(8): 2396-9.
2. Arenillas JF. Intracranial atherosclerosis: current concepts. *Stroke* 2011; **42**(1 Suppl): S20-3.
3. Qureshi AI, Caplan LR. Intracranial atherosclerosis. *Lancet* 2014; **383**(9921): 984-98.
4. Bos D, Portegies ML, van der Lugt A, et al. Intracranial carotid artery atherosclerosis and the risk of stroke in whites: the Rotterdam Study. *JAMA Neurol* 2014; **71**(4): 405-11.
5. Bos D, Ikram MA, Isaacs A, et al. Genetic loci for coronary calcification and serum lipids relate to aortic and carotid calcification. *Circ Cardiovasc Genet* 2013; **6**(1): 47-53.
6. Lopez-Cancio E, Galan A, Dorado L, et al. Biological signatures of asymptomatic extra- and intracranial atherosclerosis: the Barcelona-AsIA (Asymptomatic Intracranial Atherosclerosis) study. *Stroke* 2012; **43**(10): 2712-9.

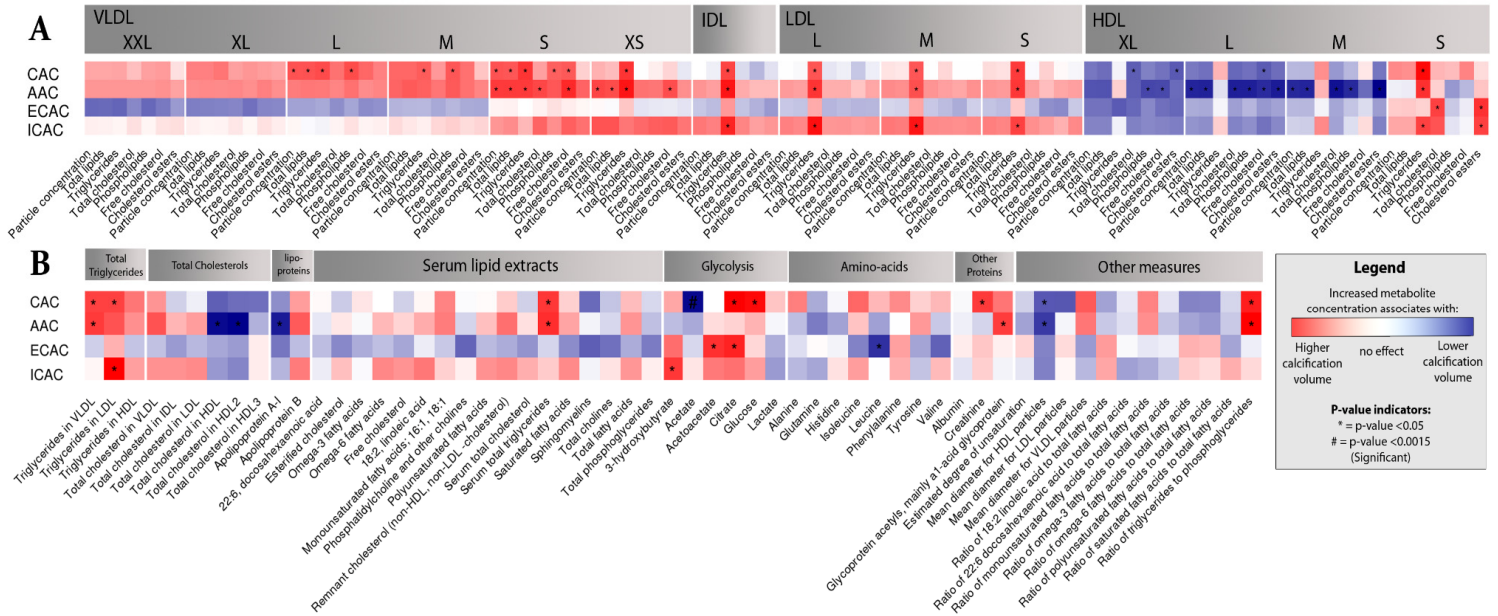
7. Allison MA, Criqui MH, Wright CM. Patterns and risk factors for systemic calcified atherosclerosis. *Arterioscler Thromb Vasc Biol* 2004; **24**(2): 331-6.
8. Bos D, van der Rijk MJ, Geeraedts TE, et al. Intracranial carotid artery atherosclerosis: prevalence and risk factors in the general population. *Stroke* 2012; **43**(7): 1878-84.
9. Mazighi M, Labreuche J, Gongora-Rivera F, Duyckaerts C, Hauw JJ, Amarenco P. Autopsy prevalence of intracranial atherosclerosis in patients with fatal stroke. *Stroke* 2008; **39**(4): 1142-7.
10. Quehenberger O, Dennis EA. The human plasma lipidome. *N Engl J Med* 2011; **365**(19): 1812-23.
11. Inouye M, Kettunen J, Soininen P, et al. Metabonomic, transcriptomic, and genomic variation of a population cohort. *Mol Syst Biol* 2010; **6**: 441.
12. Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes. *Nat Med* 2011; **17**(4): 448-53.
13. Mahendran Y, Cederberg H, Vangipurapu J, et al. Glycerol and fatty acids in serum predict the development of hyperglycemia and type 2 diabetes in Finnish men. *Diabetes Care* 2013; **36**(11): 3732-8.
14. Shah SH, Kraus WE, Newgard CB. Metabolomic profiling for the identification of novel biomarkers and mechanisms related to common cardiovascular diseases: form and function. *Circulation* 2012; **126**(9): 1110-20.
15. Wurtz P, Raiko JR, Magnussen CG, et al. High-throughput quantification of circulating metabolites improves prediction of subclinical atherosclerosis. *Eur Heart J* 2012; **33**(18): 2307-16.
16. Stegemann C, Pechlaner R, Willeit P, et al. Lipidomics profiling and risk of cardiovascular disease in the prospective population-based Bruneck study. *Circulation* 2014; **129**(18): 1821-31.
17. Roberts LD, Gerszten RE. Toward new biomarkers of cardiometabolic diseases. *Cell Metab* 2013; **18**(1): 43-50.
18. Hofman A, Brusselle GG, Darwish Murad S, et al. The Rotterdam Study: 2016 objectives and design update. *Eur J Epidemiol* 2015; **30**(8): 661-708.
19. Westendorp RG, van Heemst D, Rozing MP, et al. Nonagenarian siblings and their offspring display lower risk of mortality and morbidity than sporadic nonagenarians: The Leiden Longevity Study. *J Am Geriatr Soc* 2009; **57**(9): 1634-7.
20. Schoenmaker M, de Craen AJ, de Meijer PH, et al. Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *Eur J Hum Genet* 2006; **14**(1): 79-84.
21. Hofman A, Brusselle GGO, Darwish Murad S, et al. The Rotterdam Study: 2016 objectives and design update. *Eur J Epidemiol* 2015; **30**(8): 661-708.
22. Huan T, Esko T, Peters MJ, et al. A meta-analysis of gene expression signatures of blood pressure and hypertension. *PLoS Genet* 2015; **11**(3): e1005035.
23. Soininen P, Kangas AJ, Wurtz P, et al. High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism. *Analyst* 2009; **134**(9): 1781-5.
24. Soininen P, Kangas AJ, Wurtz P, Suna T, Ala-Korpela M. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circ Cardiovasc Genet* 2015; **8**(1): 192-206.
25. Bos D, Ikram MA, Elias-Smale SE, et al. Calcification in Major Vessel Beds Relates to Vascular Brain Disease. *Arterioscl Throm Vas* 2011; **31**(10): 2331-7.
26. Kroft LJ, van der Bijl N, van der Grond J, et al. Low computed tomography coronary artery calcium scores in familial longevity: the Leiden Longevity Study. *Age (Dordr)* 2014; **36**(4): 9668.
27. Agatston AS, Janowitz WR, Hildner FJ, Zusmer NR, Viamonte M, Detrano R. Quantification of Coronary-Artery Calcium Using Ultrafast Computed-Tomography. *J Am Coll Cardiol* 1990; **15**(4): 827-32.
28. Odink AE, van der Lugt A, Hofman A, et al. Risk factors for coronary, aortic arch and carotid calcification; The Rotterdam Study. *J Hum Hypertens* 2010; **24**(2): 86-92.

29. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* 2005; **95**(3): 221-7.
30. Owen OE, Morgan AP, Kemp HG, Sullivan JM, Herrera MG, Cahill GF, Jr. Brain metabolism during fasting. *J Clin Invest* 1967; **46**(10): 1589-95.
31. Tanegashima K, Sato-Miyata Y, Funakoshi M, Nishito Y, Aigaki T, Hara T. Epigenetic regulation of the glucose transporter gene Slc2a1 by -hydroxybutyrate underlies preferential glucose supply to the brain of fasted mice. *Genes Cells* 2017; **22**(1): 71-83.
32. Rahman M, Muhammad S, Khan MA, et al. The beta-hydroxybutyrate receptor HCA2 activates a neuroprotective subset of macrophages. *Nat Commun* 2014; **5**: 3944.
33. Bos D, Portegies MLP, van der Lugt A, et al. Intracranial Carotid Artery Atherosclerosis and the Risk of Stroke in Whites The Rotterdam Study. *Jama Neurology* 2014; **71**(4): 405-11.
34. Bos D, Vernooij MW, de Bruijn RFAG, et al. Atherosclerotic calcification is related to a higher risk of dementia and cognitive decline. *Alzheimers & Dementia* 2015; **11**(6): 639-47.
35. Kettunen J, Demirkan A, Wurtz P, et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat Commun* 2016; **7**: 11122.
36. Tregouet DA, Konig IR, Erdmann J, et al. Genome-wide haplotype association study identifies the SLC22A3-LPAL2-LPA gene cluster as a risk locus for coronary artery disease. *Nature Genetics* 2009; **41**(3): 283-5.
37. Schunkert H, Konig IR, Kathiresan S, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature Genetics* 2011; **43**(4): 333-U153.
38. Reusch RN. Poly-(R)-3-hydroxybutyrates (PHB) are atherogenic components of lipoprotein Lp(a). *Med Hypotheses* 2015; **85**(6): 1041-3.
39. Barr ELM, Zimmet PZ, Welborn TA, et al. Risk of cardiovascular and all-cause mortality in individuals with diabetes mellitus, impaired fasting glucose, and impaired glucose tolerance - The Australian diabetes, obesity, and lifestyle study (AusDiab). *Circulation* 2007; **116**(2): 151-7.
40. Kurihara O, Takano M, Yamamoto M, et al. Impact of Prediabetic Status on Coronary Atherosclerosis A multivessel angioscopic study. *Diabetes Care* 2013; **36**(3): 729-33.
41. Mahendran Y, Vangipurapu J, Cederberg H, et al. Association of Ketone Body Levels With Hyperglycemia and Type 2 Diabetes in 9,398 Finnish Men. *Diabetes* 2013; **62**(10): 3618-26.
42. Leening MJ, Ferket BS, Steyerberg EW, et al. Sex differences in lifetime risk and first manifestation of cardiovascular disease: prospective population based cohort study. *BMJ* 2014; **349**: g5992.
43. Maas AH, Appelman YE. Gender differences in coronary heart disease. *Neth Heart J* 2010; **18**(12): 598-602.
44. Fischer K, Kettunen J, Wurtz P, et al. Biomarker profiling by nuclear magnetic resonance spectroscopy for the prediction of all-cause mortality: an observational study of 17,345 persons. *PLoS Med* 2014; **11**(2): e1001606.
45. Singh-Manoux A, Shipley MJ, Bell JA, Canonico M, Elbaz A, Kivimaki M. Association between inflammatory biomarkers and all-cause, cardiovascular and cancer-related mortality. *CMAJ* 2016.
46. Harley J, Roberts R, Joyce P, et al. Orosomucoid influences the response to antidepressants in major depressive disorder. *J Psychopharmacol* 2010; **24**(4): 531-5.
47. El-Beblawy NMS, Andrawes NG, Ismail EAR, Enany BE, Abou El-Seoud HS, Erfan MA. Serum and Urinary Orosomucoid in Young Patients With Type 1 Diabetes: A Link Between Inflammation, Microvascular Complications, and Subclinical Atherosclerosis. *Clin Appl Thromb-Hem* 2016; **22**(8): 718-26.
48. Carriere I, Dupuy AM, Lacroux A, Cristol JP, Delcourt C, St POLA. Biomarkers of inflammation and malnutrition associated with early death in healthy elderly people. *Journal of the American Geriatrics Society* 2008; **56**(5): 840-6.

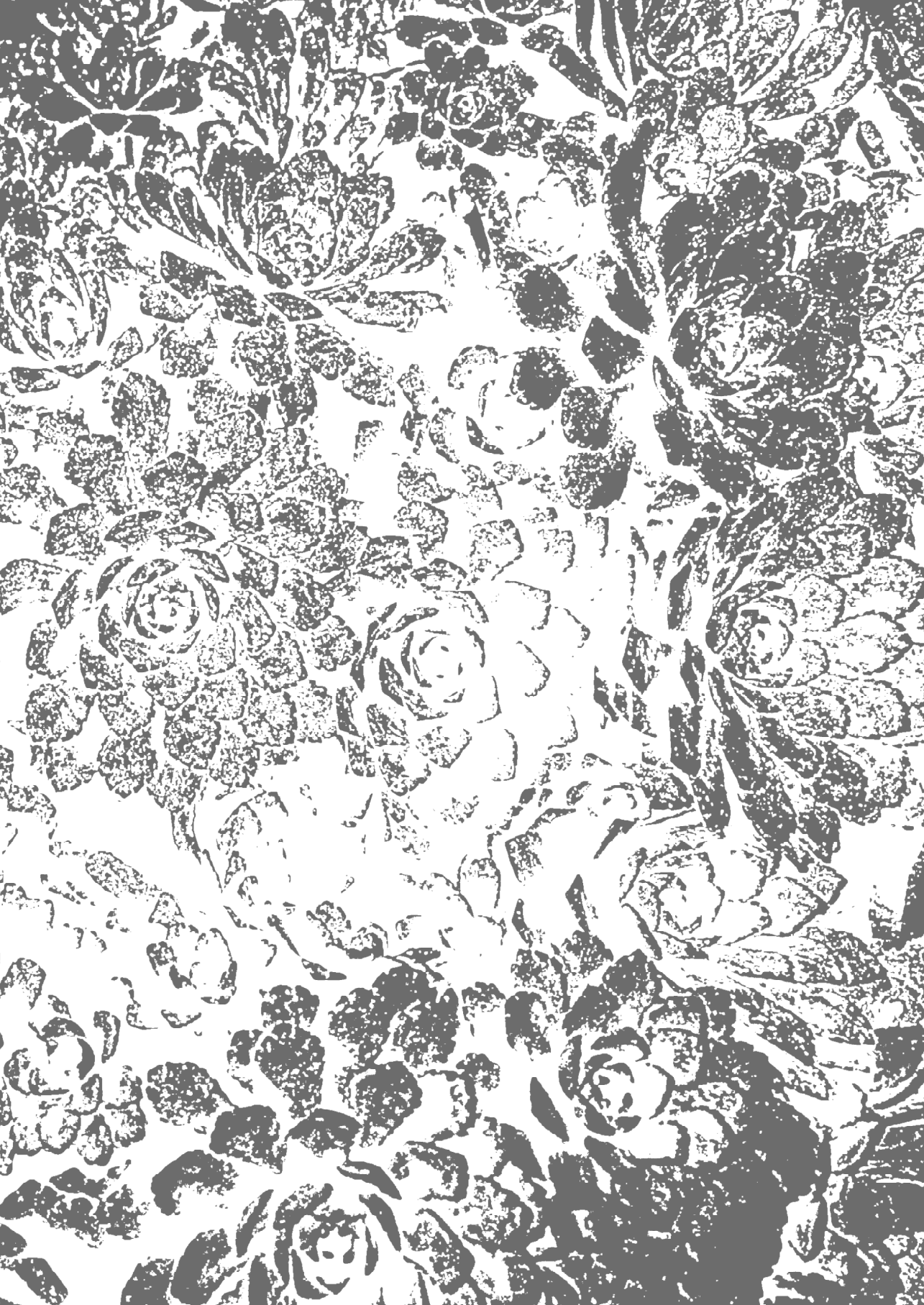
49. Bruno R, Olivares R, Berille J, et al. alpha-1-acid glycoprotein as an independent predictor for treatment effects and a prognostic factor of survival in patients with non-small cell lung cancer treated with docetaxel. *Clinical Cancer Research* 2003; **9**(3): 1077-82.
50. Bos D, Leening MJ, Kavousi M, et al. Comparison of Atherosclerotic Calcification in Major Vessel Beds on the Risk of All-Cause and Cause-Specific Mortality: The Rotterdam Study. *Circ Cardiovasc Imaging* 2015; **8**(12).
51. Wurtz P, Havulinna AS, Soininen P, et al. Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation* 2015; **131**(9): 774-85.
52. Wishart DS, Jewison T, Guo AC, et al. HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Res* 2013; **41**(Database issue): D801-7.
53. Stary HC, Chandler AB, Dinsmore RE, et al. A definition of advanced types of atherosclerotic lesions and a histological classification of atherosclerosis. A report from the Committee on Vascular Lesions of the Council on Arteriosclerosis, American Heart Association. *Arterioscler Thromb Vasc Biol* 1995; **15**(9): 1512-31.



