# A Genetic Epidemiological Study of Dementia and Cognitive Function

Carla A. Ibrahim-Verbaas

**ACKNOWLEDGMENTS**
The work presented in ths thesis was conducted at the Genetic Epidemiology Unit, Department of Epidemiology, and the Department of Neurology, Erasmus University Medical Center, Rotterdam.

The ERF study was supported by the joint grant from the Netherlands Organisation for Scientific Research (NWO, 91203014), the Center for Medical Systems Biology (CMSB), Hersenstichting Nederland (project number 12F04(2).76), Internationale Stichting Alzheimer Onderzoek (ISAO), Alzheimer Association project number 04516, and the Interuniversity Attraction Poles (IUAP) program. The ERF 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). High-throughput analysis of the ERF data was supported by a joint grant from Netherlands Organization for Scientific Research and the Russian Foundation for Basic Research (NWO-RFBR 047.017.043). Exome sequencing in ERF was supported by the ZonMw grant (project 91111025). Exome-chip genotyping was supported by BBMRI-NL.

The generation and management of GWAS genotype data for the Rotterdam Study is supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012). This study is funded by the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) project nr. 050-060-810. 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.

Financial support for publication of this thesis was kindly provided by Alzheimer Nederland, the Internationale Stichting Alzheimer Onderzoek and Erasmus University Rotterdam.

ISBN: 978-90-5335-866-5
Cover: Ridderprint BV, Ridderkerk, the Netherlands; cover idea C.A. Ibrahim-Verbaas
Printed by: Ridderprint BV, Ridderkerk, the Netherlands
© 2014 C.A. Ibrahim-Verbaas

No part of this thesis may be reproduced or transmitted in any form or by any means without prior permission of the author, C.A. Ibrahim-Verbaas, or when appropriate, of the scientific journal in which parts of this thesis have been published.

# A Genetic Epidemiological Study of Dementia and Cognitive Function

# Een genetisch epidemiologische studie naar dementie en cognitieve functie

Proefschrift

ter verkrijging van de graad van doctor aan de Erasmus Universiteit Rotterdam op gezag van de rector magnificus
Prof.dr. H.A.P. Pols
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
dinsdag 10 juni 2014 om 9.30 uur

door

Carla Antoinette Ibrahim-Verbaas
geboren te Den Haag

Promotiecommissie

| | |
|---|---|
| Promotoren: | Prof.dr.ir. C.M. van Duijn |
| | Prof.dr. J.C. van Swieten |
| Overige leden: | Prof.dr. R.M.W. Hofstra |
| | Dr. L.J. Launer |
| | Dr. M.A. Ikram |
| Copromotor: | Dr. N. Amin |

## TABLE OF CONTENTS

**Chapter 1: General introduction ........ 11**

**Chapter 2: Genetic Epidemiology of Alzheimer's Disease ........ 21**

*2. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease ........ 23*

**Chapter 3: Genetic Epidemiology of Cognitive Functioning ........ 35**

*3.1. Genome-wide linkage and fine-mapping with exonic variants identify rare deleterious mutations affecting cognitive functioning ........ 37*

*3.2. GWAS for executive function and processing speed suggests involvement of the CADM2 gene. ........ 57*

*3.3. Genome-wide studies of verbal declarative memory in non-demented older people: the CHARGE consortium ........ 77*

**Chapter 4 : Exploring vascular endophenotypes of Alzheimer's disease ........ 103**

*4.1. Linkage analysis for plasma amyloid beta levels in persons with hypertension implicates Aβ-40 levels to Presenilin 2 ........ 105*

*4.2. Predicting stroke through genetic risk functions: The CHARGE risk score project ........ 119*

**Chapter 5 : General discussion ........ 139**

**Chapter 6: Summary/Samenvatting ........ 157**

*6.1. Summary ........ 159*

*6.2. Samenvatting ........ 165*

**Chapter 7: Appendices (Supplemental Material) ........ 173**

*7.1. Supplemental material to Chapter 3.1 ........ 175*

*7.2. Supplemental material to Chapter 3.2 ........ 183*

*7.3. Supplemental material to Chapter 3.3 ........ 225*

**Chapter 8: Epilogue ........ 289**

*8.1. Dankwoord/Acknowledgments ........ 291*

*8.2 About the author ........ 299*

*8.3. List of publications ........ 303*

*8.4. PhD portfolio ........ 309*

**PUBLICATIONS AND MANUSCRIPTS BASED ON THE STUDIES DESCRIBED IN THIS THESIS**

**Chapter 2**

Jean-Charles Lambert*, Carla A Ibrahim-Verbaas*, Denise Harold*, Adam C Naj*, Rebecca Sims, Céline Bellenguez, Gyungah Jun, Anita L DeStefano, Joshua C Bis, Gary W Beecham, Benjamin Grenier-Boley, Giancarlo Russo, Tricia A Thornton-Wells, Nicola Jones, Albert V Smith, Vincent Chouraki, Charlene Thomas, M Arfan Ikram, Diana Zelenika, Badri N Vardarajan, Yoichiro Kamatani, Chiao-Feng Lin, Amy Gerrish, Helena Schmidt, Brian Kunkle, Melanie L Dunstan, Agustin Ruiz, Marie-Thérèse Bihoreau, Seung-Hoan Choi, Christiane Reitz, Florence Pasquier, Paul Hollingworth, Alfredo Ramirez, Olivier Hanon, Annette L Fitzpatrick, Joseph D Buxbaum, Dominique Campion**,** Paul K Crane, Clinton Baldwin, Tim Becker, Vilmundur Gudnason, Carlos Cruchaga, David Craig, Najaf Amin, Claudine Berr, Oscar L Lopez, Philip L De Jager, Vincent Deramecourt, Janet A Johnston, Denis Evans, Simon Lovestone, Luc Letenneur, Francisco J Morón, David C Rubinsztein, Gudny Eiriksdottir, Kristel Sleegers, Alison M Goate, Nathalie Fiévet, Matthew J Huentelman, Michael Gill, Kristelle Brown, M Ilyas Kamboh, Lina Keller, Pascale Barberger-Gateau, Bernadette McGuinness, Eric B Larson, Robert Green, Amanda J Myers, Carole Dufouil, Stephen Todd, David Wallon, Seth Love, Ekaterina Rogaeva, John Gallacher, Peter St George-Hyslop, Jordi Clarimon, Alberti Lleo, Anthony Bayer, Debby W Tsuang, Lei Yu, Magda Tsolaki, Paola Bossù, Gianfranco Spalletta, Petra Proitsi, John Collinge, Sandro Sorbi, Florentino Sanchez Garcia, Nick C Fox, John Hardy, Maria Candida Deniz Naranjo, Paolo Bosco, Robert Clarke, Carol Brayne, Daniela Galimberti, Michelangelo Mancuso, Fiona Matthews, EADI, GERAD, ADGC, CHARGE, Susanne Moebus, Patrizia Mecocci, Maria Del Zompo, Wolfgang Maier, Harald Hampel, Alberto Pilotto, Maria Bullido, Francesco Panza, Paolo Caffarra, Benedetta Nacmias, John R Gilbert, Manuel Mayhaus, Lars Lannfelt, Hakon Hakonarson, Sabrina Pichler, Minerva M Carrasquillo, Martin Ingelsson, Duane Beekly, Victoria Alvarez, Fanggeng Zou, Otto Valladares, Steven G Younkin, Eliecer Coto, Kara L Hamilton-Nelson, Wei Gu, Cristina Razquin, Pau Pastor, Ignacio Mateo, Michael J Owen, Kelley M Faber, Palmi V Jonsson, Onofre Combarros, Michael C O'Donovan, Laura B Cantwell, Hilkka Soininen, Deborah Blacker, Simon Mead, Thomas H Mosley Jr, David A Bennett, Tamara B Harris, Laura Fratiglioni, Clive Holmes, Renee FAG de Bruijn, Peter Passmore, Thomas J Montine, Karolien Bettens, Jerome I Rotter, Alexis Brice, Kevin Morgan, Tatiana M Foroud, Walter A Kukull, Didier Hannequin, John F Powell, Michael A Nalls, Karen Ritchie, Kathryn L Lunetta, John SK Kauwe, Eric Boerwinkle, Matthias Riemenschneider, Mercè Boada, Mikko Hiltunen, Eden R Martin, Reinhold Schmidt, Dan Rujescu, Li-San Wang, Jean-François Dartigues, Richard Mayeux, Christophe Tzourio, Albert Hofman, Markus M Nöthen, Caroline Graff, Bruce M Psaty, Lesley Jones, Jonathan L Haines, Peter A Holmans, Mark Lathrop, Margaret A Pericak-Vance, Lenore J Launer, Lindsay A Farrer, Cornelia M van Duijn, Christine Van Broeckhoven, Valentina Moskvina, Sudha Seshadri*, Julie Williams*, Gerard D Schellenberg*, Philippe Amouyel*.

**Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease**

*Nat Genet 2013 Dec; 45(12):1452-8*

**Chapter 3.1**

Carla A. Ibrahim-Verbaas, Maaike Schuur, M. Arfan Ikram, Eline M.L.M. Nijskens, Irina V. Zorkoltseva, Sven J. van der Lee, Anatoly Kirichenko, Inge de Koning, Tatiana I. Axenovich, Albert Hofman, Yurii S. Aulchenko, Monique M.B. Breteler, Ben A. Oostra, John C. van Swieten, Najaf Amin, Cornelia M. van Duijn

**Genome-wide linkage and fine-mapping with exonic variants identify rare deleterious mutations affecting cognitive functioning**

*Submitted*

**Chapter 3.2**

Carla A. Ibrahim Verbaas, Jan Bressler,PhD, Stephanie Debette,MD, Maaike Schuur, Albert V. Smith, Joshua C. Bis, Gail Davies, Stella Trompet, Jennifer A. Smith, Christiane Wolf, Lori B. Chibnik, Yongmei Liu, Veronique Vitart, Mirna Kirin, Katja Petrovic, Ozren Polasek, Lina Zgaga, Per Hoffmann, Juha Karjalainen, Jari Lahti, David Llewellyn, Carsten O. Schmidt, Karen A. Mather , Vincent Chouraki, Qi Sun, Susan M. Resnick, Lynda M. Rose, Christopher Oldmeadow, Marlene Stewart, Blair H. Smith, Vilmundur Gudnason, Qiong Yang, Saira S. Mirza, J. Wouter Jukema, Philip L. deJager, Tamara B. Harris, David C. Liewald, Najaf Amin, Laura H. Coker, Oliver Stegle, Oscar L. Lopez, Reinhold Schmidt, Alexander Teumer, Ian Ford, Nazanin Karbalai, James T. Becker, Maria K. Jonsdottir, Rhoda Au, Rudolf S.N. Fehrmann, Stefan Herms, Mike Nalls, Wei Zhao, Stephen T. Turner, Kristine Yaffe, Kurt Lohman, John C. van Swieten, Sharon L.R. Kardia, David S. Knopman, William Mark Meeks, Gerardo Heiss, Elizabeth G. Holliday, Peter W. Schofield, Toshiko Tanaka, David J. Stott, Jing Wang, Paul Ridker, Alan J. Gow, Alison Pattie, John M. Starr, Lynne J. Hocking, Nicola J. Armstrong, Stela McLachlan, Joshua M. Shulman, Luke C. Pilling, Gudny Eiriksdottir, Rodney J. Scott, Nicole A. Kochan, Aarno Palotie, Yi-Chen Hsieh, Johan G. Eriksson, Alan Penman, Rebecca F. Gottesman, Ben A. Oostra, Lei Yu, Anita L. DeStefano, Alexa Beiser, Melissa Garcia, Jerome I. Rotter, Markus M. Nöthen, Albert Hofman, P. Eline Slagboom, Rudi G.J. Westendorp, Brendan M. Buckley, Philip A. Wolf, André G. Uitterlinden , Bruce M. Psaty, Hans J. Grabe, Stefania Bandinelli, Daniel I. Chasman, Francine Grodstein, Katri Raikkonen, Jean-Charles Lambert, David J. Porteous, Generation Scotland, Jackie F. Price, Perminder S. Sachdev, Luigi Ferrucci, John R. Attia, Igor Rudan, Caroline Hayward, Alan F. Wright, James F. Wilson, Sven Cichon , Lude Franke, Helena Schmidt, Jingzhong Ding, Anton JM de Craen, Myriam Fornage, David A. Bennett, Ian Deary, M. Arfan Ikram, Lenore J. Launer, Annette L. Fitzpatrick, Sudha Seshadri, Cornelia M. van Duijn, Thomas H. Mosley

**GWAS for executive function and processing speed suggests involvement of the *CADM2* gene.**

*Submitted*

**Chapter 3.3**

Stéphanie Debette, Carla A. Ibrahim Verbaas, Jan Bressler, Maaike Schuur, Albert Smith, Joshua C. Bis, Gail Davies, Christiane Wolf, Vilmundur Gudnason, Lori B. Chibnik, Qiong Yang, Anita deStefano, Dominique J.F. de Quervain, Velandai Srikanth, Jari Lahti, Hans J. Grabe, Jennifer A. Smith, Lutz Priebe, Lei Yu, Nazanin Karbalai, Caroline Hayward, James F. Wilson, Harry Campbell, Katja Petrovic, Myriam Fornage, Ganesh Chauhan, Robin Yeo, Ruth Boxall, James Becker, Oliver Stegle, Karen A. Mather, Vincent Chouraki, Qi Sun, Lynda M. Rose, Susan Resnick, Christopher Oldmeadow, Mirna Kirin, Alan F. Wright, Maria K. Jonsdottir, Rhoda Au, Albert Becker, Najaf Amin, Mike A. Nalls, Stephen T. Turner, Sharon L.R. Kardia, Ben Oostra, Gwen Windham, Laura H Coker, Wei Zhao, David S. Knopman, Gerardo Heiss, Michael E. Griswold, Rebecca F. Gottesman, Veronique Vitart, Nicholas D. Hastie, Lina Zgaga, Igor Rudan, Ozren Polasek, Elizabeth G. Holliday, Peter Schofield, Seung Hoan Choi, Toshiko Tanaka, Yang An, Rodney T. Perry, Richard E. Kennedy, Michèle M., Jing Wang, Virginia G. Wadley, David C. Liewald, Paul M. Ridker, Alan J. Gow, Alison Pattie, John M. Starr, David Porteous, Xuan Liu, Russell Thomson, Nicola J. Armstrong, Gudny Eiriksdottir, Arezoo Assareh, Nicole A. Kochan, Elisabeth Widen, Aarno Palotie, Yi-Chen Hsieh, Johan G. Eriksson, Christian Vogler, John C. van Swieten, Joshua M. Shulman, Alexa Beiser, Jerome Rotter, Carsten O. Schmidt, Wolfgang Hoffmann, Markus M. Nöthen, Luigi Ferrucci, John Attia, Andre G. Uitterlinden, Philippe Amouyel, Jean-François Dartigues, Hélène, Katri Räikkönen, Melissa Garcia, Philip A. Wolf, Albert Hofman, W.T. Longstreth Jr., Bruce M. Psaty, Eric Boerwinkle, Philip L. DeJager, Perminder S. Sachdev, Reinhold Schmidt, Monique M.B. Breteler, Alexander Teumer, Oscar Lopez, Sven Cichon, Daniel I. Chasman, Francine Grodstein, Bertram Müller-Myhsok, Christophe Tzourio, Andreas Papassotiropoulos, David A. Bennett, M. Arfan Ikram, Ian J. Deary, Cornelia M. van Duijn, Lenore Launer, Annette L. Fitzpatrick, Sudha Seshadri, Thomas H. Mosley Jr, for the CHARGE Consortium

**Genome-wide studies of verbal declarative memory in non-demented older people: the CHARGE consortium**

*Submitted*

**Chapter 4.1**

Carla A Ibrahim-Verbaas, Irina V Zorkoltseva, Najaf Amin, Maaike Schuur, Antonia MW Coppus, Aaron Isaacs, Yurii S Aulchenko, Monique MB Breteler, M. Arfan Ikram, Tatiana I. Axenovich, Marcel M. Verbeek, John C van Swieten, Ben A Oostra, Cornelia M van Duijn

**Linkage analysis for plasma amyloid beta levels in persons with hypertension implicates Aβ-40 levels to Presenilin 2**

*Hum Genet 2012 Dec; 131(12):1869-76*

**Chapter 4.2**

Carla A Ibrahim-Verbaas, Myriam Fornage, Joshua C Bis, Seung Hoan Choi, Bruce M Psaty, James B Meigs, MPH Madhu Rao, Mike Nalls, Joao D Fontes, Christopher J. O'Donnell, Sekar Kathiresan, Georg B. Ehret, Caroline S Fox, Rainer Malik, Martin Dichgans, Helena Schmidt, Jari Lahti, Susan R Heckbert, Thomas Lumley, Kenneth Rice, Jerome I Rotter, Kent D Taylor, Aaron R Folsom, Eric Boerwinkle, Wayne D Rosamond, Eyal Shahar, Rebecca F. Gottesman, Peter J Koudstaal, Najaf Amin, Renske G. Wieberdink, Abbas Dehghan, Albert Hofman, André G Uitterlinden, Anita L. DeStefano, Stephanie Debette, Luting Xue, Alexa Beiser, Philip A. Wolf, Charles DeCarli, M. Arfan Ikram, Sudha Seshadri, Thomas H Mosley, Jr, WT Longstreth, Jr, Cornelia M van Duijn, Lenore J Launer

**Predicting stroke through genetic risk functions: The CHARGE risk score project**

*Stroke 2014 Feb; 45(2):403-12*

# Chapter 1: General introduction

The epidemic of dementia is one of the greatest challenges for health care systems in the Western world. Dementia is a chronic and progressive syndrome of deterioration of cognitive ability, beyond what might be expected from normal ageing,[1] and is seen in different underlying neurodegenerative diseases. The most common type of dementia among the aging population is Alzheimer's Disease (AD), characterized by an insidious but progressive impairment, often most of short-term memory, language , and visuospatial functions. Some patients also develop behavioural and personality changes. Estimates of lifetime incidence for dementia are up to one in three women and one in six men.[2-4] These estimates may vary depending on selection of cohorts , and overestimation has recently been suggested in the United Kingdom.[5] Despite such possible overestimation, AD still remains one of the most common diseases in later life, and its increase in prevalence over the last two decades has a major impact on health care facilities.[6]

Pathophysiologically, AD is associated with accumulation of neurotoxic amyloid β oligomers (amyloid plaques) and hyperphosphorylated Tau protein in the brain, found at autopsy.[7] Depositions of these two proteins can be found in the brains of AD patients after death.[8,9] More recently, the amyloid deposition can be visualized during life with PET imaging of the brain.[10] The hypothesis is that impaired splicing of the amyloid β precursor (APP) protein and clearance of the toxic products from the brain leads to oxidative stress response and apoptosis, inflammatory responses and changed lipid metabolism.[11-13] Hippocampal atrophy on MRI (atrophy), F-FDG-PET (temporoparietal hypometabolism), and SPECT (temporoparietal hypoperfusion) are biomarkers for AD, visualizing the processes that are crucial to memory dysfunction.[10]

Genetic factors play an important role in AD, with estimates of heritability based on twin studies ranging from 60[14] to 74[15] %. The identification of the genes driving this heritability is slowly progressing. In the early and mid-1990s, mutations in the *APP* gene coding for the APP protein,[16] and in presenilin 1 and 2 genes (*PSEN1* and *PSEN2*) were identified in families with autosomal dominant AD.[17-21] The corresponding mutant proteins are responsible for the cutting of APP resulting in the neurotoxic amyloid β proteins, leading to an increased burden of these toxic particles in mutation carriers. These rare mutations are associated with an early onset form of AD in an autosomal dominant inheritance pattern, in contrast to the higher onset age in sporadic AD. However, their contribution to the general prevalence of AD is small..[13,22] The identification of these three genes led to the implication that amyloid β processing plays a causal role in AD.[23] The breakthrough in the genetics of AD in the general population was the discovery of the ε4 haplotype in the apolipoprotein E gene (*APOE*ε4) which is found in roughly 25% of the general population, and roughly triples one's risk for AD with each copy carried.[24] The mechanisms through which *APOE* exerts its effect on AD pathogenesis is still unclear. A role of *APOE* as a transcription mediator has recently been proposed.[25] In many candidate gene studies, no other genetic variants were consistently associated to AD.[26] Although *APOE*ε4 remains the main genetic risk factor for

AD, estimated to explain 26% of the late-onset AD cases when present in 25% of persons,[27] large-scale genome-wide association studies (GWAS) further implicated genes including *CR1, CLU, PICALM, BIN1, EPHA1, MS4A, CD33, CD2AP*, and *ABCA7* consistently each conferring small risk, substantially smaller than that of *APOE*, of developing AD.[28,29] In combination, the genetic markers known to date explain about 30% of the variation in AD risk.[30] Chapter 2 of this thesis discusses a further extension of collaborative GWAS efforts and more newly identified genes. The most recent breakthrough in late-onset AD genetics is the discovery of a rare mutation in the *TREM2* gene giving a roughly three-fold increased risk of AD.[31,32]

In genetic research of many complex diseases, a powerful approach has been to target endophenotypes. Endophenotypes are quantitative traits associated to the disease (in this case AD) and, important in genetic studies, have a strong heritable component. Unlike presence or absence of disease, they do not depend on a case-control definition. In life, a diagnosis of Alzheimer's disease can be made only "probabilistically". This may lead to a dilution of effects if "true" cases in a cohort are mixed with patients with different pathophysiological mechanisms, thus limiting the statistical power of any analysis. Moreover, control subjects bearing the same genetic factors may well develop AD within a few years, further hampering gene discovery. Above the age of 90 years, the risk of developing Alzheimer's in the next year is 18%, and this probability rises up to 40 % after the age of 100.[33] Further of note is that of the nonagerians not diagnosed with dementia in life, 49% have neuropathological findings consistent with AD at autopsy, [34] and virtually everyone above the age of 85 has at least some plaques and tangles.[35] However, plaques and tangles are not synonumous with aging, as a 115-year old Dutch lady without AD pathology has convincingly shown.[36]

At the other end, the presence of AD pathology does not allow to diagnose the disease in the absence of a clinical diagnosis of a progressive dementia. The load of amyloid is associated to memory and global cognitive function in clinically nondemented persons.[35] These numbers give support to the concept that AD may be the extreme of a spectrum of cognitive aging.[37,38]

Given the many drawbacks in studying AD as a binary trait, endophenotypes may be a powerful method for AD gene discovery.[39-43] For AD, possible endophenotypes can be amyloid beta or Tau levels in blood or cerebrospinal fluid, measures of atrophy and vascular pathology on MRI, or measures of cognitive function. In this thesis, I have focused on plasma amyloid beta levels as well as cognitive measurements. As a spin-off of the discovery of the *APP* locus, measurements of amyloid β levels in the cerebrospinal fluid are now regularly performed as part of the diagnostic work-up of patients with possible dementia as well as classification in clinical studies of dementia.[44,45] Significant elevation of plasma Amyloid β levels was found in carriers of *APP*, *PSEN1* and *PSEN2* mutations causative

of familial AD.[46] The association between plasma amyloid beta levels and sporadic, late onset AD has been less consistent, and their use in the diagnostic work-up and individual risk prediction of dementia seems to be limited for now.[47-49] Yet there is evidence that in population studies and studies in healthy relatives of AD patients, plasma amyloid β levels are associated to AD. [50,51]

Cognitive functioning can be used as another endophenotype, as subjective memory complaints as well as performance on neuropsychological tests are a predictor of developing dementia.[52,53] Lower performance on memory, processing speed and executive function is predictive of the developing AD within a few years,[54] wheras lower scores on executive functioning and processing speed, but relatively spared memory function, are predictive for developing non-Alzheimer's dementia.[55] Especially subjects with high level of education may show subjective memory problems and normal formal test results, due to ceiling effect and the phenomenon of cognitive reserve.[53,56] Memory, visuospatial ability, language functions, cognitive processing speed and executive function (reflecting aspects like mental flexibility and planning) also have a strong heritable component.[57] Based on these observations, we focus on cognitive ability as an endophenotype in chapter 3 in order to identify gene variants associated with AD. We used a standard cognitive test battery derived from that used in the diagnostic work-up of AD on healthy participants from population-based and family-based studies. Beside their association to dementia, cognitive tests have extensively been studied as endophenotypes for psychiatric disease, particularly for schizophrenia and bipolar disorder.[58,59]

In the past decade, pharmacological interventions in AD have focused on the amyloid hypothesis. However, clearance of amyloid deposits from the brain by using amyloid antibodies (bapineuzumab) has not halted the disease process in AD or improved cognitive functioning.[60,61] The poor correlation between AD type changes  and staging of clinical AD, as already described by Braak,[8] indicates that amyloid deposition is not be the only driving process in AD. At higher age, AD and cardiovascular disease share the same risk factors, including hypertension, atherosclerosis, obesity, dyslipidaemia, smoking, hemostasis and inflammation.[62-64] Stroke is a risk factor for subsequent dementia independent of premorbid cognitive functioning.[65] Therefore vascular factors may play a key role in pathogenesis of dementia, and finding genes influencing stroke risk may be relevant for dementia. Interestingly, angiotensin converting enzyme plays a role both in blood pressure regulation as well as amyloid metabolism in the brain, raising the idea of identifying genes relevant in amyloid metabolism in hypertensive individuals. However the very few genes associated to stroke risk so far do not point towards amyloid metabolism.[66,67] We have followed an alternative approach for assessing genetic risk for stroke, which we present in Chapter 4.1. We performed a candidate gene approach testing whether there is a joint effect of the genes implicated in the major risk factors of stroke such as hypertension, cholesterol levels, kidney dysfunction, cardiac arrhythmia, atherosclerosis, diabetes, obesity and inflammation.

**AIMS OF THIS THESIS**

In this thesis we aim to discover genetic variants that determine the susceptibility to AD and cognitive ability in the general population. In Chapter 2, we present the results of the largest genome-wide association study to date, searching for variants associated to the clinical diagnosis of AD. In Chapter 3, we study genetic variation driving the normal variation in cognitive ability. In Chapter 3.1, we set out to perform linkage analysis in combination with exome chip and exome sequence analyses in order to identify rare variants with large effects. In Chapter 3.2 and 3.3, genome-wide association studies for cognitive ability were performed for executive function and processing speed (Chapter 3.2) and delayed recall (memory, Chapter 3.3). Chapter 4 focuses on the vascular aspects of AD risk. In Chapter 4.1, we search for genetic variants that influence plasma amyloid beta levels in the peripheral blood as an endophenotype for AD in a middle-aged in a subset of hypertensives. Using genome-wide linkage analysis and fine-mapping the linked regions using association analyses the aim was to discover rare variants that confer large effects on plasma amyloid beta levels. In Chapter 4.2, we explored the possibilities to translate genome-wide association findings of stroke and its risk factors into a risk score for clinical stroke. Finally, in Chapter 5 we discuss the findings of this thesis, and their implications for future research.

**REFERENCES**


1. Organization, W.H. Dementia, fact sheet No 362. (World Health Organization, 2012).
2. Ott, A., Breteler, M.M., van Harskamp, F., Stijnen, T. & Hofman, A. Incidence and risk of dementia. The Rotterdam Study. *Am J Epidemiol* **147**, 574-80 (1998).
3. Breteler, M.M., Ott, A. & Hofman, A. The new epidemic: frequency of dementia in the Rotterdam Study. *Haemostasis* **28**, 117-23 (1998).
4. Seshadri, S. & Wolf, P.A. Lifetime risk of stroke and dementia: current concepts, and estimates from the Framingham Study. *Lancet Neurol* **6**, 1106-14 (2007).
5. Matthews, F.E. et al. A two-decade comparison of prevalence of dementia in individuals aged 65 years and older from three geographical areas of England: results of the Cognitive Function and Ageing Study I and II. *Lancet* **382**, 1405-12 (2013).
6. Collaborators, U.S.B.o.D. The state of US health, 1990-2010: burden of diseases, injuries, and risk factors. *JAMA* **310**, 591-608 (2013).
7. Cummings, J.L., Vinters, H.V., Cole, G.M. & Khachaturian, Z.S. Alzheimer's disease: etiologies, pathophysiology, cognitive reserve, and treatment opportunities. *Neurology* **51**, S2-17; discussion S65-7 (1998).
8. Braak, H. & Braak, E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol* **82**, 239-59 (1991).
9. Braak, H., Alafuzoff, I., Arzberger, T., Kretzschmar, H. & Del Tredici, K. Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathol* **112**, 389-404 (2006).
10. Frisoni, G.B. et al. Imaging markers for Alzheimer disease: which vs how. *Neurology* **81**, 487-500 (2013).
11. Spies, P.E. et al. A prediction model to calculate probability of Alzheimer's disease using cerebrospinal fluid biomarkers. *Alzheimers Dement* **9**, 262-8 (2013).
12. Spies, P.E. et al. The cerebrospinal fluid amyloid beta42/40 ratio in the differentiation of Alzheimer's disease from non-Alzheimer's dementia. *Curr Alzheimer Res* **7**, 470-6 (2010).
13. Bettens, K., Sleegers, K. & Van Broeckhoven, C. Genetic insights in Alzheimer's disease. *Lancet Neurol* **12**, 92-104 (2013).
14. Bergem, A.L., Engedal, K. & Kringlen, E. The role of heredity in late-onset Alzheimer disease and vascular dementia. A twin study. *Arch Gen Psychiatry* **54**, 264-70 (1997).
15. Gatz, M. et al. Heritability for Alzheimer's disease: the study of dementia in Swedish twins. *J Gerontol A Biol Sci Med Sci* **52**, M117-25 (1997).
16. Goate, A. et al. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature* **349**, 704-6 (1991).
17. Sherrington, R. et al. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature* **375**, 754-60 (1995).
18. The structure of the presenilin 1 (S182) gene and identification of six novel mutations in early onset AD families. Alzheimer's Disease Collaborative Group. *Nat Genet* **11**, 219-22 (1995).

19. Cruts, M. et al. Molecular genetic analysis of familial early-onset Alzheimer's disease linked to chromosome 14q24.3. *Hum Mol Genet* **4**, 2363-71 (1995).
20. Levy-Lahad, E. et al. Candidate gene for the chromosome 1 familial Alzheimer's disease locus. *Science* **269**, 973-7 (1995).
21. Levy-Lahad, E. et al. A familial Alzheimer's disease locus on chromosome 1. *Science* **269**, 970-3 (1995).
22. Campion, D. et al. Early-onset autosomal dominant Alzheimer disease: prevalence, genetic heterogeneity, and mutation spectrum. *Am J Hum Genet* **65**, 664-70 (1999).
23. Kisilevsky, R. Amyloid beta threads in the fabric of Alzheimer's disease. *Nat Med* **4**, 772-3 (1998).
24. Corder, E.H. et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **261**, 921-3 (1993).
25. Rhinn, H. et al. Integrative genomics identifies APOE epsilon4 effectors in Alzheimer's disease. *Nature* **500**, 45-50 (2013).
26. Brouwers, N., Sleegers, K. & Van Broeckhoven, C. Molecular genetics of Alzheimer's disease: an update. *Ann Med* **40**, 562-83 (2008).
27. Evans, D.A. et al. Apolipoprotein E epsilon4 and incidence of Alzheimer disease in a community population of older persons. *JAMA* **277**, 822-4 (1997).
28. Hollingworth, P. et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet* **43**, 429-35 (2011).
29. 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**, 436-41 (2011).
30. Ridge, P.G., Mukherjee, S., Crane, P.K., Kauwe, J.S. & Alzheimer's Disease Genetics, C. Alzheimer's disease: analyzing the missing heritability. *PLoS One* **8**, e79771 (2013).
31. Guerreiro, R. et al. TREM2 variants in Alzheimer's disease. *N Engl J Med* **368**, 117-27 (2013).
32. Jonsson, T. et al. Variant of TREM2 associated with the risk of Alzheimer's disease. *N Engl J Med* **368**, 107-16 (2013).
33. Corrada, M.M., Brookmeyer, R., Paganini-Hill, A., Berlau, D. & Kawas, C.H. Dementia incidence continues to increase with age in the oldest old: the 90+ study. *Ann Neurol* **67**, 114-21 (2010).
34. Corrada, M.M., Berlau, D.J. & Kawas, C.H. A population-based clinicopathological study in the oldest-old: the 90+ study. *Curr Alzheimer Res* **9**, 709-17 (2012).
35. Bennett, D.A., Wilson, R.S., Boyle, P.A., Buchman, A.S. & Schneider, J.A. Relation of neuropathology to cognition in persons without cognitive impairment. *Ann Neurol* **72**, 599-609 (2012).
36. den Dunnen, W.F. et al. No disease in the brain of a 115-year-old woman. *Neurobiol Aging* **29**, 1127-32 (2008).
37. Brayne, C. & Calloway, P. Is Alzheimer's disease distinct from normal ageing? *Lancet* **2**, 514-5 (1988).

38. Brayne, C. & Calloway, P. Normal ageing, impaired cognitive function, and senile dementia of the Alzheimer's type: a continuum? *Lancet* **1**, 1265-7 (1988).
39. Kottgen, A. et al. New loci associated with kidney function and chronic kidney disease. *Nat Genet* **42**, 376-84 (2010).
40. Dupuis, J. et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* **42**, 105-16 (2010).
41. Wain, L.V. et al. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat Genet* **43**, 1005-11 (2011).
42. Benjamin, E.J. et al. Variants in ZFHX3 are associated with atrial fibrillation in individuals of European ancestry. *Nat Genet* **41**, 879-81 (2009).
43. Reilly, M.P. et al. Identification of ADAMTS7 as a novel locus for coronary atherosclerosis and association of ABO with myocardial infarction in the presence of coronary atherosclerosis: two genome-wide association studies. *Lancet* **377**, 383-92 (2011).
44. Hulstaert, F. et al. Improved discrimination of AD patients using beta-amyloid(1-42) and tau levels in CSF. *Neurology* **52**, 1555-62 (1999).
45. Dubois, B. et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurol* **6**, 734-46 (2007).
46. Scheuner, D. et al. Secreted amyloid beta-protein similar to that in the senile plaques of Alzheimer's disease is increased in vivo by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease. *Nat Med* **2**, 864-70 (1996).
47. Toledo, J.B., Shaw, L.M. & Trojanowski, J.Q. Plasma amyloid beta measurements - a desired but elusive Alzheimer's disease biomarker. *Alzheimers Res Ther* **5**, 8 (2013).
48. Hansson, O. et al. Evaluation of plasma Abeta(40) and Abeta(42) as predictors of conversion to Alzheimer's disease in patients with mild cognitive impairment. *Neurobiol Aging* **31**, 357-67 (2010).
49. Kester, M.I., Verwey, N.A., van Elk, E.J., Scheltens, P. & Blankenstein, M.A. Evaluation of plasma Abeta40 and Abeta42 as predictors of conversion to Alzheimer's disease in patients with mild cognitive impairment. *Neurobiol Aging* **31**, 539-40; author reply 541 (2010).
50. Ertekin-Taner, N. et al. Plasma amyloid beta protein is elevated in late-onset Alzheimer disease families. *Neurology* **70**, 596-606 (2008).
51. van Oijen, M., Hofman, A., Soares, H.D., Koudstaal, P.J. & Breteler, M.M. Plasma Abeta(1-40) and Abeta(1-42) and the risk of dementia: a prospective case-cohort study. *Lancet Neurol* **5**, 655-60 (2006).
52. Schmand, B., Jonker, C., Hooijer, C. & Lindeboom, J. Subjective memory complaints may announce dementia. *Neurology* **46**, 121-5 (1996).
53. van Oijen, M., de Jong, F.J., Hofman, A., Koudstaal, P.J. & Breteler, M.M. Subjective memory complaints, education, and risk of Alzheimer's disease. *Alzheimers Dement* **3**, 92-7 (2007).

54. Weinstein, G. et al. Brain imaging and cognitive predictors of stroke and Alzheimer disease in the Framingham Heart Study. *Stroke* **44**, 2787-94 (2013).
55. Kohler, S. et al. Progression to dementia in memory clinic patients without dementia: a latent profile analysis. *Neurology* **81**, 1342-9 (2013).
56. Stern, Y. Cognitive reserve in ageing and Alzheimer's disease. *Lancet Neurol* **11**, 1006-12 (2012).
57. Sleegers, K. et al. Cerebrovascular risk factors do not contribute to genetic variance of cognitive function: the ERF study. *Neurobiol Aging* **28**, 735-41 (2007).
58. Hu, M. et al. Semantic fluency and executive functions as candidate endophenotypes for the early diagnosis of schizophrenia in Han Chinese. *Neurosci Lett* **502**, 173-7 (2011).
59. Glahn, D.C. et al. Neurocognitive endophenotypes for bipolar disorder identified in multiplex multigenerational families. *Arch Gen Psychiatry* **67**, 168-77 (2010).
60. Salloway, S. et al. Two phase 3 trials of bapineuzumab in mild-to-moderate Alzheimer's disease. *N Engl J Med* **370**, 322-33 (2014).
61. Giacobini, E. & Gold, G. Alzheimer disease therapy-moving from amyloid-beta to tau. *Nat Rev Neurol* **9**, 677-86 (2013).
62. Kivipelto, M. et al. Risk score for the prediction of dementia risk in 20 years among middle aged people: a longitudinal, population-based study. *Lancet Neurol* **5**, 735-41 (2006).
63. Breteler, M.M. Vascular risk factors for Alzheimer's disease: an epidemiologic perspective. *Neurobiol Aging* **21**, 153-60 (2000).
64. van Oijen, M. et al. Atherosclerosis and risk for dementia. *Ann Neurol* **61**, 403-10 (2007).
65. Reitz, C., Bos, M.J., Hofman, A., Koudstaal, P.J. & Breteler, M.M. Prestroke cognitive performance, incident stroke, and risk of dementia: the Rotterdam Study. *Stroke* **39**, 36-41 (2008).
66. Williams, F.M. et al. Ischemic stroke is associated with the ABO locus: the EuroCLOT study. *Ann Neurol* **73**, 16-31 (2013).
67. Traylor, M. et al. Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies. *Lancet Neurol* **11**, 951-62 (2012).

# Chapter 2: Genetic Epidemiology of Alzheimer’s Disease

## *2. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease*

Nat Genet 2013 Dec; 45(12):1452-8

*The supplementary information for this paper is available at*
*http://www.nature.com/ng/journal/v45/n12/extref/ng.2802-S1.pdf*

**ABSTRACT**

Eleven susceptibility genes for late-onset Alzheimer's disease (LOAD) were identified by previous genetic studies. However, a large part of the genetic risk for this disease remains unknown. We conducted a large two-stage study based upon genome-wide association studies (GWAS) on individuals of European ancestry. In stage 1, we used genotyped and imputed data (7,055,881 single nucleotide polymorphisms (SNPs)) to meta-analyse four previously published GWAS datasets consisting of 17,008 Alzheimer's disease cases and 37,154 controls. In stage 2, 11,632 SNPs were genotyped and tested for association in an independent set of 8,572 Alzheimer's disease cases and 11,312 controls. In addition to the apolipoprotein E locus, 19 loci reached genome-wide significance (*p-value* $< 5x10^{-8}$) in the combined stage 1 and stage 2 data of which 11 are new associations with Alzheimer' disease.