Supplementary Figure 1. Metabolic association profiles of ICAC, ECAC, AAC and CAC in men. *$p < 0.05$, #$p < 1.5{\times}10^{-3}$. The colors represent standardized effect estimates of the metabolites with the calcification volume. The effect estimates of lipoprotein subfractions with calcification in all studied vessel beds are shown in **A** grouped by classes of lipoproteins. The effect estimates of other metabolite measures are shown in **B** grouped by type of metabolic measure.



Supplementary Figure 2. Metabolic association profiles of ICAC, ECAC, AAC and CAC in women. $*p < 0.05$, $\#p < 1.5\times10^{-3}$. The colors represent standardized effect estimates of the metabolites with the calcification volume. The effect estimates of lipoprotein subfractions with calcification in all studied vessel beds are shown in **A** grouped by classes of lipoproteins. The effect estimates of other metabolite measures are shown in **B** grouped by type of metabolic measure.



Chapter 6

General discussion

The aim of this thesis was to identify molecular mechanisms underlying Alzheimer's disease (AD) and its endophenotypes through genetic studies. This general discussion will highlight the main findings from this thesis and place these in the broader context of developments in genetic research. Further, the possible clinical implications and the next steps to unravel the genetics of this devastating disease are discussed.

Translation of genetic risk factors to the clinic

Since the discovery in the mid-1990s of mutations that cause AD in the gene coding for the amyloid precursor protein (*APP*) and in presenilin 1 and 2 genes (*PSEN1* and *PSEN2*), AD patients have been screened for mutations in a clinical genetics setting to elucidate the risk of relatives.[1] These genetic discoveries also benchmarked a causal role for β-amyloid processing in the pathogenesis of AD,[2] led to the discovery of amyloid-β-42 and amyloid-β-40 in CSF[3] as biomarkers and the development of medication targeting β-amyloid and tau. Apart from the Mendelian mutations in *PSEN1/2* and *APP* a common variant in the gene Apolipoprotein E (*APOE*) has been discovered that increases AD risk three fold per allele, but is not sufficient to cause the disease.[4,5] *APOE* is therefore not used for risk prediction in clinic, but is widely used to stratify risk groups in research.[6] Since these milestones two dozen common genetic variants have been discovered in the last decade through genome-wide association studies (GWAS) and sequencing studies uncovered several rare variants associate with AD. These discovered variants have relatively small effects on the risk of AD, limited predictive ability[7] and consequently are considered to have no clinical use. In the first chapter of my thesis I described studies that address the clinical utility of genetic findings.

In **chapter 1.1** I describe the combined effects of *APOE* and 23 genetic variants previously associated with AD in a genetic risk score (GRS) on the cumulative risk and onset age of AD and dementia. The GRS had significant effects on risk of AD and the magnitude of the effect was in part determined by *APOE* genotypes, i.e. there was significant evidence for interaction. Further these risk differences are accompanied by shifts in age at onset. Currently, the common risk variants are not used for risk stratification in clinical studies or intervention trails, but our results showed clinically relevant differences in age at onset between subgroups based on the GRS and *APOE*.

The identification of small subgroups at highest genetic risk of AD with an earlier onset age has important implications for preventive measures in precision medicine. Precision medicine is the emerging approach for complex disease treatment and prevention that takes into account individual variability in environment, lifestyle and genes for each person.[8] Pathological changes related to AD begin to develop decades before the earliest clinical symptoms, therefore preventive interventions, such as participation in pharmaceutical trials, should be offered earlier to the subgroups at highest risk.[9] Specially as the failures of clinical trials has led to general awareness that initiation of treatment should be in the very early, maybe even pre-clinical stages of disease for the treatments to be effective.[10] An important economic benefit for these costly trails is that selection of only the subgroups at highest risk will decrease the duration of the trails.[10] Another practical application of **chapter 2.1** is that more accurately risk estimates can be reported to customers of direct-to-consumer genetic testing. For example, one of the largest direct-to-consumer genetic testing companies "*23andMe*" only reports *APOE* genotype specific risks,[11,12] this now can be extended by showing additional subgroups based on the common risk variants.

With the increasing availability of whole exome sequencing in clinical practice it is possible to detect uncommon, highly personal exonic variants.[13,14] In current clinical practice to find AD causative variants the *APP*, *PSEN1* and *PSEN2* genes are screened. The *SORL1* gene is also known to contain rare exonic variants that associate with AD risk increase.[15-20] However, it is unclear which of these *SORL1* rare genetic variants are clinically relevant. Therefore, I studied the characteristics of *SORL1* variants in **chapter 2.2**. This is the first study that attempts to classify rare variants in a gene other than *APP*, *PSEN1* and *PSEN2*. I discovered near private variants that were predicted to disturb protein structure associated with an up to 12-fold increase in AD risk which is comparable to the risk increase of *APOE* ε4 homozygotes.[21] Moreover, protein truncating *SORL1* variants were carried only in AD-cases suggesting this type of *SORL1* variants is causal for AD and comparable to *APP*, *PSEN1* and *PSEN2* mutations. Additionally in other independent studies protein truncating variants were also exclusively reported in AD patients.[15,17,19] These findings can be useful in clinical practice to explain a part of AD that clusters in families.

In **chapter 2.3** I report the value of the most commonly used genetic test in clinic; "asking for family history of dementia". A positive family history is known to be an important risk factor for dementia, the underlying mechanisms and genes remain largely unknown to date. In this study we show that when parents were diagnosed before the age of 80 years, the risk of AD for their children was 2.5-fold increased, but not for those with parents that were diagnosed above 80 years. Again, risk stratification for preventive interventions, as well as selection of participants for research purposes may benefit from these results.

New ways to study rare variants through imputation

The Haplotype Reference Consortium (HRC) released in 2015 is a large imputation panel that allows more accurate imputation of genetic variants.[22] This reference panel of 64,976 human haplotypes includes the previously used reference panels such as the HapMap[23], 1000Genomes project (1000GP)[24], Genome of the Netherlands (GoNL)[25] and UK10K[26] project. **Chapter 3.1** describes a study that evaluates the performance of the HRC imputation panel against the 1000GP. I compared a set of directly assayed common and rare variants from the exome array with imputed genotypes from 1000GP and HRC. Imputation using the HRC panel was superior to the 1000GP for common, low-frequency variants and rare variants. The increased concordance for low-frequency variants ($0.01 < MAF < 0.05$) is relevant for gene discovery, as previously highlighted by the HRC.[22] These variants, which are commonly filtered out in meta-analyses of GWAS, can now be studied reliably using HRC imputations. For rare variants ($0.001 < MAF < 0.01$) there is also a relevant improvement in concordance that facilitates in silico validation of findings of exome array or sequencing results. I used this last property of the HRC to *in-silico* replicate variants in a large AD association study described in **chapter 3.4**.

Search for rare variants associated with AD

Even though large scale GWAS led to the discovery of over two dozen common genetic variants associated with AD[27-36], the individual effects of the identified genetic variants were small.[7,37-39] Next generation sequencing accelerated the search for rare genetic variants that could have relatively larger impact on the disease compared to the

common genetic variants. Early sequencing studies identified rare variants in *TREM2* and *PLD3* genes. While the *TREM2* variants were confirmed by subsequent independent studies, association of *PLD3* variants remained unsubstantiated as described in **chapter 3.2** and companion papers by Heilmann *et al*., Lambert *et al*. and Hooli *et al*.[40-42] Meta-analysis of all discovery and replication cohorts did not show convincing evidence that *PLD3*-Val232Met associates with AD (p=3.5×10^{-4}).[41] Besides this, the functional work from the original paper linking *PLD3* to APP expression is now being re-evaluated.[43] I further note in this chapter a high variability of the frequency of the *PLD3*-Val232Met variant across our populations. The important implication of our work for future rare variant studies is that rare variant findings preferably should be replicated in independent samples and a high variability of the frequency rare variants across populations is possible. This highlights the need for careful matching of cases and controls for ethnic background when investigating rare variants.

Chapter 3.3 and 3.4 describe two studies that aimed to identify new rare variants associated with AD. In **chapter 3.3** I performed, together with collaborators from the CHARGE consortium, an exome-wide association analysis in 1,393 late-onset AD (LOAD) cases and 8,141 controls. We discovered a rare variant (P155L) in *TM2D3*, enriched in Icelanders (~0.5% versus <0.05% in other European populations), associated with AD. In 433 LOAD cases and 3,903 controls from the Icelandic AGES sub-study, P155L was associated with 7.5-times increased risk and an earlier onset of LOAD (p=6.6x10^{-9}). Through functional work we discovered that mutation in the Drosophila *TM2D3* homolog, almondex, causes a phenotype similar to loss of Notch/Presenilin signaling. Human *TM2D3* is capable of rescuing these phenotypes, but this activity is abolished by P155L, establishing it as a functionally damaging allele. These results establish a rare *TM2D3* variant in association with LOAD susceptibility, and together with prior work suggest a possible link to the β-amyloid cascade. A limitation of this study is that we were not able to replicate the association *TM2D3* outside the Icelandic population. In **chapter 3.4** I expand the search rare variants together with collaborators from the International Genomics of Alzheimer's Project (IGAP) consortium. We combined results from four different consortia (CHARGE,

ADGC, GERAD and EADI) in a 3-stage case-control study of up to a total of 85,133 subjects. In addition to wet-lab genotyping we used imputation to the HRC imputation panel as shown in **chapter 3.1** for in-silico replication of the rare variants. We observed 3 novel genome-wide significant AD associated non-synonymous variants; a protective variant in *PLCG2 (*p.P522R), a risk variant in *ABI3* (p.S209F), and a novel GWS variant in *TREM2* (p.R62H), a known AD susceptibility gene. *TREM2*, *ABI3* and *PLCG2* have a common expression pattern in human brain cortex, with high expression in microglia cells and limited expression in neurons, oligodendrocytes, astrocytes and endothelial cells. Other known LOAD loci with the same expression pattern include *SORL1*, the *MS4A* gene cluster, and *HLA-DRB1*. *PLCG2*, *ABI3*, and *TREM2* are up-regulated in LOAD human cortex and in two *APP* mouse models. These changes in expression appear to be related to microgliosis. Together this chapter provides additional evidence that the microglial response in LOAD is directly part of a causal pathway leading to disease and is not simply a downstream consequence of neurodegeneration.[44-46] Current new developments in AD related therapies are targeted at modifying this immune response that is initiated as a reaction to aberrant amyloid-β plaques.[44-46] Our results are encouraging that these immuno-modulating therapies might be successful. Especially *PLCG2* is interesting as a drug target, as it is an enzyme, it is specifically expressed in microglia and we show a rare variant to have a strong protective effect for AD. In fact the protein product of *PLCG2*-p.P522R can be seen as a human-model of how AD could be prevented. If this 'mutated' plcγ2 protein could be carried to or expressed in the brain through medication at the right time in the disease process, it should in theory have a beneficial effect. Future biological studies of plcγ2-P522R can reveal when and how in the process that leads to AD the protein exerts its protective effect.

I have summarized all new insights in the genetics of AD described in this thesis in a single figure (**Figure 1**), the shaded genes are the ones previously reported. In this figure the frequency of genetic variants can be found on the x-axis and the impact of the variant AD is plotted on the y-axis. The previously described genetic findings are shown shaded in the figure.[27-36]

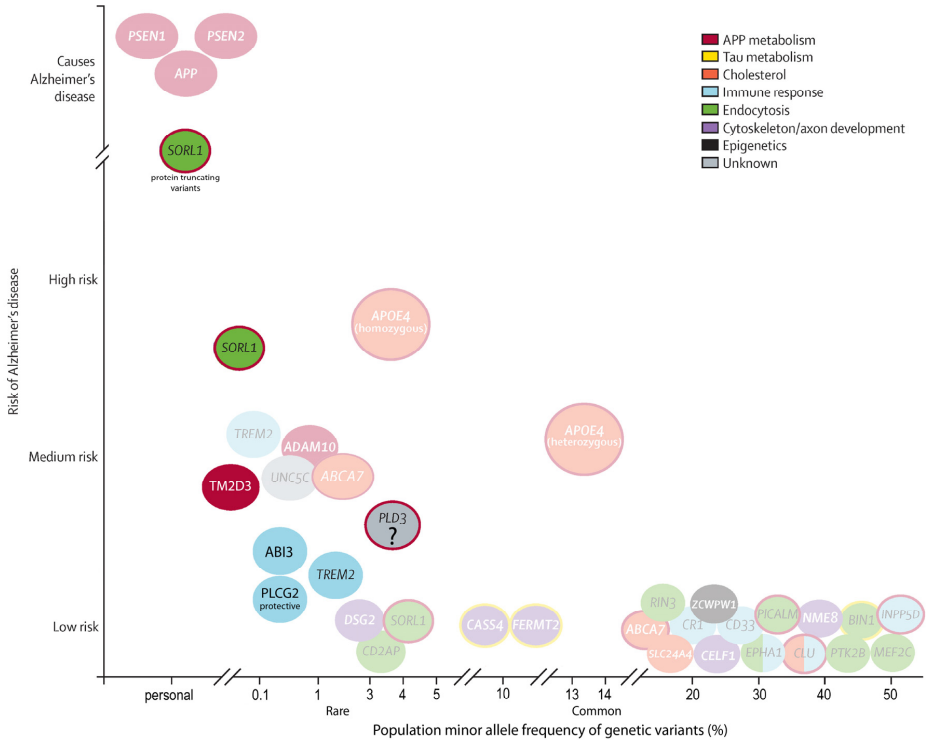


Figure 1: Summary of the contributions described in this thesis to the field of genetic epidemiology of Alzheimer's disease.

Endophenotypes of AD in neuroimaging of the brain

Brain magnetic resonance imaging allows us to visualize the structures of brain *in vivo*. Studying the genetics of brain structures could inform us on development of the brain and possibly allow insights in the pathophysiology of AD. In **chapter 4** I focused on the genetics of these MRI-based measures. In **chapter 4.1** I studied the genetic background of grey matter through studies of detailed measures of the brain obtained from high-resolution structural imaging (i.e. voxels) in unrelated subjects from the Rotterdam Study and two family-based studies. The resulting voxel-wise heritability maps provided detailed information which part of the brain has a strong genetic background. Significantly heritable voxels mapped to subcortical structures, but also to voxels in the language areas of the left hemisphere. This is in line with the findings of previous studies that the volumes of subcortical structure are among the most heritable in the brain.[47] Although I used different methods to calculate heritability in the population and in the

family data there was high regional consistency of the heritability measures across study designs (Pearson's correlation coefficient=0.73, p=2.6×10^{-13}). I subsequently showed enhancement of association signals of two genetic variants, previously linked with subcortical brain volume,[48] using only the most heritable voxels. This is the first time such an enhancement of signal is shown and if applied to future voxel-wise genome-wide association studies we do not expect statistical signals to be uniformly enhanced. High heritability estimates capture a variety of sources that can affect power to detect associations, including lower signal to noise ratios and higher genetic homogeneity.[49,50] Using these benefits to increase statistical signal is desirable, irrespective of the underlying cause and the heritability maps from this study were made freely available enabling future studies to use them.

In **chapter 4.2** we report the effects of genetic variants associated with AD[29] on brain voxel-based morphometry (VBM) in cognitively healthy elderly. We studied the genetic effect on healthy brain morphology to provide insight into disease etiology in the pre-clinical phase of AD. None of the individual voxel passed multiple testing correction suggesting there is no large effect of these genetic loci in cognitively healthy individuals. However, I did find significant spatial overlap between the effects of AD risk loci on VBM and the expression of AD related genes (*MEF2C*, *CLU*, *SLC24A4*) in the Allen Brain Atlas. Taken together this suggests this approach might be able to reliably identify voxel associations and larger samples will be necessary to clarify the identified associations.

Brain lobar volumes are highly and differentially heritable[51], but specific genetic association studies are scarce. In **chapter 4.3** we performed a genome-wide association study of the volume of the frontal, occipital, parietal and temporal lobe in 16,016 individuals from 21 cohorts from the CHARGE consortium, we identified six genetic loci associated with specific lobar volumes. We found rs1337736 (6q22.32), which is located in a region previously associated with intracranial volume (ICV). Rs61921502 (12q14.3) associated with occipital lobe volume was associated with temporal lobe volume, which is located in a region associated with hippocampal volume.[48,52,53] Four loci not previously associated with brain volume measurements were common and located in regions rich in epigenetic data (*DAAM1* and *THBS3*) or located close to genes that cause

brain related diseases with Mendelian inheritance patterns (*ZIC4* and *FGFRL1*).[54-59] There was no overlap of the significant loci with AD nor was there a genetic overlap at the genome-wide level. The findings reported in this chapter reveal part of the complex genetics underlying brain development and links brain lobe size in healthy individuals with genes that cause Mendelian syndromes that affect head shape or intelligence, but not those related with AD.

Blood-based markers of AD

In **chapter 5.1** we performed a genome-wide association study of amyloid-β concentration (Aβ) in blood to elucidate, which factors are important in modifying its levels. Concentration of Aβ in blood are interesting to study as blood is easily accessible and are associated with incident AD and dementia.[60-65] In up to 11,969 subjects and we found two genome-wide significant loci and several interesting suggestive loci that associate with Aβ in blood. The first rs429358, a non-synonymous variant responsible for the *APOE* ε4 allele, associated with Aβ42 and the ratio of Aβ42/Aβ40. *APOE* ε4 is the most important risk factor for AD and has previously been associated with levels of Aβ in CSF as well as brain pathology markers in genome-wide association studies.[66-69] The second variant rs650585 (45 kb from the *BACE1* gene) associated with lower plasma Aβ40 levels (p=3.3×10^{-8}). We further observed that variants near *APP* and *PSEN2* suggestively associated with Aβ blood concentrations. This study showed first evidence that blood Aβ levels associate with variants in and near major AD genes (Figure 1) and *BACE1*. The association with *BACE1* is especially interesting as a *BACE1* is involved in the production of Aβ. The fact that *BACE1* effects blood Aβ supports that the effect of the multitude of drugs, so called BACE inhibitors, might be effective in treating AD a that their effects might be followed in blood. These BACE inhibitors, such as the compounds MK8931, AZD-3293, JNJ-54861911 and CNP520, are currently in the later stages of drug trails for AD.[70] Our genetic study of circulating Aβ showed important insights into Aβ processing and showed that blood Aβ could be an interesting readout for clinical trials of BACE inhibitors.

In **chapter 5.2** I performed together with collaborators a metabolomics study, focusing on only incident AD and dementia in prospective studies. Eight prospective cohorts

with 22,623 participants profiled by NMR or MS metabolomics. Four cohorts served as the discovery sample with replication in four independent cohorts. Three branched-chain amino acids (BCAA) (iso-leucine, leucine, valine), creatinine and two VLDL specific lipoprotein lipid subclasses were associated with lower dementia risk. One HDL (L-HDL-CE-%) and one VLDL (XL-VLDL-C-%) specific lipoprotein lipid subclass was associated with increased dementia risk. BCAAs were also associated with decreased and L-HDL-CE-% with increased AD risk. Further studies are needed to elucidate whether these metabolites or lipoprotein lipids play a causal role in the pathogenesis of dementia or are early markers of developing cognitive impairment.

Chapter 5.3 describes the results from a metabolomics analysis in 11 independent cohorts. We studied the association of the metabolites with cognition, dementia and life style. Studying circulating metabolites that are associated with cognition and dementia may improve our understanding of the pathogenesis of dementia and provide crucial read-outs for preventive and therapeutic life style interventions. We discovered and replicated 15 metabolites associated with general cognitive ability including sub-fractions of high density lipoprotein, docosahexaenoic acid (DHA), ornithine, glutamine and glycoprotein acetyls. Six of the cognition-associated metabolites were also related to the risk of dementia and lifestyle factors. DHA and free cholesterol in HDL were most significantly associated with AD and dementia. This is the largest study exploring the association of a large array of blood based metabolites with general cognitive ability to date. Also we followed a thorough replication strategy, which is often lacking in other metabolite studies.[71,72] Our results show that circulating metabolites are consistently associated with cognition, dementia and lifestyle factors, opening new avenues for prevention of cognitive decline and dementia.

Lastly in **chapter 5.4** I studied the association of metabolites as modifiable risk factors for intra-carotid artery calcifications was studied. Intra-carotid artery calcification volume is an emerging measure of arteriosclerosis that is associated with stroke,[73] cognitive decline and dementia.[74] The etiology of arteriosclerosis in intracranial carotid arteries is thought to be different from arteriosclerosis in other main vessel beds, but its etiology is largely unknown. Therefore I studied, as described in **Chapter 5.4**, the

association of circulating metabolites with internal carotid coronary artery calcifications (ICAC), a proxy for arteriosclerosis. In total 1,111 individuals of the Rotterdam study had metabolites measured and underwent CT-scanning. There was significant evidence for association between 3-hydroxybutyrate (p-value = 2.3 x 10^{-4}) and ICAC. When comparing the metabolic association profile of ICAC with that of three other main vessel beds including; coronary arteries, aortic arch, extracranial carotid arteries. The metabolic association profile of ICAC volume was different from the metabolic association profile of three other major vessel beds suggesting distinct differences in location specific etiology of calcification across vessel beds. Especially the metabolic association profile of coronary artery calcifications was different from ICAC. ICAC showed sparse associations with lipoprotein subfractions and none of the lipid measures, the traditional risk factor targets of lipid and cholesterol lowering medication. Possibly the treatment of these plaques requires a different approach.

Methodological considerations

Using genetics to study Alzheimer's disease and its related endophenotypes has major benefits over studying other determinants of disease. The direction-of-causation in genetic studies is clear, genetic variants can be very reliably determined in different studies and genome-wide association studies are hypotheses free. Genetic studies are therefore hypothesis generating and can provide novel insights. A perquisite of GWAS studies is that is has sufficient statistical power to detect true associations. For especially the rare variant association studies described in chapter 3 this is difficult as statistical power greatly depends on the frequency of the studied genetic variations.[75,76] The most straightforward way to attain sufficient power is to increase the sample size of the study. For example in chapter 2.4 where the stage 1 had for variants of approximately 1% MAF approximately 60% statistical power to identify a genome-wide significant genetic association with an odds ratio of 1.5 and 100% power to detect an variant with an OR of 2.[77] Indeed the *TREM2* variant (OR=2.4) was genome-wide significant in stage 1 and the *PLCG2* and *ABI3* variants were significant only after adding over 30,000 samples to the study in stage 2. A major drawback of attaining a large sample size through the cheap exome arrays used in chapter 2.3 and 2.4 is that the number of variants tested is limited making it very likely additional rare variants can be discovered testing more variants in

sufficient number of samples. Other limitations of increasing the size of a study by combining the results of multiple studies is that it increases the heterogeneity among genotypes and phenotype definition. This might result in that population specific effects of variants might go unnoticed. In contrast, population specific effects can also be used to boost the power of rare variant studies by studying genetic isolates, where the allele frequency of rare variants can increase due to genetic drift. A good example is the *TM2D3* variant identified in chapter 2.3, which was enriched 10-fold in the Icelandic population.

Other options to increase the power of genetic studies are to study homogeneous AD subgroups, based on e.g. specific clinical presentation or biomarker combinations, to study families with a high rate of AD, or to compare groups with extreme phenotypes. An example of extreme phenotype analysis would be to compare early onset AD cases with controls that survived to very old age without being demented, where the latter group is expected to be enriched for protective variants, like the *PLCG2* variants described in this thesis.

Even if a genetic study is sufficiently powered many pitfalls are present. In studies of rare genetic variants technical bias is often introduced when genotyping is performed differently in cases and controls can cause false positive findings. Examples of technical bias are that genotyping of cases and controls was performed at different sites, with different machines/ techniques, exome sequencing with different capture kits or processing of the data in slightly different pipelines with different versions of software. The steps taken to avoid these biases are described in detail in the different chapters of this thesis.

I found that the methodological considerations that were encountered in genetic studies also applied to the metabolomics studies described in **chapter 5**. First the discovery analysis should be sufficiently powered. This can be attained by combining results from studies that measured the same metabolites and phenotype of interest. Second some form of correction for multiple testing should be applied followed by a replication that is again sufficiently powered. The exact definitions of these thresholds

are under continuous debate as the field of metabolomics gains progressive insights. This is the reason that the methods to declare significance varied in the metabolomics studies described in this thesis, while for all genetic studies they were generally comparable. The lessons learned in genetic studies can help us to resolve these methodological issues in metabolomics studies.

Future directions

Genetic discoveries have played a pivotal role in many of the advances in understanding AD and dementia. The search to understand this devastating disease will continue and in this section I describe the avenues that could drive research in the coming years.

Future directions for genetic research of AD and its endophenotypes

The most successful approach to reliably uncover new common and rare genetic variation has been to increase the total sample size, replication of significant results and increase the number of variants studied. This is exemplified in genetic analysis of height, historically the trait studied with largest sample sizes. A rare variant analysis included 711,428 individuals (discovery in 458,927 and validation 252,501 individuals) found over 80 new rare variants associated with height on top of over 700 loci containing common variants.[78] The same holds for AD research where there is an exponential increase in identified genes when combining results of the members of the IGAP consortium (CHARGE, EADI, GERAD, ADGC). The combined IGAP sample size now reaches approximately 85,000 samples (**chapter 2.4**), the most AD cases and controls worldwide. Inclusion of additional well characterized AD cases to the existing studies is the rate limiting step to further grow the consortiums sample size. Including additional studies that have not previously joined this worldwide effort, especially the well characterized studies from ethnicities other than European ancestry could relatively quickly increase the sample size and broaden the generalizability of the results. Another option is to genotype well characterized AD patients that have been included in clinical studies and trails in the past. For example the Europese Alzheimer's Disease DNA Biobank (EADB) is such an effort that aims to localize and genotype 30,000 and 40,000 controls for genetic studies. Another major leap forward is to use the large scale genetic biobanks that use clinical records for the follow-up their participants. Biobanks such as

the the UKbiobank (target 0.5 million), Million Veteran Program (1 million individuals), Vanderbilt University (0.25 million individuals),[79], the Kaiser Permanente Research Program on Genes, Environment, and Health (0.5 million individuals),[80] the China Kadoorie Biobank (0.5 million individuals),[81] as well as other biobanks worldwide study gather the DNA of millions of individuals in mid-life. These individuals will age over time creating an immense pool of future AD patients over time.

Another approach is to increase the number of genetic variants studied. This can be done by moving from imputation to sequencing studies as sequencing studies do not only study the variants that are known. Sequencing efforts in AD research such as the Alzheimer's Disease Sequencing Project (ADSP), a project that whole exome sequenced at present over 5,000 AD cases and 5,000 controls and is gathering whole genome sequence data of over 3,000 individuals. However these sequencing studies are relatively small compared to the large datasets that are necessary to reliably associate rare genetic variants with AD. For example, I showed in **chapter 2.4** that only after studying over 80,000 individuals we could reliably identify three rare genetic variants in IGAP. However these numbers of individuals can be studied through imputation. The real major leap forward would therefore be to combine the complementary approaches of sequencing and imputation. Following this approach, future studies should aim at imputing the new rare genetic variants discovered in ADSP by sequencing to the whole IGAP dataset. This would require the sequencing data to be combined into a reference panel or added to existing reference panels, such as the HRC panel discussed in **chapter 2.1**.

Apart from the case-control studies in unrelated populations a shift of focus should be on the unexplained autosomal dominant AD families, finding the culprit genetic variation in these families can be a powerful approach to find new variants. The ADSP already adopted this strategy and is studying over 800 whole-genome sequenced individuals form families with a high burden of AD. A last alternative approach could also be to study alternative phenotypes of AD such as the clinical states prior to manifestation of disease such as mild cognitive impairment (MCI) with whole genome sequencing and imputation.