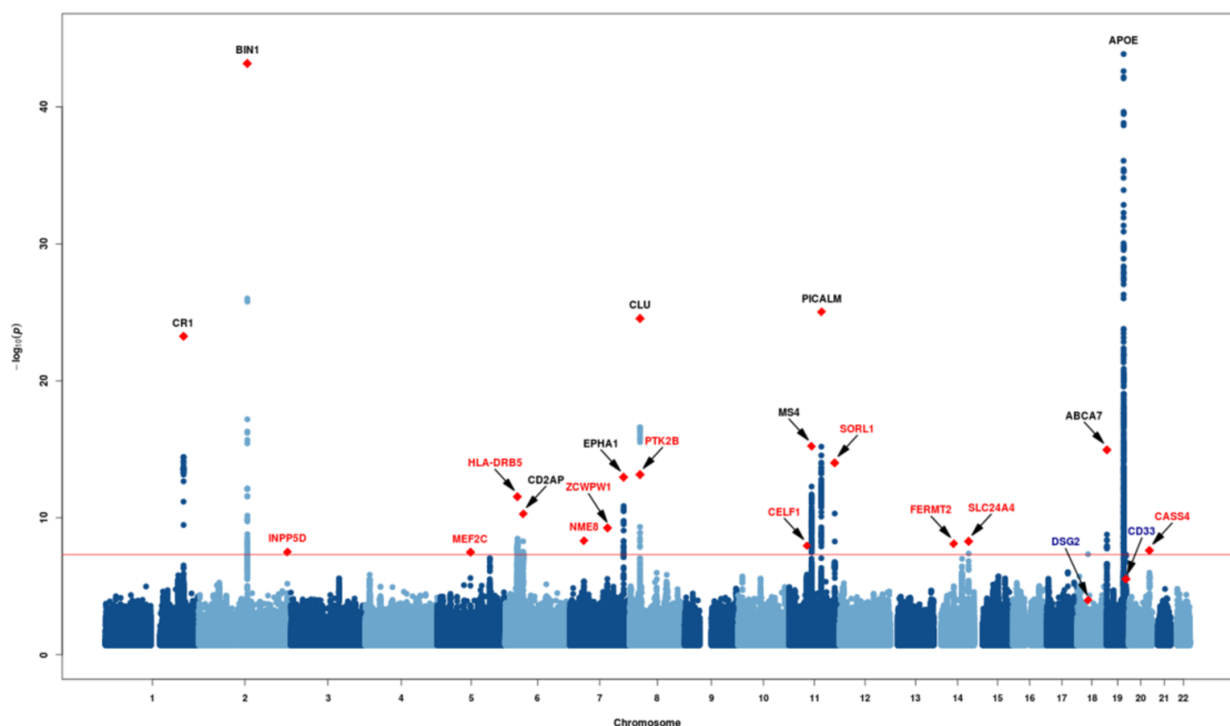
Alzheimer's disease (AD) is a devastating neurological disorder, primarily affecting the elderly. The disease manifests with progressive deterioration in cognitive functions, leading to loss of autonomy. The Apolipoprotein E (*APOE*) gene is a major genetic risk factor for AD[1,2]. Previous genome-wide association studies (GWAS) in individuals of European ancestry identified 9 other genomic regions associated with late-onset AD (LOAD)[3-7]. Recently, a rare susceptibility variant in *TREM2*, was identified[8,9]. The search for additional genetic risk factors requires large-scale meta-analysis of GWAS to increase the statistical power. Under the banner of I-GAP (International Genomics of Alzheimer's Project), we conducted a meta-analysis of four GWAS samples of European ancestry totalling 17,008 cases and 37,154 controls (stage 1), followed up by genotyping of 11,632 single nucleotide polymorphisms (SNPs) showing moderate evidence of association (stage 1, *p-value* $< 10^{-3}$) in an independent sample that included 8,572 cases and 11,312 controls (stage 2).
Stage 1 meta-analysis used data from four consortia: the Alzheimer's Disease Genetic Consortium (ADGC), the Cohorts for Heart and Ageing Research in Genomic Epidemiology (CHARGE) consortium, the European Alzheimer's Disease initiative (EADI), and the Genetic and Environmental Risk in Alzheimer's Disease (GERAD) consortium (**Table 1**, Online methods, Supplementary note and **Supplementary Table 1**). We used the European (EUR) population reference haplotype data from the 1000 Genomes project (2010 interim release based on sequence data freeze from 04 Aug 2010 and phased haplotypes from Dec 2010) to impute genotypes for as many as 11,863,202 SNPs per dataset.

**Table 1. Description of the consortium data sets used for stage 1 and stage 2**

| | | AD cases | | | Controls | | |
|---|---|---|---|---|---|---|---|
| | **Consortium** | **N** | **% women** | **Mean AAO (SD)** | **N** | **% women** | **Mean AAE** |
| Stage 1 | ADGC | 10,273 | 59.4 | 74.7 (7.7) | 10,892 | 58.6 | 76.3 (8.1) |
| | CHARGE | 1,315 | 63.6 | 82.7 (6.8) | 12,968 | 57.8 | 72.8 (8.6) |
| | EADI | 2,243 | 64.9 | 68.5 (8.9) | 6,017 | 60.7 | 74.0 (5.4) |
| | GERAD | 3,177 | 64.0 | 73.0 (8.5) | 7,277 | 51.8 | 51.0 (11.8) |
| | N | 17 008 | | | 37 154 | | |
| | **Country** | **N** | **% women** | **Mean AAO (SD)** | **N** | **% women** | **Mean AAE** |
| Stage 2 | Austria | 210 | 61.0 | 72.5 (8.1) | 829 | 43.3 | 65.5 (8.0) |
| | Belgium | 878 | 66.1 | 75.4 (8.6) | 661 | 59.5 | 65.7 (14.3) |
| | Finland | 422 | 68.0 | 71.4 (6.9) | 562 | 59.3 | 69.1 (6.2) |
| | Germany | 972 | 63.9 | 73.0 (8.6) | 2,378 | 53.1 | 69.5 (10.1) |
| | Greece | 256 | 63.3 | 69.2 (8.0) | 229 | 34.1 | 49.3 (16.4) |
| | Hungary | 125 | 68.0 | 74.9 (6.8) | 100 | 69.0 | 74.4 (6.5) |
| | Italy | 1,729 | 66.5 | 71.5 (8.7) | 720 | 55.7 | 70.0 (10.4) |
| | Spain | 2,121 | 66.3 | 75.0 (8.3) | 1,921 | 55.3 | 70.2 (10.8) |
| | Sweden | 797 | 61.7 | 76.8 (8.1) | 1,506 | 62.8 | 70.6 (8.7) |
| | UK | 490 | 57.6 | 74.6 (8.7) | 1,066 | 29.2 | 73.8 (6.5) |
| | USA | 572 | 61.9 | 83.5 (7.6) | 1,340 | 54.0 | 79.3 (6.8) |
| | N | 8 572 | | | 11 312 | | |

We excluded SNPs failing quality control in each study (**Supplementary Table 2** and Supplementary note). Our meta-analysis analyzed SNPs either genotyped or successfully imputed in at least 40% of the AD cases and 40% of the control samples across all datasets (7,055,881 SNPs, Online Methods). In each dataset, genotype dosages were analysed as described in Supplementary Note (**Supplementary Table 2**). Results were meta-analysed after applying genomic control correction to each study. The genomic control inflation factor for the meta-analysis was 1.087 for the full set of SNPs and 1.082 after withdrawing SNPs within the *APOE* locus (chr19:44,000,000-47,000,000) and within 500 kb on either side of SNPs associated with AD at a pre-specified level of genome-wide significance (*p-value* < $5x10^{-8}$) in stage 1 (see **Supplementary Figure 1** for the Quantile-Quantile plots).
In addition to the *APOE* locus, fourteen genomic regions reached genome-wide significance level (**Figure 1**). Nine had been previously identified by GWAS as genetic susceptibility factors[3-7] and five (*HLA-DRB5/HLA-DRB1, PTK2B, SORL1, SLC24A4/RIN3* and *DSG2*) represent novel loci (**Table 2**). *SORL1* had previously been identified as an AD gene through candidate gene approaches and in a GWAS combining ADGC and Asian samples[10].
Genes attributed to a signal were those closest to the most significant SNP. However, we are aware that they are potentially not the causative genes. Detailed results for each region are in **Supplementary Fig. 2-7.**

**Figure 1. Manhattan plot of stage 1 for genome-wide association with AD (17,008 cases and 37,154 controls)**




*The threshold for genome-wide significance (P<5.10$^{-8}$) is indicated by the red line. The previously GWAS-defined genes are indicated in black and the new ones are indicated in red. Red diamond symbols are the P-value SNP yielding the smallest P-value for the overall analysis.*

In stage 2, we selected for genotyping all Stage 1 SNPs with a *P*-value less than $10^{-3}$ excluding SNPs flanking *APOE* (chr19:44,000,000-47,000,000) (n=19,532, see database access URL). From the initial set of SNPs, 14,445 could be genotyped using the ILLUMINA iSelect technology. After quality control procedures (Online Methods), we considered 11,632 SNPs for association analysis. The stage 2 sample included 8,572 cases and 11,312 controls of European ancestry originating from Austria, Belgium, Finland, Germany, Greece, Hungary, Italy, Spain, Sweden, the UK and the USA (Table 1 and Supplementary Note). We observed 116 SNPs showing the same risk allele and direction of association in stages 1 and 2 and significantly associated with AD risk in stage 2 after a strict Bonferroni correction for multiple testing (*p-value* < $4.3 \times 10^{-6}$). Among these 116 SNPs, 80 were initially genome-wide significantly associated with AD risk in stage 1. Additionally, in stage 2 analyses, 2,562 SNPs were associated with AD at the nominal level (*p-value* < 0.05) and with the same risk allele and direction of association.
The results from stage 1, stage 2 and from the combined stage 1 and stage 2 datasets, which represent a secondary discovery effort, are shown in **Table 2.** With the exception of *CD33* and *DSG2*, we nominally replicated all loci that surpassed a genome-wide significance level in the stage 1. Failure to replicate *DSG2* is not surprising since evidence for this locus was from a single SNP, and was not supported by surrounding SNPs in LD (r2>0.8, **Supplementary Fig. 7b**). Moreover, seven new loci reached genome-wide significance level in the combined analysis (**Table 2**). **Supplementary Figs. 8-11** provide more detailed results for these 7 novel LOAD loci. There was no significant heterogeneity across studies at any of the loci except *DSG2* (**Table 2**) (**Supplementary Figs. 12-16**). To identify the potential causative genes, we also examined all SNPs with a P-value $<5 \times 10^{-8}$ that were within 500kb of the top SNP at each locus to identify cis-eQTL associations (**Online Methods** and **Supplementary Table 3**).
The results from the combined Stage 1 and Stage 2 datasets also indicated 13 suggestive loci ($P<10^{-6}$) (**Supplementary Table 4**). Among these, we detected a signal with rs9381040 ($P=6.3 \times 10^{-7}$), approximately 5.5kb from the 3’ end of *TREML2* and 24kb from the 5’ end of *TREM2*. This later gene was recently reported to carry a rare variant (R47H) associated with a 3-4 fold increased risk of developing AD[8,9]. We were not able to assess whether this rare variant might explain the GWAS signal we observed. This region also reached genome-wide significance in a study of cerebral spinal fluid phospho-tau levels, a biomarker for AD[11]. Beyond the already known GWAS-defined genes (*ABCA7, BIN1, CD33, CLU, CR1, CD2AP, EPHA1, MS4A4, PICALM*), the most significant new association is in the *HLA-DRB5/DRB1* region (major histocompatibility complex, class II, DR beta 5 or DR beta 1 respectively). This region is associated with immunocompetence and histocompatibility and interestingly with the risk of both multiple sclerosis and Parkinson disease[12,13]. Due to the complex genetic organisation of the HLA region on chromosome 6, we were unable to define which gene(s) is (are) responsible for this signal (**Supplementary Fig. 6a**). The second strongest signal is within the *SORL1* (sortilin-related receptor, L(DLR class) 1) gene. Our data clearly

**Table 2. Summary of stage 1, stage 2 and overall meta-analyses for SNPs reaching genome-wide significance after stages 1 and 2**

| SNP[1] | Chr. | Position[2] | Closest gene[3] | Major/Minor alleles | MAF[4] | Stage 1 | | Stage 2 | | Overall | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | OR[5] (95% CI) | Meta P-value | OR[5] (95% CI) | Meta P-value | OR[5] (95% CI) | Meta P-value | $I^2$ (%), P-value[6] |
| *Known GWAS-defined genes* | | | | | | | | | | | | |
| **rs6656401** | 1 | 207,692,049 | *CR1* | G/A | 0.197 | 1.17 (1.12-1.22) | $7.7x10^{-15}$ | 1.21 (1.14-1.28) | $7.9x10^{-11}$ | 1.18 (1.14-1.22) | $5.7x10^{-24}$ | *0*, $7.8x10^{-1}$ |
| **rs6733839** | 2 | 127,892,810 | *BIN1* | C/T | 0.409 | 1.21 (1.17-1.25) | $1.7x10^{-26}$ | 1.24 (1.18-1.29) | $3.4x10^{-19}$ | 1.22 (1.18-1.25) | $6.9x10^{-44}$ | *28*, $6.1x10^{-2}$ |
| **rs10948363** | 6 | 47,487,762 | *CD2AP* | A/G | 0.266 | 1.10 (1.07-1.14) | $3.1x10^{-8}$ | 1.09 (1.04-1.15) | $4.1x10^{-4}$ | 1.10 (1.07-1.13) | $5.2x10^{-11}$ | *0*, $9x10^{-1}$ |
| **rs11771145** | 7 | 143,110,762 | *EPHA1* | G/A | 0.338 | 0.90 (0.87-0.93) | $8.8x10^{-10}$ | 0.90 (0.86-0.95) | $2.8x10^{-5}$ | 0.90 (0.88-0.93) | $1.1x10^{-13}$ | *14*, $2.4x10^{-1}$ |
| **rs9331896** | 8 | 27,467,686 | *CLU* | T/C | 0.379 | 0.86 (0.84-0.89) | $9.6x10^{-17}$ | 0.86 (0.82-0.90) | $4.5x10^{-10}$ | 0.86 (0.84-0.89) | $2.8x10^{-25}$ | *0*, $4.9x10^{-1}$ |
| **rs983392** | 11 | 59,923,508 | *MS4A6A* | A/G | 0.403 | 0.90 (0.87-0.93) | $2.8x10^{-11}$ | 0.90 (0.86-0.94) | $4.5x10^{-6}$ | 0.90 (0.87-0.92) | $6.1x10^{-16}$ | *1*, $4.5x10^{-1}$ |
| **rs10792832** | 11 | 85,867,875 | *PICALM* | G/A | 0.358 | 0.88 (0.85-0.91) | $6.5x10^{-16}$ | 0.85 (0.81-0.89) | $1.1x10^{-11}$ | 0.87 (0.85-0.89) | $9.3x10^{-26}$ | *0*, $9.8x10^{-1}$ |
| **rs4147929** | 19 | 1,063,443 | *ABCA7* | G/A | 0.190 | 1.14 (1.10-1.20) | $1.7x10^{-9}$ | 1.17 (1.10-1.24) | $9.9x10^{-8}$ | 1.15 (1.11-1.19) | $1.1x10^{-15}$ | *0*, $9.4x10^{-1}$ |
| ***rs3865444*** | *19* | *51,727,962* | *CD33* | *C/A* | *0.307* | *0.91 (0.88-0.94)* | $5.1x10^{-8}$ | *0.99 (0.94-1.04)* | $6.9x10^{-1}$ | *0.94 (0.91-0.96)* | $3.0x10^{-6}$ | *0*, $6.9x10^{-1}$ |
| *New loci reaching genome-wide significant level in the discovery analysis* | | | | | | | | | | | | |
| **rs9271192** | 6 | 32,578,530 | *HLA-DRB5/HLA-DRB1* | A/C | 0.276 | 1.11 (1.07-1.16) | $1.6x10^{-8}$ | 1.12 (1.06-1.18) | $4.2x10^{-5}$ | 1.11 (1.08-1.15) | $2.9x10^{-12}$ | *0*, $5.4x10^{-1}$ |
| **rs28834970** | 8 | 27,195,121 | *PTK2B* | T/C | 0.366 | 1.10 (1.07-1.14) | $3.3x10^{-9}$ | 1.11 (1.06-1.17) | $4.3x10^{-6}$ | 1.10 (1.08-1.13) | $7.4x10^{-14}$ | *10*, $3.0x10^{-1}$ |
| **rs11218343** | 11 | 121,435,587 | *SORL1* | T/C | 0.039 | 0.76 (0.70-0.83) | $5.0x10^{-11}$ | 0.78 (0.70-0.88) | $4.0x10^{-5}$ | 0.77 (0.72-0.82) | $9.7x10^{-15}$ | *0*, $8.3x10^{-1}$ |
| **rs10498633** | 14 | 92,926,952 | *SLC24A4/RIN3* | G/T | 0.217 | 0.90 (0.87-0.94) | $1.5x10^{-7}$ | 0.93 (0.88-0.98) | $7.8x10^{-3}$ | 0.91 (0.88-0.94) | $5.5x10^{-9}$ | *0*, $6.3x10^{-1}$ |
| ***rs8093731*** | *18* | *29,088,958* | *DSG2* | *C/T* | *0.017* | *0.54 (0.43-0.67)* | $4.6x10^{-8}$ | *1.01 (0.80-1.28)* | $9.0x10^{-1}$ | *0.73 (0.62-0.86)* | $1.0x10^{-4}$ | *38*, $3.9x10^{-2}$ |
| *New loci reaching genome-wide significant level in the combined discovery and replication analysis* | | | | | | | | | | | | |
| **rs35349669** | 2 | 234,068,476 | *INPP5D* | C/T | 0.488 | 1.07 (1.03-1.10) | $9.6x10^{-5}$ | 1.10 (1.05-1.15) | $5.7x10^{-5}$ | 1.08 (1.05-1.11) | $3.2x10^{-8}$ | *0*, $8.0x10^{-1}$ |
| **rs190982** | 5 | 88,223,420 | *MEF2C* | A/G | 0.408 | 0.92 (0.89-0.95) | $2.5x10^{-6}$ | 0.93 (0.89-0.98) | $3.4x10^{-3}$ | 0.93 (0.90-0.95) | $3.2x10^{-8}$ | *0*, $6.4x10^{-1}$ |
| **rs2718058** | 7 | 37,841,534 | *NME8* | A/G | 0.373 | 0.93 (0.90-0.96) | $1.3x10^{-5}$ | 0.91 (0.87-0.95) | $6.3x10^{-5}$ | 0.93 (0.90-0.95) | $4.8x10^{-9}$ | *0*, $9.2x10^{-1}$ |
| **rs1476679** | 7 | 100,004,446 | *ZCWPW1* | T/C | 0.287 | 0.92 (0.89-0.96) | $7.4x10^{-6}$ | 0.89 (0.85-0.94) | $9.7x10^{-6}$ | 0.91 (0.89-0.94) | $5.6x10^{-10}$ | *0*, $7.0x10^{-1}$ |
| **rs10838725** | 11 | 47,557,871 | *CELF1* | T/C | 0.316 | 1.08 (1.04-1.11) | $6.7x10^{-6}$ | 1.09 (1.04-1.14) | $4.1x10^{-4}$ | 1.08 (1.05-1.11) | $1.1x10^{-8}$ | *0*, $7.6x10^{-1}$ |
| **rs17125944** | 14 | 53,400,629 | *FERMT2* | T/C | 0.092 | 1.13 (1.07-1.19) | $1.0x10^{-5}$ | 1.17 (1.08-1.26) | $1.6x10^{-4}$ | 1.14 (1.09-1.19) | $7.9x10^{-9}$ | 10, $3.0x10^{-1}$ |
| **rs7274581** | 20 | 55,018,260 | *CASS4* | T/C | 0.083 | 0.87 (0.82-0.92) | $1.6x10^{-6}$ | 0.89 (0.82-0.96) | $4.1x10^{-3}$ | 0.88 (0.84-0.92) | $2.5x10^{-8}$ | *0*, $9.9x10^{-1}$ |

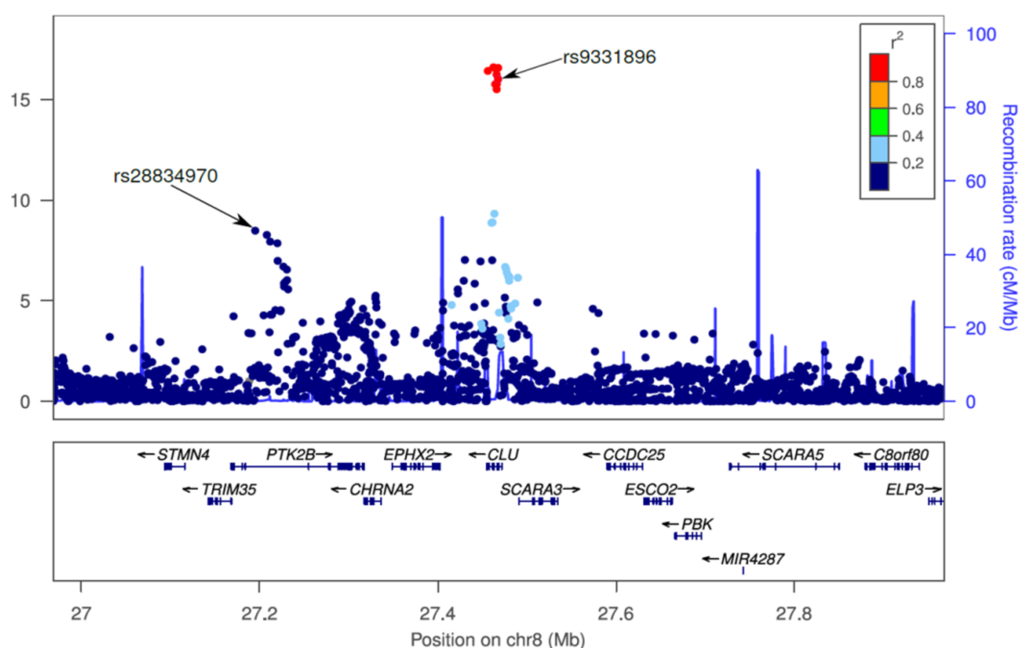
*[1] are presented the SNPs showing the best level of association after meta-analysis of Stages 1 and 2; [2] Build 37, Assembly Hg19; [3] +/- 100kb; [4] Average in the discovery sample; [5] Calculated with respect to the minor allele; [6] Cochran's Q test; rs italics indicate loci not replicated in Stage 2*

demonstrated that this gene was genome-wide significant in European samples. *SORL1* is interesting as it is be associated with an increased risk in both autosomal dominant and sporadic forms of AD[14,15] and the first LOAD gene that directly connects aberrant trafficking and metabolism of the amyloid precursor protein (APP) to LOAD[14].

The third locus, *PTK2B* (protein tyrosine kinase 2 beta), is only approximately 130 kb away from *CLU*, but we believe the two signals are independent because: (i) the two most strongly associated SNPs within each of these two genes are not in LD (D'=0.06 and $r^2$=0.003 as computed using 1000 Genomes data), (ii) a recombination peak exists between the two loci (**Figure 2**); (iii) conditional analysis in the stage 2 data confirmed the independence of the *PTK2B* association (**Supplementary Table 5** and **Supplementary Fig. 17**). The protein encoded by *PTK2B* may be an intermediate between neuropeptide-activated receptors or neurotransmitters that increase calcium flux and the downstream signals regulating neuronal activity such as MAP (mitogen activated proteins) kinase signalling[16]. *PTK2B* is involved in induction of long-term potentiation in the hippocampal CA1 (cornu ammonis 1) region, a central process in the formation of memory[17]. We cannot however exclude the possibility that there are multiple signals in the *PTK2B-CLU* region that are functionally connected to a single gene. For instance, two genome-wide significant SNPs in the *PTK2B-CLU* region are eQTLs for the gene *DPYSL2* that has been implicated in AD[18] (**Supplementary Table 3**).

The fourth locus is *SLC24A4* (solute carrier family 24 (sodium/potassium/calcium exchanger), member 4) which encodes a protein involved in iris development and hair and skin colour variation in humans as well as with the risk of developing hypertension[19,20]. *SLC24A4* is also expressed in the brain and may be involved in neural development[21]. Of note, in the vicinity of the most strongly associated SNP, is another gene called *RIN3* (Ras

**Figure 2. Regional plot for the PTK2B/CLU locus (17,008 cases and 37,154 controls).**



and Rab interactor 3) that encodes a protein that directly interacts with the *BIN1* gene product[22], a protein that may be connected to Tau pathology.[23]

In addition to these four loci reaching genome-wide significance in Stage 1, 7 new loci reached genome-wide significance in the combined analysis.

The strongest association at one of these new loci is intronic in the gene *ZCWPW1* (zinc finger, CW type with PWWP domain 1) whose corresponding protein modulates epigenetic regulation[24]. However, the region defined by all the SNPs associated with AD risk in our data is large and contains about ten genes (**Supplementary Fig. 9b**). For instance a possible causal gene in the *ZCWPW1* region is *NYAP1* whose disruption in mice affects brain size, neurite elongation and more generally, neuronal morphogenesis[25]. On a genetic basis alone, it is impossible to favour one gene over another.

A second locus is within the *CELF1* (CUGBP, Elav-like family member 1) gene, a member of a protein family that regulates pre-mRNA alternative splicing[26]. As for the *ZCWPW1* locus, the region of interest is large and contains about ten genes (**Supplementary Fig. 10a**). Among these genes, is *MADD* (MAP-kinase activating death domain), reduced expression of which may affect long-term neuronal viability in AD[27].

A discrete signal was observed adjacent to *NME8* (NME/NM23 family member 8) which is responsible for primary ciliary dyskinesia type 6[28].

The *FERMT2* (fermitin family member 2) gene is expressed in the brain. Its corresponding protein localizes to cell-matrix adhesion structures, activates integrins, is involved in orchestration of actin assembly/cell shape modulation and is an important mediator of angiogenesis[29]. An association between the fly ortholog of *FERMT2* (Fit 1/2) and Tau toxicity was recently described.[30]

We identified a fifth signal on chromosome 20 at *CASS4* (Cas scaffolding protein family member 4). Little is known about the function of the encoded protein. However, the drosophila CASS family ortholog (p130Cas) binds to CMS, the drosophila ortholog of *CD2AP* (CMS), a known AD gene (**Table 2**) involved in actin dynamics[31].

Another locus is at *INPP5D* (inositol polyphosphate-5-phosphatase, 145kDa) on chromosome 2. *INPP5D* is expressed at low levels in the brain but this protein has been shown to interact with *CD2AP,* one of the previously GWAS identified AD genes[32] and to modulate along with *GRB2,* metabolism of APP[33].

We identified a seventh signal adjacent to *MEF2C* (myocyte enhancer factor 2). Mutations at this locus are associated with severe mental retardation, stereotypic movements, epilepsy, and cerebral malformation[34]. The protein MEF2C limits excessive synapse formation during activity-dependent refinement of synaptic connectivity and thus may facilitate hippocampal-dependent learning and memory[35].

In summary, our AD GWAS meta-analysis has identified eleven new susceptibility loci in or near plausible candidate genes in addition to the already known *ABCA7, APOE, BIN1, CLU, CR1, CD2AP, EPHA1, MS4A4* and *PICALM* genes. However, we were not able to replicate the association of *CD33* in our Stage 2 analysis (P=0.61). We did not detect any biases in terms of imputation in our discovery datasets or genotyping in our replication datasets (data not

shown), suggesting a potential statistical fluctuation across our populations as an explanation for the lack of replication. However, recent data suggest that genetically determined decreased *CD33* expression might decrease AD risk and interfere with Aβ peptide clearance[36], a dysfunction thought to be central in late-onset forms of AD[37]. Further investigations in independent case-control studies will thus be required to confirm or refute the association of *CD33* with AD.

The novel genes reinforce the importance of some previously suspected pathways such as APP (*SORL1*, *CASS4*) and Tau (*CASS4*, *FERMT2*) pathology. Several of these genes are involved in pathways already shown to be enriched for association signal in AD GWAS[38,39], such as immune response and inflammation (*HLA-DRB5/DRB1*, *INPP5D*, *MEF2C*), which is also supported by the described association of AD with *CR1*[3] and *TREM2*[8,9], cell migration (*PTK2B*) and lipid transport and endocytosis (*SORL1*). Our results also suggest the existence of novel pathways underlying AD. These novel pathways could include hippocampal synaptic function (*MEF2C, PTK2B*), cytoskeletal function and axonal transport (*CELF1, NME8, CASS4*), regulation of gene expression and post-translational modification of proteins, and microglial and myeloid cell function (*INPP5D*).

Examining the amount of genetic effect attributable to all the candidate genes, the most strongly associated SNPs at each locus other than *APOE* demonstrated population attributable fractions (PAFs) or preventive fractions between 1.0%–8.0% in the stage 2 sample (**Supplementary Table 6**). Strong efforts in sequencing and post-GWAS analyses will be now required to fully characterise the functional variants responsible for the association of these GWAS-defined genes with AD risk and to understand their exact roles in the pathophysiology of AD[40].

## REFERENCES


1 Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, Pericak-Vance MA. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 261, 921-3 (1993).

2 Genin, E. et al. APOE and Alzheimer disease: a major gene with semi-dominant inheritance. *Mol Psychiatry* 16, 903-7 (2012).

3 Lambert, J.-C. et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet* 41, 1094-9 (2009).

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

5 Seshadri, S. et al. Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA* 303, 1832-40 (2010).

6 Hollingworth, P. et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet* 43, 429-35 (2011).

7 Naj, AC. et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet* 43, 436-41 (2011).

8 Guerreiro, R. et al. TREM2 Variants in Alzheimer's Disease. *N Engl J Med*, 368, 117-27 (2013).

9 Jonsson, T. et al. Variant of TREM2 Associated with the Risk of Alzheimer's Disease. *N Engl J Med* 368, 107-16 (2013).

10 Miyashita, A., et al. SORL1 is genetically associated with late-onset Alzheimer's disease in Japanese, Koreans and Caucasians. *PloS One* 8, e58618 (2013).

11 Cruchaga, C. et al. GWAS of Cerebrospinal Fluid Tau Levels Identifies Risk Variants for Alzheimer's Disease. *Neuron* 78, 256-68 (2013).

12 Sawcer, S. et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 476, 214-9 (2011).

13 Nalls, M.A. et al. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet* 377, 641-9 (2011).

14 Rogaeva, E. et al. The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. *Nat Genet* 39, 168-77 (2007).

15 Pottier, C. et al. High frequency of potentially pathogenic SORL1 mutations in autosomal dominant early-onset Alzheimer disease. *Mol Psychiatry* 17, 875-9 (2012).

16 Pandey, P. et al. Activation of p38 mitogen-activated protein kinase by PYK2/related adhesion focal tyrosine kinase-dependent mechanism. *J Biol Chem* 274, 10140-4 (1999).

17 Huang, Y. et al. CAKbeta/Pyk2 kinase is a signaling link for induction of long-term potentiation in CA1 hippocampus. *Neuron* 29, 485-96 (2001).

18 Williamson R, van Aalten L, Mann DM, Platt B, Plattner F, Bedford L, Mayer J, Howlett D, Usardi A, Sutherland C, Cole AR. CRMP2 hyperphosphorylation is characteristic of

Alzheimer's disease and not a feature common to other neurodegenerative diseases. *J Alzheimers Dis* 27, 615-25 (2011).
19 Sulem, P. et al. Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat Genet* 39, 1443-52 (2007).
20 Adeyemo, A. et al. A genome-wide association study of hypertension and blood pressure in African Americans. *PLoS Genet* 5, e1000564 (2009).
21 Larsson, M. et al. GWAS findings for human iris patterns: associations with variants in genes that influence normal neuronal pattern development. *Am J Hum Genet*. 89, 334-43 (2011).
22 Kajiho, H., Saito, K., Tsujita, K., Kontani, K., Araki, Y., Kurosu, H., Katada, T. RIN3: a novel Rab5 GEF interacting with amphiphysin II involved in the early endocytic pathway. *J Cell Sci* 116, 4159-68 (2003).
23 Chapuis, J. et al. Increased expression of BIN1 mediates Alzheimer genetic risk by modulating tau pathology. *Mol Psychiatry*. doi: 10.1038/mp.2013.1 (2013)
24 He, F., et al. Structural insight into the zinc finger CW domain as a histone modification reader. *Structure* 18, 1127-39 (2010).
25 Yokoyama, K. et al. NYAP: a phosphoprotein family that links PI3K to WAVE1 signalling in neurons. *EMBO J* 30, 4739-54 (2011).
26 Gallo, J.M., Spickett, C. The role of CELF proteins in neurological disorders. *RNA Biol* 7, 474-9 (2010).
27 Del Villar, K., Miller, C.A. Down-regulation of DENN/MADD, a TNF receptor binding protein, correlates with neuronal cell death in Alzheimer's disease brain and hippocampal neurons. *Proc Natl Acad Sci U S A* 101, 4210-5 (2004).
28 Duriez, B. et al. A common variant in combination with a nonsense mutation in a member of the thioredoxin family causes primary ciliary dyskinesia. *Proc Natl Acad Sci U S A* 104, 3336-41 (2007).
29 Pluskota, E. et al. The integrin coactivator kindlin-2 plays a critical role in angiogenesis in mice and zebrafish. *Blood* 117, 4978-87 (2011).
30 Shulman JM, Imboywa S, Diamond AE, Chipendo P, De Jager PL, Feany MB. Integrating human and fly genetics to understand Alzheimer's disease susceptibility. Poster presentation at the American Neurological Association (ANA), 137th Annual Meeting in Boston, MA. (Oct. 7-9, 2012).
31 Kirsch, K.H., Georgescu, M.M., Ishimaru, S., Hanafusa, H. CMS: an adapter molecule involved in cytoskeletal rearrangements. *Proc Natl Acad Sci U S A* 25;96, 6211-6 (1999).
32 Bao, M. et al. CD2AP/SHIP1 complex positively regulates plasmacytoid dendritic cell receptor signaling by inhibiting the E3 ubiquitin ligase Cbl. *J Immunol* 189, 786-92 (2012).
33 Brauer, H., Strauss, J., Wegner, W., Müller-Tidow, C., Horstmann, M., Jücker, M. Leukemia-associated mutations in SHIP1 inhibit its enzymatic activity, interaction with

the GM-CSF receptor and Grb2, and its ability to inactivate PI3K/AKT signaling. *Cell Signal*, 24, 2095-101 (2012).

34 Le Meur, N. *et al*. MEF2C haploinsufficiency caused by either microdeletion of the 5q14.3 region or mutation is responsible for severe mental retardation with stereotypic movements, epilepsy and/or cerebral malformations. *J Med Genet* 47, 22-9 (2010).

35 Akhtar, M.W. et al. In vivo analysis of MEF2 transcription factors in synapse regulation and neuronal survival. *PLoS One* 7, e34863 (2012).

36 Griciuc, A. et al. Alzheimer's Disease Risk Gene CD33 Inhibits Microglial Uptake of Amyloid Beta. *Neuron* 78, 631–643 (2013)

37 Lambert, J.C. & Amouyel, P. Genetics of Alzheimer's disease: new evidences for an old hypothesis? *Curr Opin Genet Dev* 21, 295-301 (2010).

38. Lambert, J.-C. et al. Implication of the immune system in Alzheimer's disease: evidence from genome-wide pathway analysis. *J Alzheimers Dis* 20, 1107-18 (2010).

39. Jones, L. et al. Genetic evidence implicates the immune system and cholesterol metabolism in the aetiology of Alzheimer's disease. *PLoS One* 5, e13950 (2010).

40. Bettens, K., Sleegers, K. & Van Broeckhoven C. Genetic insights in Alzheimer's disease. *Lancet Neurol* 12, 92-104 (2013).

# Chapter 3: Genetic Epidemiology of Cognitive Functioning

## *3.1. Genome-wide linkage and fine-mapping with exonic variants identify rare deleterious mutations affecting cognitive functioning*

Submitted

*The supplemental information for this chapter can be found in Chapter 7.1*

**ABSTRACT**

Cognitive function is a complex trait which involves multiple environmental and genetic factors. To localize genes involved in cognitive functioning, we conducted genome-wide linkage analyses in a large family from a genetically isolated population. A broad range of cognitive test measurements were available for 2882 participants. We performed non-parametric linkage analysis in participants with low cognitive test scores, defined as a score in the lowest 10% of the distribution. Genome-wide significant and suggestive thresholds for linkage were estimated empirically using simulation study. We sought for causative variants under the linkage peaks by performing single association analyses on exome sequencing and exome chip data among the same individuals, for quantitative test performance.We found significant linkage (LOD > 4.1) of cognition to chromosomes 1p13.1, 12q24.33, 19q13.43, 20p13, 21q22.13 and 21q22.3. Fine-mapping analyses under the linkage peaks revealed associations in *TRIB3*, a known gene for cognitive function, on chromosome 20, two associations in *TRIM28* and *MZF1* on chromosome 19, and in *PDE4DIP* on chromosome 1. All mutations identified have a predicted deleterious effect and are implicated in developmental or degenerative processes in the brain. Additional functional studies are needed.

**INTRODUCTION**

Cognitive function is a broad concept referring to multiple cognitive domains, among which memory, language, executive function and visuospatial ability. Impairment of cognitive function is seen in patients with various diseases including dementia, bipolar disorder, schizophrenia and attention deficit hyperactivity disorder (ADHD).[1-3] Although cognitive function has a large heritable component estimated at 20-79%[4,5], the genetic variation driving this heritability is still elusive. Many candidate genes have been studied,[6,7] and the most consistently reported genetic factor affecting cognition is apolipoprotein E (*APOE,* MIM 107741), the major genetic determinant of early and late-onset Alzheimer's disease.[8,9] The effect of *APOE* on cognitive function in healthy persons is small, especially compared to its effect in Alzheimer's disease (MIM 104300).[10] In search for unknown genes, genome wide association studies (GWAS) as well as linkage studies have been conducted. To our knowledge there are currently 7 GWAS studies published on cognitive traits among adults[11-17]. In these studies, associations were found of memory with the calsynthenin-2 (*CLSTN2*, MIM 611323) gene on chromosome 3, with the WW-and-C2-domain-containing-1 (*KIBRA,* MIM 610533) gene on chromosome 5 and with the sodium-channel-voltage-gated-type I-alpha subunit (*SCN1A*, MIM 182389) gene on chromosome 2. Further, abstract reasoning was associated with the sortilin-related-receptor (*SORL1*, MIM 602005) gene located on chromosome 11. Processing speed was suggestively associated with several loci, of which the tribbles pseudokinase 3 (*TRIB3*, MIM 607898) gene was the most important and biologically of most interest. As for executive function, Leblanc et al. identified a genome-wide significant association of a single nucleotide polymorphism (SNP) in the WD repeat-containing protein 72 (*WDR72*, MIM 613214) gene (chromosome 15) to the interference part of KEFS-D, a test similar to the classic Stroop interference test.[14] Although not much is known about the function of this gene, it was associated to kidney function in a GWAS.[18] The GWAS studies on executive function conducted to date are limited by small sample size, and there were often lenient significance thresholds used. Replication of the findings has been lacking both across and within cognitive domains. A pathway analysis suggested pathways involving cell junction and focal adhesion processes.[16] A recent GWAS on educational attainment found three genome-wide significant loci: in a noncoding region on 6q16.1, near leucine rich repeat neuronal 2(*LRRN2,* MIM 605492) and near LON peptidase N-terminal domain and ring finger 2 (*LONRF2*, not in OMIM).[19]

There is evidence for linkage, suggesting the presence of genes with a relatively large effect, of memory to chromosomes 4 and 12, mental flexibility to chromosomes 5 and 11, IQ to chromosomes 2 and 6, motor timing to chromosomes 2 and 13, processing speed to chromosome 14, reading ability to chromosome 6 and 18, and sustained attention to chromosome 12 (Table 1).[17,20-30] In a family with affective disorders with linkage to chromosome 4, the carriers of the risk haplotype performed badly on verbal memory and verbal fluency tests, even if they did not have the psychiatric phenotype.[31] These linkage studies predominantly used cognitive function as an endophenotype for psychiatric

**Table 1: Previous linkage findings on cognitive traits (LOD >= 2.5) in chronological order**

| First author | Year | Family | Trait | Chromosome | LOD |
|---|---|---|---|---|---|
| Paunio | 2004 | Schizophrenia | Verbal learning, memory | 4q21 | 3.8 |
| Posthuma | 2005 | General population | Performance IQ | 2q24.1-31.1 | 4.4 |
| | | | Verbal and full scale IQ | 6p25.3-22.3 | 3.2 |
| Buyske | 2006 | Alcohol dependence | Digit Span Test | 11q25 | 3.1 |
| | | | Digit Symbol Substitution Task | 14q11 | 6.0 |
| | | | Digit Symbol Substitution Task | 14q24.2 | 3.9 |
| Dick | 2006 | Alcohol dependence | Full scale IQ | 6p | 3.3 |
| Luciano | 2006 | General population | Cambridge Contextual Reading Test, Performance IQ | 2q24-31 | 4.2, 3.7 |
| | | Adolescent twins | Arithmetic—verbal subtest | 6p | 3.3 |
| | | | Schonell reading test | 14q13-21 | 3.2 |
| | | | Arithmetic—verbal subtest | 21q22 | 3.0 |
| Singer | 2006 | Adult twins | Prospective memory | 12q22 | 2.8 |
| Seshadri | 2007 | General population | Reading Test, Native intelligence | 18p11 | 5.1 |
| Almasy | 2008 | Schizophrenia | Abstraction, mental flexibility | 5q | 3.4 |
| Rommelse | 2008 | ADHD | Motor Timing | 2q21.1 | 3.9 |
| | | | Digit span | 13q12.11 | 4.0 |
| Lien | 2010 | Schizophrenia | Continuous performance test | 12q24.32 | 3.3 |
| Frazier-Wood | 2012 | ADHD | Aggregate of intra-individual variability measures of cognitive performance | 12q24.3 | 2.9 |

outcomes including schizophrenia (MIM 181500), alcohol dependence (MIM 103780) and ADHD (MIM 143465). Linkage analysis is a powerful method to search for large genomic regions likely harbouring rare variants associated to the outcome. We aimed to provide an in-depth analysis of rare variation driving differences in cognitive performance in individuals unselected for neuropsychiatric traits, using genome-wide linkage analyses followed by fine-mapping of the linked regions with rare deleterious mutations identified using exome-sequencing and exome-chip genotyping.