Translation of findings to clinical practice

Next to the new insights into molecular mechanisms underlying AD that are the result of new genetic discoveries it is important to translate the genetic findings into clinically useful endpoints. The most imminent question to be answered is: how do we interpret variants found by sequencing in genes that are associated with AD? Studies were the information from sequencing data is combined with and other characteristics of the variants are vital to answer this question. In **chapter 1.2** we proposed such a classification for *SORL1* variants based on bioinformatics prediction software and in-silico sequence data that can be used on any not previously described variant. An important next step will be to use bioinformatics and in silico data in combination with information from animal studies. Such an effort has been done for example for Creutzfeldt-Jakob disease (CJD), which is a rare fast progressive dementia caused by misfolded prion protein. Genetic CJD is caused by variants in the gene coding for the prion protein (*PRNP*), for decades it has been unclear what the penetrance is for some variants in the PRNP gene. In an effort combining sequencing information of over 60,000 individuals Menikel *et al.*[82] showed that some *PRNP* mutations increase genetic CJD risk, but are not causal for the disease as their frequency in the healthy population was much higher than expected. Animal studies of CJD show that animals without the prion protein can survive.[83] Translating this information to humans, Menikel *et al.*[82] showed that carriers of rare PRNP mutations that causes natural knock-out of PRNP are healthy. This provides preliminary evidence that a reduction in *PRNP* dosage, if achievable in patients by e.g. CRISPR-Cas[84], is likely to be tolerated. For CJD the protein to target is clear. However, the million-dollar question for a complex disease like AD is to find the right target, in which tissue it has an effect and lastly when to target it in the disease process. To find the right target rare genetic variations that protect against AD can be used as a roadmap, e.g. the *APOE* ε2 allele,[21] a variant in APP[33] and the *PLCG2*-P522R variant in described in **chapter 2.4**. The individuals who carry these mutations are healthy and at the same time they have a lower risk of AD. Expression databases and studies can then show where the targets are expressed[85] and lastly these mutations could be studied in animal and cell models of AD to understand when the targets influence AD risk.

In this thesis I studied the utility of the common low-risk genetic findings and showed that very high risk subgroups based on these genetic factors. This is important as there is a need for early detection and prevention of AD.[9,10] Further improvement of risk classification based on common variants is to be expected as the number of AD associated loci increases.[37,38] Using variants over the whole genome it is estimated that 53% of phenotypic variance is explained by genetics and current genetic discoveries explain approximately 31%.[86] Another future direction to predict AD based on a multi-modal approach combining genetic risk estimations with risk based on measuring other endophenotypes and risk factors of AD.[87,88] The most well-studied biomarkers include cerebrospinal fluid measurements of tau, phosphorylated tau, and amyloid-β, positron emission tomography of amyloid β and tau deposition.[8] Indeed in small samples it has been shown that AD can be better classified from controls when using a multi-modal based on a wide array of measured biomarkers.[87] Likely such classification based on these established CSF and CT can be further improved when combined with the blood measures described in **chapter 3** and the detailed measures derived from MRI described in **chapter 4**. This integrational approach where the best predictive markers or risk factors from all research fields are brought could be a promising avenue for risk prediction.[88]

Combining metabolomics and genetic findings in functional studies

In this thesis we show that large scale metabolite studies are able to reliably identify metabolites associating with cognitive ability (**chapter 3.2**) and dementia (**chapter 3.3**). Previous metabolite studies[72,89-91] were smaller and did not replicate.[92,93] The increasing sample size of metabolite studies and the possibility of thorough replication makes systematically mapping of circulating metabolites with AD feasible. The identified metabolites could then be further studied. For example, the genetic correlation between AD and the metabolite measure could imply causality.[94,95] For example, when evaluating the genetic correlation between metabolites identified in this thesis and AD using publically available GWAS.[96], we found a nominally significant genetic correlation between AD and isoleucine (R_g=-0.55, p= 0.009) and isoleucine (R_g=-0.64, p=0.016).

Even with a shared genetic background the pathways that lead to the metabolic changes associated with AD are largely unknown. It is also unknown through which tissue the pathways act, e.g. brain, immune cells in brain, or monocytes in blood. The same holds true for most of the genetic factors uncovered by GWAS. Further the common low risk genes are most likely working in concert with non-genetic circulating factors.[87] The major challenge is therefore how to translate these many low risk genes and the multitude of molecular signals. Developing classical animal models will be costly and time consuming especially as individualization is necessary by including the multiple combinations of genetic variants. Therefore this requires cost- and time efficient approaches that go beyond classical animal and cellular studies addressing effects of single genes. The rapid pace of development of induced pluripotent stem cells (iPSCs) research offers unique opportunities to overcome this problem providing possibilities to study the metabolism of neurons in iPSC derived from patients that developed AD due to a combination of genetic low risk genes. These cells can be used to build so called organ on-a-chip models of the brain.[97] These models can be used in future research to understand the metabolic profiles and genetic findings that are underlying the occurrence of AD patients and provide a model to develop and study the effects of novel therapeutic interventions.

Concluding remarks

During the time of my PhD the field of AD disease genetics has seen remarkable progress both in the translation of the genetic findings to clinically useful entities as well as for the unraveling of the pathological process that leads to this devastating disease. The final steps for the use of genetics in clinical practice are yet to be made, but with the vast load of sequencing data coming available I am positive this will happen in due time. I am convinced that the discoveries described in this theses and future genetic findings in AD and its endophenotypes will further shape our basic understanding of the disease. Especially the strong innate immunity and microglial related genes that clearly emerge from the genetic findings give hope for the future. The next steps are to translate the genetic findings to biological understanding of the immunity response, animal research and eventually development of drugs to modify the immune reaction of AD to prevent or halt disease progress.

References:

1. Cruts M, Theuns J, Van Broeckhoven C. Locus-specific mutation databases for neurodegenerative brain diseases. *Hum Mutat* 2012; **33**(9): 1340-4.
2. Hardy JA, Higgins GA. Alzheimer's disease: the amyloid cascade hypothesis. *Science* 1992; **256**(5054): 184-5.
3. McKhann GM, Knopman DS, Chertkow H, 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 & Dementia* 2011; **7**(3): 263-9.
4. Saunders AM, Strittmatter WJ, Schmechel D, et al. Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology* 1993; **43**(8): 1467-72.
5. Corder EH, Saunders AM, Strittmatter WJ, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 1993; **261**(5123): 921-3.
6. Green RC, Roberts JS, Cupples LA, et al. Disclosure of APOE genotype for risk of Alzheimer's disease. *N Engl J Med* 2009; **361**(3): 245-54.
7. Sleegers K, Bettens K, De Roeck A, et al. A 22-single nucleotide polymorphism Alzheimer's disease risk score correlates with family history, onset age, and cerebrospinal fluid Abeta42. *Alzheimers Dement* 2015; **11**(12): 1452-60.
8. Kosik KS. Personalized medicine for effective Alzheimer disease treatment. *JAMA Neurol* 2015; **72**(5): 497-8.
9. Jack CR, Jr., Knopman DS, Jagust WJ, et al. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *Lancet Neurol* 2013; **12**(2): 207-16.
10. Reiman EM, Langbaum JB, Fleisher AS, et al. Alzheimer's Prevention Initiative: a plan to accelerate the evaluation of presymptomatic treatments. *J Alzheimers Dis* 2011; **26 Suppl 3**: 321-9.
11. 23andMe. 2016. https://www.23andme.com/en-ca/health/i_alzheimers/techreport/.
12. Genin E, Hannequin D, Wallon D, et al. APOE and Alzheimer disease: a major gene with semi-dominant inheritance. *Mol Psychiatry* 2011; **16**(9): 903-7.
13. Bamshad MJ, Ng SB, Bigham AW, et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 2011; **12**(11): 745-55.
14. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016; **536**(7616): 285-91.
15. Vardarajan BN, Zhang Y, Lee JH, et al. Coding mutations in SORL1 and Alzheimer disease. *Ann Neurol* 2015; **77**(2): 215-27.
16. Verheijen J, Van den Bossche T, van der Zee J, et al. A comprehensive study of the genetic impact of rare variants in SORL1 in European early-onset Alzheimer's disease. *Acta Neuropathol* 2016; **132**(2): 213-24.
17. Pottier C, Hannequin D, Coutant S, et al. High frequency of potentially pathogenic SORL1 mutations in autosomal dominant early-onset Alzheimer disease. *Mol Psychiatry* 2012; **17**(9): 875-9.
18. Rogaeva E, Meng Y, Lee JH, et al. The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. *Nat Genet* 2007; **39**(2): 168-77.
19. Nicolas G, Charbonnier C, Wallon D, et al. SORL1 rare variants: a major risk factor for familial early-onset Alzheimer's disease. *Mol Psychiatry* 2016; **21**(6): 831-6.
20. Cuccaro ML, Carney RM, Zhang Y, et al. SORL1 mutations in early- and late-onset Alzheimer disease. *Neurol Genet* 2016; **2**(6): e116.
21. Farrer LA, Cupples LA, Haines JL, et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *JAMA* 1997; **278**(16): 1349-56.

22. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016.
23. International HapMap C, Frazer KA, Ballinger DG, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007; **449**(7164): 851-61.
24. Genomes Project C, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature* 2015; **526**(7571): 68-74.
25. Boomsma DI, Wijmenga C, Slagboom EP, et al. The Genome of the Netherlands: design, and project goals. *Eur J Hum Genet* 2014; **22**(2): 221-7.
26. Consortium UK, Walter K, Min JL, et al. The UK10K project identifies rare variants in health and disease. *Nature* 2015; **526**(7571): 82-90.
27. Lambert JC, Heath S, Even G, et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet* 2009; **41**(10): 1094-9.
28. Hollingworth P, Harold D, Sims R, et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet* 2011; **43**(5): 429-35.
29. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 2013; **45**(12): 1452-8.
30. Guerreiro R, Wojtas A, Bras J, et al. TREM2 variants in Alzheimer's disease. *N Engl J Med* 2013; **368**(2): 117-27.
31. Naj AC, Jun G, Beecham GW, 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.
32. Desikan RS, Schork AJ, Wang Y, et al. Polygenic Overlap Between C-Reactive Protein, Plasma Lipids, and Alzheimer Disease. *Circulation* 2015; **131**(23): 2061-9.
33. Jonsson T, Atwal JK, Steinberg S, et al. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* 2012; **488**(7409): 96-9.
34. Jun G, Ibrahim-Verbaas CA, Vronskaya M, et al. A novel Alzheimer disease locus located near the gene encoding tau protein. *Mol Psychiatry* 2016; **21**(1): 108-17.
35. Seshadri S, Fitzpatrick AL, Ikram MA, et al. Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA* 2010; **303**(18): 1832-40.
36. Jonsson T, Stefansson H, Steinberg S, et al. Variant of TREM2 associated with the risk of Alzheimer's disease. *N Engl J Med* 2013; **368**(2): 107-16.
37. Escott-Price V, Shoai M, Pither R, Williams J, Hardy J. Polygenic score prediction captures nearly all common genetic risk for Alzheimer's disease. *Neurobiol Aging* 2017; **49**: 214 e7- e11.
38. Escott-Price V, Sims R, Bannister C, et al. Common polygenic variation enhances risk prediction for Alzheimer's disease. *Brain* 2015; **138**(Pt 12): 3673-84.
39. Chouraki V, Reitz C, Maury F, et al. Evaluation of a Genetic Risk Score to Improve Risk Prediction for Alzheimer's Disease. *J Alzheimers Dis* 2016; **53**(3): 921-32.
40. Heilmann S, Drichel D, Clarimon J, et al. PLD3 in non-familial Alzheimer's disease. *Nature* 2015; **520**(7545): E3-E5.
41. Hooli BV, Lill C, Mullin K, et al. PLD3 gene variants and Alzheimer's disease. *Nature* 2015; **520**(7545): E7-E9.
42. Lambert JC, Grenier-Boley B, Bellenguez C, et al. PLD3 and sporadic Alzheimer's disease risk. *Nature* 2015; **520**(7545): E1-E.
43. Fazzari P, Horre K, Arranz AM, et al. PLD3 gene and processing of APP. *Nature* 2017; **541**(7638): E1-E2.
44. Ulrich JD, Burchett JM, Restivo JL, et al. In vivo measurement of apolipoprotein E from the brain interstitial fluid using microdialysis. *Mol Neurodegener* 2013; **8**: 13.
45. Wang Y, Cella M, Mallinson K, et al. TREM2 lipid sensing sustains the microglial response in an Alzheimer's disease model. *Cell* 2015; **160**(6): 1061-71.

46. Wang Y, Ulland TK, Ulrich JD, et al. TREM2-mediated early microglial response limits diffusion and toxicity of amyloid plaques. *J Exp Med* 2016; **213**(5): 667-75.
47. Blokland GA, de Zubicaray GI, McMahon KL, Wright MJ. Genetic and environmental influences on neuroimaging phenotypes: a meta-analytical perspective on twin imaging studies. *Twin Res Hum Genet* 2012; **15**(3): 351-71.
48. Hibar DP, Stein JL, Renteria ME, et al. Common genetic variants influence human subcortical brain structures. *Nature* 2015; **520**(7546): 224-9.
49. Ganjgahi H, Winkler AM, Glahn DC, Blangero J, Kochunov P, Nichols TE. Fast and powerful heritability inference for family-based neuroimaging studies. *Neuroimage* 2015; **115**: 256-68.
50. Jahanshad N, Kochunov PV, Sprooten E, et al. Multi-site genetic analysis of diffusion images and voxelwise heritability analysis: a pilot project of the ENIGMA-DTI working group. *Neuroimage* 2013; **81**: 455-69.
51. DeStefano AL, Seshadri S, Beiser A, et al. Bivariate heritability of total and regional brain volumes: the Framingham Study. *Alzheimer Dis Assoc Disord* 2009; **23**(3): 218-23.
52. Stein JL, Medland SE, Vasquez AA, et al. Identification of common variants associated with human hippocampal and intracranial volumes. *Nat Genet* 2012; **44**(5): 552-61.
53. Bis JC, DeCarli C, Smith AV, et al. Common variants at 12q14 and 12q24 are associated with hippocampal volume. *Nat Genet* 2012; **44**(5): 545-51.
54. Blank MC, Grinberg I, Aryee E, et al. Multiple developmental programs are altered by loss of Zic1 and Zic4 to cause Dandy-Walker malformation cerebellar pathogenesis. *Development* 2011; **138**(6): 1207-16.
55. Blank MC, Grinberg I, Chizhikov VV, Millen KJ. Zic1 and Zic4 are required for mammalian cerebellar patterning and growth. *Dev Biol* 2008; **319**(2): 594-5.
56. Grinberg I, Northrup H, Ardinger H, Prasad C, Dobyns WB, Millen KJ. Heterozygous deletion of the linked genes ZIC1 and ZIC4 is involved in Dandy-Walker malformation. *Nature Genetics* 2004; **36**(10): 1053-5.
57. Tohyama J, Kato M, Kawasaki S, et al. Dandy-Walker Malformation Associated With Heterozygous ZIC1 and ZIC4 Deletion: Report of a New Patient. *American Journal of Medical Genetics Part A* 2011; **155a**(1): 130-3.
58. Twigg SR, Forecki J, Goos JA, et al. Gain-of-Function Mutations in ZIC1 Are Associated with Coronal Craniosynostosis and Learning Disability. *Am J Hum Genet* 2015; **97**(3): 378-88.
59. Niu T, Liu N, Zhao M, et al. Identification of a novel FGFRL1 MicroRNA target site polymorphism for bone mineral density in meta-analyses of genome-wide association studies. *Hum Mol Genet* 2015; **24**(16): 4710-27.
60. Chouraki V, Beiser A, Younkin L, et al. Plasma amyloid-beta and risk of Alzheimer's disease in the Framingham Heart Study. *Alzheimers Dement* 2015; **11**(3): 249-57 e1.
61. Toledo JB, Shaw LM, Trojanowski JQ. Plasma amyloid beta measurements - a desired but elusive Alzheimer's disease biomarker. *Alzheimers Res Ther* 2013; **5**(2): 8.
62. Shah NS, Vidal JS, Masaki K, et al. Midlife blood pressure, plasma beta-amyloid, and the risk for Alzheimer disease: the Honolulu Asia Aging Study. *Hypertension* 2012; **59**(4): 780-6.
63. Koyama A, Okereke OI, Yang T, Blacker D, Selkoe DJ, Grodstein F. Plasma amyloid-beta as a predictor of dementia and cognitive decline: a systematic review and meta-analysis. *Arch Neurol* 2012; **69**(7): 824-31.
64. Hansson O, Stomrud E, Vanmechelen E, et al. Evaluation of plasma Abeta as predictor of Alzheimer's disease in older individuals without dementia: a population-based study. *J Alzheimers Dis* 2012; **28**(1): 231-8.
65. van Oijen M, Hofman A, Soares HD, Koudstaal PJ, Breteler MM. Plasma Abeta(1-40) and Abeta(1-42) and the risk of dementia: a prospective case-cohort study. *Lancet Neurol* 2006; **5**(8): 655-60.

66. Cruchaga C, Kauwe JS, Harari O, et al. GWAS of cerebrospinal fluid tau levels identifies risk variants for Alzheimer's disease. *Neuron* 2013; **78**(2): 256-68.
67. Ramanan VK, Risacher SL, Nho K, et al. APOE and BCHE as modulators of cerebral amyloid deposition: a florbetapir PET genome-wide association study. *Mol Psychiatry* 2014; **19**(3): 351-7.
68. Shulman JM, Chen K, Keenan BT, et al. Genetic susceptibility for Alzheimer disease neuritic plaque pathology. *JAMA Neurol* 2013; **70**(9): 1150-7.
69. Beecham GW, Hamilton K, Naj AC, et al. Genome-wide association meta-analysis of neuropathologic features of Alzheimer's disease and related dementias. *PLoS Genet* 2014; **10**(9): e1004606.
70. Vassar R. BACE1 inhibitor drugs in clinical trials for Alzheimer's disease. *Alzheimers Res Ther* 2014; **6**(9): 89.
71. Henriksen K, O'Bryant SE, Hampel H, et al. The future of blood-based biomarkers for Alzheimer's disease. *Alzheimers Dement* 2014; **10**(1): 115-31.
72. Mapstone M, Cheema AK, Fiandaca MS, et al. Plasma phospholipids identify antecedent memory impairment in older adults. *Nat Med* 2014; **20**(4): 415-8.
73. Bos D, Portegies MLP, van der Lugt A, et al. Intracranial Carotid Artery Atherosclerosis and the Risk of Stroke in Whites The Rotterdam Study. *Jama Neurology* 2014; **71**(4): 405-11.
74. Bos D, Vernooij MW, de Bruijn RFAG, et al. Atherosclerotic calcification is related to a higher risk of dementia and cognitive decline. *Alzheimers & Dementia* 2015; **11**(6): 639-47.
75. Zuk O, Schaffner SF, Samocha K, et al. Searching for missing heritability: designing rare variant association studies. *Proc Natl Acad Sci U S A* 2014; **111**(4): E455-64.
76. Page CM, Baranzini SE, Mevik BH, Bos SD, Harbo HF, Andreassen BK. Assessing the Power of Exome Chips. *PLoS One* 2015; **10**(10): e0139642.
77. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 2006; **38**(2): 209-13.
78. Marouli E, Graff M, Medina-Gomez C, et al. Rare and low-frequency coding variants alter human adult height. *Nature* 2017; **542**(7640): 186-90.
79. Gaziano JM, Concato J, Brophy M, et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol* 2016; **70**: 214-23.
80. Hedderson MM, Ferrara A, Avalos LA, et al. The Kaiser Permanente Northern California research program on genes, environment, and health (RPGEH) pregnancy cohort: study design, methodology and baseline characteristics. *BMC Pregnancy Childbirth* 2016; **16**(1): 381.
81. Chen Z, Chen J, Collins R, et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int J Epidemiol* 2011; **40**(6): 1652-66.
82. Minikel EV, Vallabh SM, Lek M, et al. Quantifying prion disease penetrance using large population control cohorts. *Sci Transl Med* 2016; **8**(322): 322ra9.
83. Steele AD, Lindquist S, Aguzzi A. The prion protein knockout mouse: a phenotype under challenge. *Prion* 2007; **1**(2): 83-93.
84. KNAW V. GENOME EDITING. 2016.
85. Consortium GT, Ardlie KG, Deluca DS, et al. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science (New York, N Y)* 2015; **348**(6235): 648-60.
86. Ridge PG, Hoyt KB, Boehme K, et al. Assessment of the genetic variance of late-onset Alzheimer's disease. *Neurobiol Aging* 2016; **41**: 200 e13-20.
87. Lista S, O'Bryant SE, Blennow K, et al. Biomarkers in Sporadic and Familial Alzheimer's Disease. *J Alzheimers Dis* 2015; **47**(2): 291-317.
88. Eckerstrom C, Olsson E, Klasson N, et al. Multimodal prediction of dementia with up to 10 years follow up: the Gothenburg MCI study. *J Alzheimers Dis* 2015; **44**(1): 205-14.
89. Soininen P, Kangas AJ, Wurtz P, et al. High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism. *Analyst* 2009; **134**(9): 1781-5.

90. Fiandaca MS, Zhong X, Cheema AK, et al. Plasma 24-metabolite Panel Predicts Preclinical Transition to Clinical Stages of Alzheimer's Disease. *Front Neurol* 2015; **6**: 237.
91. Hye A, Riddoch-Contreras J, Baird AL, et al. Plasma proteins predict conversion to dementia from prodromal disease. *Alzheimers Dement* 2014; **10**(6): 799-807 e2.
92. Casanova R, Varma S, Simpson B, et al. Blood metabolite markers of preclinical Alzheimer's disease in two longitudinally followed cohorts of older individuals. *Alzheimers Dement* 2016; **12**(7): 815-22.
93. Li D, Misialek JR, Boerwinkle E, et al. Plasma phospholipids and prevalence of mild cognitive impairment and/or dementia in the ARIC Neurocognitive Study (ARIC-NCS). *Alzheimers Dement (Amst)* 2016; **3**: 73-82.
94. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet* 2015; **47**(11): 1236-41.
95. Bulik-Sullivan BK, Loh PR, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 2015; **47**(3): 291-5.
96. Zheng J, Erzurumluoglu AM, Elsworth BL, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* 2016.
97. Yi Y, Park J, Lim J, Lee CJ, Lee SH. Central Nervous System and its Disease Models on a Chip. *Trends in Biotechnology* 2015; **33**(12): 762-76.

Chapter 7

Summary / Samenvatting

Summary

Dementia and its most common type, Alzheimer's disease (AD) is a devastating disease that afflicts millions worldwide. Research in the past decade led to major progress in understanding the epidemiology of late onset AD and dementia. Epidemiological studies suggest, that optimal prevention or treatment of both cardiovascular risk factors as well as improvement of educational level, could reduce the incidence of the dementia. Still, there are no targeted preventive interventions or pharmacological treatments for AD available. This makes further unraveling the pathophysiology of AD to find new targets for treatment a global research priority. Genetic research has played a pivotal role to identify novel insights in the pathophysiology of AD (as summarized in **chapter 1)** and has influenced clinical diagnostics and treatment developments. Therefore, the overarching aim of the studies included in this thesis is to gain novel insights in the molecular and biological mechanisms underlying AD and its related endophenotypes. The studies that make up this thesis are organized in chapters that all approach this goal from a different angle.

In **chapter 2** I studied the mechanisms of known genetic risk factors of AD, translating their effects to clinically useable outcomes. In the studies in **chapter 3** I studied genetic variants that are rare in the general population aiming to identify novel molecular targets that are involved in the pathophysiology of AD. Below is a short description of the findings of the studies in these first two chapters.

In **chapter 2.1** I studied the combine impact of the known genetic risk loci on risk of AD. I showed that common variants with small effects jointly modify the risk and onset age of AD and dementia, particularly in *APOE* ε4 carriers. These findings contributed towards efficacy improvements of clinical trials and better risk prediction for AD and dementia. In **chapter 2.2** I determined the risk of AD for carriers of rare variants in the *SORL1* gene. Rare variants in *SORL1* were previously associated with AD. This study aimed to determe the magnitude of the risk of different *SORL1* variants. The results show that the risk of AD could be classified based on allele frequency and predicted deleteriousness. *SORL1* mutations that were considered most pathogenic were only

observed in AD patients, suggesting they are causal. We proposed a first classification for *SORL1* variants that can be used for evaluation of the genetic risk of AD patients and their family members. **Chapter 2.3** studied the association of parental family history with risk of dementia. The results showed that independent of known genetic risk factors, parental history of dementia increased the risk of dementia primarily when the age at parental diagnosis was <80 years. This study underlines the importance of familial history and suggested that there are still undiscovered genetic variants. Indeed with the studies in **chapter 3** I aimed to uncover novel insights in the genetic architecture of AD using genetic association studies of rare variants. First I describe in **chapter 3.1** that rare variants up to a frequency in the population of 0.001% could be imputed, statistically inferred, with confidence using a large reference panel, the Haplotype Reference Consortium (HRC). This is important as these variants were neither present on the genome-wide genotyping arrays nor could be imputed with confidence previously, facilitating studies of rare genetic variants. In **chapter 3.2,** I provided partially conflicting evidence to a publication proposing *PLD3* as a gene that contains multiple rare risk variants for AD. The results from this paper contributed to the scientific discussion if *PLD3* should be considered an AD-risk gene. In **chapter 3.3**, using exome-chip data in the CHARGE (Cohorts for Heart and Aging Research in Genomic Epidemiology) consortium we identified a missense mutation in the *TM2D3* gene with a strong impact on LOAD risk. The *TM2D3* variant is enriched ~10-fold, compared with other European populations, and associated with both risk and age at onset of AD in the Icelandic population. We further showed that this variant is associated with a loss-of-function in the heterologous but potentially relevant context of Notch signaling in *Drosophila* (fruit fly) embryos. In **chapter 3.4** the search for rare variants continued with the largest (N=85,133) rare-variant exome chip study to date performed in the IGAP consortium. Using exome chip data and the HRC imputation panel (studied in **chapter 3.1**), this study identified new coding rare variants associated with AD in *ABI3, PLCG2* and *TREM2*. These genes are highly expressed in microglia and link directly to an immune-related protein-protein interaction network enriched for previously identified AD risk genes. These genetic findings provide additional evidence that the microglia-mediated innate immune response contributes directly to AD development.

AD is known to have a long pre-clinical phase in which there are no apparent clinical signs. Sub-clinical signs start to appear up to twenty years before the onset of clinical symptoms. These subclinical signs include amyloid-β depositions, brain changes and changes in circulating blood measures. In the following chapters I aimed to identify and understand these so called AD-related (endo) phenotypes. I focused in **chapter 4** on the genetic background of endophenotypes in brain measured with Magnetic Resonance Imaging (MRI) and studied circulating endophenotypes, namely Amyloid-β and metabolites, in **chapter 5**. Below is a short description of the findings of the studies in these two chapters.

In **chapter 4.1** I studied detailed information of grey matter, derived from high resolution imaging of the brain. This technique is called voxel-based morphometry (VBM). I created maps of the heritability, i.e. the genetics contribution, of grey matter voxels based on both two family-based studies and the large population-based Rotterdam study. I found that significantly heritable voxels located in subcortical structures and the language areas of the brain. Further, I showed enhancement of the statistical signal of two previously discovered genetic loci associated with subcortical volumes by using only the most heritable voxels. **Chapter 4.2** studied the effects of related AD genetic loci on voxel morphometry. AD risk loci affected cortical gray matter in several brain regions known to be involved in AD, as well as regions that have not been implicated in AD before. I then searched for common genetic factors that influence the lobar volumes of the brain (**chapter 4.3**). This study showed that 30-40% of the heritability of brain lobar volume was determined by common variants over the genome. On top of this I identified six genetic loci that significantly associated with brain lobar volume. Two of these loci located in genetic regions that were associated previously brain volume measures. Two loci were located in regions rich of epigenetic data linked to the genes *DAAMI* and *THBS3* and two located close to genes that cause brain related diseases with Mendelian inheritance patterns (*ZIC4* and *FGFRL1*). These findings revealed part of the complex genetics underlying brain development.

Finally, in **chapter 5,** I studied circulating endophenotypes and risk factors for AD**.** In **chapter 5.1,** I performed a genome-wide association study of plasma β-amyloid (Aβ)42,

Aβ40 and the ratio (Aβ)42/Aβ40. Amyloid-β is the one of the core pathological findings in brains of AD patients. For the first time this study linked *APOE* with Aβ42 in blood and Aβ40 in blood with variants near the Beta-secretase 1 (*BACE1*) gene. Suggestive signals were found around the *APP* and *PSEN2* genes. These findings showed that performing GWAS on plasma Aβ40 and Aβ42 levels in blood, can pick up relevant molecular actors of the β-amyloid metabolism. The next studies of **chapter 5** focussed on hundreds of circulating metabolites. The association of these metabolites with the risk of AD and dementia was studied in **chapter 5.2**. In total this study identified ten metabolites and lipoprotein lipids that associated with the risk for future dementia. Among the metabolites were leucine, valine, saturated fatty acid ratio, creatinine, two high density lipoprotein (HDL) subfractions and three very low high density lipoprotein subfractions. These associations were independent of the traditional risk factors suggesting that these metabolites are novel independent risk factors for AD and/or dementia. Next to these associations with dementia I studied in **chapter 5.3** the relation of metabolites with cognitive function. I discovered and replicated a profile of 15 metabolites that associated with general cognitive ability. This metabolic profile included sub-fractions of HDL, docosahexaenoic acid (DHA), ornithine, glutamine and glycoprotein acetyls. Six of these 15 metabolites, were additionally associated with dementia. Further lifestyle factors such as diet, smoking and physical activity, showed strong associations with the metabolites in the profile. The identification of circulating metabolites that were consistently associated with cognition, dementia and lifestyle factors, opens new avenues for prevention of cognitive decline and dementia. The same metabolites studied in **chapter 5.2 and 5.3** were used in the last chapter (**chapter 5.4)** to study the metabolite association of internal carotid artery calcifications (ICAC), a risk factor for dementia and cognitive decline. We found that 3-hydroxybutyrate was related with a larger volume of ICAC. When comparing the metabolic association profile of ICAC with that of calcification in other vessel beds, we difference in association patterns. ICAC was mainly associated with metabolic measures of glucose metabolism, whereas the other vessel beds were generally affected more by lipid and lipoprotein measures. This study led to novel insights in blood measures that associated with ICAC and confirmed the existence of location-specific differences in etiology of arterial calcification.