## MATERIALS AND METHODS

*Study population*

The study was conducted in a genetically isolated population in the South-West of the Netherlands. Participants were part of the Erasmus Rucphen Family (ERF) study, a family based cohort embedded in a program aiming to identify genetic risk factors of complex diseases and genetic associations to complex traits. Participants are all descendants of a limited number of founders living in the 19$^{th}$ century. Extensive genealogical data spanning 23 generations and consisting of over 23,000 individuals is available for this population.[32,33] The extensive phenotyping included a neuropsychological test battery. DNA was obtained through venapuncture. Participants were asked to report their highest level of education. Education was categorized in 8 categories: 1. primary education; 2. primary education plus a higher not completed education; 3. lower vocational education; 4. lower secondary

education; 5. intermediate vocational education; 6. higher secondary education; 7. higher vocational education; 8. university training. Years of education varied from 6 years in the lowest category to at least 16 years in the highest.
All participants (*N* ~ 3,000) gave written informed consent and the study was approved by the medical ethics committee at Erasmus MC University Medical Center. For the current study, participants with a history of cerebrovascular accidents, dementia, brain tumors or other conditions that could have influenced reliable neuropsychological assessment were excluded from analysis (*N* = 80). Cognitive test data were available for 2882 participants.

*Cognitive evaluation*
The test battery included tests that were applicable over a wide age-range and for which a Dutch validated version was available.[34,35] Memory was tested with a word learning test (after Rey's Auditory Verbal Learning test)[34] consisting of five exposure trials measuring immediate memory, learning, delayed recall and recognition. Immediate memory was defined as the total number of words from a 15-item word list recalled at the first trial, learning was defined as the total number of correctly recalled words in trial 2 to 5, delayed recall was defined as the number of correctly recalled words after 20 minutes delay and recognition was defined as the sum of the number of correctly recognized and correctly rejected words.
Executive function was assessed with the time-demanding Trail Making Test (part A and B), the Stroop Colour and Word Test (card I, II and III), and with a verbal fluency test. The Trail Making Test and the Stroop Colour and Word Test are time-demanding tasks, in which subjects have to connect numbers and letters (TMT) or name correct words and colours (Stroop) as quickly as possible.[36,37] Verbal fluency was assessed with an animal naming and letter naming subtask, in which subjects had to name as many items as possible of the same category with a one minute time limit.
Visuospatial ability (VSA) was assessed with the WAIS III - Block Design Test which consisted of reproducing two-dimensional patterns using cubes that have red, white, and half-red-half-white faces.[38] The total number of correctly reproduced blocks within a time limit was used as a score.
For analyses, we used the ratios of TMT-B and TMT-A, and of the Stroop card III and card II. People with missing test scores on TMT-B due to exceeding the time limit or misunderstanding of test instructions, were given a worst possible score of 300 seconds.
To derive more general measures of cognition for analyses, we also computed three composite scores. These scores were based on z-score transformations of test values.[39] A memory composite score (zmem) was derived by taking the average of z-scores for immediate memory (WLTI), learning, delayed recall (WLTD) and recognition. An executive composite score (zexec) was derived from average of z-scores for verbal fluency, stroop-ratio and TMT-ratio. Finally, a global cognitive function score (zglob) was computed by taking the average of z-scores for all tests.

*Statistical analysis*

General descriptive statistical analyses were performed with IBM SPSS for Windows (version 20) using logistic regression or Chi2-statistics for the comparison of cases and controls.

*Linkage analysis*

For the linkage analysis, we considered the subjects in the lowest 10% of the distribution of the standardized residuals from the regression of the respective cognitive scores onto age, sex and education as "cases". All study subjects were part of one large pedigree containing 23,612 individuals spanning 23 generations. For analysis, we constructed smaller subpedigrees with a maximum bit size of 18 using the software PEDCUT [40] because of the linkage software restraints.

Samples were genotyped on the Illumina HumanHap 6k Beadchip linkage panel. Markers with a minor allele frequency greater than 5% and callrate higher than 95% were used in the analyses. Genotyping errors were checked with MERLIN and PEDCHECK.[41,42] In case of sporadic errors, inconsistent variants were set to missing. For analyses, 5,250 autosomal SNPs were available. The linkage analysis included 233 to 260 cases (subjects with the lowest 10% of scores) depending on the cognitive trait that was analyzed.

Before running linkage analysis, the data was reformatted with the software MEGA2 to derive the correct input format.[43] We performed affected-only genome-wide nonparametric linkage analyses in MERLIN [41] using a pair-wise approach of estimating IBD allele sharing [44-46].

Thresholds for genome-wide significant and suggestive findings were estimated by performing 500 genome-wide simulations on the global cognitive trait (zglob) within the complete pedigree.[47] The typed marker set was used for simulation of the number of markers and intermarker distances using GENEDROP.[48] Simulation linkage analyses were done using the same files containing allele frequencies, pedigrees and genetic model as were used in the original linkage analysis. Per simulation, the highest log of odds (LOD) was extracted and combining the 500 simulations resulted in a LOD score of 4.1 corresponding to a genome-wide type 1 error rate of 5% (significant threshold) and of 2.8 corresponding to a type 1 error rate of 50% (suggestive threshold).

*Exome sequencing*

1,336 subjects from the ERF study were sequenced "in-house" at the Center for Biomics of the Cell Biology department of the Erasmus MC, The Netherlands. These subjects were selected for having good quality phenotype information on a wide range of topics, and therefore random with regards to cognitive test scores. The sequencing was performed using the Agilent version V4 capture kit on an Illumina Hiseq2000 sequencer using the TruSeq Version 3 protocol. The sequence reads were aligned to the human genome build 19 (hg19) using BWA and the NARWHAL pipeline. [49,50] Subsequently, the aligned reads were processed further using the IndelRealigner, MarkDuplicates and TableRecalibration tools from the Genome Analysis Toolkit (GATK) [51] and Picard (http://picard.sourceforge.net) to

remove systematic biases and to recalibrate the PHRED quality scores in the alignments. After processing, genetic variants were called using the Unified Genotyper tool from the GATK. The effects of the called variants on the respective protein sequences were determined with a custom variant annotation script. For each sample, at least 4 Gigabases of sequence was aligned to the genome. Further, for comparison and prediction of the functionality of the variants, annotations were also performed using the dbNSFP[52] (database of human non-synonymous SNPs and their functional predictions) and Seattle databases available at: http://snp.gs.washington.edu/SeattleSeqAnnotation131/. These databases gave functional prediction results from five different programs including polyPhen2, MutationTaster, SIFT, MutationAssessor and LRT, apart from gene and variant annotations, as well as conservation scores (Grantham score, PhyloP). In total 1,415,9934 Single Nucleotide Variants (SNVs) were called. After removing the low quality variants (QUAL < 150) we retrieved 578,018 high-quality SNVs which were included in the analysis in this study. Of these, there were 1,743 truncating mutations and ~ 93,000 missense mutations, ~ 30,000 of these are predicted to be damaging.

*Genotyping on the Exome Array*

Study participants from the ERF cohort whose exomes were not sequenced (*N* = 1,527) were genotyped on the Illumina Infinium HumanExome BeadChip, version 1.1, which contains over 240,000 exonic variants selected from multiple sources together spanning 12,000 samples from multiple ethnicities. Calling was performed with GenomeStudio and the ZCall variant calling tool (Broad Institute) [53]. We removed subjects with a call rate < 0.95, IBS > 0.99 and heterozygote ratio > 0.60, and SNPs that were monomorphic in our sample or had a call rate < 0.95. After QC we retrieved about 70,000 polymorphic SNVs, and 1,512 subjects to be included in the analysis.

*Association analysis*

For each of the areas under the linkage peak (LOD-1) we selected exonic SNVs that had a minor allele frequency of 0.15% (~ 5 carriers) to 15%. In our genetically isolated population, a relatively rare mutation can be enhanced up to roughly three-fold due to genetic drift, meaning that a MAF of maximum 15% corresponds to a MAF of maximum 5 percent in the general population [32]. Furthermore, selected SNPs (1) were truncating, splice site or missense mutations and (2) if missense, were predicted to be (possibly) damaging to the protein function by at least three of the five prediction software consulted. For the exome-chip data we used the annotations provided by the chip's manufacturer. For the SNVs that were available in both exome chip and exome sequencing, and met the inclusion criteria, we performed association analyses in Merlin using the quantitative trait outcomes and correcting for age, sex and education.

*Bioinformatics analysis*

We consulted two bioinformatics tools to assess functionality of our main findings. The first, the genotype-tissue expression portal (GTex, www.broadinstitute.org/gtex), contains detailed tissue expression and expression quantitative trait locus (eQTL) data on many SNPs in a broad variety of human tissues obtained postmortem [54]. The other is the genenetwork tool containing co-expression data and summarizing functional prediction from gene ontology, Biocarta and the Kyoto Encyclopedia of Genes and Genomes (KEGG) (www.genenetwork.nl; University Medical Center Groningen, Department of Medical Genetics).

## RESULTS

The mean age of the cases (subjects with a score in the lowest 10 percent of the distribution) per trait ranged from 46 to 54 years (Table S1). Descriptive statistics of the samples included in association analysis are presented in the Table S2.

Table 2 shows the significant linkage regions with LOD scores greater than 4.1. Significant linkage was observed at chromosomes 1p13.1, 12q24.33, 19q13.43, 20p13, 21q22.13 and 21q22.3. An overview of LOD scores per trait for these five regions is shown in Figure 1. Regions on chromosome 12, 20 and 21 were significantly linked to at least 2 different cognitive tests and that there was a large overlap between the different cognitive domains (Figure 1). Fine-mapping in the LOD-1 interval of the significantly and suggestively linked regions with the overlapping SNVs from exome-sequence and exome-chip yielded 171 SNVs meeting our selection criteria (Table S3). Association analysis of these variants revealed significant association in the 19q13.43 of a rare *C→T* variant rs138661227 in the *TRIM28* (MIM 601742)gene with recognition (*p-value* = $3.7*10^{-5}$, $\beta$ = 1.66) (Table 3). Another *G→A* variant rs150630080 in the gene *MZF1* (MIM 194550) in the same region was associated to DART (*p-value* = $1.3*10^{-3}$, $\beta$ = 0.28) (Table 3). Figure S1 gives a more detailed view of the region, including the *APOE* gene and the *PLD3* (not in OMIM) gene recently implicated in Alzheimer's disease. In the 1p13.1 region, which was linked to several executive traits, association was observed between the composite score for executive function (zexec) and rs76199660, a damaging but relatively common *G→A* missense mutation in *PDE4DIP* (*p-value* = $3.1*10^{-4}$). Of note is that there were no homozygous individuals for the variant allele despite an excellent genotyping quality (callrate > 0.99), and good clustering of genotypes on the exome chip (Figure S2) and good read depth (mean site depth = 115x) in the sequencing data. In the 20p13 region association was observed for a *G→A* variant rs41281850 (*p-value* = 0.0056) in the *TRIB3* gene with zexec. All four variants were highly conserved (Conservation score GERP > 3) and predicted to be severely damaging (Polyphen score > 0.97) (Table 3). All other nominally associated variants are presented in the Table S4. None of the top SNPs had eQTL data available in the GTex database. Gene network database shows that *PDE4DIP* (MIM 608117) is highly expressed in several cortical areas including the prefrontal cortex (AUC = 0.97, p-value = $6*10^{-27}$) and the hippocampus (AUC =

**Table 2: Genome-wide significant (LOD>4.1) results of linkage analysis in the ERF Study**

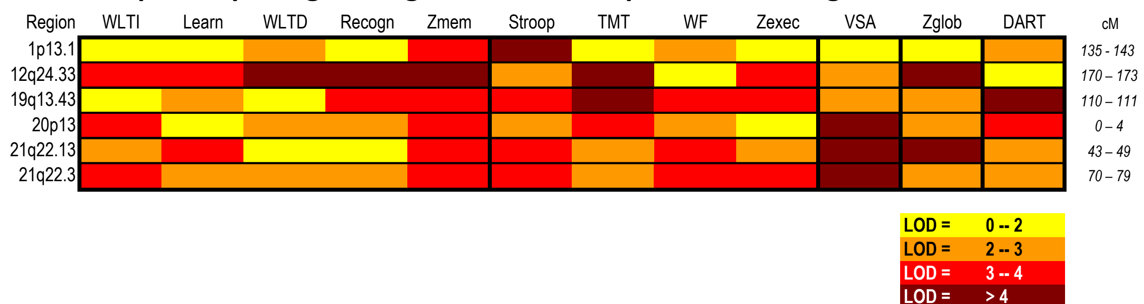
| Chromosome | Variant | cM* | Trait | Domain | Non-parametric model LOD |
|---|---|---|---|---|---|
| 1p13.1 | rs1555793 | 135 – 142 | Stroop | Executive | **4.21** |
| 12q24.33 | rs2270928 | 170 – 173 | Zglob | Global | **6.36** |
| | | | Zmem | Memory | **5.70** |
| | | | WLTD | Memory | **4.89** |
| | | | TMT | Executive | **4.37** |
| 19q13.43 | rs893186 | 110 – 111 | DART | Global | **5.26** |
| 20p13 | rs751596 | 0 – 4 | VSA | Visuospatial | 4.10 |
| 21q22.13 | rs2835629 | 43 – 49 | VSA | Visuospatial | **4.26** |
| | | | Zglob | Global | **4.14** |
| 21q22.3 | rs2256207 | 70 – 79 | VSA | Visuospatial | **4.52** |

*LOD scores in bold pass the significance threshold.*

** boundaries of the linkage peak defined as the highest (H)LOD +/- 1 (H)LOD.*

*Stroop: stroop-ratio; zglob: composite score for global cognition; zmem: composite score for memory; WLTD: AVLT delayed recall; TMT: trail making test ratio; VSA: Block Design Test; WLTI: AVLT immediate recall; WF: verbal fluency. ^Recognition for chromosome 19 is added for comparison to the association results.*

**Figure 1: Heatplot depicting the highest LOD score per trait and region.**




*WLTI: AVLT immediate recall; Learn: word learning test trial 2-5; WLTD: AVLT delayed recall; Recogn: AVLT recognition; zmem: composite score for memory; Stroop: stroop-ratio; TMT: trail making test ratio; WF: verbal fluency; zexec: composite score for executive function; VSA: Block Design Test; zglob: composite score for global cognition*

**Table 3. Association results from pooled analysis of exome chip and exome sequencing, surviving Bonferroni correction within the peak.**

Linkage and association within the same trait

| Trait | MarkerName | rsID | MAF seq | MAF chip | Mutation | Gene | Function | Effect | LOD | *p-value* | PP | GERP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| recognition | 19:59061795 | rs138661227 | 0.005 | 0.006 | p.R795C, g.2672C>T | *TRIM28* | missense | -1.66 | 3.7 | $3.7\times10^{-5}$ | 1 | 5.06 |
| DART | 19:59073740 | rs150630080 | 0.022 | 0.014 | p.T635R, g.2147G>C | *MZF1* | missense | -0.28 | 2.26 | $1.3\times10^{-3}$ | 0.97 | 4.23 |

Association on a composite score

| Trait | MarkerName | rsID | MAF seq | MAF chip | Mutation | Gene | Function | Effect | LOD | *p-value* | PP | GERP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Zexec | 1:145075775 | rs76199660 | 0.039 | 0.108 | p.P30S, g.412G>A | *PDE4DIP* | missense | -0.007 | 2.83 | $3.1\times10^{-4}$ | 1 | 3.25 |
| Zexec | 20:372076 | rs41281850 | 0.003 | 0.005 | p.S146N, g.943G>A | *TRIB3* | missense | -0.436 | 1.66 | $5.6\times10^{-3}$ | 0.998 | 5.24 |

*DART: Dutch Adult Reading Test; Zexec: composite score for executive function; MAF: minor allele frequency; Effect: effect per copy of the variant allele; LOD: log of odds for linkage; PP: polyphen2 score; GERP: GERP conservation score*

0.93, *p-value* = $2*10^{-28}$), *TRIM28* is expressed in neural stem cells (AUC = 0.89, *p-value* = $6*10^{-6}$), while *MZF1* shows expression in the putamen (AUC = 0.80; p-value = $2*10^{-5}$), entorhinal cortex and hippocampus (AUC = 0.76 for both; p-value for entorhinal cortex = $2*10^{-15}$, p-value for hippocampus = $2*10^{-10}$).

**DISCUSSION**

Our linkage analysis of cognitive function in the ERF Study yielded a total of 6 genome-wide significant regions at chromosomes 1p13.1, 12q24.33, 19q13.43, 20p13, 21q22.13 and 21q22.3. Most regions were linked to multiple cognitive domains, fitting the 'generalist gene' hypothesis indicating that there are pleiotropic pathways involved in cognitive function [5]. Fine-mapping using association in the LOD-1 intervals of these linked regions identified four highly conserved and deleterious variants. This includes rs138661227 in the *TRIM28* gene associated with recognition and rs150630080 in the gene *MZF1* associated with DART in the 19q13.43 region, rs76199660 in the *PDE4DIP* gene in the chromosome 1p13.1 and rs41281850 in the *TRIB3* gene in the 20p13 region associated with the composite score of executive function. Chromosome 1p13.1 was previously linked to Alzheimer's disease in a large linkage meta-analysis [55] but not to cognitive function. Linkage of cognitive traits to chromosome 12q24 was also observed earlier in a linkage analysis in schizophrenia affected families.[29]. The linkage peak reaching its maximum at 12q24.33 overlaps with the 12q22 region which was found to be suggestively linked to prospective memory in a previous study [27]. The top linked markers are only 5 MB apart. In our study, the region was significantly linked to delayed recall, Trailmaking and memory and overall compound scores. Luciano et al. found evidence for linkage of verbal IQ to chromosome 21q22 that we found linked to visuospatial ability.[23] The regions on chromosome 21include the so-called Down's syndrome critical region, which is an intensively studied region within the human genome and is recognized to be crucial for Alzheimer's disease as it also harbors the amyloid-precursor-protein (*APP*, MIM 104760) gene. However, *APP* is located 3 Mb from our linkage peak.[56-58]. Although no earlier linkage of cognitive traits has been described for the chromosome 20 region, the *TRIB3* gene in this region, associated to composite executive function in our study, was identified in a GWAS of processing speed [16]. *TRIB3* is a specific inducer of neuronal apoptosis in endoplasmatic reticulum stress, for example when nerve growth factor is withdrawn.[59-61] It is also important in glucose homeostasis, and variants in *TRIB3* are associated to metabolic syndrome and atherosclerosis.[62] The rare variant we identified has not been associated with any trait before. It is a coding variant, predicted to be damaging.

None of the variants identified by GWAS for cognition related phenotype (Alzheimer's Disease, *N* = 21; cognitive GWASes (unpublished), *N* = 5; educational attainment, *N* = 3; schizophrenia, *N* = 36) were in the regions of our linkage peaks [19,63-76].

Of the novel loci, in19q13 region we identified two plausible candidate genes including *TRIM28* associated with recognition and *MZF1* associated with DART. *TRIM28*, tripartite motif containing 28, also known as *KAP1* (kinesin –II-associated protein) or *KRIP-1* (KRAB-

interacting protein 1), is an essential cofactor of the KRAB-zinc finger proteins and plays a role in gene expression through regulation of gene acetylation [77,78]. It is expressed throughout the brain and especially in the hippocampus and cerebellum in mice, and mice in which *TRIM28* is deleted in the forebrain in early postnatal life, which affected expression levels mainly in the dentate gyrus and CA1 regions in the hippocampus, showed increased anxiety behavior, and impaired learning when exposed to subchronic stress [78]. The other identified gene *MZF1*, that is the myeloid zinc finger protein 1 gene (alias: Zinc Finger Protein 42, *ZNF42*) is a zinc finger protein regulating DNA transcription. It influences the expression of cathepsin B [79]. Cathepsin B, in turn, is a β-secretase involved in the production of the neurotoxic β-amyloid peptides that constitute the amyloid plaques in Alzheimer's dementia, and its knockout in mouse models improved memory deficits and β-amyloid related biomarker levels [80,81]. Moreover, *MZF1* has been identified as a possible central factor in an Alzheimer's disease pathway involving inflammatory response and cholesterol metabolism pathways [82].

Another interesting finding of our study is the relatively common missense mutation in *PDE4DIP* (rs76199660) associated with the composite executive function score. Despite a frequency of 7% in our sample we did not observe a single homozygous carrier. Of note, no homozygotes for the variant allele have been observed among 4298 European American subjects in the NHLBI Exome Sequencing Project (ESP6500) either (MAF = 0.026, Hardy Weinberg equilibrium *p-value* = 0.08). (see http://browser.1000genomes.org/Homo_sapiens/Variation/Population?db=core;r=1:145075275-145076275;v=rs76199660;vdb=variation;vf=20060420, accessed December 31, 2013) Flies homozygous for the mutation are known to die during late embryogenesis or soon after hatching [83,84]. This may suggest a complete selection against the homozygous mutant types in humans also. The variant causes a proline-to-serine amino acid change in codon 30. *PDE4DIP* shows high expression in the cerebral cortex and the *PDE4D* (MIM 600129) gene anchored by *PDE4DIP* has a very dynamic expression pattern during embryogenesis,[85] which, combined with the findings on brain development in mutant flies, suggests that *PDE4DIP* plays a key role in the early developmental period. *PDE4DIP* anchors the phosphodiesterase 4D to the Golgi/centrosome region of the cell, and thus plays a role in the cell division cycle [86]. A *PDE4DIP* paralogue, *CDK5RAP2* (MIM 608201), is known for pathogenic mutations causing autosomal recessive primary microcephaly (MIM 604804), a developmental disorder characterized by reduced head size (mainly due to impaired growth of the cortex) and mental retardation without other symptoms [83]. Also in cognitively healthy adults, the intracranial volume, a proxy for the maximal brain volume in lifetime, is related to executive performance [87]. *PDE4DIP* is an orthologue of the centrosomin (Cnn) gene in drosophila, and Cnn mutant drosophilas had reduced cell numbers in their central and peripheral nervous system, as well as impaired development in the gut. Nevertheless, PDE4D, the primary cAMP hydrolyzing enzyme in cells, is under study as a drug target for, among other diseases, Alzheimer's Disease and schizophrenia, as inhibition of PDE4D has a beneficial effect on memory [88,89].

Our study presents a comprehensive genetic dissection of cognitive traits. The linkage analysis using extreme phenotypes was designed to capture loci harboring rare variants that have large effects on these cognitive traits. The structure of the population where rare variants are enriched and the extensive availability of rare deleterious exonic variants for this sample provided us a unique opportunity to dissect each of the linked regions for causal variants. As a consequence of investigating rare variants, even in our large family cohort, we had limited power to detect association of these variants and therefore, despite strong linkage signals we could not explain the linkage at chromosomes 12q24.33, 21q22.13 and 21q22.3.

To summarize our study demonstrates evidence for significant linkage of chromosomes 1p13.1, 12q24.33, 19q13.43, 20p13, 21q22.13 and 21q22.3 to cognitive function. We confirmed previous findings, specifically the linkage region on chromosome 12q24.33. With fine-mapping we indict variants in four plausible candidate genes including *TRIB3, TRIM28, MZF1* and *PDE4DIP*. While *TRIB3* has been previously implicated in cognitive functioning, the other three genes are novel and warrant further studies.

**ACKNOWLEDGMENTS**

The ERF Study was supported by the joint grant from the Netherlands Organization for Scientific Research (NWO, 91203014), the Center of Medical Systems Biology (CMSB), Hersenstichting Nederland, Internationale Stichting Alzheimer Onderzoek (ISAO), Alzheimer Association project number 04516, Hersenstichting Nederland project number 12F04(2).76, and the Interuniversity Attraction Poles (IUAP) program.
The ERF 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). High-throughput analysis of the ERF data was supported by joint grant from Netherlands Organization for Scientific Research and the Russian Foundation for Basic Research (NWO-RFBR 047.017.043). Exome sequencing analysis in ERF was supported by the ZonMw grant (project 91111025). Exome-chip genotyping was supported by BBMRI-NL. We are grateful to all study participants and their relatives, general practitioners and neurologists for their contributions, to P. Veraart for her help in genealogy, J. Vergeer for the supervision of the laboratory work, P. Snijders for his help in data collection and L. Karssen for performing variant calling of the exome-chip.

**CONFLICTS OF INTEREST**

The authors do not have conflicts of interest.

**REFERENCES**


1. Green, M.F. Cognitive impairment and functional outcome in schizophrenia and bipolar disorder. *J Clin Psychiatry* **67 Suppl 9**, 3-8; discussion 36-42 (2006).
2. Seidman, L.J., Biederman, J., Weber, W., Hatch, M. & Faraone, S.V. Neuropsychological function in adults with attention-deficit hyperactivity disorder. *Biol Psychiatry* **44**, 260-8 (1998).
3. Huntley, J.D. & Howard, R.J. Working memory in early Alzheimer's disease: a neuropsychological review. *Int J Geriatr Psychiatry* (2009).
4. Sleegers, K. et al. Cerebrovascular risk factors do not contribute to genetic variance of cognitive function: the ERF study. *Neurobiol Aging* **28**, 735-41 (2007).
5. Butcher, L.M., Kennedy, J.K. & Plomin, R. Generalist genes and cognitive neuroscience. *Curr Opin Neurobiol* **16**, 145-51 (2006).
6. Deary, I.J., Johnson, W. & Houlihan, L.M. Genetic foundations of human intelligence. *Hum Genet* **126**, 215-32 (2009).
7. Payton, A. The Impact of Genetic Research on our Understanding of Normal Cognitive Ageing: 1995 to 2009. *Neuropsychol Rev* (2009).
8. Corder, E.H., Lannfelt, L., Bogdanovic, N., Fratiglioni, L. & Mori, H. The role of APOE polymorphisms in late-onset dementias. *Cell Mol Life Sci* **54**, 928-34 (1998).
9. van Duijn, C.M. et al. Apolipoprotein E4 allele in a population-based study of early-onset Alzheimer's disease. *Nat Genet* **7**, 74-8 (1994).
10. Wisdom, N.M., Callahan, J.L. & Hawkins, K.A. The effects of apolipoprotein E on non-impaired cognitive functioning: A meta-analysis. *Neurobiol Aging* (2009).
11. Papassotiropoulos, A. et al. A genome-wide survey of human short-term memory. *Mol Psychiatry* (2009).
12. Papassotiropoulos, A. et al. Common Kibra alleles are associated with human memory performance. *Science* **314**, 475-8 (2006).
13. Cirulli, E.T. et al. Common genetic variation and performance on standardized cognitive tests. *Eur J Hum Genet* (2010).
14. Leblanc, M. et al. Genome-wide study identifies PTPRO and WDR72 and FOXQ1-SUMO1P1 interaction associated with neurocognitive function. *J Psychiatr Res* **46**, 271-8 (2012).
15. Need, A.C. et al. A genome-wide study of common SNPs and CNVs in cognitive performance in the CANTAB. *Hum Mol Genet* **18**, 4650-61 (2009).
16. Luciano, M. et al. Whole genome association scan for genetic polymorphisms influencing information processing speed. *Biol Psychol* **86**, 193-202 (2011).
17. Seshadri, S. et al. Genetic correlates of brain aging on MRI and cognitive test measures: a genome-wide association and linkage analysis in the Framingham Study. *BMC Med Genet* **8 Suppl 1**, S15 (2007).
18. Kottgen, A. et al. New loci associated with kidney function and chronic kidney disease. *Nat Genet* **42**, 376-84 (2010).

19. Rietveld, C.A. et al. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **340**, 1467-71 (2013).
20. Rommelse, N.N. et al. Neuropsychological endophenotype approach to genome-wide linkage analysis identifies susceptibility loci for ADHD on 2q21.1 and 13q12.11. *Am J Hum Genet* **83**, 99-105 (2008).
21. Buyske, S. et al. Cognitive traits link to human chromosomal regions. *Behav Genet* **36**, 65-76 (2006).
22. Dick, D.M. et al. Linkage analyses of IQ in the collaborative study on the genetics of alcoholism (COGA) sample. *Behav Genet* **36**, 77-86 (2006).
23. Luciano, M. et al. Genome-wide scan of IQ finds significant linkage to a quantitative trait locus on 2q. *Behav Genet* **36**, 45-55 (2006).
24. Posthuma, D. et al. A genomewide scan for intelligence identifies quantitative trait loci on 2q and 6p. *Am J Hum Genet* **77**, 318-26 (2005).
25. Almasy, L. et al. A genome screen for quantitative trait loci influencing schizophrenia and neurocognitive phenotypes. *Am J Psychiatry* **165**, 1185-92 (2008).
26. Paunio, T. et al. Genome-wide scan in a nationwide study sample of schizophrenia families in Finland reveals susceptibility loci on chromosomes 2q and 5q. *Hum Mol Genet* **10**, 3037-48 (2001).
27. Singer, J.J., Falchi, M., Macgregor, A.J., Cherkas, L.F. & Spector, T.D. Genome-wide scan for prospective memory suggests linkage to chromosome 12q22. *Behav Genet* **36**, 18-28 (2006).
28. Konig, I.R. et al. Mapping for dyslexia and related cognitive trait loci provides strong evidence for further risk genes on chromosome 6p21. *Am J Med Genet B Neuropsychiatr Genet* **156B**, 36-43.
29. Lien, Y.-J. A genome-wide quantitative trait loci scan of neurocognitive performances in families with schizophrenia. *Genes Brain Behav* **9**, 695-702 (2010).
30. Frazier-Wood, A.C. et al. Neuropsychological intra-individual variability explains unique genetic variance of ADHD and shows suggestive linkage to chromosomes 12, 13, and 17. *Am J Med Genet B Neuropsychiatr Genet* **159B**, 131-40 (2012).
31. Drysdale, E., Knight, H.M., McIntosh, A.M. & Blackwood, D.H. Cognitive endophenotypes in a family with bipolar disorder with a risk locus on chromosome 4. *Bipolar Disord* **15**, 215-22 (2013).
32. Pardo, L.M., MacKay, I., Oostra, B., van Duijn, C.M. & Aulchenko, Y.S. The effect of genetic drift in a young genetically isolated population. *Ann Hum Genet* **69**, 288-95 (2005).
33. Aulchenko, Y.S. et al. Linkage disequilibrium in young genetically isolated Dutch population. *Eur J Hum Genet* **12**, 527-34 (2004).
34. Sleegers, K. et al. Familial clustering and genetic risk for dementia in a genetically isolated Dutch population. *Brain* **127**, 1641-9 (2004).
35. Liu, F. et al. The apolipoprotein E gene and its age-specific effects on cognitive function. *Neurobiol Aging* (2008).

36. Golden, C.J. Identification of brain disorders by the Stroop Color and Word Test. *J Clin Psychol* **32**, 654-8 (1976).
37. *Army Individual Test Battery Manual of directions and scoring*, (Washington, DC: United States War Department, Adjutant General's Office, 1944).
38. Wechsler, D. *WAIS-IIIUK administration and scoring manual* (London, UK, Psychological Corporation, 1998).
39. Prins, N.D. et al. Cerebral small-vessel disease and decline in information processing speed, executive function and memory. *Brain* **128**, 2034-41 (2005).
40. Liu, F., Kirichenko, A., Axenovich, T.I., van Duijn, C.M. & Aulchenko, Y.S. An approach for cutting large and complex pedigrees for linkage analysis. *European Journal of Human Genetics* **16**, 854-60 (2008).
41. Abecasis, G.R., Cherny, S.S., Cookson, W.O. & Cardon, L.R. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* **30**, 97-101 (2002).
42. O'Connell, J.R. & Weeks, D.E. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* **63**, 259-66 (1998).
43. Mukhopadhyay, N., Almasy, L., Schroeder, M., Mulvihill, W.P. & Weeks, D.E. Mega2: data-handling for facilitating genetic linkage and association analyses. *Bioinformatics* **21**, 2556-7 (2005).
44. Kong, A. & Cox, N.J. Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet* **61**, 1179-88 (1997).
45. Kruglyak, L., Daly, M.J., Reeve-Daly, M.P. & Lander, E.S. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* **58**, 1347-63 (1996).
46. Whittemore, A.S. & Halpern, J. A class of tests for linkage using affected pedigree members. *Biometrics* **50**, 118-27 (1994).
47. Amin, N. et al. Suggestive linkage of ADHD to chromosome 18q22 in a young genetically isolated Dutch population. *Eur J Hum Genet* **17**, 958-66 (2009).
48. Thomas, A. A method and program for estimating graphical models for linkage disequilibrium that scale linearly with the number of loci, and their application to gene drop simulation. *Bioinformatics* **25**, 1287-92 (2009).
49. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60 (2009).
50. Brouwer, R.W., van den Hout, M.C., Grosveld, F.G. & van Ijcken, W.F. NARWHAL, a primary analysis pipeline for NGS data. *Bioinformatics* **28**, 284-5 (2012).
51. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-303 (2010).
52. Liu, X., Jian, X. & Boerwinkle, E. dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum Mutat* **32**, 894-9 (2011).
53. Goldstein, J.I. et al. zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics* **28**, 2543-5 (2012).

54. Consortium, G.T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-5 (2013).
55. Butler, A.W. et al. Meta-analysis of linkage studies for Alzheimer's disease-A web resource. *Neurobiol Aging* (2009).
56. Van Broeckhoven, C. & Kumar-Singh, S. Genetics and pathology of alpha-secretase site AbetaPP mutations in the understanding of Alzheimer's disease. *J Alzheimers Dis* **9**, 389-98 (2006).
57. Harris, S.E. et al. A genetic association analysis of cognitive ability and cognitive ageing using 325 markers for 109 genes associated with oxidative stress or cognition. *BMC Genet* **8**, 43 (2007).
58. Chumakov, I. et al. Continuum of overlapping clones spanning the entire human chromosome 21q. *Nature* **359**, 380-7 (1992).
59. Kristiansen, M., Menghi, F., Hughes, R., Hubank, M. & Ham, J. Global analysis of gene expression in NGF-deprived sympathetic neurons identifies molecular pathways associated with cell death. *BMC Genomics* **12**, 551 (2011).
60. Zou, C.G. et al. The molecular mechanism of endoplasmic reticulum stress-induced apoptosis in PC-12 neuronal cells: the protective effect of insulin-like growth factor I. *Endocrinology* **150**, 277-85 (2009).
61. Zareen, N., Biswas, S.C. & Greene, L.A. A feed-forward loop involving Trib3, Akt and FoxO mediates death of NGF-deprived neurons. *Cell Death Differ* **20**, 1719-30 (2013).
62. Prudente, S. et al. The mammalian tribbles homolog TRIB3, glucose homeostasis, and cardiovascular diseases. *Endocr Rev* **33**, 526-46 (2012).
63. Cross-Disorder Group of the Psychiatric Genomics, C. et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* **45**, 984-94 (2013).
64. Aberg, K.A. et al. A comprehensive family-based replication study of schizophrenia genes. *JAMA Psychiatry* **70**, 573-81 (2013).
65. Betcheva, E.T. et al. Whole-genome-wide association study in the Bulgarian population reveals HHAT as schizophrenia susceptibility gene. *Psychiatr Genet* **23**, 11-9 (2013).
66. Irish Schizophrenia Genomics, C. & the Wellcome Trust Case Control, C. Genome-wide association study implicates HLA-C*01:02 as a risk factor at the major histocompatibility complex locus in schizophrenia. *Biol Psychiatry* **72**, 620-8 (2012).
67. Bergen, S.E. et al. Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder. *Mol Psychiatry* **17**, 880-6 (2012).
68. Shi, Y. et al. Common variants on 8p12 and 1q24.2 confer risk of schizophrenia. *Nat Genet* **43**, 1224-7 (2011).
69. Yue, W.H. et al. Genome-wide association study identifies a susceptibility locus for schizophrenia in Han Chinese at 11p11.2. *Nat Genet* **43**, 1228-31 (2011).