In the last chapter I discuss the findings of this thesis in the context of developments in AD research and give directions for future research. In conclusion the discoveries described in this chapters of this thesis improve our understanding of AD and its endophenotypes, and I hope they will enable or facilitate the so necessary development of treatments, or interventions for this devastating disease.

Samenvatting

Dementie, waarvan de ziekte van Alzheimer de meest voorkomende vorm is, is een ernstige ziekte waar wereldwijd miljoenen mensen aan lijden. In het afgelopen decennium heeft wetenschappelijk onderzoek geleid tot belangrijke vooruitgang in begrip van de epidemiologie van de ziekte van Alzheimer en dementie. Zo suggereren epidemiologische studies dat, optimale preventie van cardiovasculaire risicofactoren en het bevorderen van opleidingsniveau, de incidentie van dementie zouden kunnen verlagen. Er zijn echter nog steeds geen gerichte preventieve interventies of medicijnen beschikbaar om de ziekte van Alzheimer te voorkomen of te behandelen. Dit maakt het verder ontrafelen van de pathofysiologie van de ziekte van Alzheimer, om zo nieuwe aanknopingspunten voor behandelopties te vinden, een wereldwijde onderzoeksprioriteit. Onderzoek naar de genetica van de ziekte van Alzheimer speelt hierbij een belangrijke rol. In het verleden leidde genetisch onderzoek tot nieuwe inzichten in de pathofysiologie van AD, de ontwikkeling van diagnostische testen en de ontwikkeling van medicijnen (samengevat in **hoofdstuk 1**). Het doel van dit proefschrift is dan ook, met behulp van genetische studies, nieuwe inzichten te verkrijgen in de moleculaire mechanismen onderliggend aan de ziekte van Alzheimer en aan de ziekte gerelateerde endofenotypen. De studies waaruit dit proefschrift bestaat benaderen dit doel vanuit aparte invalshoeken en zijn daarop georganiseerd in hoofstukken.

In **hoofdstuk 2** bestudeer ik de effecten van, de tot nu toe bekende, genetische risicofactoren van de ziekte van Alzheimer met als doel deze te vertalen naar klinisch bruikbare uitkomsten. In **hoofdstuk 3** zocht ik naar nieuwe zeldzame genetische varianten die associëren met de ziekte van Alzheimer met als doel nieuwe moleculaire inzichten te verkrijgen in de ziekte van Alzheimer. Hieronder volgt een korte samenvatting van de belangrijkste bevindingen van de studies die zijn opgenomen in de eerste twee hoofdstukken van dit proefschrift.

In **hoofdstuk 2.1** heb ik de impact onderzocht van eerder ontdekte veel voorkomende genetische varianten op het risico dat iemand in zijn leven de ziekte van Alzheimer

krijgt. De resultaten tonen aan dat deze veelvoorkomende genetische varianten, welk individueel slechts een klein effect hebben, gezamenlijk een groot effect hebben op het risico en leeftijd van aanvang van de ziekte van Alzheimer. Deze effecten blijken groter in dragers van het *APOE* $\varepsilon 4$ genotype.Deze bevindingen kunnen bijdragen aan de doeltreffendheid van klinische studies en het beter kunnen schatten van het risico op de ziekte van Alzheimer. In **hoofdstuk 2.2** heb ik het risico op de ziekte van Alzheimer voor dragers van zeer zeldzame genetische varianten in het *SORL1* gen onderzocht. Het risico van individuele varianten kan worden geclassificeerd op basis van de frequentie van de varianten in de populatie en de mate van voorspelde schadelijkheid van deze varianten. Varianten die het eiwit transcript van *SORL1* verkorten of verstoren lijken de ziekte zelfs te veroorzaken, omdat ze enkel voorkomen bij patiënten en niet bij cognitief gezonde deelnemers. De voorgestelde classificatie voor *SORL1* varianten kan worden gebruikt voor evaluatie van het genetisch risico van patiënten met de ziekte van Alzheimer en hun familieleden. **Hoofdstuk 2.3** beschrijft de waarde van een positieve familie geschiedenis voor dementie, in de anamnese. De resultaten tonen aan dat een positieve familie geschiedenis de kans op dementie bij familieleden vergroot, vooral indien de ziekte bij de ouders begon voor het 80ste levensjaar. Dit onderzoek onderstreept het belang van een goede familie anamnese. Bekende genetische risicofactoren verklaren het verhoogde risico op dementie bij een positieve familiegeschiedenis slechts gedeeltelijk, dit suggereert dat er nog veel genetische varianten ontdekt moeten worden. In **hoofdstuk 3** zocht ik daarom naar nieuwe genetische varianten die associëren met de ziekte van Alzheimer. Ik heb me hierbij gefocust op genetische varianten die niet vaak voorkomen, omdat hier nog weinig over bekend is. Eerst beschrijf ik, in **hoofdstuk 3.1,** dat met behulp van een nieuw referentie panel, het Haplotype Reference Consortium (HRC), zeldzame genetische varianten met een zeer lage frequentie in de populatie (tot 0.001% allel frequentie) met zekerheid geïmputeerd kunnen worden. Dit is een belangrijke bevinding omdat imputeren goedkoper en sneller is dan direct genotyperen, waardoor onderzoek sneller kan plaatsvinden. In **hoofdstuk 3.2** beschrijf ik een replicatiestudie van zeldzame varianten in het *PLD3* gen. Onze resultaten ondersteunen de resultaten van de originele publicatie slechts gedeeltelijk. Dit hoofdstuk heeft bijgedragen aan de wetenschappelijke discussie over *PLD3* als risico gen voor de ziekte van Alzheimer. In **hoofdstuk 3.3** heb ik met

behulp van exome-chip data in het CHARGE (Cohorts for Heart and Aging Research in Genomic Epidemiology) consortium een missense-mutatie in het *TM2D3* gen ontdekt die associeert met een sterk verhoogd risico op de ziekte van Alzheimer. Deze *TM2D3* variant komt 10 keer zo vaak voor in de IJslandse populatie, ten opzichte van andere Europeanen en associeert met zowel het risico, als leeftijd van diagnose van de ziekte van Alzheimer. Verder laten we zien dat deze *TM2D3* variant verlies-van-functie van het TM2D3 eiwit veroorzaakt in embryo's van *Drosophila*, fruitvliegjes. In **hoofdstuk 3.4** heb ik de zoektocht naar zeldzame varianten voortgezet in de grootste genetische (N=85,133) studie naar de ziekte van Alzheimer tot nu toe, uitgevoerd in het IGAP (International Genomics of Alzheimer's Project) consortium. Door gebruik te maken van exome-chip data en de HRC imputaties (**hoofdstuk 3.1**) hebben we coderende zeldzame varianten gevonden in de genen *ABI3* en *PLCG2* en *TREM2* die associëren met de ziekte van Alzheimer. Deze genen komen in hoge mate tot expressie in microglia en leveren aanvullend bewijs dat het aangeboren immuunsysteem rechtstreeks bijdraagt aan de pathofysiologie van de ziekte van Alzheimer.

Van de ziekte van Alzheimer is bekend dat er een lange pre-klinische fase is waarbij subklinische signalen van de ziekte bestaan zonder dat er klinische symptomen zichtbaar zijn. Deze subklinische signalen ontstaan tot twintig jaar voor aanvang van de klinische symptomen. Subklinische signalen zijn bijvoorbeeld: amyloid-β afzettingen in het brein, hersenveranderingen en veranderingen in compositie van het bloed. In de laatste twee hoofdstukken richt ik me op het identificeren en begrijpen van deze subklinische signalen, zogenaamde endofenotypen van de ziekte van Alzheimer. In **hoofdstuk 4** bestudeer ik de genetische achtergrond van endofenotypen in de hersenen gemeten met behulp van MRI-scans (Magnetic Resonance Imaging) en ik bestudeer endofenotypen in het bloed, namelijk amyloid-β en metabolieten in **hoofdstuk 5**. Hieronder volgt een korte beschrijving van de bevindingen uit deze twee hoofdstukken.

In **hoofdstuk 4.1** heb ik de erfelijkheid van de grijze stof in detail bestudeerd met hoge resolutie beeldvorming van de hersenen, zogenaamde voxel-based morfometrie (VBM). Ik heb zo een atlas gecreëerd van de mate van erfelijkheid van de grijze stof. In zowel familie als populatie studies zijn de subcorticale structuren en de taalgebieden het

meest erfelijk bepaald. Daarnaast kan met de mate van erfelijkheid het associatie signaal van eerder ontdekte genetische varianten versterkt worden. **Hoofdstuk 4.2** beschrijft het effect van genetische varianten die associëren met de ziekte van Alzheimer in de grijze stof. De varianten hadden invloed op grijze stof in meerdere hersengebieden. Enkele van deze gebieden zijn betrokken bij de ziekte van Alzheimer. Vervolgens zocht ik naar de genetische factoren die invloed hebben op het volume van hersenkwabben (**hoofdstuk 4.3**). Uit deze studie bleek dat 30-40% van de erfelijkheid van hersenkwabvolume wordt bepaald door de veel voorkomende varianten in het genoom. Daarnaast associëren zes genetische regio's statistisch significant met het volume van verschillende hersenkwabben. Twee van deze genetische regio's waren al eerder geassocieerd aan het volume van andere delen van de hersenen, twee liggen in regio's rijk aan epigenetische signalen, en twee regio's bevatten genen die gelinkt zijn aan aangeboren hersenziekten. Deze studie toont nieuwe moleculaire inzichten in de complexe genetische achtergrond van de hersenen.

Tot slot heb ik, in **hoofdstuk 5,** onderzoek gedaan naar in het bloed circulerende endofenotypen van der ziekte van Alzheimer. In **hoofdstuk 5.1** heb ik een genoom-wijde associatie studie verricht naar amyloid-β (Aβ)42, Aβ40 en de ratio van Aβ42/40 in het bloed. Voor het eerst is het *APOE ε4* genotype in verband gebracht met Aβ42 en de ratio Aβ 42/40 in bloed. Daarnaast associëren varianten nabij het Beta-secretase 1 (*BACE1*) gen met Aβ40. Bijna significante signalen werden gevonden rond de genen *APP* en *PSEN2*. De genen *APOE, BACE1, APP en PSEN2* coderen voor proteïnen die belangrijk zijn in het amyloid-β metabolisme en varianten in *APOE, APP en PSEN2* associeren met de ziekte van Alzheimer. Studies van amyloid-β in bloed kunnen daarom dienen om nieuwe moleculaire inzichten te krijgen in amyloid-β metabolisme en mogelijk de ziekte van Alzheimer. In de rest van de studies van **hoofdstuk 5** bestudeer ik, in het bloed circulerende, metabolieten. Metabolieten zijn stoffen die nodig zijn tijdens, of een product zijn van, de stofwisseling en weerspiegelen de processen in het lichaam. Door technologische vooruitgang is het nu mogelijk deze metabolieten in grote studies te onderzoeken. In **hoofdstuk 5.2** associëren we, in acht verschillende studies, honderden van deze metabolieten met het toekomstig risico op dementie. In totaal worden in deze studie tien metabolieten en lipoproteïne vetten gevonden die significant associëren met

het risico op dementie. Daarna beschrijf ik in **hoofdstuk 5.3** een profiel van 15 metabolieten die significant associëren met algemeen cognitief functioneren door resultaten van 11 studies te combineren. Dit profiel bestaat uit sub-fracties van HDL, DHA, ornithine, glutamine en glycoproteïnen. In aanvullende analyses associëren een deel van deze metabolieten met dementie en met leefstijlfactoren zoals dieet, roken en lichamelijke activiteit. Deze twee studies zijn de grootste studies tot nu toe naar de relatie tussen metabolieten, dementie en cognitief functioneren. De verkregen inzichten in het metabole profiel van dementie en cognitie geeft nieuwe aangrijpingspunten voor gerichte preventie van de ziekte van Alzheimer. In **hoofdstuk 5.4** onderzocht ik de relatie tussen metabolieten in het bloed en calcificaties in de carotis interna arterie (ICAC). ICAC is een risicofactor voor dementie en cognitieve achteruitgang. Ik stelde vast dat 3-hydroxybutyrate, een ketonlichaam, correleerde met een groter calcificatie volume in de carotis interna. Daarnaast werd het metabole associatie profiel van carotis interna vergeleken met dat van andere belangrijke arteriën. Hieruit kon ik concluderen dat ICAC voornamelijk associeerde metabolieten uit het glucose metabolisme en dat de lipiden en lipoproteïne waardes meer van invloed waren op de andere bloedvaten. Deze studie vergroot het inzicht in calcificaties in de carotis interna en bevestig het bestaan van locatie-specifieke verschillen in de etiologie van arteriële calcificaties.

In het laatste hoofdstuk plaats ik de bevindingen van dit proefschrift in de context van de ontwikkelingen in het onderzoeksveld van ziekte van Alzheimer en geef aanbevelingen voor toekomstig onderzoek. Concluderend, beschrijven de studies uit dit proefschrift belangrijke ontdekkingen die onze kennis over de moleculaire achtergrond van de ziekte van Alzheimer vergroot hebben en ik hoop dat ze zullen bijdragen aan de ontwikkeling van de zo nodige interventies of medicijnen tegen deze verschrikkelijke ziekte.

Chapter 8

Acknowledgments / Dankwoord

PhD Portfolio

Other publications

About the Author

Bijna klaar, almost done. Onderzoeker zijn is een leuk, spannend en soms een emotioneel vak. Het is werk dat je alleen maar samen kan volbrengen, daarom wil ik iedereen bedanken die een bijdrage heeft geleverd aan de totstandkoming van dit proefschrift.

First and foremost I would like to thank all study participants, researchers, employees and funders of all the studies used in this thesis. Their generous contribution of time and devotion to build and enforce scientific studies of high quality are at the basis of all my work and will undoubtedly help to further unravel diseases in the future. Of all these studies I first especially would like to mention studies I had the pleasure to work for during my PhD training, the Erasmus Rucphen Family (ERF) study and the Rotterdam study (RS). Secondly, I would like to thank all studies of the different consortia I collaborated with, the Cohorts for Heart and Aging in Genomic Epidemiology (CHARGE), the Alzheimer's Disease Genetics Consortium (ADGC), the European Alzheimer's Disease Initiative (EADI) and the Genetic and Environmental Risk in Alzheimer's disease (GERAD).