70. Schizophrenia Psychiatric Genome-Wide Association Study, C. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet* **43**, 969-76 (2011).
71. Wang, K.S., Liu, X.F. & Aragam, N. A genome-wide meta-analysis identifies novel loci associated with schizophrenia and bipolar disorder. *Schizophr Res* **124**, 192-9 (2010).
72. Huang, J. et al. Cross-disorder genomewide analysis of schizophrenia, bipolar disorder, and depression. *Am J Psychiatry* **167**, 1254-63 (2010).
73. International Schizophrenia, C. et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748-52 (2009).
74. Stefansson, H. et al. Common variants conferring risk of schizophrenia. *Nature* **460**, 744-7 (2009).
75. Jonsson, T. et al. Variant of TREM2 associated with the risk of Alzheimer's disease. *N Engl J Med* **368**, 107-16 (2013).
76. Lambert, J.C. et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* **45**, 1452-8 (2013).
77. Groner, A.C. et al. KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. *PLoS Genet* **6**, e1000869 (2010).
78. Jakobsson, J. et al. KAP1-mediated epigenetic repression in the forebrain modulates behavioral vulnerability to stress. *Neuron* **60**, 818-31 (2008).
79. Rafn, B. et al. ErbB2-driven breast cancer cell invasion depends on a complex signaling network activating myeloid zinc finger-1-dependent cathepsin B expression. *Mol Cell* **45**, 764-76 (2012).
80. Hook, V. et al. Cysteine Cathepsins in the secretory vesicle produce active peptides: Cathepsin L generates peptide neurotransmitters and cathepsin B produces beta-amyloid of Alzheimer's disease. *Biochim Biophys Acta* **1824**, 89-104 (2012).
81. Kindy, M.S. et al. Deletion of the cathepsin B gene improves memory deficits in a transgenic ALZHeimer's disease mouse model expressing AbetaPP containing the wild-type beta-secretase site sequence. *J Alzheimers Dis* **29**, 827-40 (2012).
82. Porcellini, E. et al. Haplotype of single nucleotide polymorphisms in exon 6 of the MZF-1 gene and Alzheimer's disease. *J Alzheimers Dis* **34**, 439-47 (2013).
83. Bond, J. & Woods, C.G. Cytoskeletal genes regulating brain size. *Curr Opin Cell Biol* **18**, 95-101 (2006).
84. Li, K. & Kaufman, T.C. The homeotic target gene centrosomin encodes an essential centrosomal component. *Cell* **85**, 585-96 (1996).
85. Huang, Z. et al. Dynamic expression pattern of Pde4d and its relationship with CpG methylation in the promoter during mouse embryo development. *Biochem Biophys Res Commun* (2013).
86. Verde, I. et al. Myomegalin is a novel protein of the golgi/centrosome that interacts with a cyclic nucleotide phosphodiesterase. *J Biol Chem* **276**, 11189-98 (2001).
87. Farias, S.T. et al. Maximal brain size remains an important predictor of cognition in old age, independent of current brain pathology. *Neurobiol Aging* **33**, 1758-68 (2012).

88. Burgin, A.B. et al. Design of phosphodiesterase 4D (PDE4D) allosteric modulators for enhancing cognition with improved safety. *Nat Biotechnol* **28**, 63-70 (2010).
89. Li, Y.F. et al. Phosphodiesterase-4D knock-out and RNA interference-mediated knock-down enhance memory and increase hippocampal neurogenesis via increased cAMP signaling. *J Neurosci* **31**, 172-83 (2011).

## *3.2. GWAS for executive function and processing speed suggests involvement of the CADM2 gene.*

Submitted

*Supplemental information for this paper is available in Chapter 7.2 of this thesis*

**ABSTRACT**

To identify common variants contributing to the normal variation in executive cognitive performance and speed of information processing in non-demented older individuals, we conducted a genome-wide association study (GWAS) in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. Neuropsychological testing was available for 5 429 to 32 080 subjects of European ancestry aged 45 years or older, free of dementia and clinical stroke at the time of cognitive testing from 20 cohorts in the discovery phase. We analyzed performance on the Trail Making Test parts A and B, the Letter Digit Substitution Test (LDST), the Digit Symbol Substitution Task (DSST), semantic and phonemic fluency tests, and the Stroop Color and Word Test. Replication was sought in 2 650 to 9 770 subjects from 20 independent cohorts.

Significant association was observed for the SNP rs17518584 (discovery *p-value* = $2.33 \times 10^{-8}$, replication *p-value* = 0.046) and the joint discovery and replication meta-analysis (*p-value* = $3.91 \times 10^{-9}$) in an intron of the gene *CADM2* with LDST/DSST. Rs17518584 is located about 170 kb upstream of the transcription start site of the major transcript for the *CADM2* gene, but is within an intron of a variant transcript that includes an alternative first exon. The variant is associated with expression of *CADM2* in the cingulate cortex (*p-value* = $4 \times 10^{-4}$). The protein encoded by *CADM2* is involved in glutamate signaling (*p-value* = $7.22 \times 10^{-15}$), gamma-aminobutyric acid (GABA) transport (*p-value* = $1.36 \times 10^{-11}$) and neuron cell-cell adhesion (*p-value* = $1.48 \times 10^{-13}$). Our findings suggest that genetic variation in the *CADM2* gene is associated with individual differences in processing speed measured by the LDST/DSST.

**INTRODUCTION**

Cognitive function is a broad concept referring to multiple dissociable but inter-correlated cognitive domains, which include memory, language, executive function, and visuospatial ability. Having unimpaired cognitive abilities is an important determinant of quality of life. Impairment of these functions is seen in patients with dementia, bipolar disorder, schizophrenia and attention deficit hyperactivity disorder (ADHD) [1-3]. Among the cognitive domains, processing speed is sometimes characterised as a fundamental cognitive process, reflecting the speed at which many processing operations can take place, allowing one to perform cognitive tasks[4,5].

Executive function is a second feature of all cognitive domains and addresses the level of difficulty of cognitive tasks, presumably challenging the frontal lobes, to a range of activities including response inhibition, attention, cognitive flexibility, and planning[6]. Both processing speed and executive function are in part genetically determined; heritability of performance on cognitive tests is estimated to be between 35 and 62%[7,8] for processing speed and between 16 and 63%[9,10] for executive function; the underlying cognitive processes are estimated by some to have heritability rates as high as 99%[11]. Processing speed and executive function are domains of cognitive ability that, in addition to their domain-specific variance, are in part explained by a person's general cognitive ability. The same holds true for the genetic variance of the performance on cognitive tests; some of it is attributable to domain and test-specific variation, and some of it to variation in general cognitive ability[12]. A genome-wide association study (GWAS) for human intelligence differences found that about 50% of variance in general, fluid-type intelligence can be explained by genetic factors [11,13-15]. Finding susceptibility genes for processing speed and executive function tests may provide insight into the normal variation in these important cognitive functions, and may also increase the knowledge of diseases that are associated with impairments in those cognitive domains.

Although various genes have been identified as potential candidates for a role in the different dimensions of cognitive function, analyses of associations between genetic variants and these phenotypes have yielded inconsistent results[16]. Candidate gene meta-analyses have shown associations of the apolipoprotein E (*APOE*) gene [17] and the dystrobrevin binding protein 1 (*DTNBP1)* gene [18] to general cognitive ability, although these findings do not meet the current standard of genome-wide significance (*p-value* between 0.01 and 0.05 for *APOE*, *p-value* = 0.003 for *DTNBP1*). Linkage analyses of executive function tasks found regions on chromosomes 2q, 5q, 11q, 13q and 14q [19-21]. To our knowledge there are currently 5 GWAS studies published on processing speed and executive traits among adults [5,22-25]. Processing speed was suggestively associated with several loci, of which the tribbles homolog 3 (*TRIB3*) gene was the strongest and biologically of most interest [5]. As for executive function, one study [25] identified a genome-wide significant association (*p-value* = $4.32x10^{-8}$) of a single nucleotide polymorphism (SNP) in the *WDR72* gene (chromosome 15) to the interference part of D-KEFS, a test similar to the classic Stroop interference test. The *WDR72* gene has also been associated with kidney function [26]. The studies on executive

function and processing speed conducted to date are limited by small sample size between 700 and 4 000 subjects, and were often lenient in the significance thresholds that were used. Replication of the findings has been lacking both across and within cognitive domains.
In this study we performed a large-scale meta-analysis combining GWAS from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium [27] focusing on the cognitive domains of processing speed (Trail Making test part A, hereafter referred to as Trails A, Digit-Symbol Substitution Test or Letter-Digit Substitution Task (LDST/DSST)) and executive function (Stroop test, semantic and phonemic fluency, Trail Making test part B (TrailsB)) in non-demented subjects.

## MATERIALS AND METHODS

*Study populations*

The discovery phase included 20 cohorts contributing to one or more test (*N* per test = 5 429 to 32 080) (Supplementary Table 1). The number of discovery cohorts and subjects included varied based on availability of each test. Each cohort had extensive phenotypic data on one or more traits, and genome-wide SNP data available. Details for each cohort are given in Supplementary Table 1. Subjects aged 45 and older who were free of stroke and dementia and of European ancestry were eligible for the study.
Twenty replication cohorts (Total *N* = 2 650 to 9 770) (Supplementary Table 1) were selected based upon comparable study populations and availability of genotype data, and invited to share data from a GWAS run according to the same protocols. Whereas the same inclusion criteria regarding age (≥ 45 years) and absence of stroke or dementia applied, we also included cohorts of African American ancestry (*N* = 1 004 to 3 164 depending on the trait) in the replication phase, partly to evaluate whether findings extrapolated to those of other ethnicities. Data from the discovery and replication studies were further meta-analyzed. Additional details are provided in the Supplement. Each participant provided informed consent and all studies were approved by their local Institutional Review Boards.

*Executive function and speed of processing tests*

Each cohort included some executive function and/or processing speed tests. The test batteries differed substantially across cohorts. Neurocognitive tests included the Stroop test card 2 (color card) and 3 (color word card), Trails A and B, fluency tests for both phonemic (variable letters used, see supplements) and semantic (animals) fluency, and the LDST/DSST. Of these, the Stroop card 2, Trails A and the LDST/DSST are generally considered to be tests of processing speed, whereas the Stroop card 3, Trails B and fluency tests are believed to be tests of executive function, reflecting mental flexibility, attention and planning. However, there is overlap in the cognitive processes needed to perform each of these tests. Tests were administered using a standardized method by an investigator unaware of any genetic information on the subjects. Details for the test administration and raw scores are given in Supplementary Table 1 and in the Supplement Methods text.

*Genotyping and imputation*

The genotyping was independently performed by each cohort using commercially available arrays ranging from Illumina 300k to 610k and Affymetrix GeneChip SNP Array 6.0 (Supplementary Table 2). Each cohort applied standard quality control filtering prior to genetic imputations. These included a SNP call rate of at least 90%, sample call rate of at least 92%, minor allele frequency (MAF) of at least 0.01 and Hardy-Weinberg deviation p-value of at most $10^{-3}$ (Supplementary Table 3). Genetic data imputations were performed in each cohort using HapMap II CEPH (Utah residents with ancestry from northern and western Europe) (CEU) as the reference panel, with the exceptions of ARIC and GENOA African American cohorts who imputed their genetic data using both HapMap II CEU and Yoruban in Ibadan (YRI) populations as reference. Details on genotyping and imputations are provided in Supplementary Tables 2-4.

*Genome-wide association analysis*

Each cohort performed a linear regression model of test scores against the dosage of coded alleles for each SNP testing an additive effect of the genetic variants. Skewness and kurtosis were evaluated before the analyses.

We used two models of association analysis; with and without the level of education as a confounder in addition to age, gender and other study specific confounders e.g. study site, familial relations or population substructure. This is because there is a dynamic, two-way relationship between cognitive function and level of education. The ability to score high on cognitive tests is influenced by background familiarity with, for example, the numbers and alphabet (Trails tests) or a large vocabulary (fluency tests), which are typically trained during one's formal education. Estimates are that, even in the Netherlands (10%) and the USA (22%), functional illiteracy is common among adults aged 16-65 [28,29]. Conversely, there is a genetic correlation between education and cognitive ability.[30]

*Meta-analysis*

The meta-analysis was performed in METAL [31]. For all tests except the Stroop test, meta-analysis was performed using the inverse variance method. For the Stroop test, the sample size weighted meta-analysis was performed because of the differences in the test methodology and measurement units that impeded the pooling of the beta coefficients. The z-statistic was weighted by the effective sample size (sample size * (observed dosage variance/expected dosage variance)) for each SNP. Genomic control was applied within each cohort prior to meta-analysis. The meta-analyses were restricted to autosomal SNPs common to all studies for each neurocognitive test.

*Additional analyses*

eQTL analyses

All variants with a discovery *p-value* $< 5\times10^{-6}$ were analyzed further to test whether they were associated with RNA expression. For this, we used the Genotype Tissue expression

portal (GTex, Broad Institute, Boston, MA, USA; www.broadinstitute.org/gtex) to assess the variants for their influence on expression of their closest genes in brain tissue.[32] We also performed eQTL experiments in 138 human hippocampal cell lines obtained *in vivo* from patients undergoing surgery for treatment-resistant epilepsy. Details on the methods of this functional follow-up are given in the Supplement. Some tests require the use of short-term memory, e.g. those words already mentioned in fluency. Thus, some hippocampal involvement might be expected along with the predominant frontal processes of executive function. These associations were assessed by correlating whole-genome SNP (Illumina Human 660W array) and RNA expression data (Illumina HumanHT-12v3). The Bonferroni corrected level of eQTL significance for a given SNP can be assumed to be 0.05/n of genes for a given SNP and a given test, leading to a p-value of $0.05/20\,000=2.5*10^{-6}$ for the hypothesis-free, genome-wide search for eQTL associations.

Gene network and functional prediction analysis

A gene network analysis was performed using the Gene Network database (Fehrmann et al, manuscript in preparation), which incorporates gene expression data from 77 840 human, mouse and rat Affymetrix microarrays from the Gene Expression Omnibus (GEO). Using principal components analysis on the probeset correlation matrices, so-called transcriptional components were identified, that describe major biological pathways. We combined this data into a multi-species gene network with 19 997 unique human genes. Predictions of gene function were made using biological databases such as the Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO) Database, and Reactome pathway database (http://www.genenetwork.nl/genenetwork, by Groningen University, The Netherlands). Additionally, we consulted the GTex website for tissue expression data of the genes of interest. More detailed information on the gene network approach is given in the Supplement.

## RESULTS

### *GWAS and eQTL analysis*

Baseline characteristics of the study populations and mean test results for each trait are provided in Supplementary Table 1. Supplementary Figure 1 shows the quantile-quantile (QQ) plots for the discovery-phase GWAS meta-analysis of each trait. No inflation due to hidden substructure or cryptic relatedness was observed for any cohort.

In the meta-analysis for processing speed (LDST/DSST tests), only one genome wide significant association was observed at an intronic variant (rs17518584, *p-value* = $2.33\text{x}10^{-8}$ in the model adjusted for age, sex and education) in the gene encoding cell adhesion molecule 2 (*CADM2*) on chromosome 3 (Supplementary Figure 2 & Table 1). This gene is also known as synaptic cell adhesion molecule 2 (*SYNCAM2*). This variant explains 0.05% (age, sex, and education adjusted model) of the variance in LDST/DSST in the Rotterdam Study, one of the largest population-based cohorts.

Additional information was sought for all SNPs with a discovery p-value below $5\times10^{-6}$ and a minor allele frequency (MAF) > 0.05. At this phase, 7 SNPs were analyzed for LDST/DSST, 15 SNPs for the Stroop test, 12 SNPs for Trails A, 17 SNPs for Trails B, 9 SNPs for letter fluency, and 34 SNPs for semantic fluency. Of all the loci tested in the independent cohorts, including both subjects of European and African American ancestries, a nominally significant association in the same direction as in the discovery analyses was observed only for the *CADM2* variant rs17518584 with LDST/DSST (nominal *p-value* = 0.05 for meta-analyses of results of the additional cohorts, adjusting for age and sex). Meta-analysis of discovery and replication cohorts yielded genome wide significant evidence for association of rs17518584 in *CADM2* with scores for the LDST/DSST processing speed tests, in both the age- and sex-adjusted model (*p-value* = $3.92\times10^{-8}$) and in the fully-adjusted model that included age, sex and education (*p-value* = $3.91\times10^{-9}$) (Table 1). The findings of the individual studies (both discovery and replication) are shown in Figure 1.
When looking for association signals for earlier identified loci, rs7412 and rs429358 in the *APOE* ε4 locus associated with general cognition [17] were not present in the analyzed SNP sets. However, a single SNP proxy for *APOE* ε4 (rs4420638) [33] was associated to LDST/DSST (*p-value* = $2.11\times10^{-4}$) in the fully adjusted model. There was no evidence for association of SNPs in any of the other tested candidate genes that have previously been associated with executive and speed functions. (*DTNBP1:* lowest *p-value =* 0.019 for rs2619522 to Stroop*, TRIB3:* lowest *p-value =* 0.155 for phonemic fluency*, WDR72* lowest *p-value =* 0.149) in our GWAS meta-analysis for the fully adjusted model. Also, regions for which linkage was reported (2q, 5q, 11q, 13q and 14q) were not genome wide significant. The strongest association was seen in the 11q.25 region in which we found a SNP rs2734839 to be suggestively associated with performance on the LDST/DSST (*p-value* = $4.39\times10^{-7}$ for the age- and sex-adjusted model and $3.08\times10^{-6}$ for the model adjusted for age, sex and education). Rs2734839 is an intronic SNP in the *DRD2* gene encoding the D2 subtype of the dopamine receptor.

*Bioinformatics analysis*

The GTex tissue expression data (www.broadinstitute.org/gtex) shows that *CADM2* is expressed in different areas of the brain more abundantly than in any other tissue, and most specifically in the frontal and anterior cingulate cortex (Supplementary Figure 3). In the GTex eQTL analysis, the top hit from the LDST/DSST GWAS, rs17518584, showed association with RNA expression levels of *CADM2* in the cingulate cortex (*p-value* = $4\times10^{-4}$), general hemispheral cortex (*p-value* = 0.004), substantia nigra (*p-value* = 0.007), frontal cortex (*p-value* = 0.02), and cerebellum (*p-value* = 0.03). No significant *cis* or *trans* associations were identified in the hippocampal data. Gene network analysis shows that *CADM2* is significantly expressed in human brain (*p-value* for expression in cerebral cortex = $2.47\times10^{-171}$, AUC=0.99; p-value for expression in the prefrontal cortex specifically = $1.29\times10^{-31}$, AUC=1.0). The gene network analyses suggests that *CADM2* is likely involved in biological processes that include

## Table 1. SNPs with p-value < $10^{-6}$ in original meta-analysis

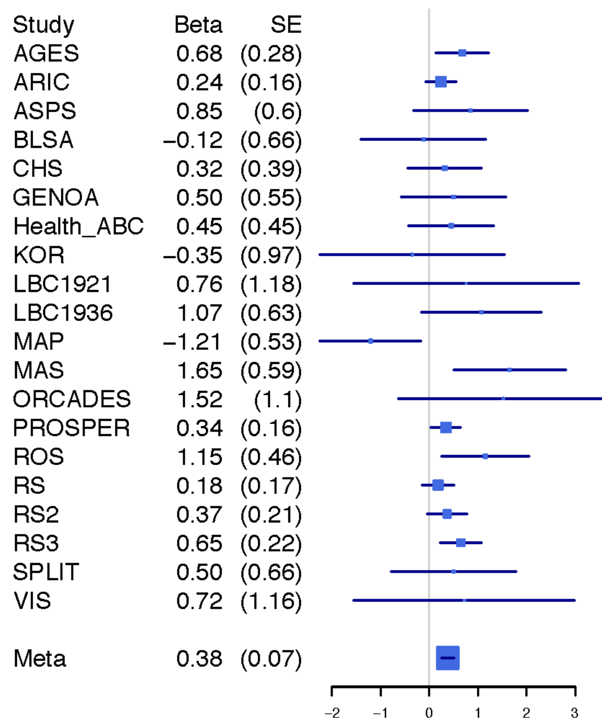
| Trait | SNP | Chr | Position | Gene | feature | A1/A2 | average EAF | Discovery | | | | | Replication | | | | | Combined | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | N | Model1 | | Model2 | | N | Model1 | | Model2 | | Model1 | | Model2 | |
| | | | | | | | | | *Beta (SE)* | *P-value* | *Beta (SE)* | *P-value* | | *Beta (SE)* | *P-value* | *Beta (SE)* | *P-value* | *Beta (SE)* | *P-value* | *Beta (SE)* | *P-value* |
| **Trails A** | rs9514964 | 13 | 109679886 | MYO16 | intron | a/g | 0.17 | 5,429 | 0.048 (0.009) | $3.0x10^{-07}$ | 0.044 (0.009) | $2.3x10^{-06}$ | 9,553 | -0.006 (0.007) | 0.370 | 0.007 (0.007) | 0.320 | 0.029 (0.007) | $1.5x10^{-04}$ | 0.020 (0.006) | $3.0x10^{-04}$ |
| | rs9559465 | 13 | 109683883 | MYO16 | intron | a/g | 0.84 | | -0.048 (0.010) | $3.3x10^{-07}$ | -0.043 (0.009) | $2.6x10^{-06}$ | | 0.006 (0.008) | 0.430 | -0.007 (0.008) | 0.360 | -0.035 (0.008) | $1.6x10^{-04}$ | -0.021 (0.006) | $2.4x10^{-04}$ |
| | rs1230154 | 4 | 99988659 | METAP1, ADH5 | intergenic | t/c | 0.71 | | -0.038 (0.007) | $7.7x10^{-07}$ | -0.036 (0.008) | $1.6x10^{-06}$ | | 0.003 (0.006) | 0.630 | 0.001 (0.006) | 0.930 | -0.030 (0.007) | $9.3x10-^{-04}$ | -0.013 (0.005) | $4.8x10^{-03}$ |
| **Trails B** | rs11082233 | 18 | 37265119 | PIK3C3, KC6 | intergenic | t/c | 0.11 | 6,210 | 0.073 (0.013) | $6.9x10^{-08}$ | 0.061 (0.013) | $2.3x10^{-06}$ | 10,817 | - 0.012 (0.010) | 0.250 | -0.017 (0.010) | 0.090 | 0.020 (0.008) | $1.7x10^{-02}$ | 0.0123 (0.0079) | 0.12 |
| **Stroop** | | | | | | | | 12,866 | | | | | 2,650 | | | | | | | | |
| **Phonemic fluency** | rs6583634 | 8 | 143546590 | BAI1 | intron | t/c | 0.15 | 13,454 | 1.046 (0.209) | $5.7x10^{-07}$ | 0.914 (0.194) | $2.5x10^{-06}$ | 26,222 | -0.162 (0.230) | 0.480 | -0.009 (0.164) | 0.960 | 0.498 (0.155) | $1.0x10^{-03}$ | 0.376 (0.125) | $2.7x10^{-03}$ |
| | rs10481393 | 8 | 143556168 | BAI1 | intron | t/c | 0.15 | | 1.050 (0.210) | $5.9x10^{-07}$ | 0.914 (0.195) | $2.8x10^{-06}$ | | 0.068 (0.479) | 0.880 | -0.015 (0.183) | 0.937 | 0.891 (0.193) | $3.7x10^{-06}$ | 0.420 (0.134) | $1.6x10^{-03}$ |
| **Semantic Fluency** | rs6587905 | 1 | 61093597 | C1orf87, NFIA | intergenic | t/c | 0.16 | 6,383 | 0.664 (0.129) | $2.8x10^{-07}$ | 0.656 (0.126) | $2.1x10^{-07}$ | 15,826 | 0.091 (0.099) | 0.358 | 0.138 (0.077) | 0.071 | 0.303 (0.079) | $1.1x10^{-04}$ | 0.280 (0.065) | $1.9x10^{-05}$ |
| | rs10115337 | 9 | 88711574 | LOC392358, GAS1 | intergenic | t/c | 0.18 | | -0.494 (0.100) | $8.0x10^{-07}$ | -0.478 (0.098) | $9.9x10^{-07}$ | | -0.058 (0.079) | 0.462 | -0.083 (0.059) | 0.159 | -0.224 (0.062) | $2.9x10^{-04}$ | -0.189 (0.050) | $1.6x10^{-04}$ |
| | rs10780801 | 9 | 88710941 | LOC392358, GAS1 | intergenic | t/c | 0.82 | | 0.493 (0.100) | $8.2x10^{-07}$ | 0.477 (0.976) | $1.0x10^{-06}$ | | 0.055 (0.079) | 0.485 | 0.081 (0.059) | 0.166 | 0.223 (0.062) | $3.2x10^{-04}$ | 0.189 (0.050) | $1.8x10^{-04}$ |
| **LDST/DSST** | rs664154 | 6 | 19099538 | RP1-239K6.1 | intergenic | t/c | 0.24 | 32,080 | 0.417 (0.079) | $1.1x10^{-07}$ | 0.367 (0.074) | $7.9x10^{-07}$ | 1,311 | -0.526 (0.486) | 0.279 | -0.571 (0.486) | 0.240 | 0.393 (0.076) | $2.5x10^{-07}$ | 0.345 (0.072) | $1.8x10^{-06}$ |
| | rs17518584 | 3 | 85687613 | CADM2 | intron | t/c | 0.64 | | 0.366 (0.071) | $2.7x10^{-07}$ | 0.376 (0.067) | $2.3x10^{-08}$ | | 0.875 (0.439) | 0.046 | 0.763 (0.439) | 0.082 | 0.379 (0.069) | $3.9x10^{-08}$ | 0.385 (0.065) | $3.9x10^{-09}$ |
| | rs2734839 | 11 | 112791700 | DRD2 | intron | t/c | 0.61 | | 0.332 (0.068) | $9.9x10^{-07}$ | 0.294 (0.064) | $4.9x10^{-06}$ | | 0.284 (0.442) | 0.520 | 0.118 (0.44) | 0.789 | 0.333 (0.066) | $4.4x10^{-07}$ | 0.292 (0.063) | $3.0x10^{-06}$ |

*Model 1: age, sex adjusted*

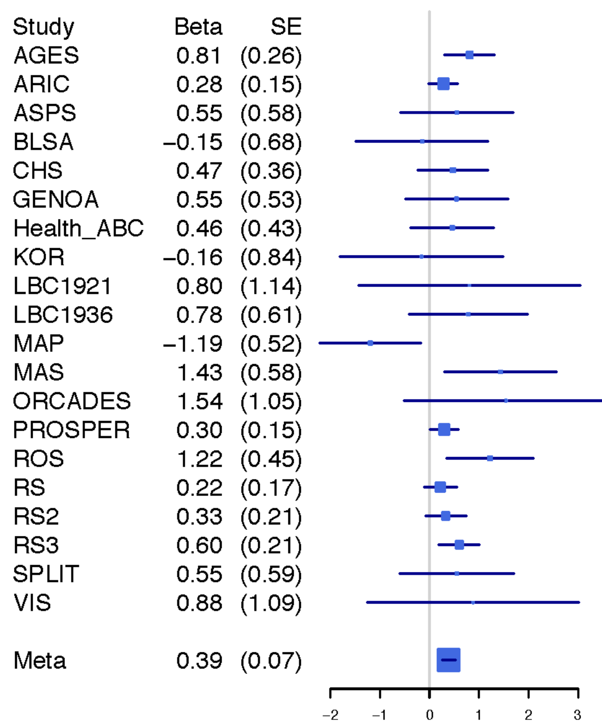
*Model 2: age, sex, education adjusted*

**Figure 1. Forest plot for rs17518584 (*CADM2*), LDST/DSST.**

Panel 1A: Adjusted for age and sex.




Panel 1B: adjusted for age, sex and education




*The magnitude and direction of effect of rs17518584 for each cohort are shown, as well as the summary statistics (effect size and standard error)*

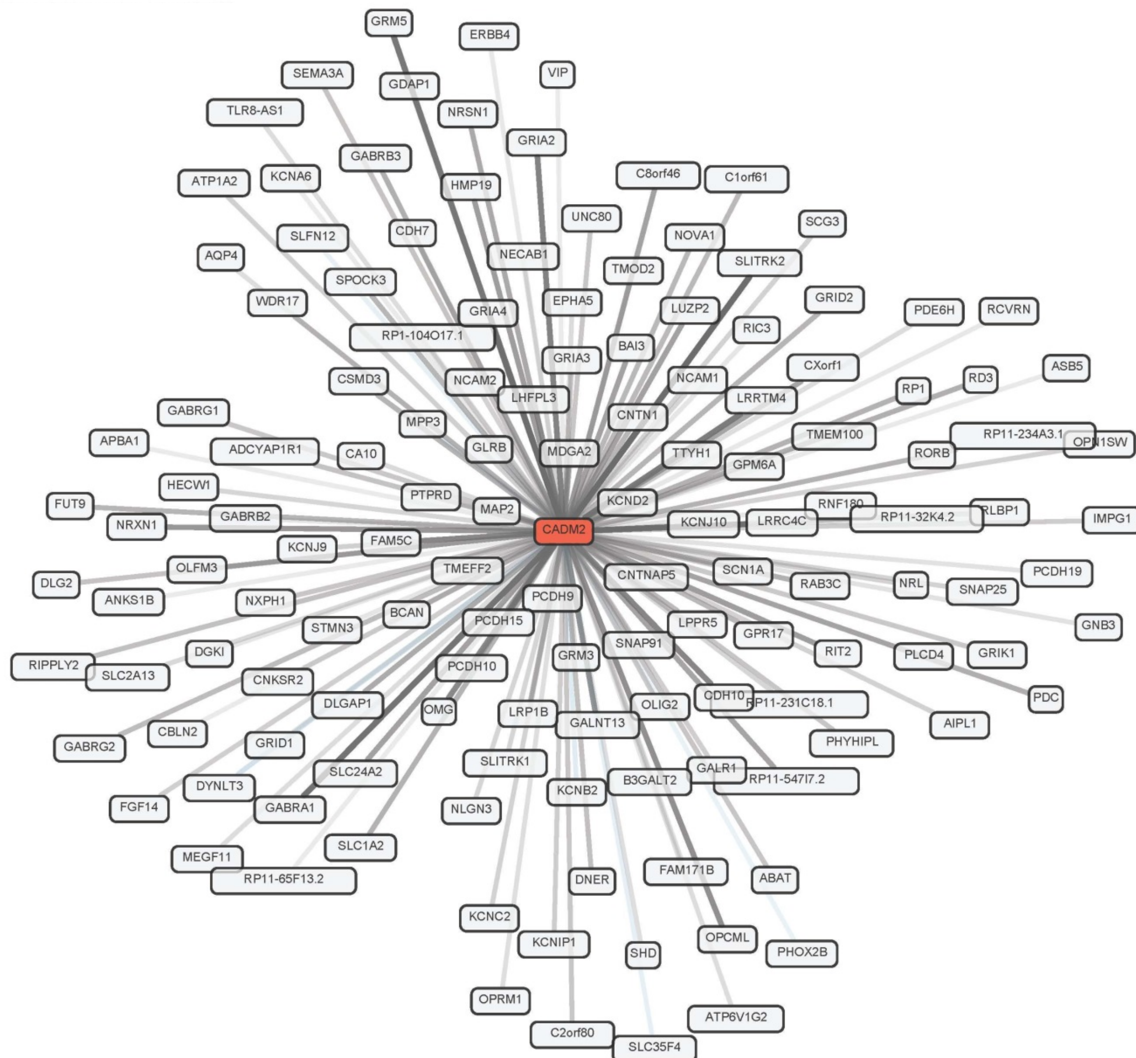
the glutamate signaling pathway (*p-value* = $7.22 \times 10^{-15}$), gamma-aminobutyric acid (GABA) transport (*p-value* = $1.36 \times 10^{-11}$) and neuron cell-cell adhesion (*p-value* = $1.48 \times 10^{-13}$). *CADM2* shows strong positive co-expression with several genes involved in the GABA and other glutamate neurosignaling pathways, such as GABA A receptor alpha 1 and beta 2 (*GABRA1, GABRB2*), and glutamate receptor metabotropic 5 (*GRM5*). Further, there is strong positive co-expression with many members of the voltage gated potassium channel group including *KCNJ9, KCNJ10*, *KCNB2* and *KCNC*, and with the opioid binding protein/cell adhesion molecule-like (*OPCML*) gene and EPH receptor A5 (*EPHA5*). (Figure 2). Supplementary Table 5 and Supplementary Figure 3 provide the expression data in neuronal tissues. As expected, Gene network and GTex show that *DRD2* is predominantly expressed in the pituitary (GTex reads per kilobase per million (RPKM) = 50), putamen (genenetwork AUC = 1, *p-value* = $5 \times 10^{-12}$, GTex RPKM = 40), substantia nigra (genenetwork AUC = 0.98, *p-value* = $4 \times 10^{-15}$, GTex RPKM = 4), and nucleus accumbens (GTex RPKM = 28). However, the intronic variant rs2734839 is not associated with expression at nominal significance in any region reported (cingulate, frontal or general cortex, amygdala, caudate nucleus, nucleus accumbens, putamen, substantia nigra, cerebellum, hippocampus, hypothalamus, pituitary or spinal cord).

**DISCUSSION**

The discovery phase of our study yielded genome-wide significant evidence for association of processing speed tests (LDST/DSST) to r rs17518584. *In silico* replication in additional independent cohorts yielded nominally significant support for association. In the combined discovery-replication analysis both the model adjusted for age and sex and the model adjusted for age, sex and education resulted in genome-wide significant association between rs17518584 and LDST/DSST performance.  The variant is associated with *CADM2* expression, at least at nominal significance.

Rs17518584 is located about 170 kb upstream of the transcription start site of the major transcript for the *CADM2* gene, but is within an intron of a variant transcript that includes an alternative first exon. Variants in the *CADM2* gene have been previously associated with body mass index [33], but not to cognitive function. The gene has been studied as a candidate for autism spectrum disorders [34], and showed some suggestive association with scores on the persistence items on a personality scale [35]. The protein is likely involved in long-term signal depression and potentiation, neuroactive ligand-receptor interaction (http://www.genome.jp/kegg/) and is a member of the immunoglobulin superfamily. The gene encodes a neuronal adhesion molecule that has been shown to be widely expressed in the developing and adult brain in laboratory mice [36], as well as in human, mouse and rat brain tissue, with highest AUCs for the prefrontal cortex. In the GTex analyses, rs17518584 significantly influences *CADM2* expression levels in the brain, most specifically in the cingulate cortex. This finding is of interest in the background of diffusion tensor imaging experiments that have shown an association between fractional anisotropy in the cingulum and performance on executive and processing speed tasks [37,38]. Moreover, the cingulum

**Figure 2. Gene network plot for *CADM2*, based on biological processes.**




*Grey lines indicate positive co-expression, blue lines indicate negative co-expression, with the density of the line reflecting the strength of the co-expression.*

may be critical for development of executive functioning, which supports a possible direct role of *CADM2* in processing speed as well as executive function development through an effect on the development of prefrontal cortex and cingulum. The co-expression data from gene network also revealed some interesting links to Alzheimer's disease (Figure 2). *OPCML* was associated with a variety of cognitive domains in a follow-up of linkage regions for Alzheimer's disease [39]. *EPHA5* is a member of the same family of ephrin receptors as the *EPHA1* gene, which is associated with Alzheimer's disease [40,41].

We identified only one genome wide significant variant explaining a small part of the variance ($R^2$ = 0.005), leaving most of the variance unexplained. These findings should be interpreted in light of the sample size studied. The number of persons analyzed varied considerably between outcomes, there being from 5 555 to 32 900 subjects per individual test. Our analyses have lacked the power to address genes with small effects affecting tests

for which we had only a limited sample size. The lack of test-specific power should also be taken into account when interpreting the lack of association to the candidate genes beyond *APOE* (*DTNBP1, TRIB3, and WDR72).* Of the candidate genes, there is only one other intronic variant (rs2734839) in the *DRD2* gene encoding the D2 subtype of the dopamine receptor that approaches genome wide significance but the bioinformatics analyses does not support a functional effect of the expression of the gene in the substantia nigra. This variant therefore remains to be replicated. A further limitation is that the cognitive phenotypes consisted of single assessments and thus we are not able to study age-related cognitive decline. Despite of these limitations, this is the first study to discover and replicate a genetic variant involved in the executive function.

The present report provides the most comprehensive meta-analysis of processing speed and executive function GWAS to date. We show a genome wide significant association between widely-used tests of processing speed (LDST/DSST) and a SNP in the *CADM2* gene, which is involved in glutamatergic and GABA-ergic transmission, in middle-aged and older non-demented individuals. This gene is a candidate for autism and personality, but based on the pathway analyses it may also be relevant for a broad range of dementias and psychiatric disorders.

**ACKNOWLEDGMENTS**

3CS: The work was made possible by the generous participation of the control subjects and their families. This work was supported by the National Foundation for Alzheimer's disease and related disorders, the Institut Pasteur de Lille, the Centre National de Génotypage, Inserm, FRC (fondation pour la recherche sur le cerveau) and Rotary. This work has been developed and supported by the LABEX (laboratory of excellence program investment for the future) DISTALZ grant (Development of Innovative Strategies for a Transdisciplinary approach to ALZheimer's disease).

J.C.L. was funded by the MEDIALZ Project (Grant 11001003) financed by ERDF (European Regional Development Fund) and Conseil Régional Nord Pas de Calais.

The Three-City Study was performed as part of a collaboration between the Institut National de la Santé et de la Recherche Médicale (Inserm), the Victor Segalen Bordeaux II University and Sanofi-Synthélabo. The Fondation pour la Recherche Médicale funded the preparation and initiation of the study. The 3C Study was also funded by the Caisse Nationale Maladie des Travailleurs Salariés, Direction Générale de la Santé, MGEN, Institut de la Longévité, Agence Française de Sécurité Sanitaire des Produits de Santé, the Aquitaine and Bourgogne Regional Councils, Fondation de France and the joint French Ministry of Research/INSERM "Cohortes et collections de données biologiques" programme. Lille Génopôle received an unconditional grant from Eisai.

AAA: We thank the cohort participants and team members who contributed to this study. Phenotype collection and DNA extraction was supported by the Wellcome Trust, the British Heart Foundation and the Chief Scientist Office of the Scottish Executive. The AAA Trial was performed and the database is maintained by members of the University of Edinburgh Molecular Epidemiology Research Group in the Centre for Population Health Sciences. We also thank staff at the Wellcome Trust Clinical Research Facility in Edinburgh where some of the research clinics and genotyping were undertaken.

Aging Gene-Environment Susceptibility-Reykjavik Study: The research has been funded by NIA contract N01-AG-12100 with contributions from NEI, NIDCD and NHLBI, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament).

The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL70825, R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.

ASPS: We thank Ing. Johann Semmler and Irmgard Pölzl for creating the DNA bank and for supervising the quality managementof the biobanking and DNA analyses.The ASPS is funded

by the Austrian Science Fond (FWF) grant number P20545-P05 and P13180.
Baltimore Longitudinal Study of Aging (BLSA): The Baltimore Longitudinal Study of Aging is supported by the Intramural Research Program of the NIH, National Institute on Aging.
This CHS research was supported by NHLBI contracts N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85084, N01HC85085, N01HC85086; N01HC35129, N01HC15103, N01HC55222, N01HC75150, N01HC45133, N01HC85239, and by HHSN268201200036C and NHLBI grants HL080295, HL087652, HL105756 with additional contribution from NINDS. Additional support was provided through AG023629, AG15928, AG20098, and AG027058 from the NIA. See also http://www.chs-nhlbi.org/pi.htm.DNA handling and genotyping at Cedars-Sinai Medical Center was supported in part by the National Center for Research Resources, grant UL1RR033176, and is now at the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124; in addition to the National Institute of Diabetes and Digestive and Kidney Disease grant DK063491 to the Southern California Diabetes Endocrinology Research Center.
The CROATIA-Korcula study was funded by grants from the Medical Research Council (UK), European Commission Framework 6 project EUROSPAN (Contract No. LSHG-CT-2006-018947) and Republic of Croatia Ministry of Science, Education and Sports research grants to I.R. (108-1080315-0302). We would like to acknowledge the invaluable contributions of the recruitment team in Korcula, the administrative teams in Croatia and Edinburgh and the people of Korcula.
The SNP genotyping for the CROATIA-Korcula cohort was performed in Helmholtz Zentrum München, Neuherberg, Germany
The CROATIA-Split study is funded by grants from the Medical Research Council (UK), European Commission Framework 6 project EUROSPAN (Contract No. LSHG-CT-2006-018947) and Republic of Croatia Ministry of Science, Education and Sports research grants to I.R. (108-1080315-0302). We would like to acknowledge the invaluable contributions of the recruitment team in Split, the administrative teams in Croatia and Edinburgh and the people of Split The SNP genotyping for the CROATIA-Split cohort was performed by AROS Applied Biotechnology, Aarhus, Denmark.
The CROATIA-Vis study was funded by grants from the Medical Research Council (UK) and Republic of Croatia Ministry of Science, Education and Sports research grants to I.R. (108-1080315-0302). We would like to acknowledge the staff of several institutions in Croatia that supported the field work, including but not limited to The University of Split and Zagreb Medical Schools, the Institute for Anthropological Research in Zagreb and Croatian Institute for Public Health.
The SNP genotyping for the CROATIA-Vis cohort was performed in the core genotyping laboratory of the Wellcome Trust Clinical Research Facility at the Western General Hospital, Edinburgh, Scotland
ERF: This study was financially supported by the Netherlands Organization for Scientific Research (NWO), the Internationale Stichting Alzheimer Onderzoek (ISAO), the Hersenstichting Nederland (HSN) and the Centre for Medical Systems Biology (CMSB) in the

framework of the Netherlands Genomics Initiative (NGI) and by the Russian Foundation for Basic Research (RFBR). We thank the participants from the Genetic Research in Isolated Populations, Erasmus Rucphen Family, who made this work possible. Also, we thank Petra Veraart for collecting all genealogical data.
FHS: This work was supported by the National Heart, Lung and Blood Institute's Framingham Heart Study (Contract No. N01-HC-25195) and its contract with Affymetrix, Inc for genotyping services (Contract No. N02-HL-6-4278) and grants (U01 HL096917 and R01 HL093029). A portion of this research utilized the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. Analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. This study was also supported by grants from the National Institute of Neurological Disorders and Stroke (NS17950) and the National Institute of Aging (AG033193, AG081220, AG16495). The content is solely the responsibility of the authors and does not necessarily represent the official views of NINDS, NHLBI, NIA, NIH or AHA. Dr. Debette is a recipient of a Chaire d'Excellence grant from the Agence National de la Recherche.
Support for the Genetic Epidemiology Network of Arteriopathy (GENOA) was provided by the National Heart, Lung and Blood Institute (HL054464, HL054457, HL054481, HL071917, and HL87660) and the National Institute of Neurological Disorders and Stroke (NS041558) of the National Institutes of Health. Genotyping was performed at the Mayo Clinic (S.T.T., Mariza de Andrade, Julie Cunningham) and was made possible by the University of Texas Health Sciences Center (Eric Boerwinkle, Megan L. Grove-Gaona). We would also like to thank the families that participated in the GENOA study.
We are grateful to the GS Executive Committee Professors Andrew D. Morris, Blair H. Smith, Anna F. Dominiczak, David J. Porteous and Drs Lynne J. Hocking and Sandosh Padmanabhan and all the families who took part, the general practitioners and the Scottish School of Primary Care for their help in recruiting them, and the whole Generation Scotland team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, healthcare assistants and nurses. Generation Scotland receives core support from the Chief Scientist Office of the Scottish Government Health Directorates CZD/16/6 and the Scottish Funding Council.
We thank all study participants as well as everybody involved in the Helsinki Birth Cohort Study. Helsinki Birth Cohort Study has been supported by grants from the Academy of Finland, the Finnish Diabetes Research Society, Folkhälsan Research Foundation, Novo Nordisk Foundation, Finska Läkaresällskapet, Signe and Ane Gyllenberg Foundation, University of Helsinki, Ministry of Education, Ahokas Foundation, Emil Aaltonen Foundation, Juho Vainio Foundation, and Wellcome Trust (grant number WT089062).
Health ABC: This research was supported by NIA contracts N01AG62101, N01AG62103, and N01AG62106. The genome-wide association study was funded by NIA grant 1R01AG032098-01A1 to Wake Forest University Health Sciences and genotyping services were provided by

the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C. This research was supported in part by the Intramural Research Program of the NIH, National Institute on Aging.
Hunter: The authors would like to thank the men and women participating in the HCS as well as all the staff, investigators and collaborators who have supported or been involved in the project to date. The cohort was made possible with support from the University of Newcastle's Strategic Initiative Fund, the Vincent Fairfax Family Foundation, and the Hunter Medical Research Institute.
The Invechhiare in Chianti (InCHIANTI) Study was supported as a targeted project (ICS 110.1RS97.71) by the Italian Ministry of Health, by the U.S. National Institute on Aging (Contracts N01]AG]916413, N01]AG] 821336, 263 MD 9164 13, and 263 MD 821336), and, in part, by the Intramural Research Program, National Institute on Aging, National Institutes of Health.
LBC: We thank the cohort participants and team members who contributed to these studies. Phenotype collection in the Lothian Birth Cohort 1921 was supported by the BBSRC, The Royal Society and The Chief Scientist Office of the Scottish Government. Phenotype collection in the Lothian Birth Cohort 1936 was supported by Research Into Ageing (continues as part of Age UK The Disconnected Mind project). Genotyping of the cohorts was funded by the UK Biotechnology and Biological Sciences Research Council (BBSRC). The work was undertaken by The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (G0700704/84698). Funding from the BBSRC, Engineering and Physical Sciences Research Council (EPSRC), Economic and Social Research Council (ESRC), and MRC is gratefully acknowledged.
We would like to acknowledge and thank the Sydney MAS participants and the Research Team for their contributions and assistance. We would like to specifically acknowledge the support and contributions of Professor Henry Brodaty (Chief Investigator), Dr. Simone Reppermund (Study Co-ordinator), Professor Peter Schofield, Dr Arezoo Assareh and Dr John Kwok to this work. DNA was extracted by Genetic Repositories Australia, an Enabling Facility supported by NHMRC Grant 401184. DNA sample preparation was undertaken in the laboratory of Professor Peter Schofield and Dr John Kwok, Neuroscience Research Australia, with the assistance of Dr. Arezoo Assareh. Genotyping was performed by the Ramaciotti Centre, University of New South Wales. Sydney MAS is supported by the Australian National Health & Medical Research Council Program Grants 350833 &568969. Karen Mather is supported by the Capacity Building Grant 568940. Nicola Armstrong is supported by the NHMRC Project grant 525453.
The hippocampal gene expression study was supported by the German Federal Ministry of Education and Research (BMBF) through the Integrated Genome Research Network (IG) MooDS (Systematic Investigation of the Molecular Causes of Major Mood Disorders and Schizophrenia, under the auspices of the National Genome Research Network

plus (NGFNplus).
NHS: This study was supported by research grants CA87969, CA49449, HL34594, U01HG004399, DK058845, CA65725, CA67262, CA50385, 5UO1CA098233, EY09611, EY015473, HG004728, HL35464, CA55075, CA134958, and DK070756 from the National Institutes of Health. The genotyping was partly supported by an unrestricted grant from Merck Research Laboratories. Dr. Sun is supported by career development award K99HL098459 from the National Heart, Lung, and Blood Institute.Supported in part by NIH
ORCADES was supported by the Chief Scientist Office of the Scottish Government, the Royal Society, the MRC Human Genetics Unit, Arthritis Research UK and the European Union framework program 6 EUROSPAN project (contract no. LSHG-CT-2006-018947). DNA extractions were performed at the Wellcome Trust Clinical Research Facility in Edinburgh. We would like to acknowledge the invaluable contributions of Lorraine Anderson and the research nurses in Orkney, the administrative team in Edinburgh and the people of Orkney.
The PROSPER study was supported by an investigator initiated grant obtained from Bristol-Myers Squibb. Prof. Dr. J. W. Jukema is an Established Clinical Investigator of the Netherlands Heart Foundation (grant 2001 D 032). Support for genotyping was provided by the seventh framework program of the European commission (grant 223004) and by the Netherlands Genomics Initiative (Netherlands Consortium for Healthy Aging grant 050-060-810).
The Rotterdam Study is sponsored by the Erasmus Medical Center and Erasmus University Rotterdam, The Netherlands Organization for Scientific Research (I), The Netherlands Organization for Health Research and Development (ZonMW), the Research Institute for Diseases in the Elderly (RIDE), The Netherlands Genomics Initiative, the Ministry of Education, Culture and Science, the Ministry of Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. Further support was obtained from the Netherlands Consortium for Healthy Ageing. Dr. Ikram was supported by a ZonMW Veni grant: 916.13.054
RUSH: Supported in part by NIA grants P30AG10161, R01AG15819, R01AG17917, R01AG30146, K08AG34290 and K25AG41906.
SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403). The SHIP authors are grateful to Holger Prokisch and Thomas Meitinger (Helmholtz Zentrum München) for the genotyping of the SHIP-TREND cohort.
The WGHS is supported by HL043851 and HL080467 from the National Heart, Lung, and Blood Institute and CA047988 from the National Cancer Institute, the Donald W. Reynolds Foundation and the Fondation Leducq, with collaborative scientific support and funding for genotyping provided by Amgen.
The study sponsors had no role in the study design, the collection, analysis, and interpretation of data; writing the report, or the decision to submit the report for publication.

**REFERENCES**


1. Green, M.F. Cognitive impairment and functional outcome in schizophrenia and bipolar disorder. *J Clin Psychiatry* **67 Suppl 9**, 3-8; discussion 36-42 (2006).
2. Seidman, L.J., Biederman, J., Weber, W., Hatch, M. & Faraone, S.V. Neuropsychological function in adults with attention-deficit hyperactivity disorder. *Biol Psychiatry* **44**, 260-8 (1998).
3. Huntley, J.D. & Howard, R.J. Working memory in early Alzheimer's disease: a neuropsychological review. *Int J Geriatr Psychiatry* **25**, 121-32 (2009).
4. Salthouse, T.A. The processing-speed theory of adult age differences in cognition. *Psychol Rev* **103**, 403-28 (1996).
5. Luciano, M. et al. Whole genome association scan for genetic polymorphisms influencing information processing speed. *Biol Psychol* **86**, 193-202 (2011).
6. Smith, E.E. & Jonides, J. Storage and executive processes in the frontal lobes. *Science* **283**, 1657-61 (1999).
7. Wright, M. et al. Genetics of cognition: outline of a collaborative twin study. *Twin Res* **4**, 48-56 (2001).
8. Lee, T. et al. Genetic influences on five measures of processing speed and their covariation with general cognitive ability in the elderly: the older Australian twins study. *Behav Genet* **42**, 96-106 (2012).
9. Lee, T. et al. Genetic influences on four measures of executive functions and their covariation with general cognitive ability: the Older Australian Twins Study. *Behav Genet* **42**, 528-38 (2012).
10. Sleegers, K. et al. Cerebrovascular risk factors do not contribute to genetic variance of cognitive function: the ERF study. *Neurobiol Aging* **28**, 735-41 (2007).
11. Friedman, N.P. et al. Individual differences in executive functions are almost entirely genetic in origin. *J Exp Psychol Gen* **137**, 201-25 (2008).
12. Deary, I.J., Penke, L. & Johnson, W. The neuroscience of human intelligence differences. *Nat Rev Neurosci* **11**, 201-11 (2010).
13. Butcher, L.M., Kennedy, J.K. & Plomin, R. Generalist genes and cognitive neuroscience. *Curr Opin Neurobiol* **16**, 145-51 (2006).
14. Deary, I.J., Johnson, W. & Houlihan, L.M. Genetic foundations of human intelligence. *Hum Genet* **126**, 215-32 (2009).
15. Davies, G. et al. Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Mol Psychiatry* (2011).
16. Payton, A. The Impact of Genetic Research on our Understanding of Normal Cognitive Ageing: 1995 to 2009. *Neuropsychol Rev* (2009).
17. Wisdom, N.M., Callahan, J.L. & Hawkins, K.A. The effects of apolipoprotein E on non-impaired cognitive functioning: A meta-analysis. *Neurobiol Aging* (2009).
18. Zhang, J.P., Burdick, K.E., Lencz, T. & Malhotra, A.K. Meta-analysis of genetic variation in DTNBP1 and general cognitive ability. *Biol Psychiatry* **68**, 1126-33 (2010).

19. Buyske, S. et al. Cognitive traits link to human chromosomal regions. *Behav Genet* **36**, 65-76 (2006).
20. Rommelse, N.N. et al. Neuropsychological endophenotype approach to genome-wide linkage analysis identifies susceptibility loci for ADHD on 2q21.1 and 13q12.11. *Am J Hum Genet* **83**, 99-105 (2008).
21. Almasy, L. et al. A genome screen for quantitative trait loci influencing schizophrenia and neurocognitive phenotypes. *Am J Psychiatry* **165**, 1185-92 (2008).
22. Seshadri, S. et al. Genetic correlates of brain aging on MRI and cognitive test measures: a genome-wide association and linkage analysis in the Framingham Study. *BMC Med Genet* **8 Suppl 1**, S15 (2007).
23. Cirulli, E.T. et al. Common genetic variation and performance on standardized cognitive tests. *Eur J Hum Genet* (2010).
24. Need, A.C. et al. A genome-wide study of common SNPs and CNVs in cognitive performance in the CANTAB. *Hum Mol Genet* **18**, 4650-61 (2009).
25. Leblanc, M. et al. Genome-wide study identifies PTPRO and WDR72 and FOXQ1-SUMO1P1 interaction associated with neurocognitive function. *J Psychiatr Res* **46**, 271-8 (2012).
26. Kottgen, A. et al. New loci associated with kidney function and chronic kidney disease. *Nat Genet* **42**, 376-84 (2010).
27. Psaty, B.M. 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* **2**, 73-80 (2009).
28. A, K.I.J.A.J.L.K. Adult literacy in America. (National Center for Education Statistics, U.S. Department of Education
Office of Educational Research and Improvement, Washington, DC, USA, 2002).
29. R, F.D.H.W.V.d.V. Laaggeletterdheid in Nederland. (Expertisecentrum Beroepsonderwijs, Utrecht/'s-Hertogenbosch, The Netherlands, 2011).
30. Calvin, C.M. et al. Multivariate genetic analyses of cognition and academic achievement from two population samples of 174,000 and 166,000 school children. *Behav Genet* **42**, 699-710 (2012).
31. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
32. Consortium, G.T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-5 (2013).
33. Speliotes, E.K. et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* **42**, 937-48 (2010).
34. Casey, J.P. et al. A novel approach of homozygous haplotype sharing identifies candidate genes in autism spectrum disorder. *Hum Genet* **131**, 565-79 (2012).
35. Service, S.K. et al. A genome-wide meta-analysis of association studies of Cloninger's Temperament Scales. *Transl Psychiatry* **2**, e116 (2012).

36. Thomas, L.A., Akins, M.R. & Biederer, T. Expression and adhesion profiles of SynCAM molecules indicate distinct neuronal functions. *J Comp Neurol* **510**, 47-67 (2008).
37. Peters, B.D. et al. Age-Related Differences in White Matter Tract Microstructure Are Associated with Cognitive Performance from Childhood to Adulthood. *Biol Psychiatry* (2013).
38. Sasson, E., Doniger, G.M., Pasternak, O., Tarrasch, R. & Assaf, Y. White matter correlates of cognitive domains in normal aging with diffusion tensor imaging. *Front Neurosci* **7**, 32 (2013).
39. 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 (2007).
40. 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**, 436-41 (2011).
41. Hollingworth, P. et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet* **43**, 429-35 (2011).

## *3.3. Genome-wide studies of verbal declarative memory in non-demented older people: the CHARGE consortium*

Submitted

*Also see the co-investigator list at the end of this paper.*

*The supplemental material for this paper can be found in Chapter 7.*

**ABSTRACT**

Background: Memory performance in older persons can reflect genetic influences on cognitive function and dementing processes. We aimed to identify genetic contributions to verbal declarative memory in a community setting.

Methods: We conducted genome-wide association studies (GWAS) for paragraph or word-list delayed recall in 19 cohorts from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium, comprising 29,076 dementia- and stroke-free individuals of European descent, aged ≥45 years. Replication of suggestive associations ($p<5x10^{-6}$) was sought in 10,617 participants of European descent, 3,811 African-Americans and 1,561 young adults.

Results: Rs4420638, near *APOE*, was associated with poorer delayed recall performance in discovery ($p=5.57x10^{-10}$) and replication cohorts ($p=5.65x10^{-8}$). This association was stronger for paragraph than word-list delayed recall and in the oldest persons. Two associations with specific tests, in subsets of the total sample, reached genome-wide significance in combined analyses of discovery and replication (rs11074779 [*HS3ST4*], $p=3.11x10^{-8}$, and rs6813517 [*SPOCK3*], $p=2.58x10^{-8}$), near genes involved in immune response. A genetic score combining 58 independent suggestive memory risk variants was associated with increasing Alzheimer disease pathology in 725 autopsy samples. Association of memory risk loci with gene expression in 138 human hippocampus samples showed cis-associations with *WDR48* ($p=6.59x10^{-7}$) and *CLDN5* ($p=7.63x10^{-3}$), both related to ubiquitin metabolism.

Conclusions: This largest study to date exploring the genetics of memory function in ~40,000 older individuals, revealed genome-wide associations and suggested an involvement of immune and ubiquitin pathways.

## INTRODUCTION

The ability to form and retrieve memories is one of the most fundamental and complex aspects of human cognition. Decline in memory performance is a prominent marker of cognitive decline that occurs in late-life, and is one of the earliest signs of dementia.[1,2] Verbal declarative memory, the conscious recall of information that can be retrieved verbally, can be measured using word list and paragraph recall tests. The delayed recall performance of these tests is a powerful predictor of Alzheimer disease (AD).[3] Cognitive ability and memory performance were shown to be highly heritable.[4-7] However, few consistent genetic associations have been described, mostly assessed by candidate gene association studies.[8,9] Three genome-wide association studies (GWAS) of verbal declarative memory, on overlapping samples of 333 to 1,073 young adults in their twenties, have identified associations of genetic variants in the *KIBRA* and *CTNNBL1* genes with delayed recall.[10,11] No GWAS of verbal declarative memory delayed recall performance has been performed in older individuals to our knowledge.

Genetic determinants of verbal declarative memory are likely to differ between young and old individuals, although some may be shared across age groups.[12] In young adults, developmental genes determining the neural networks required for learning, storage and retrieval, or genes involved in the molecular mechanisms of memory storage,[13] could be expected to harbor most susceptibility variants. In older individuals, variants in genes involved in brain aging and neurodegenerative disease may be more likely revealed.[14] Our aim was to identify genetic variants associated with memory performance occurring in middle and later life. We conducted a meta-analysis of GWAS for delayed recall performance in tests of verbal declarative memory, in 29,076 older community-based individuals, and sought replication and extension of findings in 13,998 independent older participants (10,617 of European descent and 3,381 African-Americans) and 1,561 young adults.

## METHODS AND MATERIALS

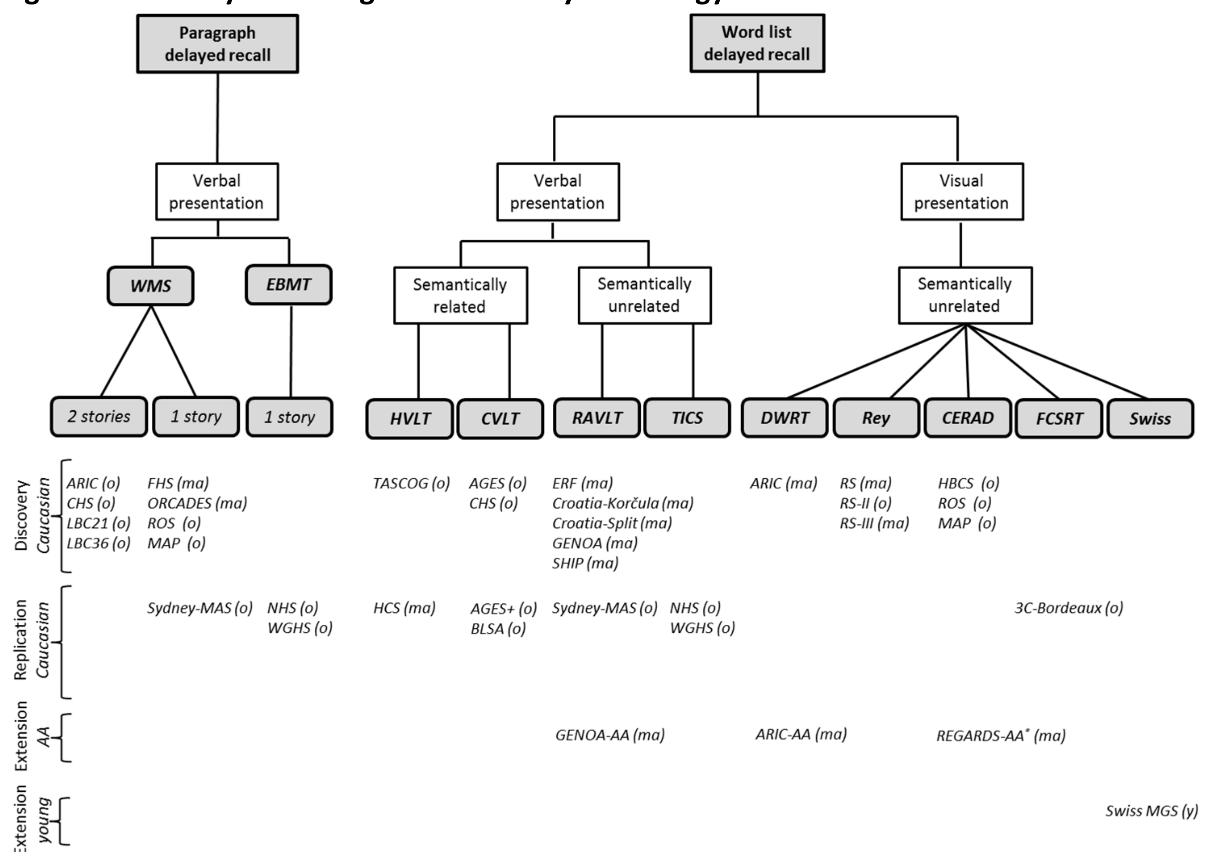
### *GWAS study population*

Analyses were performed in 19 population-based cohorts participating in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium (Supplemental Section 2, Table S1). All subjects were aged ≥45 years, and dementia- and stroke-free at cognitive assessment. The study population comprised 29,076 participants of European ancestry, including 6,674 participants with paragraph recall and 24,604 participants with word list recall tests. Each cohort secured approval from institutional review boards, and all participants provided written informed consent for study participation, cognitive testing, and use of DNA for genetic research. None of these studies have published GWAS for delayed recall performance in tests of verbal declarative memory before.

*Memory Tests*

Participants were administered one of two, or both types of verbal declarative memory tests: word list delayed recall (WL-dr) and paragraph delayed recall (PAR-dr). WL-dr comprised tests using visually or verbally presented word lists, with or without semantic relatedness between the words; PAR-dr comprised tests using one or two verbally presented stories (Figure 1, Table S2). Participants were asked to remember as many words or paragraph elements as possible after a specified delay interval, preceded by an immediate recall task (Supplemental Section 3). We decided *a priori* to run both global meta-analyses combining all tests and meta-analyses combining similar tests. Indeed, different memory tests may involve partly distinct neural networks and mechanisms (Supplemental Section 3). Meta-analyses thus comprised a combination of: all measures of delayed recall (ALL-dr, N=29,076), PAR-dr (N=6,674), WL-dr (N=24,604), and various subtypes of WL-dr tests, including Consortium to Establish a Registry for Alzheimer's Disease delayed recall (CERAD-dr, N=4,274), Delayed Word Recall Test (DWRT-dr, N=9,188), Rey Auditory Verbal Learning Test (RAVLT-dr, N=4,274), California Verbal Learning Test (CVLT-dr, N=2,950), and Hopkins Verbal Learning Test (HVLT-dr, N=331) (Figure 1).

**Figure 1: Memory test categories and analysis strategy**




*ma: "middle-aged" cohorts aged >45 years, with an average age <65 years; o: "old" cohorts aged >65 years; y: "young" cohorts aged <45 years; * in REGARDS the CERAD-dr test was administered by phone*

*Genotyping and imputation*

The consortium was formed after the individual studies had finalized their GWAS platforms, hence the studies included used different platforms. Genotyping platforms are described in Table S4. Imputation to non-monomorphic, autosomal SNPs from HapMap's CEU panel was performed with standard quality control filters (Supplemental Section 4, Table S5-S6). *APOE-ε* genotypes are not available on the GWAS arrays, however *APOEε* had been genotyped separately in most cohorts.

*Discovery GWAS*

Within each cohort, a linear regression model was used to evaluate the association of raw scores for delayed recall (number of words or story elements recalled) with the number of minor alleles (0 to 2) at each SNP. Analyses were adjusted for age and sex, and, if relevant, study site, familial structure and population substructure (Supplemental Section 5). We additionally adjusted for educational achievement (Table S1) in a secondary model only, as this could weaken associations with developmental genes that may impact educational attainment.

We undertook meta-analyses with METAL,[15] using inverse-variance weighted meta-analysis to combine GWAS for the same memory test, and effective sample size weighted meta-analysis to combine GWAS for non-identical memory tests (Supplemental Section 5). Effective sample size weighted meta-analysis is recommended when the dependent variable is measured on different scales between cohorts and does not yield any directly comparable effect estimate. For each SNP the z-statistic was weighted by the effective sample size (product of the sample size and the ratio of the empirically observed dosage variance to the expected binomial dosage variance for imputed SNPs). A combined estimate was obtained by summing the weighted z-statistics and dividing by the summed weights. Genomic control was used to remove residual population stratification within cohorts and in the meta-analysis.

*Replication and extension*

We sought to replicate our strongest association signals ($p<5x10^{-6}$) in seven independent population-based cohorts including 10,617 participants of European descent. We also attempted to extend our findings to 3,811 individuals of African-American ancestry and to 1,561 young adults in their twenties (Supplemental Section 2, Table S1; for genotyping methods, Supplemental Section 4, Tables S4-S6).

*Functional studies*

<u>Expression quantitative trait loci (eQTL)</u>

We analyzed the association of suggestive memory risk loci from the discovery GWAS ($p<5x10^{-6}$) with hippocampus cell line whole-genome gene expression profiles. This information was derived from 138 human pre-mortem hippocampus samples.[16] Exploratory

p-value thresholds of 0.05 were used for *cis*-eQTL (distance between SNP and probe <2 Mb) and $10^{-4}$ for *trans*-eQTLs (Supplemental Section 6).

Pathway analysis

Pathway analyses were carried out using the core analysis function of the Ingenuity Pathway Analysis software (IPA©, Ingenuity Systems). We performed gene-based tests for association based on results from the PAR-dr and WL-dr discovery GWAS, using the Versatile Gene-based Association Study (VEGAS) software.[17] The full list of genes and gene-based p-values generated by VEGAS was uploaded into IPA for use as a reference set (16,965 genes were available for the PAR-dr analysis and 16,953 for the WL-dr analysis). From this list p-value cut-offs of 0.01 or 0.05 were used to identify IPA focus molecules (Supplemental Section 7). Networks generated by IPA provide insight into the molecular interactions of the focus molecules, independent of any predictions of biological function. For this analysis only "direct" interactions were used, that is, where there is physical contact between the two molecules. The network building algorithm ranks focus molecules by interconnectedness (the number of triangular connections with other pairs of genes). The top molecule is taken as a seed gene and additional focus molecules are added, prioritizing those that have most overlap with the existing network. The default IPA network size of 35 nodes was used (Supplementary Section 7).

*Association with AD pathology*

We examined the relation of memory risk alleles and a memory genetic score with AD pathology (intracellular neurofibrillary tangles and extracellular amyloid plaques) in the Religious Orders Study and the Rush Memory and Aging Project (N=725, Supplemental Section 8).[18] The memory genetic score comprised all independent SNPs ($r^2$<0.25) associated with memory performance at $p<5x10^{-6}$ (Supplemental Section 9).

*Candidate genes*

*APOE* locus

Because *APOEε4* is the major known genetic risk factor for AD, we explored its association with memory performance in discovery and replication studies where this genotype was available (N=33 403 for WL-dr, N=13,170 for PAR-dr), comparing *APOEε4*-carriers to non-carriers. Secondary analyses stratified on the cohorts' mean age ($\geq$ vs. <65) were run.

Published memory and AD risk alleles

In the discovery GWAS, we tested whether memory performance was associated with published memory susceptibility SNPs and confirmed AD risk variants other than *APOEε4* (Tables S11-S12). Bonferroni correction for the number of independent SNPs tested was used, corresponding to significance thresholds of p<0.0029 for memory risk variants and p<0.0056 for AD risk variants. We extracted gene-based p-values (obtained with VEGAS[17]) for genes closest to the aforementioned candidate SNPs.

**RESULTS**

*GWAS of verbal declarative memory*

GWAS meta-analyses of verbal declarative memory comprised in total 29,076 participants (mean age 63.6±7.0 years, 56·0% women, see Table S1 for detailed demographic characteristics). Quantile-quantile (QQ) plots showed no evidence of spurious inflation of p-values or significant population stratification (Figure S1). Genome-wide plots of p-values for SNPs against their genomic position are shown in Figure S2. For replication, 10,617 participants (72.8±5.4 years, 78.3% women) were available.

Two loci reached genome-wide significance (Tables 1 and 2): on chromosome 19q12 (rs4420638, $p=1.94 \times 10^{-10}$) with PAR-dr, and on chromosome 5q11 (rs13358049, $p=9.69 \times 10^{-9}$) with CVLT-dr. The top SNPs on chromosome 19q12 are in linkage disequilibrium (LD) with *APOEε4* and are known to be associated with an increased risk of AD. They were significantly replicated ($p=5.65 \times 10^{-8}$ for rs4420638). Although the direction of effect was consistent, the chromosome 5q11 locus was not confirmed in the replication sample.

In total, 174 SNPs reached a suggestive p-value ($p<5 \times 10^{-6}$) in at least one GWAS meta-analysis (Table S7), representing 58 independent loci ($r^2<0.25$). The association between RAVLT-dr and rs11074779 (chromosome 16p12, near *HS3ST4*) was replicated ($p=1.08 \times 10^{-3}$) and reached genome-wide significance when combining discovery and replication cohorts ($p=3.11 \times 10^{-8}$, Table 3, Figure 2). Association of CERAD-dr with rs6813517 (chromosome 4q32.3, near *SPOCK3*) was significant in an African-American extension sample (p=0.020), and reached genome-wide significance when combining with the discovery GWAS ($p=2.58 \times 10^{-8}$, Table 4, Figure 2). No European replication sample was available for CERAD-dr. Three additional SNPs were replicated at $p<0.05$ with effects in the same direction, without reaching genome-wide significance (Tables 1-2): rs9528384 (chromosome 13q21), rs1633735 (chromosome 5p15) with PAR-dr, and rs13166268 (chromosome 5q23) with CVLT-dr. The above findings were not significant in a sample of 1,561 young adults (22.3±3.4 years, 69.1% women, Table S7).

*APOE and delayed recall performance*

The *APOE* locus (i.e. *APOEε4* and GWAS SNPs in LD) yielded genome-wide significant associations with PAR-dr, while associations with WL-dr were much less significant ($p>5 \times 10^{-6}$), despite a larger sample size (Table 1 and 5). When restricting analyses to older cohorts (mean age >65), associations of *APOEε4* with memory performance were significantly strengthened ($p=1.22 \times 10^{-5}$ and $p=2.64 \times 10^{-3}$ for difference with WL-dr and PAR-dr in young cohorts), still reaching weaker significance levels for WL-dr ($p=6.25 \times 10^{-11}$) than for PAR-dr ($p=3.84 \times 10^{-20}$). This remained true when restricting analyses to studies where the same participants underwent both tests of WL-dr ($p=1.34 \times 10^{-17}$) and PAR-dr ($p=6.80 \times 10^{-7}$). Associations of *APOEε4* with memory performance were substantially weakened after adjustment for the most significant GWAS proxy (rs4420638), suggesting no additional independent signal at this locus (Table 5).

## Table 1: Most significant genetic associations with *Paragraph delayed recall (PAR-dr)*

| | | | | | | | | | | Discovery | | Replication | | Meta | | Extension AA | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | *N= 6,674* | | *N= 6,496* | | *N= 13,170* | | *N= 416* | |
| **SNP** | **chr** | **Position** | **function** | **gene** | **dist[‡]** | **EA** | **EAF** | **SNPs[*]** | **dir[§]** | ***p-value*** | **p(het)** | **dir[||]** | ***p-value*** | **dir[¶]** | ***p-value*** | **dir** | ***p-value*** |
| *Adjusted for age and gender* | | | | | | | | | | | | | | | | | |
| **rs4420638** | 19 | 50114786 | downstream | *APOC1* | 340.0 | A | 0.82 | 1 | ++++++++ | $5.6x10^{-10}$ | 0.150 | +++ | $5.7x10^{-8}$ | ++ | $1.4x10^{-16}$ | + | 0.710 |
| **rs6857** | 19 | 50084094 | 3'UTR | *PVRL2* | wg | T | 0.16 | 1 | ------+- | $3.1x10^{-7}$ | $7.7x10^{-4}$ | --- | $3.9x10^{-7}$ | -- | $5.3x10^{-13}$ | - | 0.067 |
| **rs2075650** | 19 | 50087459 | intronic | *TOMM40* | wg | A | 0.86 | 1 | ++++++-+ | $3.2x10^{-7}$ | $9.7x10^{-4}$ | +++ | $8.5x10^{-6}$ | ++ | $1.5x10^{-11}$ | n.a. | n.a. |
| **rs13172717** | 5 | 21807809 | intronic | *CDH12* | wg | T | 0.46 | 3 | -------- | $6.2x10^{-7}$ | 0.420 | +-+ | 0.390 | +- | $1.5x10^{-3}$ | + | 0.390 |
| **rs9528384** | 13 | 61200566 | intergenic | *PCDH20* | 313.3 | A | 0.71 | 1 | ++++++-+ | $1.5x10^{-6}$ | 0.540 | -++ | 0.040 | ++ | $1.1x10^{-6}$ | - | 0.130 |
| **rs1633735** | 5 | 8596190 | intergenic | *SEMA5A* | 491.9 | T | 0.25 | 1 | -+------ | $4.9x10^{-6}$ | 0.590 | --- | 0.046 | -- | $3.1x10^{-6}$ | + | 0.750 |
| *Adjusted for age, gender and education* | | | | | | | | | | | | | | | | | |
| **rs4420638** | 19 | 50114786 | downstream | *APOC1* | 0.3 | A | 0.82 | 1 | ++++++++ | $1.9x10^{-10}$ | 0.370 | +++ | $1.7x10^{-7}$ | ++ | $1.6x10^{-16}$ | + | 0.620 |
| **rs6857** | 19 | 50084094 | 3'UTR | *PVRL2* | wg | T | 0.16 | 1 | ------+- | $5.4x10^{-8}$ | $3.9x10^{-3}$ | --- | $1.4x10^{-6}$ | -- | $4.2x10^{-13}$ | - | 0.043 |
| **rs2075650** | 19 | 50087459 | intronic | *TOMM40* | wg | A | 0.86 | 1 | ++++++-+ | $6.8x10^{-8}$ | $2.9x10^{-3}$ | +++ | $2.7x10^{-5}$ | ++ | $1.4x10^{-11}$ | n.a. | n.a. |
| **rs11720125** | 3 | 39177345 | intergenic | *AXUD1* | 7.2 | A | 0.78 | 8 | ++++++++ | $4.5x10^{-7}$ | 0.360 | ++- | 0.640 | ++ | $7.5x10^{-5}$ | n.a. | n.a. |
| **rs11711871** | 3 | 39203047 | NS-coding | *XIRP1* | wg | T | 0.78 | 4 [†] | ++++++++ | $9.6x10^{-7}$ | 0.340 | ++- | 0.500 | ++ | $6.4x10^{-5}$ | n.a. | n.a. |
| **rs1913243** | 3 | 39239499 | intergenic | *XIRP1* | 30.4 | T | 0.29 | 2 | -------- | $2.3x10^{-6}$ | 0.470 | +-+ | 0.390 | -+ | $6.2x10^{-3}$ | + | 0.170 |
| **rs2280630** | 3 | 39170968 | upstream | *AXUD1* | 0.9 | T | 0.32 | 1 | ----+--- | $2.6x10^{-6}$ | 0.290 | --- | 0.058 | -- | $2.5x10^{-6}$ | - | 0.700 |
| **rs9870795** | 3 | 39218299 | intergenic | *XIRP1* | 9.2 | T | 0.23 | 1 | -------- | $4.9x10^{-6}$ | 0.600 | --- | 0.320 | -- | $7.5x10^{-5}$ | + | 0.011 |
| **rs13172717** | 5 | 21807809 | intronic | *CDH12* | wg | T | 0.46 | 3 | -------- | $1.4x10^{-6}$ | 0.340 | +-+ | 0.450 | -+ | $1.8x10^{-3}$ | + | 0.380 |
| **rs9907597** | 17 | 56790279 | intronic | *BCAS3* | wg | A | 0.28 | 1 | -------- | $1.8x10^{-6}$ | 0.900 | --- | 0.640 | -- | $1.8x10^{-4}$ | + | 0.370 |
| **rs1990292** | 17 | 56799540 | intronic | *BCAS3* | wg | A | 0.79 | 1 | +++-++-+ | $4.6x10^{-6}$ | 0.190 | ++- | 0.940 | +- | $1.6x10^{-3}$ | - | 0.780 |
| **rs4555854** | 5 | 38752487 | intergenic | *LIFR* | 121.2 | T | 0.73 | 3 | ++------ | $1.9x10^{-6}$ | 0.060 | -++ | 0.550 | -+ | $2.7x10^{-3}$ | - | 0.680 |
| **rs6672300** | 1 | 51696814 | syn-coding | *EPS15* | wg | T | 0.29 | 1 | ++++++++ | $3.9x10^{-6}$ | 0.780 | --- | 0.024 | +- | 0.085 | + | 0.059 |
| **rs10493155** | 1 | 51711785 | intronic | *EPS15* | wg | T | 0.24 | 1 | ++++++++ | $4.4x10^{-6}$ | 0.830 | --- | 0.064 | +- | 0.044 | + | 0.100 |

*Only SNPs with $p<5x10^{-6}$ are shown (also see Table S7); AA: African-Americans; dir: direction, "-" means EA is associated with lower score, "+" means EA is associated with higher score; EA: effect allele; EAF: effect allele frequency; wg: within gene; [*] number of SNPs in locus (top SNP + SNPs in LD r2 > 0·80); [†] of which one other non-synonymous coding (rs3732383); ‡ distance in kilobase; [§] directions are in the following order: ARIC, CHS, FHS, LBC21, LBC36, MAP, ORCADES, ROS; [||] Directions are in the following order: NHS, Sydney MAS, WGHS*