Daarna is de eerste persoon die ik wil bedanken **prof. Cornelia van Duijn**. Cornelia, jij was diegene wie de afgelopen jaren strak de lijnen uitzette en regelmatig het roer radicaal omgooide. Ik diep respect heb voor je wetenschappelijke prestaties, ben gevangen door je aanstekelijke liefde voor onderzoek en daarnaast heb ik ontzettend veel van je geleerd. Je verwacht van jouw studenten dat ze je net als jij kunnen onderzoeken, nadenken, vooruitplannen, politiek bedrijven en observeren. Dus, vanaf de eerste maanden verwachtte je veel van mij, stelde je doelen en deadlines. Dat ik dat nodig had, heb je goed geobserveerd en ik ben je daar erg dankbaar voor. Als laatste wil ik je bedanken voor de toewijding waarmee je met mij aan mijn thesis hebt gewerkt en voor de solide basis voor de toekomst die je me daarmee gegeven hebt. De tweede persoon die ik wil bedanken is **prof. Arfan Ikram**. Arfan, in veel opzichten ben je een voorbeeld. Schoolvoorbeeld epidemioloog, in wetenschappelijke discussies zijn je bijdragen bijna altijd doorslaggevend, carrière tijger en daarnaast een bevlogen vader. De laatste jaren heb je een stempel gedrukt op mijn artikelen, zeker op de introducties, en ik ben blij dat je nu mijn tweede promotor kunt zijn. Ik weet zeker dat de Rotterdam

Studie de volgende 25 jaar onder jouw leiding verder groeit en belangrijke bijdragen van hoog wetenschappelijk niveau zal leveren.

Third, but vital to this thesis is my, **co-promotor dr. Najaf Amin.** Dear Najaf, you are one of a kind. I clearly remember my job interview, were you explicitly made clear what you expected from a PhD student: work hard, listen well (to you) and most importantly don't come in crying every week about girlfriends. I kept my promise, that the last would not be an issue ;). In the years I worked with you, you grew enormously as co-promotor, solved every question I had, gave sensible advice when things went south and occasionally made Cornelia listen. You always say I lack a word filter, when I say things that come to mind (which is true). So let me say; Najaf you are a formidable person and example co-promotor, I hope to make you proud on my defence!

Vervolgens wil ik bedanken de leden van mijn leescommissie: **Prof. Monique Breteler,** geachte Monique. Ik heb in de vier jaar aan het Erasmus MC enkel woorden van lof over u vernomen. Georganiseerd, gepassioneerd en kritisch. Als hoofd van de Rhineland Study, waar zelfs Angela Merkel voor langskomt, staat u midden in het veld en zal u, naar verwachting, nog jaren een grote impact hebben op het onderzoek naar dementie. Daarom ben ik vereerd dat u deel uitmaakt van mijn leescommissie. Veel dank gaat ook uit naar **prof. Philip Scheltens**. Beste Philip, bedankt dat je plaats wilde nemen in de leescommissie van mijn proefschrift en daarnaast dat jij mij aangenomen hebt als post-doc voor de 100-plus studie. Ik kijk erg uit naar mijn komende tijd bij het VUmc Alzheimer Centrum waarin ik tegelijkertijd patiënten kan zien en de genetica van de ziekte van Alzheimer verder mag ontrafelen. **Prof. Thomas Hankemeier**, beste Thomas, bedankt voor het plaatsnemen in de leescommissie als secretaris. Ik bewonder je noviteit in het veld van metabolomics. De micro- en nano technologie en het organ-on-chip principe waar je aan werkt zijn gamechangers waar we nog veel van zullen vernemen, zowel in pathofysiologisch als in medicijn onderzoek. Op de afdeling noemden we je soms de nieuwe Ben Oostra, omdat je positieve invloed op de afdeling voelbaar was. Na de verdediging zien we zien elkaar vast weer fietsend in Leiden.

Ik wil ook graag de overige leden **dr. Kristel Sleegers** en **Prof. Wiro Niessen** bedanken dat zij plaats wilden nemen in mijn commissie. Beste Kristel, in de afgelopen jaren heb je legio zeer mooie, solide artikelen geschreven met daarin nieuwe inzichten in de genetica van de ziekte van Alzheimer. Ik heb veel geleerd van je werk en kijk ernaar uit met je van gedachten te wisselen tijdens mijn verdediging. Wiro, in je werk sla je een brug tussen de technici en clinici, dit legt je geen windeieren, gezien de vele honors en prijzen die je hiervoor ontvangt. Ik neem een voorbeeld aan je, hoe je dit hebt aangepakt en daarnaast wil ik je bedanken voor je waardevolle bijdragen aan meerdere hoofdstukken van mijn proefschrift.

My time at the Erasmus was so pleasant, mainly because of the nice colleagues on the genetic epidemiology department. First **Dina**, we started on the same day and shared an office for three years. Where to start Dina? I will miss you. You are the most likable person, extremely kind to everyone, very patient, loyal, fierce when needed and only angry over injustice or incompetence. Thank you for the wonderful time and I could not have wished for better paranimf. Second, **Adriana**, cheerful, small, full of doubt, but decisive in the end. I only noticed you after you submerged from your fish-work, but what a time we had afterwards. We wasted days and sometimes nights working on analysis, in our jungle of plants, until our bones ached and all our songs were played. And what will happen with the magic number? Why couldn't we identify its true meaning☺? I think it says, you can keep Olli. Please stay in touch and I hope to visit you in Colombia when Luc is older. **Ashley**. I really enjoyed having you as a fellow "Dutchie" on the department and your everyday chats with coffee. Of course our greatest endeavor was getting the follow-up of Rucphen. I calculated we went ~35 times to Rucphen and surroundings, driving around 8,000 kilometers (Bootjes!), drank 150 cups of coffee (Huisartsenteam Willebrord wins) and reviewing 2,200 patient files. Sooo many patients, that you could not stand the smell of oranges and we hád to get fries with a 'kroket' afterwards, just to recover. I am very proud of our achievements in Rucphen, glad you finished it with Hata and I know the department will reap the benefits of if for years to come. You do have an unique mix of laboratory knowledge, scientific knowledge and no-nonsense attitude with which you can sail any sea. **Shahzad,** you have a specific knock on the door, look through the window routine, that I recognize in

an instance ☺. Followed by: "Can I ask you a question?". Always looking disappointed when I said no ;-). I really got to know you on our trip to Charlottesville, the expression; calm waters run deep applies to you. I am glad you kept talking to keep us awake driving to the hotel. Keep up with the ADSP families, the signals on chromosome 4 and 5 will undoubtedly reveal their secrets to you, 'Fee aman'nillah'. **Jun,** Dreamy, a.k.a. the new Sven, more stubborn, but much more sweet. I remember the first day you came all dressed up and the second day in baggy jeans and a sweater :-O. Just keep working on your dream to get your impact of 10, first author paper, and let me know when you do so we can celebrate. Also, keep working on your Chinglish and drive safely. **Ivana**, you are a very warm person and basically the mother of most foreign students in the epidemiology department, please also think of yourself when finishing your thesis! If I ever decide I like to eat chocolate, the first thing I will do is ask for the recipe of your chocolate cake! Take care, I try not to say "Idi u pičku materinu" anymore. Sweet **Henriët.** Thanks for being my wife-at-work, it's good to have someone standing for you when you forget your ERGO-shift ☺. You are a mix of worlds in one person, "Nick en Simon", specialist in training and genetic epidemiologist. It was great to be your paranimf. Luckily we got divorced before we both got children. I get to meet your little Smurf Ninthe soon. Dear **Hata**, finally they are hiring a Dutch person I thought when you came to the department. On top of being organized and extremely kind, you are a bit old-fashioned, in the good way ☺. Thank you for the smooth transition of the CJD work and the Rucphen visits. You picked it up fast and took over before I noticed. **Elisa** (and Pim), thanks for all the tea and endless chatter on whatever happened at the department. I will never forget that you said: 'I don't talk a lot, I just listen very well", a priceless quote. You have a big heart and Sophie thanks you for the pregnancy clothes she borrowed from you, they were very stylish. They say trouble, doubles, but for you it would be good if it does not anymore, but together with Pim and the kids you will be ok! Dear **Ayse**, you are the coolest mix of persons I know. Thank you for all the help with the metabolomics projects, I would have never understood the basics without you. Please stay as you are (maybe with a different hair color every now and then), and I am super curious what your little baby will look like. I also thank Sara and Carla for the CJD help and all the rice waffles. Aaron, Linda, Maarten and Lennart for their company and help accommodating during the first years of my stay at the department. Lastly the wet-

lab. Jeannette, you are the history book of the department, on top of that super organized and accurate. Thank you for your help with the CJD and unraveling the mysteries of the ADSP swaps. Andy, Bernadet en Andrea, bedankt voor alle gezelligheid en dat jullie altijd voor me klaar stonden met de CJD bezoekjes, hoe lang ik ook in de file stond.

I had the pleasure to work with the epidemiology department. Here I had the pleasure to word with **Hieab.** Dear Hieab, you are without any doubt the most puzzeling genius I know of. Extremely productive and still making time for everyone. I honestly don't know how you do it. Thank you for all the help on the MRI projects and the Abeta-paper. I wish you good fortune with the medical studies and I look forward reading your groundbreaking work. The same amount of thanks goes to **Gennady**. I always wondered, how did a Russian genius with physical superpowers end up in Rotterdam? Thanks for the insightful scientific discussions we had and the help with the VBM analyses. They definitely would not have happened without you!! Then comes the time to thank **Frank**. Dear Frankie, we have the same taste humor and therefore jokes always land well ☺. At work you are a very skilled epidemiologist and researcher with an eye for detail. After work you are more fun ;). In the future I hope to keep in touch, trough work and after work. Please make time to go to future AAIC's! And I would like to be invited on your wedding in some Spanish speaking country. Dear **Sanaz**, Leiden trains, trains, trains, Rotterdam, we spend over 2000 hours in them :O. Thanks for sharing some of these train hours chatting over life and R-figures. Ow, how we both like to make them perfect ☺. My thanks also goes to the others in the department: Rens, Hoyan, Hazal (see you in A'dam!), Marileen, Silvan, Unal, Sirwan. From the Epidemiology department I also thank Prof. Albert Hofman for setting up the Rotterdam Study and prof. Aad van der Lugt, thank you for the good advice. Frank van Rooij for taking care of all the data. Nano for his password and fixing my computer. Cris voor alle koffie die ik gebietst heb. De dames van het ERGO centrum voor hun goede zorgen en de taart. As well as the others I worked with scattered in and around the Erasmus MC. Tsz and Jeroen van Rooij. Good luck met het afronden van jullie PhDs we zien elkaar weer in ADES of in het ziekenhuis. Sara Galle, van Amsterdam, naar Rucphen, naar Rotterdam, succes met alles. Ik wil alle huisartsen in Rucphen en omstreken bedanken voor de

hartelijke ontvangst en alle hulp bij de data verzameling. In het bijzonder wil ik de bijdrage van Pieter Snijders, en *in memoriam*, Peter van Wouw noemen.

Hoewel het in mijn proefschrift niet aan bod komt, was het werk voor de Landelijke registratie prionziekten een belangrijk onderdeel van mij promotie. Het prionlab in Utrecht met pathologen Annemieke Rozemuller en Wim Spliet, ondersteunend Will Hermsen, bedankt voor alle terugkoppelingen ik heb er veel van geleerd. Corien Swaan van het RIVM, bedankt voor de steun bij de jaarlijkse moeilijke casus en analyse van mogelijke besmettingsrisico's. In dit kader wil ik ook alle GGD-artsen, waaronder enkele vrienden en vriendinnen, bedanken. Jullie kwamen voor me op als verpleegkundigen mij het "rare mannetje van de prionziekten" noemden. Bob Will and colleagues of the University of Edinbrugh, thank you for always having my back when a MRI needed your judgment and the RT-Quic tests you performed. I sincerely hope this will soon be implemented in the Netherlands as standard test option. Evelien Lemstra van het VUmc Alzheimer centrum voor de meerdere, overigens altijd juiste, second opinions. Ik dank ook Gabrielle, Erica en Riet voor de jarenlange telefonische ondersteuning, het lab voor de ondersteuning, Frank Wolters, Hoyan Wen en Sara Willems voor de back-up zodat ik soms op vakantie en congres mocht gaan en Hata Comic voor de soepele opvolging. Als laatste dank ik ook alle patiënten en familieleden die mij hartelijk ontvingen om deel te nemen aan dit zo nodige wetenschappelijk onderzoek naar prionziekten, in de tijd dat hun dierbare ernstig ziek was.

All but two studies in my thesis incorporated results from multiple study sites, this means I have to send gratitude to the colleagues I worked with in the other institutions in Europe and the United States. The analysts I worked close with and learned a great deal from; Josha Bis, Ganesh Chauhan, Vincent Chouraki, Claudia Satizabal, Rebecca Sims, Adam Naj, Céline Bellenguez, Nandini Badarinarayan, Albert Smith, Edith Hofer, Alexander Teumer, Johanna Jakobsdottir, Juho Tynkkynen, Jussi Hernesniemi, Martin Shipley and Krista Fischer. Of these I would especially like to mention some. Johanna, thank you for your work on the exome chip project and teaching me most of the tricks how to work with seqMeta. Big, big thanks to my AAIC buddies Vincent and Claudia. Staying in an AirBNB in Copenhagen was so awesome, working in candlelight going to

the pub (not you Vincent ;)), that we made it a tradition. In Washington there is always a nice Mexican restaurant just around the corner, in Toronto (crowded apartment!) the elevators broke down and the 54 stairs almost killed us, in London we went by cable and there was a surprise party! Juho, Jussi, Martin and Krista, you are example collaborators, thanks for making chapter 5 such a success. I am sure our papers will make it through the endless review. I also thank the senior researchers from these studies: Sudha Seshadri, Gerard Schellenberg, Alfredo Ramirez, Agustin Ruiz, Henne Holstege, Li-San Wang, Jean-Charles Lambert, Julie Williams, Charles DeCarli, Lenore Launer, Myriam Fornage, Helena Schmidt, Paul Nyquist, Stephanie Debette, Joanna Wardlaw, Reinhold Schmidt, Brian W. Kunkle, Stephanie Debette, Eline Slagboom, Mika Ala-Korpela, Mika Kivimäki, Veikko Salomaa, Dorret Boomsma, Wiesje van der Flier, Philippe Amouyel. Of these I especially thank **Prof. Sudha Seshadri.** Dear Sudha, the NeuroCHARGE working group could not have a more passionate shepherd. You know all of us and of all our projects, in this way assuring that the right researchers meet to get the best out of the data through collaborations. Thank you for the multiple diners at various meetings, I will always cancel my agenda for them as you make sure you get us the best food in town. Hopefully we can stay in touch in future collaborations. Dear **Prof. Gerard Schellenberg** and **Adam Naj**; Adam and Jerry, thank you for receiving me in the wonderful city of Philadelphia, it was a learning experience to work on the HRC data in ADGC and the genetics of PSP. If you are in the Netherlands, drop by for a BBQ, although I am definitely not such a grill master as you. Beste Henne, twee papers samengewerkt en ontzettend leuk dat we nu weer samen op kunnen trekken. Ik dank jou, het 100-plus team en de rest van het Alzheimer centrum voor het warme welkom in Amsterdam.