## Table 2: Most significant genetic associations with performance on *California Verbal Learning Test delayed recall* (*CVLT-dr*)

| | | | | | | | | | Discovery | | | | | Replication | | All | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | ***N= 2,950*** | | | | | ***N= 2,237*** | | ***N= 5,187*** | |
| **SNP** | **chr** | **position** | **function** | **gene** | **dist†** | **EA** | **EAF** | **SNPs*** | **dir‡** | **beta** | **SE** | ***p-value*** | **p(het)** | **dir§** | ***p-value*** | **dir‖** | ***p-value*** |
| ***Adjusted for age and gender*** | | | | | | | | | | | | | | | | | |
| rs13358049 | 5 | 50578216 | intergenic | *ISL1* | 136.5 | T | 0.07 | 15 | -- | -0.85 | 0.15 | 2.29x10$^{-8}$ | 0.58 | +- | 0.97 | -- | 2.71x10$^{-5}$ |
| rs13360092 | 5 | 50564932 | intergenic | *ISL1* | 149.8 | A | 0.07 | 15 | -- | -0.85 | 0.15 | 2.36x10$^{-8}$ | 0.55 | ?- | 0.73 | -- | 1.25x10$^{-7}$ |
| rs157092 | 20 | 55685666 | intronic | *TMEPAI* | wg | T | 0.26 | 4 | -- | -0.46 | 0.09 | 8.50x10$^{-7}$ | 0.29 | +- | 0.82 | -+ | 5.88x10$^{-4}$ |
| rs13177865 | 5 | 120494952 | intergenic | *PRR16* | 444.1 | A | 0.09 | 2 | +? | 0.74 | 0.15 | 1.29x10$^{-6}$ | 1 | ?+ | 0.79 | ++ | 1.03x10$^{-5}$ |
| rs13166268 | 5 | 120387817 | intergenic | *PRR16* | 337.0 | C | 0.90 | 2 | -+ | -0.60 | 0.13 | 7.20x10$^{-6}$ | 0.006 | -- | 0.016 | -- | 8.28x10$^{-7}$ |
| rs1445765 | 5 | 103572875 | intergenic | *NUDT12* | 646.5 | T | 0.69 | 1 | -- | -0.40 | 0.09 | 4.29x10$^{-6}$ | 0.58 | -+ | 0.45 | -- | 9.42x10$^{-5}$ |
| ***Adjusted for age, gender and education*** | | | | | | | | | | | | | | | | | |
| rs13358049 | 5 | 50578216 | intergenic | *ISL1* | 136.5 | T | 0.07 | 16 | -- | -0.85 | 0.15 | 9.69x10$^{-9}$ | 0.66 | +- | 0.89 | -- | 1.09x10$^{-5}$ |
| rs13360092 | 5 | 50564932 | intergenic | *ISL1* | 149.8 | A | 0.07 | 16 | -- | -0.85 | 0.15 | 1.00x10$^{-8}$ | 0.63 | ?- | 0.51 | -- | 2.71x10$^{-8}$ |
| rs157092 | 20 | 55685666 | intronic | *TMEPAI* | wg | T | 0.26 | 5 | -- | -0.46 | 0.09 | 3.69x10$^{-7}$ | 0.26 | +- | 0.81 | -+ | 3.50x10$^{-4}$ |
| rs1445765 | 5 | 103572875 | intergenic | *NUDT12* | 646.5 | T | 0.69 | 2 | -- | -0.39 | 0.08 | 3.29x10$^{-6}$ | 0.52 | -+ | 0.55 | -- | 1.13x10$^{-4}$ |

*Only SNPs with p<5x10-6 are shown (also see Table S7); dir: direction, "-" means EA is associated with lower score on CVLT-dr, "+" means EA is associated with higher score on CLVT-dr; "?" means this SNP was not available in the corresponding dataset; EA: effect allele; EAF: effect allele frequency; wg: within gene; * number of SNPs in locus (top SNP + SNPs in LD r2 > 0·80); † distance in kilobase; ‡ directions are in the following order: AGES, CHS; § directions are in the following order: AGES-plus (N=1,525), BLSA (N=712); || directions are in the following order: Discovery, Replication; ¶ rs13166268 was added to the list of SNPs for replication despite a p-value slightly above the 5x10-6 threshold to serve as a proxy for rs13177865 (r2=1), with data available in both discovery and both replication cohorts (rs13177865 is available only in one discovery and one replication cohort).*

**Table 3: Most significant genetic associations with *Rey's Auditory Verbal Learning Test* delayed recall (*RAVLT-dr*)**

| | | | | | | | | | | Discovery | | | | Replication | | Meta | | Extension AA | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | ***N= 4,274*** | | | | ***N= 880*** | | ***N= 5,154*** | | ***N= 708*** | |
| **SNP** | **chr** | **position** | **function** | **gene** | **dist**[†] | **EA** | **EAF** | **dir**[‡] | **SNPs**[*] | **beta** | **SE** | ***p-value*** | **p(het)** | **dir**[§] | ***p-value*** | **dir**[¶] | ***p-value*** | **dir**[||] | ***p-value*** |
| ***Adjusted for age and gender*** | | | | | | | | | | | | | | | | | | | |
| rs963798 | 13 | 38287693 | intronic | *FREM2* | wg | A | 0.52 | +++++ | 3 | 0.29 | 0.06 | $1.46x10^{-6}$ | 0.65 | - | 0.34 | +- | $2.00x10^{-5}$ | + | 0.92 |
| rs11074779 | 16 | 26358944 | intergenic | *HS3ST4* | 302.4 | T | 0.81 | +++++ | 1 | 0.36 | 0.08 | $3.05x10^{-6}$ | 0.93 | + | $1.08x10^{-3}$ | ++ | $3.11x10^{-8}$ | - | 0.48 |
| rs5747035 | 22 | 16098606 | intergenic | *CECR1* | 27.8 | T | 0.93 | ----- | 1 | -0.65 | 0.14 | $3.50x10^{-6}$ | 0.89 | - | 0.19 | -- | $1.38x10^{-6}$ | + | 0.93 |
| rs11237982 | 11 | 79119342 | intergenic | *ODZ4* | 918.5 | T | 0.85 | +++++ | 1 | 0.40 | 0.09 | $4.33x10^{-6}$ | 0.29 | + | 0.099 | ++ | $9.73x10^{-7}$ | + | 0.31 |
| ***Adjusted for age, gender and education*** | | | | | | | | | | | | | | | | | | | |
| rs5747035 | 22 | 16098606 | intergenic | *CECR1* | 27.8 | T | 0.93 | ----- | 1 | -0.65 | 0.14 | $2.61x10^{-6}$ | 0.92 | - | 0.18 | -- | $1.03x10^{-6}$ | - | 0.97 |
| rs963798 | 13 | 38287693 | intronic | *FREM2* | wg | A | 0.52 | +++++ | 1 | 0.28 | 0.06 | $2.71x10^{-6}$ | 0.47 | - | 0.27 | +- | $4.15x10^{-5}$ | - | 0.87 |
| rs11237982 | 11 | 79119342 | intergenic | *ODZ4* | 918.5 | T | 0.85 | +++++ | 1 | 0.39 | 0.08 | $3.68x10^{-6}$ | 0.22 | + | 0.087 | ++ | $7.63x10^{-7}$ | + | 0.29 |
| rs16991213 | 20 | 44402546 | intergenic | *SLC35C2* | 9.0 | A | 0.07 | +++++ | 1 | 0.57 | 0.12 | $4.81x10^{-6}$ | 0.93 | - | $8.63x10^{-4}$ | +- | $1.10x10^{-3}$ | - | 0.84 |

*Only SNPs with $p<5x10^{-6}$ are shown (also see Table S7); AA: African-Americans (GENOA-AA); dir: direction, "-" means EA is associated with lower score on RAVLT-dr, "+" means EA is associated with higher score on RALVT-dr; EA: effect allele; EAF: effect allele frequency; wg: within gene; [*] number of SNPs in locus (top SNP + SNPs in LD r2 > 0·80); [†] distance in kilobase; [‡] directions are in the following order: ERF, GENOA, Croatia-Korcula, SHIP, Croatia-Split; [§] Sydney MAS; [||] GENOA-AA; [¶] directions are in the following order: Discovery, Replication in European cohorts (meta-analysis results combining all European and AA samples are given in Table S7)*

**Table 4: Most significant genetic associations with *Consortium to Establish a Registry for Alzheimer's Disease* delayed recall (CERAD-dr)**

| | | | | | | | | | | Meta Discovery *N= 4,274* | | | | Extension AA *N= 627* | | | Meta All *N= 4,901* | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SNP** | **chr** | **position** | **function** | **gene** | **dist**† | **EA** | **EAF** | **dir**‡ | **SNPs*** | **beta** | **SE** | ***p-value*** | **p(het)** | **beta** | **SE** | ***p-value*** | **beta** | **SE** | ***p-value*** |
| ***Adjusted for age and gender*** | | | | | | | | | | | | | | | | | | | |
| rs6813517 | 4 | 168759326 | intergenic | *SPOCK3* | 367.0 | T | 0.79 | +++ | 5 | 0.36 | 0.07 | $4.96 \times 10^{-7}$ | 0.57 | 0.34 | 0.15 | 0.026 | 0.36 | 0.07 | $3.49 \times 10^{-8}$ |
| rs298210 | 8 | 65452923 | upstream | *BHLHB5* | 202.4 | A | 0.05 | --- | 1 | -0.67 | 0.14 | $1.04 \times 10^{-6}$ | 0.84 | 0.52 | 0.52 | 0.32 | -0.59 | 0.13 | $7.05 \times 10^{-6}$ |
| rs6046393 | 20 | 19800250 | intergenic | *RIN2* | 18.0 | T | 0.8 | +++ | 4 | 0.35 | 0.07 | $1.47 \times 10^{-6}$ | 0.11 | 0.14 | 0.12 | 0.24 | 0.30 | 0.06 | $1.96 \times 10^{-6}$ |
| rs4292676 | 8 | 21564528 | intergenic | *GFRA2* | 29.3 | T | 0.61 | +++ | 3 | 0.28 | 0.06 | $3.48 \times 10^{-6}$ | 0.72 | 0.20 | 0.17 | 0.23 | 0.27 | 0.06 | $1.69 \times 10^{-6}$ |
| rs1890709 | 14 | 48171583 | intergenic | *RPS29* | 942.2 | A | 0.29 | +++ | 1 | 0.28 | 0.06 | $4.28 \times 10^{-6}$ | 0.73 | -0.03 | 0.14 | 0.83 | 0.23 | 0.06 | $3.49 \times 10^{-5}$ |
| rs10894804 | 11 | 133753052 | downstream | *B3GAT1* | 0.6 | A | 0.53 | --- | 1 | -0.31 | 0.07 | $4.31 \times 10^{-6}$ | 0.22 | -0.11 | 0.16 | 0.50 | -0.28 | 0.06 | $6.25 \times 10^{-6}$ |
| rs17053482 | 4 | 168836226 | intergenic | *ANXA10* | 414.1 | C | 0.09 | --- | 4 | -0.48 | 0.10 | $4.79 \times 10^{-6}$ | 0.38 | 0.10 | 0.14 | 0.47 | -0.28 | 0.08 | $1.06 \times 10^{-3}$ |
| ***Adjusted for age, gender and education*** | | | | | | | | | | | | | | | | | | | |
| rs6813517 | 4 | 168759326 | intergenic | *SPOCK3* | 367.0 | T | 0.79 | +++ | 5 | 0.37 | 0.07 | $4.12 \times 10^{-7}$ | 0.62 | 0.35 | 0.15 | 0.020 | 0.37 | 0.07 | $2.58 \times 10^{-8}$ |
| rs298210 | 8 | 65452923 | upstream | *BHLHB5* | 202.4 | A | 0.05 | --- | 1 | -0.69 | 0.14 | $6.83 \times 10^{-7}$ | 0.95 | 0.44 | 0.52 | 0.40 | -0.62 | 0.13 | $4.45 \times 10^{-6}$ |
| rs17053482 | 4 | 168836226 | intergenic | *ANXA10* | 414.1 | C | 0.09 | --- | 6 | -0.51 | 0.11 | $162 \times 10^{-6}$ | 0.34 | 0.10 | 0.14 | 0.48 | -0.29 | 0.08 | $6.27 \times 10^{-4}$ |
| rs10894804 | 11 | 133753052 | downstream | *B3GAT1* | 0.6 | A | 0.53 | --- | 1 | -0.32 | 0.07 | $2.78 \times 10^{-6}$ | 0.19 | -0.10 | 0.16 | 0.54 | -0.28 | 0.06 | $5.35 \times 10^{-6}$ |

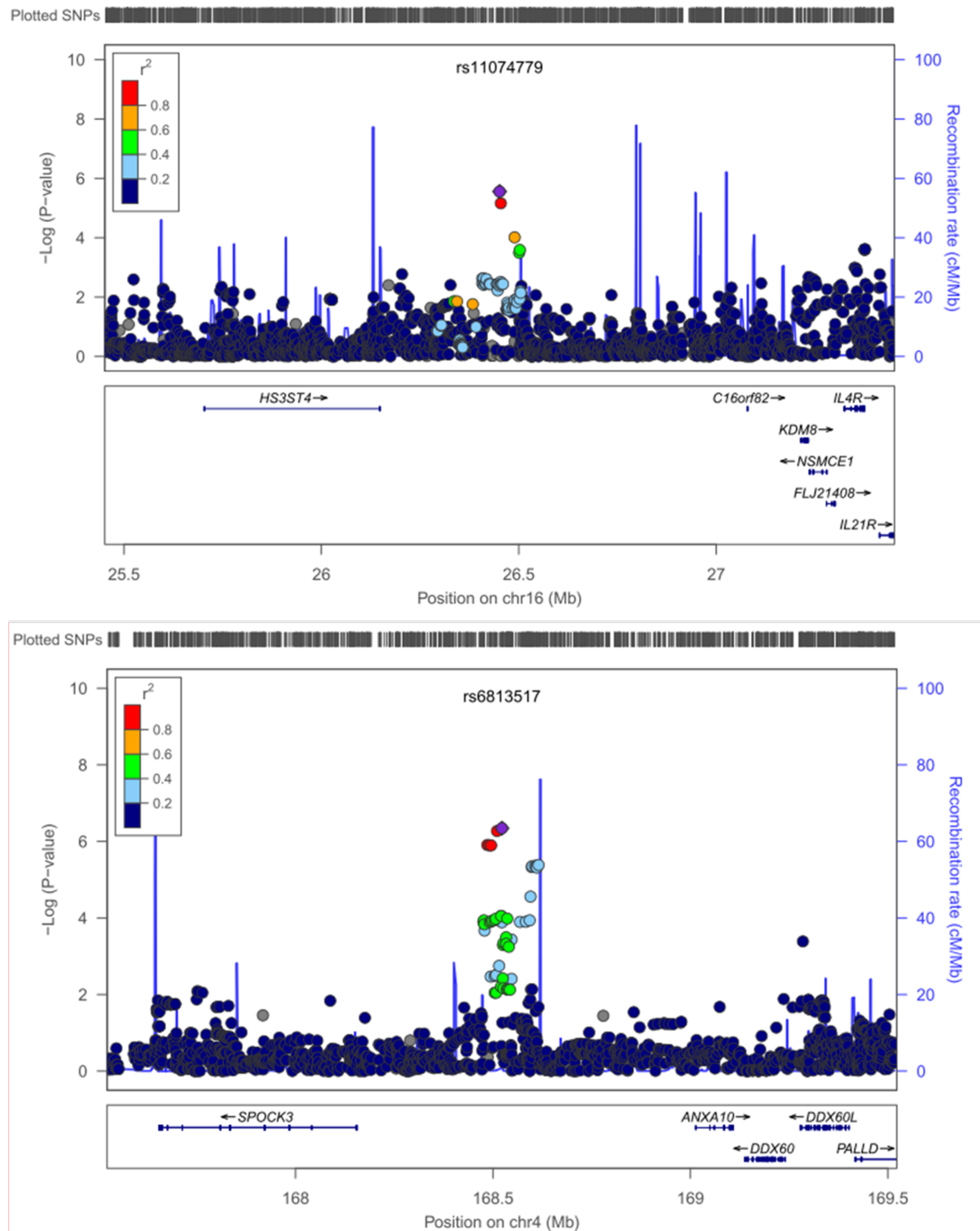
*Only SNPs with p<5x10-6 and available in the AA extension sample are shown (also see Table S7); AA: African-Americans (REGARDS-AA); dir: direction, "-" means EA is associated with lower score on CERAD-dr, "+" means EA is associated with higher score on CERAD-dr; EA: effect allele; EAF: effect allele frequency; wg: within gene; *number of SNPs in locus (top SNP + SNPs in LD r2 > 0·80); †distance in kilobase; ‡directions are in the following order: HBCS, MAP, ROS*

## Table 5: Association of *APOE* variants with memory performance according to test-type, age, and ethnicity

| | | | | All European | | Old[*] European | | Old[1] European: overlapping samples for PAR and WL[§] | | Middle-aged European | | African-American | | Young European* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | **PAR-dr** | **WL-dr** | **PAR-dr** | **WL-dr** | **PAR-dr** | **WL-dr** | **PAR-dr** | **WL-dr** | **PAR-dr[†]** | **WL-dr[‡]** | **WL-dr** |
| | | | | **Age: 71.9** | **Age: 65.4** | **Age: 74.0** | **Age: 73.3** | **Age: 74.3** | **Age: 74.4** | **Age: 64.6** | **Age: 58.8** | **Age: 71.8** | **Age: 57.6** | **Age: 22.3** |
| | | | | **N: 13,170** | **N: 33,403** | **N: 10,258** | **N: 15,118** | **N: 7,934** | **N: 7,641** | **N: 2,912** | **N: 18,285** | **N: 416** | **N: 3,184** | **N: 1,561** |
| **AD Risk Allele** | **Position** | **Gene** | **Model** | **p (PAR)** | **p (WL)** | **p (PAR)** | **p (WL)** | **p (PAR)** | **p (WL)** | **p (PAR)** | **p (WL)** | **p (PAR)** | **p (WL)** | **p (WL)** |
| rs4420638-G | 19:50114786 | *APOC1* | A | $4.18x10^{-10}$ | $2.57x10^{-4}$ | $8.05x10^{-17}$ | $6.26x10^{-10}$ | $1.68x10^{-13}$ | $3.97x10^{-6}$ | 0.23 | 0.10 | 0.71 | 0.60 | 0.79 |
| rs6857-T | 19:50084094 | *PVRL2* | A | $2.51x10^{-7}$ | $1.24x10^{-3}$ | $2.41x10^{-14}$ | $5.09x10^{-7}$ | $8.08x10^{-12}$ | $2.03x10^{-4}$ | 0.83 | 0.13 | 0.07 | 0.62 | n.a. |
| rs2075650-G | 19:50087459 | *TOMM40* | A | $2.58x10^{-7}$ | $3.03x10^{-3}$ | $2.52x10^{-12}$ | $1.88x10^{-5}$ | $3.55x10^{-10}$ | $1.57x10^{-3}$ | 0.56 | 0.10 | n.a. | 0.39 | n.a. |
| Epsilon-4[\|\|] | 19:50103781/ 50103919 | *APOE* | A | $2.67x10^{-21}$ | $1.28x10^{-7}$ | $3.84x10^{-20}$ | $6.25x10^{-11}$ | $1.34x10^{-17}$ | $6.80x10^{-7}$ | $2.64x10^{-3}$ | 0.28 | 0.08 | 0.23 | n.a. |
| | | | B | $6.58x10^{-23}$ | $6.53x10^{-9}$ | $6.55x10^{-21}$ | $2.07x10^{-12}$ | $2.48x10^{-18}$ | $4.07x10^{-7}$ | $5.77x10^{-4}$ | 0.18 | 0.12 | 0.23 | n.a. |
| | | | C | $6.44x10^{-7}$ | 0.47 | $3.91x10^{-5}$ | $3.33x10^{-3}$ | $4.06x10^{-5}$ | 0.04 | $4.85x10^{-3}$ | 0.49 | 0.07 | 0.17 | n.a. |
| | | | D | $1.00x10^{-7}$ | 0.09 | $2.09x10^{-5}$ | $2.36x10^{-3}$ | $2.47x10^{-5}$ | 0.03 | $1.14x10^{-3}$ | 0.72 | 0.12 | 0.14 | n.a. |

*These results include both discovery and replication cohorts, and in all analyses, except the young European sample, the AD risk allele was associated with worse memory performance; Model A: adjusted for age and gender; Model B: adjusted for age, gender, and educational achievement; Model C: adjusted for age, gender, and rs4420638; Model D: adjusted for age, gender, rs4420638, and educational achievement; AD: Alzheimer disease; PAR: paragraph delayed recall; WL: word list delayed recall; both discovery and replication cohorts with available genotypes for rs4420638, rs6857 and rs2075650 are included in this analysis; * "Young" corresponds to cohorts aged <45 years (Swiss MGS), "Middle-Aged" to cohorts aged >45 years, with an average age <65 years (PAR: FHS, ORCADES; WL: ERF, Croatia-Split, Croatia-Korcula, GENOA, SHIP, HCS, ARIC, RS, RS-III), and "Old" by an average age >65 years (PAR: ARIC, CHS, LBC21, LBC36, ROS, MAP, NHS, WGHS, Sydney MAS; WL: AGES, CHS, TASCOG, BLSA, RS-II, HBCS, 3C-Bordeaux, ROS, MAP, NHS, WGHS, Sydney MAS); † ARIC-AA; ‡ ARIC-AA and GENOA-AA; § studies with measures for both paragraph and word list delayed recall on largely overlapping samples were included in this analysis (ROS, MAP, Sydney MAS, NHS, WGHS); || the epsilon polymorphism corresponds to a haplotype of rs429358 and rs7412; ¶ some cohorts did not have epsilon genotypes available (NHS, Croatia-Split, SHIP, HBCS, HCS, ORCADES, TASCOG), conversely, AGES-repli had genotypes for the epsilon polymorphism, but not for rs4420638, rs6857, and rs2075650; thus sample sizes for these analyses were 10,685 (mean age 71.8 years) for PAR-dr and 29,669 (61.8 years) for WL-dr*

**Figure 2: Regional association plots in the discovery GWAS centered on rs11074779 (RAVLT-dr), and rs6813517 (CERAD-dr)**




*Regional plot for associations in region centered on rs11074779 (RAVLT-dr), and rs6813517 (CERAD-dr), drawn using the LocusZoom software.[19] All SNPs based on imputed results (dots) are plotted with their GWAS meta-analysis p-values against their genomic position. The color of the dots represents the linkage disequilibrium between SNPs. Purple line represents estimated recombination rates. Genes and exons are shown as dark blue arrows and vertical lines respectively*

*Functional studies*

Hippocampal eQTL

Among associations of suggestive memory risk variants with RNA expression in the human hippocampus (Table S9), the most significant *cis*-associations included rs2280630 (PAR-dr, p[eQTL]=$6.59 \times 10^{-7}$) with *WDR48*, encoding a ubiquitin-specific protease associated protein belonging to a family of deubiquitinating enzymes, and rs5747035 (RAVLT-dr, p[eQTL]=$7.63 \times 10^{-3}$) with *CLDN5*, a membranal tight junction protein, which plays a critical role in determining the permeability of endothelial barriers, and whose degradation is regulated by the ubiquitin-proteasome pathway.[20] Associations of rs2280630 and rs5747035 with PAR-dr and RAVLT-dr were in the same direction as in the discovery GWAS in the in the follow-up studies, but did not reach significance (p=0.058 and 0.19, Tables 1 and 3).

Pathway analysis

The networks reaching the highest score, i.e. the lowest chance of randomly finding the selected number of focus molecules in a network of the selected size, are shown in Figure S3. In the PAR-dr pathway analysis using a gene-based p-value cut-off of 0.01, the top 3 networks and networks 5 and 6 (ordered by decreasing scores, respectively of 33, 28, 26, 25, 24, 21) all included UBC (Ubiquitin C) as a central hub. UBC was also a central hub in networks 3, 4, 5, and 6 (respective scores 27, 26, 23 and 23) of the WL-dr pathway analysis using a gene-based p-value cut-off of 0.01. When using a gene-based p-value cut-off of 0.05 several PAR-dr and WL-dr networks also included UBC as a central hub. As UBC has a large number (N=8332) of directly related molecules, we calculated whether there are more UBC interactions in the whole focus molecule set than would be expected by chance, given the number of UBC interactions in the reference set, using the hypergeometric distribution test. The latter suggested that this overrepresentation was not due to chance (Table S10).

*Association with AD pathology*

The *APOE* locus and four intronic SNPs in LD within the *KIAA1797* gene on chromosome 9 were associated with increasing amyloid plaque burden and neurofibrillary tangle density (FDR-corrected p-value <0.05, Table S8). The memory genetic score, combining 58 independent variants associated with memory at $p<5 \times 10^{-6}$, was significantly associated with increasing amyloid plaque burden (effect estimate [β] ± standard error [SE]: 0.0103±0.0045, p=0.022) and neurofibrillary tangle density (β±SE: 0.0106±0.0035, p=0.0028). After removing the *APOE* locus from the score, the association was still significant for neurofibrillary tangle density (β±SE: 0.0079±0.0036, p=0.027).

*Candidate gene analysis*

Published memory risk variants

None of the SNPs previously reported to be associated with memory performance, mostly in young cohorts, reached a statistically significant level of association after correction for multiple testing in our middle-aged to older samples (Table S11).

Published risk variants for AD

In addition to the *APOE* locus, rs11136000-C in *CLU* was associated with weaker performance on ALL-dr (p=0.0054), and rs11767557-T near *EPHA1* with weaker performance on PAR-dr (p=0.0040, Table S12).

**DISCUSSION**

In this first GWAS of verbal declarative memory in almost 30,000 older non-demented community adults, we observed a genome-wide significant association of the *APOE* locus with poorer memory performance, especially for paragraph delayed recall. Two additional associations in subsets of the total sample and for specific tests, i.e. of rs11074779 near *HS3ST4* with RAVLT-dr, and of rs6813517 near *SPOCK3* with CERAD-dr, were replicated and reached genome-wide significance after combining discovery and replication samples. Although *APOEε4* is a well-established risk factor for AD, its association with memory performance is controversial;[9,21] some studies on relatively small samples have suggested an age-dependent detrimental effect of *APOEε4* on memory performance,[22] and even a protective effect in young adults.[23] The present findings confirm a highly significant association of *APOEε4* with poorer memory performance in the oldest cohorts, while the association did not or barely reached significance in the young and middle-aged cohorts, both in European and African-American samples. Whereas this age-dependent effect could be partly ascribed to the higher prevalence of subclinical AD with increasing age, *APOEε4* could also influence cognitive ageing independently of the mechanisms underlying AD.[14] The much stronger detrimental effect of *APOEε4* on PAR-dr than on WL-dr is intriguing. Word list and story listening tasks were shown to activate different brain regions, with unequal involvement of the right hemisphere,[24,25] and performance on these tests is significantly but not very strongly correlated (Table S3). Our data suggest that PAR-dr may perhaps better capture *APOEε4* –related decline in memory performance than WL-dr. More largely, these discrepancies have important implications when planning future genetic studies of cognition as it highlights how distinct memory tasks are, even when focusing on delayed recall performance.[26]

The association of rs11074779 with RAVLT-dr was replicated and reached genome-wide significance in the combined analysis of European discovery and replication samples. Rs11074779 is located at 302 kilobases (kb) from *HS3ST4*, which is strongly expressed in the hippocampus and is thought to play a role in Herpes Simplex Virus (HSV)-1 pathogenesis;[27] it is intriguing that HSV-1 infection of the brain (herpes simplex encephalitis) preferentially affects the hippocampus and can result in profound memory loss. Further, AD-related plaques and tangles were shown to be enriched in HSV-1 binding proteins[28] and, although controversial, a possible role of HSV-1 has been suggested in AD occurrence.[29-31] The association of rs6813517 with CERAD-dr was significant in an African-American extension cohort and reached genome-wide significance in the combined analysis of discovery and extension samples. Rs6813517 is located at 367 kilobases (kb) from *SPOCK3*, encoding a member of a novel family of calcium-binding proteoglycan proteins, which is strongly

expressed in cerebral cortex and hippocampus. Another variant near *SPOCK3* (rs13111850, $r^2$=0.003 with rs6813517) was recently found to be associated with variations in cytokine secretion in response to smallpox vaccine.[32] Interestingly, an immune system dysfunction has been suggested in AD, based on findings from recent AD GWAS and genome-wide pathway analyses.[33-35] Functional prediction analyses using the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) suggest that *HS3ST4* is implicated in synaptic transmission, neurotransmitter receptor activity, while *SPOCK3* appears to be involved in regulation of action potential in neurons, neurotransmitter uptake, and memory (http://genenetwork.nl). Two suggestive intergenic variants associated with PAR-dr (near *Protocadherin 20* [*PCDH20*] and *Semaphorin 5A* [*SEMA5A*]) reached nominal significance in the replication analysis. *PCDH20* belongs to the subfamily of non-clustered protocadherins, which likely contribute to the establishment and remodeling of selective synaptic connections, and to the maintenance and plasticity of adult hippocampal circuitry.[36] *SEMA5A* belongs to the semaphorin gene family, which is involved in axonal guidance during neural development,[37] and has been associated with autism.[38]
Results were overall similar with and without education adjustment. Associations with the *APOE* locus tended to be slightly more significant in the education-adjusted model, whereas SNPs near genes involved in neuronal development and synaptic function were more significantly associated without education adjustment, in line with the hypothesis that effects of developmental genes may be masked by correcting for educational achievement.[39] We did not confirm previously-reported associations with memory performance; however, previous studies were conducted mostly in young adults and did not focus exclusively on the delayed recall component of memory.
Besides *APOE*, another AD susceptibility gene (*CLU*) was associated with poorer delayed recall performance. CLU participates in amyloid β (Aβ) peptide clearance from the brain, as does APOE.[40,41] Associations of *APOE* and *CLU* variants with memory performance may suggest that, even in non-demented older adults, diminished performance in verbal declarative memory may be partly mediated by early pre-clinical neurodegenerative processes. This was further supported by the association of the memory genetic score with a larger burden of AD pathology.
In the pathway analyses, a larger proportion of genes appeared to bind *polyubiquitin C* than predicted by chance, and several eQTL associations with genes involved in ubiquitin metabolism were observed. Converging evidence suggests a major role of impaired protein degradation by the ubiquitin proteasome system in neurodegenerative disorders including AD.[42,43] Ubiquitination was also demonstrated to facilitate hippocampal plasticity and hippocampal-dependent memory storage by modulating *CPEB3* activity and *CPEB3*-dependent protein synthesis and synapse formation.[44]

Strengths of this study include the large sample size and the diversity of populations studied. We provide a wide array of functional analyses including assessments of shared genetic variation with AD pathology and hippocampal gene expression analyses. Limitations include

the intrinsic complexity and heterogeneity of mechanisms influencing memory performance, a major challenge in deciphering the genetics of human verbal memory.[26] We attempted to harmonize tests by close examination of each test selected for inclusion and by performing test-specific meta-analyses, although this implied smaller sample sizes, thereby reducing statistical power. Mean age sometimes differed between discovery and replication, which may have reduced our ability to replicate findings. Cognitive tests are influenced by medication, anxiety, and mood, inducing variance in the verbal memory phenotype that is not attributable to genetic variation, possibly hampering power to detect memory susceptibility variants. We used a single measurement of memory for each subject, as only one baseline examination was available in some studies. Although dementia-free, some older community-dwelling participants may have suffered from mild cognitive impairment (MCI), and some of our findings may partly reflect associations with the latter. We did not remove participants with MCI from the analyses, as MCI is part of the spectrum of cognitive brain aging, and does not necessarily reflect prodromal dementia. Moreover, given the distinct neuropsychological test batteries implemented in the different studies, defining a cut-off for MCI would have resulted in substantial between study heterogeneity. Finally, when considering a conservative Bonferroni correction for the 58 independent loci tested for replication, only the *APOE* locus reached a p-value $< 8.62 \times 10^{-4}$. While two additional loci (near *HS3ST4* and *SPOCK3*) reached genome-wide significance after combining discovery and replication, future independent studies will need to confirm these findings.

In conclusion, in a large sample of older community-dwelling adults, the *APOE* locus was associated with weaker verbal memory performance, especially in those above age 65 years. Two additional genome-wide associations, near *HS3ST4* and *SPOCK3,* were identified and other putative modulators of memory performance were revealed by a pathway approach and hippocampal gene expression analyses, warranting further exploration in independent cohorts. The differential associations according to memory test characteristics and age should be accounted for in future studies. Finally, exploring other types of genetic variation, including rare variants and epigenetic modifications, will be crucial to decipher the full spectrum of memory heritability.

**ACKNOWLEDGMENTS**

Aging Gene-Environment Susceptibility-Reykjavik Study: The research has been funded by NIA contract N01-AG-12100 with contributions from NEI, NIDCD and NHLBI, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament).

The Atherosclerosis Risk in Communities Study: The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL70825, R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.

The Austrian Stroke Prevention Study: The authors thank the staff and the participants of the ASPS for their valuable contributions. We thank Birgit Reinhart for her long-term administrative commitment and Ing Johann Semmler for the technical assistance at creating the DNA-bank.

The Cardiovascular Health Study: This CHS research was supported by NHLBI contracts HHSN268201200036C, N01HC85239, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants HL080295, HL087652, HL105756 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through AG023629, AG20098, and AG05133 from the National Institute on Aging (NIA). A full list of CHS investigators and institutions can be found at http://www.chs-nhlbi.org/pi.htm. DNA handling and genotyping at Cedars-Sinai Medical Center was supported in part by the National Center for Research Resources, grant UL1RR033176, and is now at the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124; in addition to the National Institute of Diabetes and Digestive and Kidney Disease grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

Croatian Cohorts: Split and Korčula: The CROATIA-Korčula and CROATIA-Split studies were funded by grants from the Medical Research Council (UK), European Commission Framework 6 project EUROSPAN (Contract No. LSHG-CT-2006-018947) and Republic of Croatia Ministry of Science, Education and Sports research grants to I.R. (108-1080315-0302). We would like to acknowledge the invaluable contributions of the recruitment teams in Korčula and Split, the administrative teams in Croatia and Edinburgh and the people of Korčula and Split. The SNP genotyping for the CROATIA-Korčula cohort was performed in Helmholtz Zentrum München, Neuherberg, Germany. The SNP genotyping for the CROATIA-Split cohort was performed by AROS Applied Biotechnology, Aarhus, Denmark.

Erasmus Rucphen Family Study: This study is financially supported by the Netherlands Organization for Scientific Research (NWO), the Internationale Stichting Alzheimer Onderzoek (ISAO), the Hersenstichting Nederland (HSN), and the Centre for Medical Systems Biology (CMSB *1 and 2*) in the framework of the Netherlands Genomics Initiative (NGI). We thank the participants from the Genetic Research in Isolated Populations, Erasmus Rucphen Family, who made this work possible.
Framingham Heart Study: From the Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine. This work was supported by the National Heart, Lung and Blood Institute's Framingham Heart Study (Contract No. N01-HC-25195) and its contract with Affymetrix, Inc for genotyping services (Contract No. N02-HL-6-4278). A portion of this research utilized the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. Analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. This study was also supported by grants from the National Institute of Neurological Disorders and Stroke (NS17950) and the National Institute of Aging (AG08122, AG16495, AG033193, AG031287).
Genetic Epidemiology Network of Arteriopathy (GENOA): Support for the Genetic Epidemiology Network of Arteriopathy (GENOA) was provided by the National Heart, Lung and Blood Institute (HL054464, HL054457, HL054481, HL071917, and HL87660) and the National Institute of Neurological Disorders and Stroke (NS041558) of the National Institutes of Health. Genotyping was performed at the Mayo Clinic (S.T.T., Mariza de Andrade, Julie Cunningham) and was made possible by the University of Texas Health Sciences Center (E.B., Megan L. Grove-Gaona). We would also like to thank the families that participated in the GENOA study.
Helsinki Birth Cohort Study (HBCS): We thank all study participants as well as everybody involved in the Helsinki Birth Cohort Study. Helsinki Birth Cohort Study has been supported by grants from the Academy of Finland, the Finnish Diabetes Research Society, Folkhälsan Research Foundation, Novo Nordisk Foundation, Finska Läkaresällskapet, Signe and Ane Gyllenberg Foundation, University of Helsinki, Ministry of Education, Ahokas Foundation, Emil Aaltonen Foundation, Juho Vainio Foundation, and Wellcome Trust (grant number WT089062).
Lothian Birth Cohort 1921 (LBC1921) and 1936 (LBC1936): We thank the cohort participants and team members who contributed to these studies. Phenotype collection in the Lothian Birth Cohort 1921 was supported by the BBSRC, The Royal Society, and The Chief Scientist Office of the Scottish Government. Phenotype collection in the Lothian Birth Cohort 1936 was supported by Research IntoAgeing (continues as part of Age UK The Disconnected Mind project). Genotyping of the cohorts was funded by the UK Biotechnology and Biological Sciences Research Council (BBSRC). The work was undertaken by The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council

Lifelong Health and Wellbeing Initiative (G0700704/84698). Funding from the BBSRC, Engineering and Physical Sciences Research Council (EPSRC), Economic and Social Research Council (ESRC), and MRC is gratefully acknowledged.
Orkney Complex Disease Study (ORCADES):
ORCADES was supported by the Chief Scientist Office of the Scottish Government, the Royal Society, the MRC Human Genetics Unit, Arthritis Research UK and the European Union framework program 6 EUROSPAN project (contract no. LSHG-CT-2006-018947). DNA extractions were performed at the Wellcome Trust Clinical Research Facility in Edinburgh. We would like to acknowledge the invaluable contributions of Lorraine Anderson and the research nurses in Orkney, the administrative team in Edinburgh and the people of Orkney.
The Rotterdam Study: The generation and management of GWAS genotype data for the Rotterdam Study is supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012). This study is funded by the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) project nr. 050-060-810. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera and Marjolein Peters for their help in creating the GWAS database, and Karol Estrada and Maksim V. Struchalin for their support in creation and analysis of imputed data.
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 Rotterdam Scan Study is supported by the Netherlands Organization of Scientific Research (NWO) project nrs. 918-46-615, 904-61-096, 904-61-133, 948-00-010, and 916-13-054 (ZonMW), and Internationaal Parkinson Fonds. Dr. Ikram was supported by a ZonMW Veni grant: 916.13.054. The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists.
The Religious Order Study (ROS) and Rush Memory and Aging Project (MAP): The ROS and MAP Study are supported in part by NIA grants P30AG10161, R01AG15819, R01AG17917, R01AG30146, K08AG34290, and K25AG41906.
Study of Health in Pomerania (SHIP): SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs, and the Social Ministry of the Federal State of Mecklenburg-West Pomerania. Genome-wide data have been supported by the Federal Ministry of Education and Research (grant no. 03ZIK012) and a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg-West Pomerania. The University of Greifswald is a member of the 'Center of Knowledge Interchange' program of the Siemens AG. This work was also funded by the German Research Foundation (DFG: GR 1912/5-1).

The Tasmanian Study of Gait and Cognition (TASCOG) is supported by Project Grants from the National Health and Medical Research Council (NHMRC IDs 403000, 491109, 606543), and a grant from the Wicking Dementia Education and Research Centre, Hobart. Velandai Srikanth is supported by an NHMRC/National Heart Foundation Career Development Fellowship (ID 606544). Matthew Brown is supported by an NHMRC Principal Research Fellowship.
Baltimore Longitudinal Study of Aging (BLSA): The Baltimore Longitudinal Study of Aging is supported by the Intramural Research Program of the NIH, National Institute on Aging.
Hunter Community Study (HCS): The authors would like to thank the men and women participating in the HCS as well as all the staff, investigators and collaborators who have supported or been involved in the project to date. The cohort was made possible with support from the University of Newcastle's Strategic Initiative Fund, the Vincent Fairfax Family Foundation, and the Hunter Medical Research Institute.
Nurses' Health Study (NHS): This study was supported by research grants CA87969, CA49449, HL34594, U01HG004399, DK058845, CA65725, CA67262, CA50385, 5UO1CA098233, EY09611, EY015473, HG004728, HL35464, CA55075, CA134958, and DK070756 from the National Institutes of Health. The genotyping was partly supported by an unrestricted grant from Merck Research Laboratories. Dr. Sun is supported by a career development award K99HL098459 from the National Heart, Lung, and Blood Institute.
REasons for Geographic and Racial Differences in Stroke (REGARDS): This research project is supported by a cooperative agreement U01 NS041588 (G. Howard, PI) from the National Institute of Neurological Disorders and Stroke, National Institutes of Health, Department of Health and Human Service. Genotyping was performed under grant R01 DK084350 (Michele Sale, PI). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Neurological Disorders and Stroke, the National Institute of Diabetes and Digestive and Kidney Diseases, or the National Institutes of Health. Representatives of the funding agency have been involved in the review of the manuscript but not directly involved in the collection, management, analysis or interpretation of the data. The authors thank the other investigators, the staff, and the participants of the REGARDS study for their valuable contributions. A full list of participating REGARDS investigators and institutions can be found at http://www.regardsstudy.org.
Sydney Memory and Ageing Study (Sydney MAS): We would like to acknowledge and thank the Sydney MAS participants and the Sydney MAS Research Team. DNA was extracted by Genetic Repositories Australia, an Enabling Facility, supported by NHMRC Grant 401184. Preparation of the DNA samples was undertaken in the laboratory of Peter Schofield and John Kwok, Neuroscience Research Australia. Genome-wide genotyping was performed by the Ramaciotti Centre, University of New South Wales. The Sydney MAS is supported by Australian National Health & Medical Research Council Program Grant 350833 & Capacity Building Grant 568940. Henry Brodaty is supported by the Australian Government-funded Dementia Collaborative Research Centre at the University of New South Wales. Nicola Armstrong is supported by the NHMRC Project grant 525453.

Three City Study (3C): We thank the staff and the participants of the 3C Study for their important contributions. The 3C Study is conducted under a partnership agreement between the Institut National de la Santé et de la Recherche Médicale (INSERM), the Victor Segalen–Bordeaux II University, and Sanofi-Aventis. The Fondation pour la Recherche Médicale funded the preparation and initiation of the study. The 3C Study is also supported by the Caisse Nationale Maladie des Travailleurs Salariés, Direction Générale de la Santé, Mutuelle Générale de l'Education Nationale (MGEN), Institut de la Longévité, Conseils Régionaux of Aquitaine and Bourgogne, Fondation de France, and Ministry of Research–INSERM Programme "Cohortes et collections de données biologiques." Lille Génopole received an unconditional grant from Eisai. We thank A. Boland (Centre National de Génotypage) for her technical help in preparing the DNA samples for analyses. This work was supported by the National Foundation for Alzheimer's Disease and Related Disorders, the Institut Pasteur de Lille and the Centre National de Génotypage. Stéphanie Debette is a recipient of a Chaire d'Excellence grant from the French national research agency (Agence Nationale de la Recherche).
Women's Genome Health Study (WGHS): The WGHS is supported by HL043851 and HL080467 from the National Heart, Lung, and Blood Institute and CA047988 from the National Cancer Institute, the Donald W. Reynolds Foundation and the Fondation Leducq, with collaborative scientific support and funding for genotyping provided by Amgen.
Swiss Memory Genetics Study (Swiss MGS): This work was funded by the Swiss National Science Foundation (Sinergia grant CRSI33_130080 to D.Q. and A.P.).
The hippocampal gene expression study was supported by the German Federal Ministry of Education and Research (BMBF) through the Integrated Genome Research Network (IG) MooDS (Systematic Investigation of the Molecular Causes of Major Mood Disorders and Schizophrenia, under the auspices of the National Genome Research Network plus (NGFNplus).

**CO-INVESTIGATOR LIST**

**TASCOG**: Jim Stankovich (PhD, Menzies Research Institute Tasmania, University of Tasmania, Hobart, Australia), Matthew Brown (MD, University of Queensland Diamantina Institute, Princess Alexandra Hospital, Woolloongabba, Brisbane, Australia); **Sydney Memory and Ageing Study**: John B Kwok (PhD, Neuroscience Research Australia and School of Medical Sciences, University of New South Wales, Sydney, New South Wales, Australia), Peter Schofield (PhD, Neuroscience Research Australia and School of Medical Sciences, University of New South Wales, Sydney, New South Wales, Australia), Henry Brodaty (MD, Primary Dementia Collaborative Research Centre, School of Psychiatry and Centre for Healthy Brain Ageing, School of Psychiatry, UNSW Medicine, University of New South Wales; Aged Care Psychiatry, Prince of Wales Hospital, Sydney, NSW, Australia); **Hunter Community Study**: Mark McEvoy (MMedSc, School of Medicine and Public Health, Faculty of Health, University of Newcastle, Australia), Rodney Scott (PhD, School of Biomedical Sciences and Pharmacy, University of Newcastle, Australia), Roseanne Peel (MPH, School of Medicine and Public

Health, Faculty of Health, University of Newcastle, Australia); **Hippocampal gene expression study**: Katharina Pernhorst (MS, Institute of Neuropathology, Universitätsklinikum Bonn, Bonn, Germany), Benno Pütz (PhD, Max Planck Institute of Psychiatry, Munich, Germany), Karsten Borgwardt (PhD, Max Planck Institutes for Developmental Biology and for Intelligent Systems, Tübingen, Germany), Bernhard Schölkopf (PhD, Max Planck Institute for Intelligent Systems, Tübingen, Germany), Johannes Schramm (MD, Department of Neurosurgery, Universitätsklinikum Bonn, Bonn, Germany), Per Hoffmann (PhD, Institute of Human Genetics, Universitätsklinikum Bonn, Bonn, Germany); **3C-Dijon:** Sabrina Schilling (MS, Inserm U740 and University of Bordeaux-Segalen, France)

**REFERENCES**


1. Elias, M.F. et al. The preclinical phase of alzheimer disease: A 22-year prospective study of the Framingham Cohort. *Arch Neurol* **57**, 808-13 (2000).
2. Amieva, H. et al. The 9 year cognitive decline before dementia of the Alzheimer type: a prospective population-based study. *Brain* **128**, 1093-101 (2005).
3. Dubois, B. et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurol* **6**, 734-46 (2007).
4. Deary, I.J., Johnson, W. & Houlihan, L.M. Genetic foundations of human intelligence. *Hum Genet* **126**, 215-32 (2009).
5. Pedersen, N.L., Plomin, R., Nesselroade, J.R. & McClearn, G.E. Aquantitative genetic analysis of cognitive abilities during thesecond half of the lifespan. *Psychol Sci* **3**, 346–353 (1992).
6. Finkel, D., Pedersen, N.L., McLearn, G.E., Plomin, R. & Berg, S. Cross-sequential analysis of genetic influences on cognitive ability inthe Swedish adoption/twin study of ageing. *Aging NeuropsycholCogn* **3**, 84–99 (1996).
7. Reynolds, C.A. et al. Quantitative genetic analysis of latent growth curve models of cognitive abilities in adulthood. *Dev Psychol* **41**, 3-16 (2005).
8. Payton, A. The impact of genetic research on our understanding of normal cognitive ageing: 1995 to 2009. *Neuropsychol Rev* **19**, 451-77 (2009).
9. Wisdom, N.M., Callahan, J.L. & Hawkins, K.A. The effects of apolipoprotein E on non-impaired cognitive functioning: a meta-analysis. *Neurobiol Aging* **32**, 63-74 (2011).
10. Papassotiropoulos, A. et al. A genome-wide survey and functional brain imaging study identify CTNNBL1 as a memory-related gene. *Mol Psychiatry* (2011).
11. Papassotiropoulos, A. et al. Common Kibra alleles are associated with human memory performance. *Science* **314**, 475-8 (2006).
12. Glahn, D.C. et al. Genetic basis of neurocognitive decline and reduced white-matter integrity in normal human brain aging. *Proc Natl Acad Sci U S A* **110**, 19006-11 (2013).
13. Kandel, E.R. The molecular biology of memory storage: a dialogue between genes and synapses. *Science* **294**, 1030-8 (2001).
14. Deary, I.J. et al. Cognitive change and the APOE epsilon 4 allele. *Nature* **418**, 932 (2002).
15. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
16. Kohli, M.A. et al. The neuronal transporter gene SLC6A15 confers risk to major depression. *Neuron* **70**, 252-65 (2011).
17. Liu, J.Z. et al. A versatile gene-based test for genome-wide association studies. *Am J Hum Genet* **87**, 139-45 (2010).
18. Chibnik, L.B. et al. CR1 is associated with amyloid plaque burden and age-related cognitive decline. *Ann Neurol* **69**, 560-9 (2011).
19. Pruim, R.J. et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336-7 (2010).

20. Mandel, I. et al. The ubiquitin-proteasome pathway regulates claudin 5 degradation. *J Cell Biochem* **113**, 2415-23 (2012).
21. Small, B.J., Rosnick, C.B., Fratiglioni, L. & Backman, L. Apolipoprotein E and cognitive performance: a meta-analysis. *Psychol Aging* **19**, 592-600 (2004).
22. Liu, F. et al. The apolipoprotein E gene and its age-specific effects on cognitive function. *Neurobiol Aging* **31**, 1831-3 (2010).
23. Mondadori, C.R. et al. Better memory and neural efficiency in young apolipoprotein E epsilon4 carriers. *Cereb Cortex* **17**, 1934-47 (2007).
24. Vigneau, M. et al. What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? Insights from a meta-analysis. *Neuroimage* **54**, 577-93 (2011).
25. Giraud, A.L. et al. Differential recruitment of the speech processing system in healthy subjects and rehabilitated cochlear implant patients. *Brain* **123 ( Pt 7)**, 1391-402 (2000).
26. Papassotiropoulos, A. & de Quervain, D.J. Genetics of human episodic memory: dealing with complexity. *Trends Cogn Sci* **15**, 381-7 (2011).
27. Tiwari, V., O'Donnell, C.D., Oh, M.J., Valyi-Nagy, T. & Shukla, D. A role for 3-O-sulfotransferase isoform-4 in assisting HSV-1 entry and spread. *Biochem Biophys Res Commun* **338**, 930-7 (2005).
28. Carter, C.J. Alzheimer's disease plaques and tangles: cemeteries of a pyrrhic victory of the immune defence network against herpes simplex infection at the expense of complement and inflammation-mediated neuronal destruction. *Neurochem Int* **58**, 301-20 (2011).
29. Letenneur, L. et al. Seropositivity to herpes simplex virus antibodies and risk of Alzheimer's disease: a population-based cohort study. *PLoS One* **3**, e3637 (2008).
30. Zimmer, C. Microbiology. Do chronic diseases have an infectious root? *Science* **293**, 1974-7 (2001).
31. Itzhaki, R.F. & Wozniak, M.A. Herpes simplex virus type 1 in Alzheimer's disease: the enemy within. *J Alzheimers Dis* **13**, 393-405 (2008).
32. Kennedy, R.B. et al. Genome-wide analysis of polymorphisms associated with cytokine responses in smallpox vaccine recipients. *Hum Genet* **131**, 1403-21 (2012).
33. Lambert, J.C. et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet* **41**, 1094-9 (2009).
34. Harold, D. et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* **41**, 1088-93 (2009).
35. Lambert, J.C. et al. Implication of the immune system in Alzheimer's disease: evidence from genome-wide pathway analysis. *J Alzheimers Dis* **20**, 1107-18 (2010).
36. Kim, S.Y. et al. The expression of non-clustered protocadherins in adult rat hippocampal formation and the connecting brain regions. *Neuroscience* **170**, 189-99 (2010).

37. Fiore, R., Rahim, B., Christoffels, V.M., Moorman, A.F. & Puschel, A.W. Inactivation of the Sema5a gene results in embryonic lethality and defective remodeling of the cranial vascular system. *Mol Cell Biol* **25**, 2310-9 (2005).
38. Weiss, L.A., Arking, D.E., Daly, M.J. & Chakravarti, A. A genome-wide linkage and association scan reveals novel loci for autism. *Nature* **461**, 802-8 (2009).
39. Deary, I.J. & Johnson, W. Intelligence and education: causal perceptions drive analytic processes and therefore conclusions. *Int J Epidemiol* **39**, 1362-9 (2010).
40. Bertram, L. & Tanzi, R.E. Alzheimer disease: New light on an old CLU. *Nat Rev Neurol* **6**, 11-3 (2010).
41. Lambert, J.C. & Amouyel, P. Genetics of Alzheimer's disease: new evidences for an old hypothesis? *Curr Opin Genet Dev* **21**, 295-301 (2011).
42. Ciechanover, A. & Brundin, P. The ubiquitin proteasome system in neurodegenerative diseases: sometimes the chicken, sometimes the egg. *Neuron* **40**, 427-46 (2003).
43. Bertram, L. et al. Family-based association between Alzheimer's disease and variants in UBQLN1. *N Engl J Med* **352**, 884-94 (2005).
44. Pavlopoulos, E. et al. Neuralized1 activates CPEB3: a function for nonproteolytic ubiquitin in synaptic plasticity and memory storage. *Cell* **147**, 1369-83 (2011).

# Chapter 4 : Exploring vascular endophenotypes of Alzheimer’s disease

## *4.1. Linkage analysis for plasma amyloid beta levels in persons with hypertension implicates Aβ-40 levels to Presenilin 2*

Hum Genet 2012 Dec; 131(12):1869-76

The supplemental information for this paper is available at
*http://link.springer.com/article/10.1007%2Fs00439-012-1210-2*

**ABSTRACT**

Plasma concentrations of Aβ40 and Aβ42 rise with age and are increased in people with mutations that cause early-onset Alzheimer's disease (AD). Amyloid beta (Aβ) plasma levels were successfully used as an (endo)phenotype for gene discovery using a linkage approach in families with dominant forms of disease. Here we searched for loci involved in Aβ plasma levels in a series of nondemented patients with hypertension in the Erasmus Rucphen Family (ERF) study. Aβ40 and Aβ42 levels were determined in 125 subjects with severe hypertension. All patients were genotyped with a 6000 single nucleotide polymorphisms (SNPs) Illumina array designed for linkage analysis. We conducted linkage analysis of plasma Aβ levels. None of the linkage analyses yielded genome-wide significant logarithm of odds (LOD) score over 3.3, but there was suggestive evidence for linkage (LOD > 1.9) for 2 regions: 1q41 (LOD = 2.07) and 11q14.3 (LOD=2.97), both for Aβ40. These regions were followed up with association analysis in the study subjects and in 320 subjects from a population-based cohort. For the Aβ40 region on chromosome 1, association of several SNPs was observed at the presenilin 2 gene (*PSEN2*) (*p-value* = $2.58*10^{-4}$ for rs6703170). On chromosome 11q14-21 we found some association (*p-value* = $3.1*10^{-3}$ for rs2514299). This linkage study of plasma concentrations of Aβ40 and Aβ42 yielded two suggestive regions, of which one points towards a known locus for familial AD.

## INTRODUCTION

Together with neurofibrillary tangles, amyloid beta (Aβ) plaques in the brain are a pathological hallmark of Alzheimer's disease (AD).[1] The formation of the extracellular peptide Aβ results from the cleavage of the amyloid precursor protein (APP) by the enzymes β and γ secretase, as opposed to α and γ secretase.[2] Aβ is partially degraded within the brain and partly cleared from brain to plasma through the blood-brain barrier, a process in which several of the recently discovered Alzheimer risk genes are involved.[3] Excess production or poor clearance of amyloid beta may lead to neurotoxicity, and formation of amyloid beta plaques. This is also true for truncated Aβ forms, Aβn-42 and Aβn40. The latter are more likely to be spliced into pathogenic forms. Aβn-42 accounts for 60% of all Aβ species in pre-clinical AD stages.[4] High plasma concentrations of Aβ are associated with an increased risk of AD and there is revived interest in the use of this marker in clinical care, despite the fact that there may not be a causal relationship.[5] Recent research has shown a decreased clearance of Aβ from the brain in those with mild cognitive dysfunction.[6]

Aβ plasma levels were successfully used as an (endo)phenotype for gene discovery using a linkage approach in families with dominant forms of disease, leading to the discovery of presenilin 1 *(PSEN1)*.[7] From a genetic perspective, it is of interest that the gene encoding the angiotensin converting enzyme (ACE) is consistently associated with AD but also plays a role in the degradation of Aβ in the brain.[8] Since hypertension in early life to middle age has now been well established as a risk factor for AD, and plasma Aβ levels are associated with hypertension,[9] we have conducted a genetic study aiming to identify genes involved in plasma Aβ levels in persons with hypertension.

## MATERIALS AND METHODS

### *Study population*

This study was embedded in the Erasmus Rucphen Family (ERF) study, a population-based study in a genetically isolated population. All approximately 3000 participants in this study are living descendants of 22 couples who, at the end of the 19th century, had at least six children baptized in the community church. Extensive genealogy data are available from the year 1600 AD. Extensive data on the participants, including cardiovascular risk factors, family history, body composition, health habits, cognitive function, blood chemistry and genotyping are available.

For this study, hypertensive subjects aged 55 to 75 years who did not have a history of stroke or dementia were selected from the study population. Hypertension was defined as a systolic blood pressure of ≥ 160 mmHg, a diastolic blood pressure of ≥ 100 mmHg, and/or the use of antihypertensive medication. We chose to limit ourselves to subjects with this high threshold for hypertension to get the maximally informative subjects out of our cohort. Shah et al provide a good insight in the interaction between amyloid beta and hypertension in the risk of AD and vascular brain changes.[10] Of the 261 eligible individuals invited for this study, 135 agreed to participate. The participants and non-participants were similar with

respect to cardiovascular risk factors, but the participants had a slightly higher duration of education. (9 years in participants as opposed to 7.5 years in non-participants). All Medical Center Medical Ethics Committee.

A replication sample was obtained from the Rotterdam Study (RS), a population-based cohort from the Rotterdam region.[11] There were 320 individuals available who met the same inclusion criteria as our discovery subjects (age 55-75 years, hypertensive, and free from stroke and dementia) and had genotype information and amyloid beta measurements available.

*Aβ measurements*

Non-fasting blood samples were obtained in EDTA tubes and immediately cooled on ice. Plasma was extracted and stored at -80 degrees Celsius. Plasma Aβ concentrations were measured using a fluorimetric bead-based immunoassay using xMAP ® technology (Innogenetics ®). Aβ40, Aβ42, and the truncated forms Aβn-42 and Aβn-40 were measured. From these measurements we also calculated ratios for Aβ40/Aβ42 and Aβ42/Aβn-42.[12,13]

*Genotyping*

Genotyping and pedigree data were available for 125 of the 129 subjects with full phenotype data. For all participants genomic DNA was extracted from peripheral venous blood utilizing the salting out method.[14] For genome-wide linkage analysis genotyping was performed using the Illumina 6K linkage panel. Of the 6000 single nucleotide polymorphisms (SNPs) on the array, 5250 were used for analysis after quality control and excluding X-chromosomal SNPs. The genotyping was performed at the Centre National de Génotypage in France according to the manufacturer's protocol.

For association analysis we used SNPs from dense genotyping platforms that included Illumina 318K, Illumina 370K, Illumina 610K and Affymetrix 250K, which were merged as previously decribed.[15] Genotyping on these platforms was performed according to the described protocols. Additionally, the array data was used to impute genotypes (using MACH version 1.0.16) to the HapMap rel. 22 CEU panel for meta-analysis.[16,17] The genome-wide imputed dataset consisted of approximately 2.5 million SNPs.

For the replication cohort, we extracted genotype data for the regions of interest from the imputed genotype dataset. This dataset consists of approximately 2.5 million SNPs, genotyped on the Illumina HumanHap550-Duo BeadChip® and imputed with MACH.

*Statistical analysis*

Statistical analysis of quantitative traits was performed using PASW Statistics version 17.0 (SPSS Inc.) for Windows. The Aβ measurements were normally distributed, but, kurtosis was increased for all distributions (from 1.98 to 3.29). Inverse-normal transformation of ranks was applied to all individual traits for linkage analysis, and ln transformation was used for the association analyses for the ratios as these were not normally distributed. The rank-transformation was performed using the GenABEL package for R.[18] The family based design

of the ERF study makes it possible to evaluate the heritability of Aβ levels (Aβ40, Aβ42, and truncated forms Aβn40 and Aβn-42) using the pedigrees in the SOLAR program.[19]

Linkage

We performed linkage analysis for the different amyloid beta subtypes as defined above with age and sex as covariates, using the variance components method as implemented in Merlin.[20] For computational reasons (software limitations), the study population was divided into 34 sub-pedigrees each consisting of 2 to 7 subjects. A significant linkage result was defined as a LOD peak of ≥3.3, a suggestive linkage result was defined as a LOD peak of ≥1.9.[21] Additionally, we investigated all remaining regions with a LOD peak ≥1.5. Linkage regions were defined as the region bounded by the maximum LOD score minus 1.

Fine mapping with association analysis

With the same traits and covariates, we performed association analysis with a polygenic model using GenABEL package for R.[18] The SNPs for the regions under the linkage peaks were taken from the dense panel of 700 000 SNPs. Association analysis was also performed for the SNPs of interest in the Rotterdam Study.
A meta-analysis of the results from the two studies for the regions under the peak was performed using the Metal software package for meta-analysis.[22] In this meta-analysis, we used the 2.5 million imputed SNP dataset for ERF.

eQTL analyses

Information on the power of the identified SNP as expression quantitative trait loci (eQTL), i.e. the association of the SNPs found in our analyses to gene expression, was extracted from the mRNA by SNP browser (MRBS) by Liang et al.[23] (www.sph.umich.edu/csg/liang/asthma). We applied the genome-wide significance threshold for eQTLs defined by the authors of this database: a LOD of 6.076, corresponding to a p-value of $1.2*10^{-7}$. We also extracted data on eQTL associations from the ScanDB database (www.scandb.org), which is based on the eQTL analyses by Zhang et al.[24]

**RESULTS**

*Descriptives*

Table 1 provides a description of the study population. Six of the 135 subjects were excluded from analysis due to unsuccessful phenotype collection, four more because of insufficient genotype or pedigree data. Plasma Aβ40 was significantly correlated with BMI (*p-value* = 0.007).
We assessed the role of possible confounding factors by performing linear regression analyses on Aβ values with and without suspected confounders. We found no evidence for confounding from either any medication acting on the RAS system, neither for Aβ40 (p-value=0.683) nor for Aβ42 (*p-value* = 0.481), or for diabetes (*p-values*: diabetes status in Aβ40=0.681, HbA1C in Aβ40 = 0.588; diabetes in Aβ42 = 0.067, HbA1C in Aβ42 = 0.133).

**Table 1. Baseline characteristics of 125 participants with Aβ measurements**

| Characteristics | Mean (SD) |
|---|---|
| Age | 64.42 (4.57) |
| Female (n,%) | 65 (52) |
| Only primary education (n,%) | 42 (34) |
| Systolic blood pressure | 146 (18) |
| Diastolic blood pressure | 84 (10) |
| *APOE* e4 carriers (n,%) | 50 (40)* |
| Current smoker (n,%) | 36 (29) |
| Body Mass Index | 29.1 (4.4) |
| Total cholesterol | 5.1 (1.12) |
| HbA1c | 5.84 (0.59) |
| Aβ40 (pg/ml) | 179 (39) |
| Aβ42 (pg/ml) | 41 (14) |
| Aβn40 (pg/ml) | 176 (36) |
| Aβn-42 (pg/ml) | 28 (7) |
| Aβ42/ Aβ40 | 0.23 (0.07) |
| Aβ42/ Aβn-42 | 1.55 (0.46) |

** 38 subjects (30%) carried 1 APOE e4 allele, 12 subjects (10%) carried two APOE e4 alleles*

We also assessed partial correlations between Aβ levels and cognitive functioning in a neuropsychological test battery including Stroop, Trailmaking test, Block Design, 15 word memory test, and word fluency. There was a nominally significant (*p-value* = 0.005, we ran 11 correlations for 2 Aβ measurements) partial correlation, corrected for age, sex and eduction, between Aβ42 levels and the Z-score for word fluency, but not for any other individual or composite measure to Aβ42 or Aβ40.

*Heritability*

First we estimated the heritabilities of Aβ levels (Aβ40, Aβ42, and truncated forms Aβn40 and Aβn-42) using the full ERF pedigree. We estimated a heritability of 0.23 (*p-value* = 0.19), 0.30 (*p-value* = 0.20), 0.12 (*p-value* = 0.31) and 0.55 (*p-value* = 0.07) for Aβ40, Aβ42, Aβn40 and Aβn-42, respectively. None of the heritabilities were significant. However, in a rank transformed analysis to adjust for non-normality in pedigree fragments, the heritability estimates for Aβ40 (*p-value* = 0.04) and Aβ42 (*p-value* = 0.046) were significant.

*Linkage analysis*

None of the linkage analyses yielded genome wide significant Log Odd (LOD) scores over 3.3 but there was suggestive linkage (LOD > 1.9) to Aβ40 in 2 regions: 1q41 (LOD = 2.07) and 11q14.3 (LOD = 2.97) (Table 2). The other individual traits had maximum LODs between 1.5 and 1.9: 2p22.2 (LOD = 1.86) for Aβn40, 15q13.3 (LOD = 1.63) and 15q26.1 (LOD = 1.7) for Aβ42 and 1q31.1 (LOD 1.6) and 17q25.1 (LOD=1.61) for Aβn-42. For the ratios we found no

evidence for linkage. Online resource 1 shows the genome-wide linkage plots for all traits and zooms in on the regions with LOD > 1.9. Although the Aβ40 and Aβ42 levels are highly correlated (partial correlation = 0.433, *p-value* = <0.001), there is no overlap in linkage peaks between the two. In the chromosome 1 region for Aβ40, the maximum LOD for Aβ42 was 0.35. In the chromosome 11 region, there was no evidence for any linkage for Aβ42 (LOD = 0). Also vice versa, the chr15 regions linked to Aβ42 showed no evidence for linkage to Aβ40 (LOD max 0.05).

**Table 2. Identified linkage regions for the different measurements**

| Chromosome | SNP with highest LOD | Physical position | Total region LOD-1 (cM) | Trait | LOD |
|---|---|---|---|---|---|
| **1** | **rs11584610** | **221898214** | **228.695-237.703** | **Aβ40** | **2.07** |
| 1 | rs11584662 | 184847447 | 187.881-230.45 | Aβn-42 | 1.60 |
| 2 | rs2691123 | 37070531 | 55.057-68.632 | Aβn40 | 1.86 |
| **11** | **rs10830888** | **91591095** | **96.234-99.186** | **Aβ40** | **2.97** |
| 15 | rs1399073 | 31331558 | 21.917-38.757 | Aβ42 | 1.63 |
| 15 | rs6497019 | 91428643 | 96.069-117.909 | Aβ42 | 1.70 |
| 17 | rs11869620 | 70326171 | 99.289-115.245 | Aβn-42 | 1.61 |

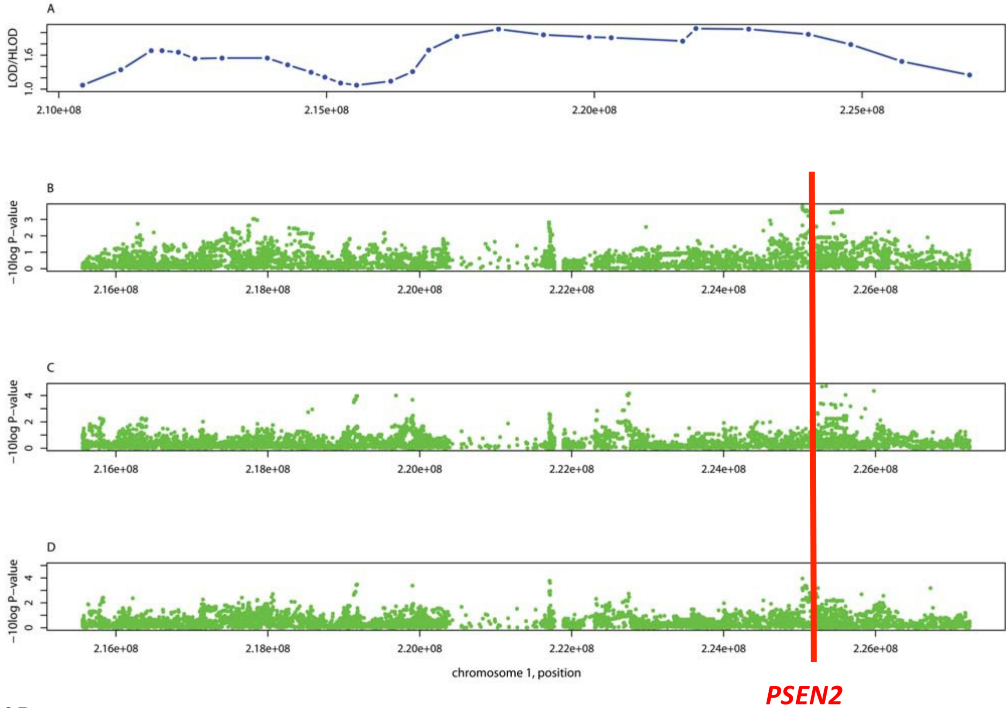
*Association and eQTL analyses*

We further explored the suggestive regions on chromosome 1 and 11 using the dense SNP genotypes from the microarrays. There were 1216 directly genotyped SNPs available for association analysis in the region identified on chromosome 1, and 562 SNPs for the region on chromosome 11. For the meta-analysis in the imputed SNP sets, there were 14986 SNPs for the region on chromosome 1, and 3981 SNPs for the region on chromosome 11. Figure 1A and Figure 1B show the local linkage and association plots for these regions. Results for the top hits in the directly genotyped dataset are provided in Online Resource 2.

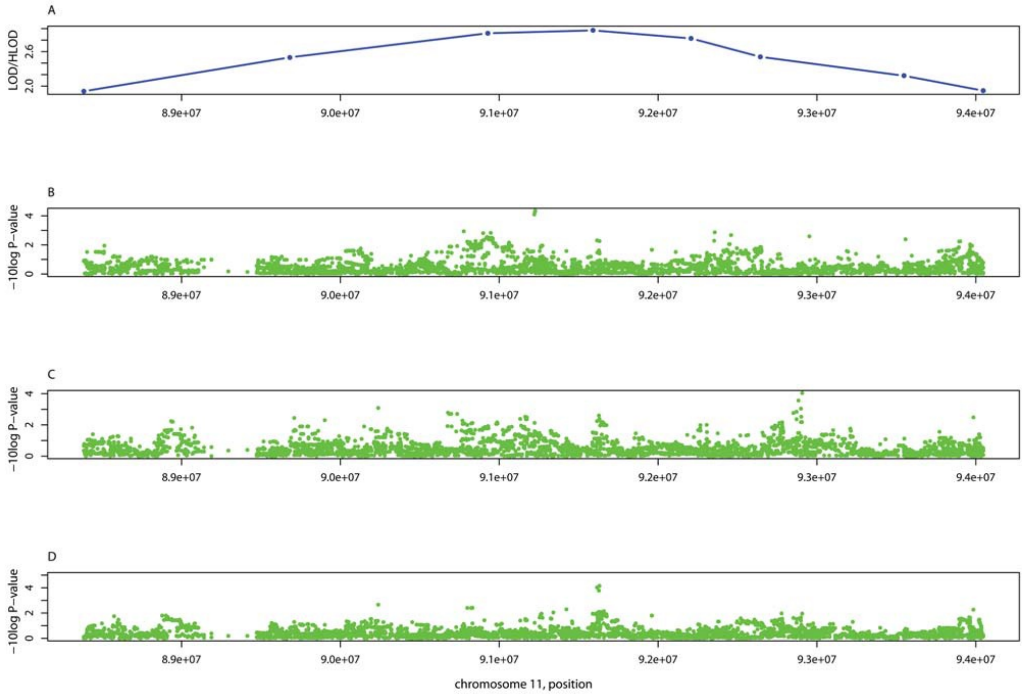
For the Aβ40 region on chromosome 1, nominal association was observed for a SNP lying 82 kb upstream from the presenilin 2 gene (*PSEN2*) (rs6703170: *p-value* = $2.58*10^{-4}$) (Online Resource 2). In the meta-analysis of ERF and the Rotterdam Study, rs6697254 had the lowest p-value but the allele frequency was low (0.0057) making the finding unreliable. Rs12409752 had the lowest p-value with a common risk allele (MAF 0.27, *p-value* =$1.1*10^{-4}$) which is an intergenic SNP between *ITPKB* and *PSEN2*. There was a large block of *PSEN2* intronic SNPs spanning from 225130294 to 225149349 kb (NCBI build 36.3) which were associated with Aβ40. When considering eQTLs, several SNPs were associated (lowest *p-value* = $2.40*10^{-10}$ for rs2236914) with *PSEN2* expression levels. The meta analysis results and eQTL results are given in Online Resource 3.
When analysing the chromosome 11 region in more detail there were no strong association results. The SNPs showing some association in ERF (top hit rs2514299, *p-value* = $3.2*10^{-3}$)

**Figure 1: Local linkage and association plots for the Aβ40 regions**

***1A***




***1B***




*This figure shows the linkage and association results for the linkage regions defined on chromosome 1 and 11. In panel A, a detail of the linkage plot is shown. Panel B and C, respectively, show the association results under the peak from the imputed datasets for ERF and RS. Panel D gives the meta-analysis results for the two. Figure 1A: Chromosome 1. A consistent signal can be seen in the PSEN2 region at $2.25*10^{8}$ bp. Figure 1B: Chromosome 11. There is no consistent signal across the two cohorts within the region.*

(Online Resource 2) did not replicate in the RS. The two SNPs most strongly associated with Aβ40 levels emerging from a meta-analysis of RS and ERF were rs947937 (*p-value* = $7.3*10^{-5}$) and rs947935 (*p-value* = $9.6*10^{-5}$) (Supplementary table 2B). eQTL analysis in this region showed no convincing results.

**DISCUSSION**

In this family-based study, we found the highest heritability for Aβ42 and Aβn-42. This heritability is lower than that found in an extended family affected with late-onset Alzheimer's disease (LOAD).[25] There was no genome-wide significant linkage of Aβ levels as none of the regions reached a LOD score of 3.3 or higher, although the region on chromosome 11 approached this genome wide significance level at a LOD score of 2.97. In total, 2 regions showed suggestive linkage with a LOD of >1.9. Of these two regions, the chromosome 11 region showed the highest LOD score. In this region, the ERF and RS association analyses were not consistent. In the second region on chromosome 1 with an LOD score of 2.07, *PSEN2* is the most remarkable gene, showing evidence both for association and an effect on expression levels of PSEN2.

When comparing our data with the literature there are a number of remarkable findings. First, the *PSEN2* gene is a known causative gene for some cases of familial AD.[26] *PSEN2* was identified due to its homology to *PSEN1*.[27] Its penetrance is lower than that of *PSEN1*. In its turn, *PSEN1* was identified in a study using Aβ plasma levels as an (endo)phenotype for gene discovery using a linkage approach in families with dominant forms of AD.[28] The presenilins are the proteases in the gamma secretase complex in the cell membrane responsible for the cleavage of APP into amyloid beta. Additionally, the presenilin 2 protein has been proven to downregulate cytokine-induced inflammatory responses in the brain which can lead to neurodegeneration.[29] When comparing the meta-analysis association results with the eQTL databases, there were several top SNP that show significant association with *PSEN2* expression levels, i.e. are eQTLs for this gene. This supports our theory that it is a variant in, or in the regulatory region of, *PSEN2* causing this signal, and pinpoints familial AD gene *PSEN2* as playing a possible role in the multifactorial LOAD pathogenesis. Our study connects a common variant near to *PSEN2* to Aβ metabolism relatively early in life.
Second, it is interesting that we found a positive association of *PSEN2* to Aβ40 while there was no evidence of any linkage or association with Aβ42 (best LOD for the region is 0.32). These findings support several recent reports that indicate that Aβ40 is a determinant at least as important as Aβ42, although whether its effect is protective or risk-increasing is currently being debated.[5,30] Animal experiments have shown that all tested familial *PSEN* and *APP* mutations resulted in decreased Aβ40 production with an accumulation of APP C-terminal fragments, a sign of decreased PSEN activity, but only some mutations including the *PSEN2* N141I (Volga German mutation) affected Aβ42 levels.[30] Our method does not allow us to pinpoint a specific mutation in the gene or its promoter regions, but it is possible that the

variant underlying our signal selectively affects Aβ40 levels. We did not see evidence for association to the Aβ42/Aβ40 ratio.
The region on chromosome 11 is of particular interest for two reasons. First, the LOD score in this region approaches genome wide significance. Second, this region has been associated with various neuropsychiatric disease, including autism and schizophrenia and it is close to the region 11q25 previously found associated with depressive disorder, autism and with late-onset Alzheimer's disease.[31-36] Given the evidence of co-occurrence of depression and Alzheimer's disease, this region is of particular interest from a clinical perspective.[37] Within this linkage region, we cannot clearly identify a likely candidate gene based upon the association or eQTL analyses.
There was no evidence of linkage with the *APOE* region on chromosome 19. *APOE* is the best-established risk gene for sporadic Alzheimer's disease, plasma ApoE levels have been associated with amyloid beta burden in the brain, and ApoE is believed to play a role in the clearance of Aβ from the brain.[3,38] A lack of power in the current study may be a possible explanation for this finding. Another explanation is that plasma Aβ42 and Aβ40 levels show no correlation to *APOE*. Only the Aβ42/Aβ40 ratio was correlated to *APOE* E4 carrier status in our sample ($R^2$ = -21; *p-value* = 0.023). Finally, compared to association, linkage analysis may not be as powerful an approach to identify a common susceptibility gene such as *APOE*. The strongest association signal within 1 Mb of the *APOE* gene was seen for rs1661197 (*p-value* = 0.001) with Aβ42. This SNP is located 312 kb away from the gene.
The main limitation of this linkage project is its limited sample size, due to the narrow inclusion criteria for this study and financial and logistics issues. Also, plasma amyloid beta is a rough estimate of the Aβ load in the brain and little is known about the variations and the changes of clearance from brain to plasma over time. There is no circadian variation in plasma Aβ levels.[39] However, our heritability studies show that Aβ42 and Aβ40 plasma levels are stable enough to yield significant evidence for familial clustering of increased Aβ levels at early age before the onset of dementia. Several studies have shown its value as an easily obtainable biomarker for risk of AD,[5,40] although there are also negative studies,[12,41] and it is the best method available for population-based research, as it is ethically not feasible to perform lumbar punctures on healthy volunteers at a large scale in a population-based setting.
Plasma amyloid beta 40 levels in healthy middle-aged subjects are associated with the locus containing the *PSEN2* gene associated with early-onset Alzheimer's disease, and with a locus on chromosome 11. Our findings support the involvement of these regions in the development of sporadic late-onset Alzheimer's disease. Additionally, the identification of a known gene involved in the plasma Aβ levels in this hypothesis-free experiment can be considered a sign of robustness for this method in our inbred population. Lastly, our eQTL analysis underlines the new associations of interest in interpreting the results of an association analysis. Next generation sequencing and expression analyses will hopefully allow us to investigate the linkage peaks identified in much more detail in the future.

**ACKNOWLEDGMENTS**
This study is financially supported by the Netherlands Organization for Scientific Research (NWO), the Internationale Stichting Alzheimer Onderzoek (ISAO), the Hersenstichting Nederland (HSN), and the Centre for Medical Systems Biology (CMSB) in the framework of the Netherlands Genomics Initiative (NGI), and by the Russian Foundation for Basic Research (RFBR). We thank the participants from the Genetic Research in Isolated Populations, Erasmus Rucphen Family, who made this work possible. Also, we thank Petra Veraart for collecting all genealogical data.