Naast onderzoek is het hele taak om vrienden te blijven zien. Lieve Danielle, lang geleden in het IFMSA bestuur en daarna altijd blijven plakken. Je voortdurende blijheid, Danielle- grapjes en luisterend oor zijn onvervangbaar. Je bent een steun en toeverlaat en altijd beschikbaar voor spontane etentjes en andere acties. Ik hoop dat we volgend jaar op jóuw PhD mogen toasten. Lieve andere friends en bijbehorende aanhang, Rayan & Bart, Zeen, Jacquelien, Daisy en Annelies. Bedankt voor alle support, af en toe aanhoren van PhD-geklaag en onvergetelijke etentjes. Hetzelfde geldt voor de

Geneeskunde buddy's en vaste aanhang; Astrid en Sebastiaan, Michelle en Lucy, Veronique en Pascal, en Irene (met wie nu weer?). Wanneer gaan we weer naar Texel? Vincent en Marloes, Julia en Lisa. Via, via zijn we nu al lang vrienden. Daarnaast verhuizen elkaar altijd achterna ☺, dus voor ons is het weer tijd om te verhuizen. Kay en Marjolein. Kay, getrouwd en gesetteld, wie had dat ooit gedacht. Hopelijk kan ik altijd op random tijden blijven aanwaaien. Dear Melinda, your perseverance will make you one of the best South African researchers. I expect the invitation to your PhD defence soon. Onderzoek is topsport: Lianne en Tessa, als ik jullie niet had als stok achter de deur om te tennissen was ik allang een slappe kantoorworst geworden. Blijf zoals jullie zijn, dan kan ik blijven lachen én kan er eindelijk is over nagedacht worden, … ofzo!

Lieve schoonfamilie, Eva, John, Suzan, Thijs, Lisa, Nina, Luce, Sterre, Rik en Distel. Aan alle Mooijs, Haasjes, Bossen, Bollen, Hammen, bedankt voor alle oprechte interesse in mijn onderzoek. Daarnaast wil ik jullie bedanken voor de warme ontvangst in de familie en voor de daarop volgende gezellige eerste 10 jaar being-in-the-family. We delen inmiddels meer dan Sophie en het is altijd gezellig op alle etentjes, uitjes en vakanties.

Als laatste mijn familie. Geweldigste zusje, leukste tante, favorit barnbarn, lieve Gea. Jij weet als geen ander dat zonder jou alles saai en grijs was en is. We hebben zo veel gemeen, van onze liefde voor Gagies en Pukjes, reisjes maken, de helft van de genen ☺, interesse in groen en onderzoek. Succes met je eigen PhD, maar dat komt goed, want je bent een van de scherpste en slimste mensen die ik ken! Ik ben enorm dankbaar dat jij mij achter me staat als paranimf.

Mama en Papa, Joop en Klaudia. Jullie zijn er altijd en geven onvoorwaardelijke steun. Een luisterend oor, een handje helpen, lekker eten, zonder jullie was dit boekje er nooit geweest. Lieve Mama, het is pas een paar jaar geleden heb jij je masterscriptie verdedigd op dezelfde plaats als waar ik nu mijn thesis mag verdedigen. Ik was supertrots dat mijn moeder haar studie afgerond had, maar snapte niets van je antwoorden. Bedankt voor alle aanmoedigingen en ik hoop je net zo trots te maken ☺. Lieve Pappa. Je roept al jaren dat Gea en ik ons niet zo uit moeten sloven, maar ondertussen weet ik dat je apetrots

bent. Bedankt voor al je oneliners die altijd waar blijken te zijn en natuurlijk de mooie planten waar ik op mag passen! Lieve Mama en Papa, ik hou van jullie.

Sophie, de mooiste, knapste, liefste en geduldigste vriendin van de wereld ben jij. Ik meen het elke keer als ik het zeg. Weet je nog dat ik je vroeg: “Mag ik een PhD doen?”. Jij antwoordde: “Natuurlijk, en ik zorg voor de catering”. Mijn liefde gaat door de maag ❤. Bedankt dat je er bent, dat je met mij een genetische studie (N=1) bent begonnen en dat ik iedere dag naast je mag wakker worden.

Lieve, lieve, lieve Luc! Wat ben je nog klein als ik dit schrijf, maar ik ben al zo trots als je iedere avond geïnteresseerd op schoot meekijkt naar het scherm van mijn laptop. Later ben je groot en kun je in dit dankwoord lezen hoeveel ik van je hou. Lees je naast het dankwoord, net als alle anderen die dit dankwoord lezen, ook de rest van mijn thesis ☺?

PhD Portfolio

Name PhD student:	Sven J. van der Lee
Erasmus MC department:	Epidemiology
Research school:	Netherlands Institute of Health Sciences (NIHES)
PhD Period:	August 2012 – Oktober 2017
Promotors:	Prof. C.M. van Duijn
	Prof. M.A. Ikram
Copromotor:	Dr. N. Amin

EDUCATION	YEAR	ECTS
Master of Science in Health Sciences, Genetic epidemiology		
Core Curriculum		
Study Design	2013	4.3
Biostatistical Methods I: Basic Principles	2012	5.7
Biostatistical Methods II: Classical Regression Models	2012	4.3
Genetic-Epidemiologic Research Methods	2012	5.1
SNP's and Human Diseases	2013	1.4
Advanced Short Courses		
Working with Linux	2012	0.6
Courses for the Quantitative Researcher	2012	1.4
Advances in Genome-Wide Association Studies	2013	1.4
Family-based Genetic Analysis	2013	1.4
A first encounter with next-generation sequencing data	2013	1.4
CMSB	2014	0.4
EURON Workshop: New Targets in Neurodegenerative diseases	2014	1.4
Introduction to Neuroinformatics	2014	0.6
Introduction to Medical Writing	2014	1.1
Erasmus Summer Programme		
Principles of Research in Medicine	2012	0.7

Genome Wide Association Analysis	2012	1.4
Principles of Genetic Epidemiology	2012	0.7
Genomics in Molecular Medicine	2012	1.4
Advances in Genomics Research	2012	0.4

Attended Seminars and workshops

Seminars at the department of Epidemiology	2012-2016	2.0
Genetic epidemiology working group meetings	2012-2017	3.0

(INTER)NATIONAL CONFERENCES AND MEETINGS	**YEAR**	**ECTS**
ENGAGE consortium meeting (Utrecht, the Netherlands)	2012	0.3
IGAP consortium meeting (Paris, France)	2013	0.6
CHARGE consortium meeting (Rotterdam, the Netherlands)	2013	0.9
AAIC (Copenhagen, Denmark), *Oral presentation, Poster presentation*	2013	1.2
IGAP consortium meeting (Copenhagen, Denmark), *Oral presentation*	2013	0.3
CHARGE consortium meeting (Los Angeles, USA)	2014	1.2
CHARGE consortium meeting (Washington, USA), *Poster presentation*	2014	0.9
HD-Ready consortium (Rotterdam, the Netherlands)	2014	0.6
International Biobanking Summit III, *Session leader*	2014	0.6
"Epidemiology of Prion diseases", *invited Masterclass on Prion diseases*	2014	0.6
ImaGene (Utrecht, the Netherlands)	2014	0.4
HD-Ready consortium (Rotterdam, the Netherlands)	2015	0.6
BBMRI-NL, multiple consortium meetings (The Netherlands)	2015	0.8
AAIC (Washington, USA), Poster presentation	2015	1.2
IGAP consortium meeting (Washington, USA)	2015	0.3
AAIC (Toronto, Canada), *Poster presentation*	2016	1.2
BBMRI-NL, multiple consortium meetings (The Netherlands)	2016	0.8

AWARDS	**YEAR**	**ECTS**
"Early Career" CHARGE Tiger Award	2014	
Travel award CHARGE-meeting	2016	

TEACHING	**YEAR**	**ECTS**
Teaching in NIHES Advance genomics course, topic: Exome chip analysis	2014-2016	1.2
Supervision research internship, Josue Gonzalez	2013	1.0
Supervision research internship, Ilse Geraedts	2016	1.0

OTHER		
Registration physician for Creutzfeldt-Jakob disease (CJD) in the Netherlands	2012-2016	20

* ECTS (European Credit Transfer System) equals a workload of 28 hours.

Other publications

1. Huang KL, et al. *A common haplotype lowers PU1 expression in myeloid cells and delays onset of Alzheimer's disease*. Nat Neurosci. 2017 Jun 19.
2. Hägg S, et al. *Short telomere length is associated with impaired cognitive performance in European ancestry cohorts.* Transl Psychiatry. 2017 Apr 18;7(4):e1100.
3. Amin N, et al. *Exome-sequencing in a large population-based study reveals a rare Asn396Ser variant in the LIPG gene associated with depressive symptoms.* Mol Psychiatry. 2017 Apr;22(4):537-543.
4. Toledo JB, et al. *Metabolic network failures in Alzheimer's disease-A biochemical road map.* Alzheimers Dement. 2017 Mar 21.
5. Hibar DP, et al. *Novel genetic loci associated with hippocampal volume.* Nat Commun. 2017 Jan 18;8:13624.
6. Springelkamp H, et al. *New insights into the genetics of primary open-angle glaucoma based on meta-analyses of intraocular pressure and optic disc characteristics.* Hum Mol Genet. 2017 Jan 15;26(2):438-453.
7. Roshchupkin GV, et al. *Heritability of the shape of subcortical brain structures in the general population*. Nat Commun. 2016 Dec 15;7:13738.
8. Adams HH, et al. *Novel genetic loci underlying human intracranial volume identified through genome-wide association.* Nat Neurosci. 2016 Dec;19(12):1569-1582.
9. Moreno-Gordaliza E, et al. *A novel method for serum lipoprotein profiling using high performance capillary isotachophoresis.* Anal Chim Acta. 2016 Nov 9;944:57-69.
10. Chouraki V, et al. *Evaluation of a Genetic Risk Score to Improve Risk Prediction for Alzheimer's Disease.* J Alzheimers Dis. 2016 Jun 18;53(3):921-32.
11. Okbay A, et al. *Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses.* Nat Genet. 2016 Jun;48(6):624-33.
12. Scott RA, et al. *A genomic approach to therapeutic target validation identifies a glucose-lowering GLP1R variant protective for coronary heart disease.* Sci Transl Med. 2016 Jun 1;8(341):341ra76.
13. Okbay A, et al. *Genome-wide association study identifies 74 loci associated with educational attainment.* Nature. 2016 May 26;533(7604):539-42.
14. Jensen RA, et al. *Novel Genetic Loci Associated With Retinal Microvascular Diameter.* Circ Cardiovasc Genet. 2016 Feb;9(1):45-54.
15. Minikel EV, et al. *Quantifying prion disease penetrance using large population control cohorts.* Sci Transl Med. 2016 Jan 20;8(322):322ra9.
16. Jun G, et al. *A novel Alzheimer disease locus located near the gene encoding tau protein.* Mol Psychiatry. 2016 Jan;21(1):108-17.
17. Adams HH, et al. *A priori collaboration in population imaging: The Uniform Neuro-Imaging of Virchow-Robin Spaces Enlargement consortium.* Alzheimers Dement (Amst). 2015 Dec;1(4):513-20.
18. Desikan RS, et al. *Polygenic Overlap Between C-Reactive Protein, Plasma Lipids, and Alzheimer Disease.* Circulation. 2015 Jun 9;131(23):2061-9.
19. Vojinovic D, et al. *The dystrophin gene and cognitive function in the general population.* Eur J Hum Genet. 2015 Jun;23(6):837-43.
20. Springelkamp H, et al. *ARHGEF12 influences the risk of glaucoma by increasing intraocular pressure.* Hum Mol Genet. 2015 May 1;24(9):2689-99.

21. Chaker L, et al. *The global impact of non-communicable diseases on macro-economic productivity: a systematic review.* Eur J Epidemiol. 2015 May;30(5):357-95.
22. Hibar DP, et al. *Common genetic variants influence human subcortical brain structures.* Nature. 2015 Apr 9;520(7546):224-9.
23. Muka T, et al. *The global impact of non-communicable diseases on healthcare spending and national income: a systematic review.* Eur J Epidemiol. 2015 Apr;30(4):251-77.
24. Chauhan G, et al. *Association of Alzheimer's disease GWAS loci with MRI markers of brain aging.* Neurobiol Aging. 2015 Apr;36(4):1765.e7-16.
25. Springelkamp H, et al. *Meta-analysis of Genome-Wide Association Studies Identifies Novel Loci Associated With Optic Disc Morphology.* Genet Epidemiol. 2015 Mar;39(3):207-16.
26. Jaspers L, et al. *The global impact of non-communicable diseases on households and impoverishment: a systematic review.* Eur J Epidemiol. 2015 Mar;30(3):163-88.
27. Davies G, et al. *Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53949).* Mol Psychiatry. 2015 Feb;20(2):183-92.
28. Rietveld CA, et al. *Common genetic variants associated with cognitive performance identified using the proxy-phenotype method.* Proc Natl Acad Sci U S A. 2014 Sep 23;111(38):13790-4.
29. Sanchez-Juan P, et al. *A genome wide association study links glutamate receptor pathway to sporadic Creutzfeldt-Jakob disease risk.* PLoS One. 2014;10(4):e0123654.

About the Author

Sven Johan van der Lee was born in Delft, the Netherlands, on the 2nd of October 1987. He completed his Gymnasium (pre-scientific) education *Cum Laude* at the Grotius College in Delft, the Netherlands, in 2005 in which year he also started his medical studies at Leiden University, Leiden, the Netherlands. In 2011 he finished his medical studies started working as a resident not in training in the Alrijne Hospital in Leiderdorp, the Netherlands. After a year he came to the Erasmus University Medical Center (ErasmusMC) and started working on this thesis under supervision of prof. Cornelia van Duijn at the genetic epidemiology unit and prof. Arfan Ikram, head of the department of epidemiology. During the work that cumulated in this thesis he worked together with leading consortia in multiple research fields. In Alzheimer's genetic research Sven was part of the Cohorts of Heart and Ageing Research in Genomic Epidemiology (CHARGE), International Genomics of Alzheimer's Project (IGAP) and the Alzheimer's Disease Sequencing Project (ADSP). For his work on imaging genetics he worked in the Population imaging genetics (ImaGene) project and with the Biobanking and BioMolecular resources Research Infrastructure The Netherlands (BBMRI-NL) for the metabolomics studies in this thesis. During his time at the ErasmusMC he was part of the official Dutch registration for prion diseases and for his thesis he finished the master Genetic epidemiology at the Netherlands Institute for Health Sciences (NIHES). In 2015 he visited the Department of Pathology and Laboratory Medicine of Prof. Gerard D. Schellenberg for a three months research fellowship at the University of Pennsylvania, Philadelphia, United states. Sven was awarded a golden tiger for his early career achievements in CHARGE. In the future Sven aims to combine clinical work with research. As first step he works now at the Alzheimer center and the 100-plus study at the VU University Medical Center in Amsterdam, the Netherlands, were will continue his work to unravel the genetics of Alzheimer's disease.