**REFERENCES**


1. Takashima, A. Amyloid-beta, tau, and dementia. *J Alzheimers Dis* **17**, 729-36 (2009).
2. Murphy, M.P. & LeVine, H., 3rd. Alzheimer's disease and the amyloid-beta peptide. *J Alzheimers Dis* **19**, 311-23 (2010).
3. Bettens, K., Sleegers, K. & Van Broeckhoven, C. Current status on Alzheimer disease molecular genetics: from past, to present, to future. *Hum Mol Genet* **19**, R4-R11 (2010).
4. Sergeant, N. et al. Truncated beta-amyloid peptide species in pre-clinical Alzheimer's disease as new targets for the vaccination approach. *J Neurochem* **85**, 1581-91 (2003).
5. van Oijen, M., Hofman, A., Soares, H.D., Koudstaal, P.J. & Breteler, M.M. Plasma Abeta(1-40) and Abeta(1-42) and the risk of dementia: a prospective case-cohort study. *Lancet Neurol* **5**, 655-60 (2006).
6. Mawuenyega, K.G. et al. Decreased clearance of CNS beta-amyloid in Alzheimer's disease. *Science* **330**, 1774 (2010).
7. Ertekin-Taner, N. et al. Linkage of plasma Abeta42 to a quantitative locus on chromosome 10 in late-onset Alzheimer's disease pedigrees. *Science* **290**, 2303-4 (2000).
8. Kehoe, P.G., Miners, S. & Love, S. Angiotensins in Alzheimer's disease - friend or foe? *Trends Neurosci* **32**, 619-28 (2009).
9. Lambert, J.C. et al. Association of plasma Ass peptides with blood pressure in the elderly. *PLoS One* **6**, e18536 (2011).
10. Shah, N.S. et al. Midlife blood pressure, plasma beta-amyloid, and the risk for Alzheimer disease: the Honolulu Asia Aging Study. *Hypertension* **59**, 780-6 (2012).
11. Hofman, A. et al. The Rotterdam Study: objectives and design update. *Eur J Epidemiol* **22**, 819-29 (2007).
12. Hansson, O. et al. Evaluation of plasma Abeta(40) and Abeta(42) as predictors of conversion to Alzheimer's disease in patients with mild cognitive impairment. *Neurobiol Aging* **31**, 357-67 (2010).
13. Lambert, J.C. et al. Association of plasma amyloid beta with risk of dementia: the prospective Three-City Study. *Neurology* **73**, 847-53 (2009).
14. Miller, S.A., Dykes, D.D. & Polesky, H.F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* **16**, 1215 (1988).
15. Demirkan, A. et al. Genetic risk profiles for depression and anxiety in adult and elderly cohorts. *Mol Psychiatry* **16**, 773-83 (2011).
16. Li, Y., Willer, C.J., Ding, J., Scheet, P. & Abecasis, G.R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* **34**, 816-34 (2010).
17. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype imputation. *Annu Rev Genomics Hum Genet* **10**, 387-406 (2009).

18. Aulchenko, Y.S., Ripke, S., Isaacs, A. & van Duijn, C.M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294-6 (2007).
19. Almasy, L. & Blangero, J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* **62**, 1198-211 (1998).
20. Abecasis, G.R., Cherny, S.S., Cookson, W.O. & Cardon, L.R. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* **30**, 97-101 (2002).
21. Lander, E. & Kruglyak, L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* **11**, 241-7 (1995).
22. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
23. Dixon, A.L. et al. A genome-wide association study of global gene expression. *Nat Genet* **39**, 1202-7 (2007).
24. Zhang, W. et al. Evaluation of genetic variation contributing to differences in gene expression between populations. *Am J Hum Genet* **82**, 631-40 (2008).
25. Ertekin-Taner, N. et al. Heritability of plasma amyloid beta in typical late-onset Alzheimer's disease pedigrees. *Genet Epidemiol* **21**, 19-30 (2001).
26. Van Broeckhoven, C. Presenilins and Alzheimer disease. *Nat Genet* **11**, 230-2 (1995).
27. Rogaev, E.I. et al. Familial Alzheimer's disease in kindreds with missense mutations in a gene on chromosome 1 related to the Alzheimer's disease type 3 gene. *Nature* **376**, 775-8 (1995).
28. St George-Hyslop, P. et al. Genetic evidence for a novel familial Alzheimer's disease locus on chromosome 14. *Nat Genet* **2**, 330-4 (1992).
29. Jayadev, S. et al. Presenilin 2 is the predominant gamma-secretase in microglia and modulates cytokine release. *PLoS One* **5**, e15743 (2010).
30. Kumar-Singh, S. et al. Mean age-of-onset of familial alzheimer disease caused by presenilin mutations correlates with both increased Abeta42 and decreased Abeta40. *Hum Mutat* **27**, 686-95 (2006).
31. Anderson, B.M. et al. Examination of association of genes in the serotonin system to autism. *Neurogenetics* **10**, 209-16 (2009).
32. Gurling, H.M. et al. Genomewide genetic linkage analysis confirms the presence of susceptibility loci for schizophrenia, on chromosomes 1q32.2, 5q33.2, and 8p21-22 and provides support for linkage to schizophrenia, on chromosomes 11q23.3-24 and 20q12.1-11.23. *Am J Hum Genet* **68**, 661-73 (2001).
33. Petit, J., Boisseau, P., Taine, L., Gauthier, B. & Arveiler, B. A YAC contig encompassing the 11q14.3 breakpoint of a translocation associated with schizophrenia, and including the tyrosinase gene. *Mamm Genome* **10**, 649-52 (1999).
34. Schol-Gelok, S. et al. A genome-wide screen for depression in two independent Dutch populations. *Biol Psychiatry* **68**, 187-96 (2010).
35. Vorstman, J.A. et al. Identification of novel autism candidate regions through analysis of reported cytogenetic abnormalities associated with autism. *Mol Psychiatry* **11**, 1, 18-28 (2006).

36. 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 (2007).
37. Aznar, S. & Knudsen, G.M. Depression and Alzheimer's disease: is stress the initiating factor in a common neuropathological cascade? *J Alzheimers Dis* **23**, 177-93 (2011).
38. Thambisetty, M. et al. Proteome-based plasma markers of brain amyloid-beta deposition in non-demented older individuals. *J Alzheimers Dis* **22**, 1099-109 (2010).
39. Lachno, D.R. et al. The influence of matrix type, diurnal rhythm and sample collection and processing on the measurement of plasma beta-amyloid isoforms using the INNO-BIA plasma Abeta forms multiplex assay. *J Nutr Health Aging* **13**, 220-5 (2009).
40. Ertekin-Taner, N. et al. Plasma amyloid beta protein is elevated in late-onset Alzheimer disease families. *Neurology* **70**, 596-606 (2008).
41. Kester, M.I., Verwey, N.A., van Elk, E.J., Scheltens, P. & Blankenstein, M.A. Evaluation of plasma Abeta40 and Abeta42 as predictors of conversion to Alzheimer's disease in patients with mild cognitive impairment. *Neurobiol Aging* **31**, 539-40; author reply 541 (2010).

## *4.2. Predicting stroke through genetic risk functions: The CHARGE risk score project*

Stroke 2014 Feb; 45(2):403-12

*The supplementary information for this paper is available online at*
http://stroke.ahajournals.org/lookup/suppl/doi:10.1161/STROKEAHA.113.003044/-/DC1

**ABSTRACT**

Background and Purpose

Beyond the Framingham Stroke Risk Score (FSRS), prediction of future stroke may improve with a genetic risk score (GRS) based on Single nucleotide polymorphisms (SNPs) associated with stroke and its risk factors.

Methods

The study includes four population-based cohorts with 2,047 first incident strokes from 22,720 initially stroke-free European origin participants aged 55 years and older, who were followed for up to 20 years. GRS were constructed with 324 SNPs implicated in stroke and 9 risk factors. The association of the GRS to first incident stroke was tested using Cox regression; the GRS predictive properties were assessed with Area under the curve (AUC) statistics comparing the GRS to age sex, and FSRS models, and with reclassification statistics. These analyses were performed per cohort and in a meta-analysis of pooled data. Replication was sought in a case-control study of ischemic stroke (IS).

Results

In the meta-analysis, adding the GRS to the FSRS, age and sex model resulted in a significant improvement in discrimination (All stroke: Δjoint AUC =0.016, p-value=2.3*10-6; IS: Δ joint AUC =0.021, p-value=$3.7*10^{-7}$), although the overall AUC remained low. In all studies there was a highly significantly improved net reclassification index (p-values $<10^{-4}$).

Conclusions

The SNPs associated with stroke and its risk factors result only in a small improvement in prediction of future stroke compared to the classical epidemiological risk factors for stroke.

**INTRODUCTION**

Stroke is a major and debilitating neurological disease that increases in frequency with age; it is estimated that in 2030, 23 million persons will have a first-ever stroke resulting in 7.8 million deaths[1,2]. Stroke is a complex disease with many modifiable risk factors, and a substantial genetic component, with heritability estimates varying from 17%[3] to 38%[4]. The genetic architecture of stroke has been difficult to unravel. Although recently some findings have been replicated for specific stroke subtypes,[5,6] initial discoveries of genetic variants from genome-wide association studies (GWAS) of all stroke (sub-types combined) outcome[7] have failed to replicate. This has led to the concept that different genes may be involved in different subtypes of stroke.

In contrast, many modifiable clinical and epidemiological risk factors consistently have been shown to increase the risk for stroke, and also have well replicated risk-associated genetic variants (single nucleotide polymorphisms – SNPs). Several modifiable risk factors have been combined into validated clinical prediction tools such as the Framingham Stroke Risk Score (FSRS), which incorporates systolic blood pressure, diabetes mellitus, cigarette smoking, prior cardiovascular disease, atrial fibrillation, left ventricular hypertrophy, and the use of antihypertensive medications[8,9]. The FSRS measures traits that may fluctuate in the short or medium term, thus affecting its predictive properties in any one individual. Insufficiency of the FSRS has been demonstrated in earlier studies[10]. Given the increasing availability of genotyping technology and the promise of using the information in a more personalized medicine approach, it is timely to investigate whether risk scores incorporating genetic information will add to the power to predict an individual's future risk for stroke.
Here we examine the predictive properties of a genetic risk score (GRS) to predict future stroke in community dwelling stroke-free individuals. We hypothesized that the combined effect of individual SNPs with small effects would improve prediction. A previous study by Kathiresan et al. found a GRS based on SNPs from a single class of risk factors, lipids, did have some value in reclassification, but not in improved discrimination of persons at future risk for cardiovascular disease[11]. However, cardiovascular diseases such as stroke have a complex pathophysiology, which can partially be accounted for in GRS, as they are in clinical risk scores. Here, we take the approach of including in a risk score, genetic variants associated with stroke *and* its multiple risk factors, with the goals of: assessing the potential of a score, based on SNPs associated with stroke and its risk factors, to predict stroke in general populations; and investigating whether the score could potentially add to the predictability of a score based on established stroke clinical and epidemiological risk factors. As far as we know, we are the first to try to combine not only a disease specific or risk factor specific set of SNPs into a risk score, but a comprehensive set of risk SNPs from the whole spectrum of non-behavioral risk factors for stroke. We also investigated the performance of the GRS in a higher risk population captured in a clinic-based case-control study of ischemic stroke (IS).

**Table 1. Participants included in the sample to develop the CHARGE Genetic Risk Score for Stroke and the Replication set**

| | | ARIC (n=9349) | CHS (n=3268) | FHS (n=4340) | RS (n=5763) | WTCCC (n=1581) |
|---|---|---|---|---|---|---|
| | | **Mean(SD)** | **Mean(SD)** | **Mean(SD)** | **Mean(SD)** | **Mean(SD)** |
| Cases (N (%)) | All stroke | 498 (5.3%) | 560 (17.1%) | 206(4.8%) | 783 (13.5%) | - |
| | Ischemic stroke | 437 (4.7%) | 453 (13.8%) | 166 (4.1%) | 467 (8.2%) | 985 (62%) |
| Age (baseline) | Cases (All stroke) | 57.17 (5.27) | 73.44 (5.45) | 75.14(9.93) | 72.12 (8.97) | 71.2 (8.7) |
| | Non-cases | 54.13 (5.67) | 72.1 (5.34) | 66.16(11.69) | 68.65 (8.96) | 66.8 (7.9) |
| Age (end) | Cases (All stroke) | 68.55 (7.19) | 81.87 (6.19) | 80.59(9.64) | 83.95 (7.21) | - |
| | Non-cases | 72.88 (6.25) | 85.69 (4.97) | 73.27(11.01) | 81.44 (7.46) | - |
| Sex (females, N (%)) | Cases (All stroke) | 218 (43.8%) | 351 (62.7%) | 109(55.6%) | 469 (58.6%) | 383 (38.9%) |
| | Non-cases | 4721 (53.3%) | 1639 (60.5%) | 2143(55.3%) | 2965 (59.5%) | 291 (48.7%) |

## MATERIALS AND METHODS

Our analyses are based on incident cases and stroke-free participants characterized in 4 cohorts participating in the Cohorts for Heart and Ageing Research in Genomic Epidemiology (CHARGE) consortium. CHARGE is a large consortium of major population-based prospective cohort studies of cardiovascular health that aims to identify new genetic variants for multiple quantitative, sub- and clinical factors contributing to health and disease in older persons[12]. The individual cohorts and the combined CHARGE genome wide association study of stroke genes have been previously described[7].

*Cohorts and case definition*

This analysis is based on the following CHARGE cohorts: the Atherosclerosis Risk in Communities (ARIC) study[13], the Cardiovascular Health Study (CHS)[14], the Framingham Heart Study (FHS)[15,16] and the first cohort of the Rotterdam Study (RS)[17]. From these cohorts, we included persons who were stroke-free at the age of 55 or older, of European descent, and who had complete outcome and genotype data. (Table 1, Supplemental Table I). For all cohorts, the baseline was established in the late 1980's and early 1990's, and all studies are ongoing. All participants provided informed consent and all studies were approved by their governing institutional review boards.

All cohorts defined stroke as a focal neurological deficit of presumed vascular cause with a sudden onset and lasting for at least 24 hours or until death if the participant died less than 24 hours after the onset of symptoms. All suspected events were adjudicated by stroke experts who reviewed medical records, death certificates, imaging studies, or some combination of these sources. We report on "All" stroke, which includes ischemic, hemorrhagic, and unknown sub-type, and separately on ischemic stroke, which is of presumed cardio-embolic/large vessel/small vessel origin. Subarachnoid hemorrhages were excluded from all analyses.

*Genotyping*

Each study separately genotyped or imputed SNPs to the same reference panel (see Supplemental Table II for methods) and provided data on imputation quality. Due to imputation, there were no missing genotypes in the datasets. Genotypes for each SNP were coded in terms of the number of risk alleles.

*Identifying risk factors, associated SNPs, and selecting SNPs for inclusion in the risk score*

<u>SNP selection</u>

Based on a literature review, as well as clinical and neurological expert opinion, we identified 9 domains of established risk factors for stroke that have also been studied in GWAS: high blood pressure, atherosclerosis, arrhythmia, diabetes, inflammation, blood constituents, hematologic changes, obesity, elevated lipids, and impaired kidney function. Within each of these risk factor domains we identified 3-5 traits that contribute to the overall domain (Table 2), resulting in a total of 33 traits. For each of the 33 traits we

identified from published, in press, and under review genome-wide association studies, SNP variants that associated with the trait at the standard GWAS significance level (*p-value* < $5*10^{-8}$) and for which there was evidence of independent replication (see Supplemental Table III for a complete list of references). We also included two ischemic stroke-associated SNPs identified and replicated by the International Stroke Genetics Consortium[5]. Although these 4 CHARGE cohorts were often included in the above referenced trait GWAS, in general the meta-analyses were based on many more subjects, with the proportion of subjects from these studies ranging between 0[18] and ~86%[19].
In total we identified 334 autosomal SNPs for 34 traits (including stroke). When several SNPs for a class of risk factors were located in the same loci, we selected the top SNP with the lowest p–value, from the most recent meta-analyses that included the largest number and thus had the largest power and highest precision. In some cases the same SNP was associated with multiple traits. For this, we assigned the SNP to the clinical rather than the sub-clinical trait (for example, diabetes over fasting glucose levels). After these exclusions we included 324 SNPs (see Supplemental Table III), including 28 traits with multiple SNPs and 6 with single SNPs.

**Table 2. Traits included in the CHARGE Genetic Risk Score for Stroke model**

| Risk score group | SNPs implicated in |
|---|---|
| Arrythmia | Atrial fibrillation, PR-interval, RR-interval |
| Atherosclerosis | Intima-media thickness, subclinical coronary artery disease, clinical coronary artery disease |
| Blood pressure | Hypertension, pulse pressure, mean arterial pressure |
| Diabetes | Type 2 diabetes, fasting glucose levels, insulin levels |
| Hematology | White blood cell count, hemoglobin, hematocrit, platelet count, mean platelet volume |
| Inflammation | Fibrinogen levels, C-reactive protein |
| Lipids | Total cholesterol, high-density lipoprotein, low-density lipoprotein, triglycerides |
| Nephrology | Estimated glomerular filtration rate, albumin/creatinine ratio, creatinine, end-stage renal disease |
| Obesity | Body mass index, waist-to-hip ratio, waist circumference |
| Stroke | Stroke genome-wide association studies |

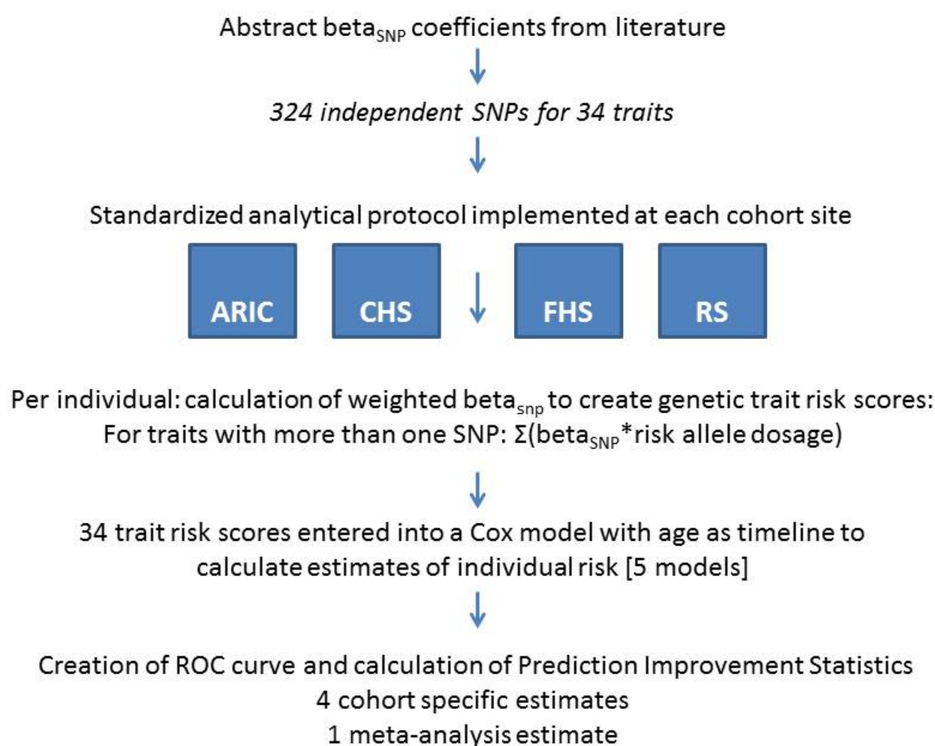
Statistics: Construction of the risk score

The construction of the risk score (Figure 1) is described in detail in the Supplemental Methods. In short, weighted risk scores were created for each of the 28 traits with multiple SNPs. Weights were the effects, or log odds ratios, of the risk allele on the outcome it was originally associated to. This effect was multiplied by the number of risk alleles (allele dosage: 0,1 or 2) the individual carried; for imputed SNPs, where the dosage is estimated, this value can obtain any fraction between 0 and 2. The risk score was the sum of the weight*dosage product for all SNPs within the trait. For the traits that had only one SNP associated (singletons), the allele dosage was used. Risk scores and singletons entered as covariates in Cox regressions with age as the time-to-event variable. Each cohort calculated a modified FSRS, which differed from the original FSRS only in the omission of the age variable.

*Analysis*

The individual risk predictions were calculated from five regression models: model 1 included sex only, model 2 included the GRS only, model 3 included sex and the GRS, model 4 included sex and the mFSRS, and model 5 included sex, the GRS and the mFSRS. CHS models also included a variable for study site and the FHS made adjustments for family relationships[15,16].

**Figure 1. Flow diagram: Construction of the Genetic Risk Score for Stroke.**



*Calculation and comparison of the ROC curves*

Receiver-operated characteristic (ROC) curves with corresponding areas under the curve (AUC) [95% confidence interval for each model] were created from the predicted risks derived from the regression models. The model AUCs were tested against the null model (AUC=0.50) and compared to each other. The latter comparisons were tested with the Hanley-McNeil test for comparison of correlated AUCs[20]. The Hanley-McNeil test requires a correlation between the AUCs, which was obtained using the metacor package for R[21]. Finally, we estimated meta-AUCs [95% confidence intervals] based on all cohorts combined using an inverse variance weighted meta-analysis.

*Reclassification statistics*

In secondary analyses we assessed the added clinical value of the genetic risk score over the FSRS using two statistics: the net reclassification improvement (NRI) without cut-off values (continuous NRI), which gives a summary of the number of subjects with reclassified predicted case status based on the new score; and the integrated discrimination improvement (IDI)[22,23], which describes the ability of the score to discriminate between cases and non-cases.

*External replication*

We investigated the predictive power of the same GRS in a German case-control set of IS previously described.[5,6] Cases (N=985) were the German samples of the Wellcome Trust Case Control Consortium 2 (WTCCC2), recruited in the Klinikum Grosshadern, Department of Neurology, Ludwig-Maximilians-University in Munich, Germany. Controls (N=596) were participants in the KORAgen study, residing in the Augsburg area in Germany (www.gsf.de/kora/en/english.html), with no history of stroke or transient ischemic attack. The studies were approved by the respective local Institutional Review Boards and all subjects gave informed consent. Details on genotyping and imputation are given in Supplemental Table II. To account for the case-control design we used a logistic regression model, adjusted for age and sex to estimate the trait scores and their association with stroke. As details on the elements of the FSRS were not generally available for the subjects in this cohort, we were unable to perform replication efforts for the FSRS comparisons in this cohort.

**RESULTS**

During follow-up, 2047 participants from the 22,720 developed first ever strokes, including 1523 ischemic strokes. The RS (N of cases=783, 13.5%) and CHS (N of cases=560, 17.1%) participants were older and had more incident stroke events (Table 1) than the FHS (N of cases=206, 4.8%) and ARIC (N of cases=498, 5.3%). participants. These cohorts also had a higher prevalence of hypertension than the other two. Most smokers were found in ARIC and RS (Supplemental Table I). The number of risk alleles had similar distribution in all cohorts (Supplemental Figure I) reflecting the population-based study designs of all cohorts.

Predicted risks from the models including the GRS and/or mFSRS were significantly higher in cases than in individuals who remained stroke-free (*p-value* < 0.001, Supplemental Table IV). Across cohorts, the AUC including only the GRS (Table 3, Supplemental Figure II) ranged from 0.563 to 0.617 for all stroke. When combining the findings of the cohorts, the meta-AUC was 0.578 (*p-value* = $9*10^{-10}$ compared to the model with only sex). For the model with sex and the GRS, the all stroke meta-AUC was 0.572, which was statistically different from the model including only sex (*p-value* = $9*10^{-18}$). For IS, the AUC of the model including only the GRS, ranged from 0.585 to 0.627 across cohorts; the meta-AUC was 0.592. For the combined model with sex and the GRS, meta-AUC was 0.597 (*p-value* compared to sex only = $2*10^{-19}$ ). (Table 4, Supplemental Figure II)

*Comparison between the genetic and clinical risk scores*

In all cohorts, the AUC for the mFSRS was higher than those seen for the GRS. For all stroke, the mFSRS AUC ranged from 0.587 in the RS to 0.709 in the FHS, with a highly significant meta-AUC of 0.621 (Table 3, Supplemental Figure III). In all cohorts, there was a low correlation between the absolute predicted risks derived from mFSRS and those derived from the genetic risk score (meta-correlation 0.012, meta *p-value* = 0.13). The full model (sex, mFSRS and the GRS) improved prediction significantly in all cohorts except the FHS, the cohort that was used to develop the original FSRS. The meta-AUC for the full model was 0.637. When the full model is compared to the model with sex and mFSRS, the GRS gives an improvement of 0.016 (*p-value* = $2*10^{-6}$) over classical risk factors. For IS, the meta-AUC for the full model was 0.654, with an improvement of 0.021 (*p-value* = $4*10^{-7}$) between the full model and the sex and mFSRS model (Table 4, Figure 2). Compared to the model based on sex and the FSRS, the GRS yielded a significant NRI (improvement ranging from 0.18 to 0.32 for all stroke and from 0.24 to 0.28 for IS, *p-values* ≤ $1.1*10^{-4}$) and IDI (improvement ranging from 0.005 to 0.02 for all stroke and from 0.008 to 0.021 for ischemic stroke, *p-values* ≤ $5*10^{-5}$) (Supplemental Table V).

*Replication*

We replicated our findings in the case-control sample of the WTCCC2. The AUC of the sex, age and GRS model was higher than the model with age and sex alone (difference between the two models = 0.014 (*p-value* = 0.04); Supplemental Table VI, Supplemental Figure IV). The reclassification statistics (continuous NRI = 0.309, *p-value* < $1*10^{-5}$, and IDI = 0.018, *p-value* < $1*10^{-5}$) showed a small but highly significant improvement in prediction of ischemic stroke when the GRS was added to models based on age and sex.

**Table 3. Area under the Curve for all stroke: CHARGE Genetic Risk Score for Stroke**

| Model | | AUC | 95%CI lower | 95%CI upper | *p-value* (compared to null) | *p-value* (compared to sex only model) |
|---|---|---|---|---|---|---|
| Sex only | ARIC | 0,548 | 0,525 | 0,570 | $2{,}908*10^{-5}$ | |
| | CHS | 0,535 | 0,509 | 0,561 | $8{,}328*10^{-3}$ | |
| | FHS | 0,499 | 0,463 | 0,534 | $9{,}387*10^{-1}$ | |
| | RS | 0,505 | 0,486 | 0,523 | $6{,}060*10^{-1}$ | |
| | **Meta** | **0,523** | **0,511** | **0,535** | $\mathbf{1{,}722*10^{-4}}$ | |
| Riskscore only | ARIC | 0,566 | 0,540 | 0,592 | $9{,}181*10^{-7}$ | $3{,}049*10^{-1}$ |
| | CHS | 0,595 | 0,570 | 0,620 | $9{,}477*10^{-14}$ | $6{,}892*10^{-4}$ |
| | FHS | 0,617 | 0,577 | 0,658 | $1{,}301*10^{-8}$ | $1{,}963*10^{-5}$ |
| | RS | 0,563 | 0,542 | 0,584 | $4{,}103*10^{-9}$ | $6{,}503*10^{-5}$ |
| | **Meta** | **0,578** | **0,565** | **0,591** | $\mathbf{6{,}273*10^{-32}}$ | $\mathbf{8{,}993*10^{-10}}$ |
| Sex + Riskscore | ARIC | 0,584 | 0,558 | 0,609 | $1{,}102*10^{-10}$ | $3{,}852*10^{-2}$ |
| | CHS | 0,587 | 0,562 | 0,613 | $9{,}053*10^{-12}$ | $8{,}019*10^{-5}$ |
| | FHS | 0,612 | 0,571 | 0,653 | $8{,}596*10^{-8}$ | $2{,}747*10^{-6}$ |
| | RS | 0,545 | 0,525 | 0,566 | $1{,}302*10^{-5}$ | $7{,}736*10^{-9}$ |
| | **Meta** | **0,572** | **0,560** | **0,585** | $\mathbf{6{,}273*10^{-32}}$ | $\mathbf{9{,}133*10^{-18}}$ |
| | | **AUC** | **95%CI lower** | **95%CI upper** | ***p-value*** **(compared to null)** | ***p-value*** **(compared to sex + FSRS model)** |
| Sex + FSRS | ARIC | 0,645 | 0,618 | 0,671 | $2{,}26*10^{-26}$ | |
| | CHS | 0,602 | 0,576 | 0,628 | $1{,}33*10^{-14}$ | |
| | FHS | 0,709 | 0,672 | 0,746 | $2{,}66*10^{-22}$ | |
| | RS | 0,587 | 0,565 | 0,608 | $1{,}37*10^{-14}$ | |
| | **Meta** | **0,621** | **0,608** | **0,634** | $\mathbf{6{,}62*10^{-73}}$ | |
| Sex + FSRS + Riskscore | ARIC | 0,664 | 0,639 | 0,690 | $2{,}63*10^{-33}$ | $1{,}56*10^{-3}$ |
| | CHS | 0,628 | 0,603 | 0,653 | $6{,}70*10^{-22}$ | $5{,}46*10^{-4}$ |
| | FHS | 0,707 | 0,667 | 0,747 | $6{,}97*10^{-22}$ | 0.91 |
| | RS | 0,601 | 0,579 | 0,622 | $5{,}57*10^{-19}$ | $7{,}70*10^{-3}$ |
| | **Meta** | **0,637** | **0,624** | **0,650** | $\mathbf{6{,}29*10^{-93}}$ | $\mathbf{2{,}31*10^{-6}}$ |

*All predictions result from Cox regressions with age as the time scale*

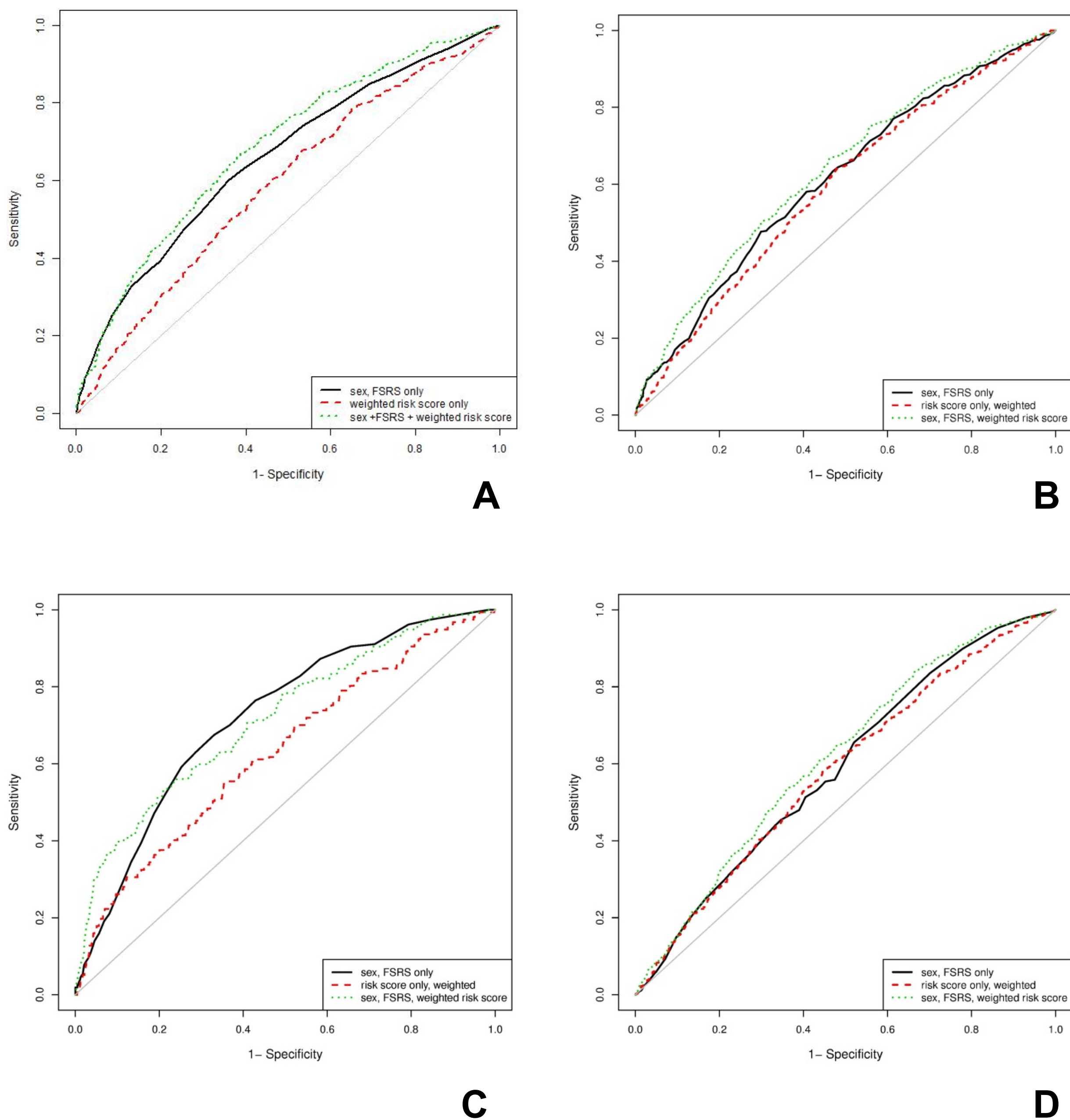
**Table 4. Area under the Curve for ischemic stroke: CHARGE Genetic Risk Score for Stroke**

| **Model** | | **AUC** | **95%CI lower** | **95%CI upper** | ***p-value* (compared to null)** | ***p-value* (compared to sex only model)** |
|---|---|---|---|---|---|---|
| Sex only | ARIC | 0,560 | 0,536 | 0,584 | $4,950*10^{-7}$ | |
| | CHS | 0,527 | 0,499 | 0,555 | $3,018*10^{-2}$ | |
| | FHS | 0,497 | 0,458 | 0,535 | $5,713*10^{-1}$ | |
| | RS | 0,521 | 0,498 | 0,545 | $3,561*10^{-2}$ | |
| | **Meta** | **0,532** | **0,518** | **0,545** | $\mathbf{7,464*10^{-6}}$ | |
| Riskscore only | ARIC | 0,585 | 0,558 | 0,612 | $8,561*10^{-10}$ | 0.17 |
| | CHS | 0,592 | 0,565 | 0,620 | $2,414*10^{-11}$ | $6,511*10^{-4}$ |
| | FHS | 0,627 | 0,583 | 0,671 | $1,135*10^{-8}$ | $1,302*10^{-5}$ |
| | RS | 0,586 | 0,560 | 0,613 | $8,987*10^{-11}$ | $2,681*10^{-4}$ |
| | **Meta** | **0,592** | **0,578** | **0,607** | $\mathbf{5,834*10^{-38}}$ | $\mathbf{2,493*10^{-9}}$ |
| Sex + Riskscore | ARIC | 0,607 | 0,581 | 0,634 | $2,291*10^{-15}$ | $1,596*10^{-5}$ |
| | CHS | 0,597 | 0,570 | 0,624 | $1,902*10^{-12}$ | $8,510*10^{-7}$ |
| | FHS | 0,622 | 0,579 | 0,666 | $4,430*10^{-8}$ | $1,969*10^{-6}$ |
| | RS | 0,578 | 0,552 | 0,604 | $4,103*10^{-9}$ | $9,108*10^{-9}$ |
| | **Meta** | **0,597** | **0,582** | **0,611** | $\mathbf{8,171*10^{-37}}$ | $\mathbf{1,551*10^{-19}}$ |

| **Model** | | **AUC** | **95%CI lower** | **95%CI upper** | ***p-value* (compared to null)** | ***p-value* (compared to sex + FSRS model)** |
|---|---|---|---|---|---|---|
| Sex + FSRS | ARIC | 0,658 | 0,629 | 0,686 | $2,25*10^{-27}$ | |
| | CHS | 0,613 | 0,585 | 0,640 | $6.56*10^{-15}$ | |
| | FHS | 0,721 | 0,682 | 0,759 | $3,12*10^{-21}$ | |
| | RS | 0,590 | 0,564 | 0,616 | $1,90*10^{-10}$ | |
| | **Meta** | **0,633** | **0,618** | **0,648** | $\mathbf{2,32*10^{-70}}$ | |
| Sex + FSRS + Riskscore | ARIC | 0,684 | 0,657 | 0,711 | $2,00*10^{-36}$ | $2,35*10^{-4}$ |
| | CHS | 0,637 | 0,610 | 0,665 | $2.87*10^{-21}$ | 0.002 |
| | FHS | 0,716 | 0,673 | 0,759 | $1,90*10^{-20}$ | 0.804 |
| | RS | 0,617 | 0,591 | 0,642 | $2,06*10^{-16}$ | $7,76*10^{-4}$ |
| | **Meta** | **0,654** | **0,639** | **0,669** | $\mathbf{3,45*10^{-96}}$ | $\mathbf{3,66*10^{-7}}$ |

*All predictions result from Cox regressions with age as the time scale*

**Figure 2. ROC curves for the discovery cohorts, ischemic stroke: CHARGE Genetic Risk Score for Stroke.**




*Each lettered panel gives the sensitivity*(1-specificity) curves for the clinical prediction (sex+FSRS, black line), prediction based on the GRS (red line), and on the two combined (green line). Panels: A=ARIC, B=CHS, C=FHS, D=RS.*

**DISCUSSION**

We assessed the predictive properties of a genetic risk score based on stroke risk factors in a population-based sample of 2,047 well-characterized incident stroke cases among 22,720 initially stroke-free individuals. We found that a genetic risk score that included genome-wide associated SNPs for 9 domains of risk factors plus stroke provided a small but very significant and consistent improvement in the prediction of an individual's risk for future stroke. This small difference was observed when the GRS was compared to sex adjusted models as well as to a widely used clinical-epidemiological risk score. Similar results were found when the score was applied in a large clinic-based case control sample.

Age is an important predictor of many diseases, in particular a late onset outcome as stroke. A large part of the discrimination of a prediction model is therefore determined by age, biasing the interpretation of the value of the other variables, in our case genetic predictors. In this study, we controlled for age as the follow-up time so we could clearly assess whether genetic information could improve prediction conditional on the person's age It is important to note that the AUCs are therefore in general lower than reported in the literature. For instance, in the Rotterdam Study, we find much lower AUCs in the Cox approach with age as the timescale (AUC for FSRS=0,59; p-value=$1*10^{-14}$) than in the logistic model where age is a covariate, essentially examining the GRC with all subjects of the mean age (AUC for FSRS =0.65; p-value=$2*10^{-39}$).

The GRS performs significantly worse than the age+sex model in the replication, but this was not the primary comparison we wished to test. Our goal was to test whether the GRS has added value over and above the clinical risk factors, and we do indeed find modest (Δ AUC (discovery) = 0.021, Δ AUC (replication) = 0.024) but significant (*p-value* (discovery) = $3.66*10^{-7}$, p-value (replication) = 0.03) (Table 4 & Supplemental Table 6) improvement in the AUCs in both the discovery and replication analyses. The difference in p-values most likely reflects the differences in sample size of the discovery and replication cohorts.

Before interpreting the findings it is important to note the assumptions underlying the construction of the score and consider the limitations: 1) The relationships between variants and risk factors, and between risk factor and outcome are always the same, ie (log)linear ; 2) Each variant associated with the *stroke risk factor*, will also be a risk variant for stroke; 3) The effect is proportional to the number of risk variants (0, 1, or 2 copies) (Whereas in reality, the variant may have dominant or recessive effect on stroke); 4) There is no interaction between genetic variants (we did not find evidence for interaction between the loci in this study). Each of these assumptions is necessary when creating such a score, but they may be simplifications of the true underlying biological model.

Previous studies have examined the association of clinical disease to SNPs that associate with risk factors for the disease. For example studies by Ehret et al and Wain et al [24,25] showed genetic variants associated with blood pressure traits were also associated with

stroke. Paynter et al.[26] assessed the performance of a GRS based on a small number of SNPs associated to cardiovascular disease (CVD) to predict CVD outcomes including stroke. They did not find their genetic risk score improved prediction over the traditional risk factors. A similar conclusion has been reached by other investigators examining risk SNPs for a single class of risk factors[11,27,28], although recently some authors have found risk differences between GRS quintiles for stroke[29,30]. Here, we present a different approach: we test a single outcome (stroke) to a risk score based on SNPs associated with a variety of risk factors. This approach takes into account the complexity of disease and our results suggest incorporating such genetic information into risk scores may be fruitful even in a very complex disease such as stroke.

Our analyses showed that the genetic risk score improved the discrimination over the mFSRS, although the absolute increase in prediction was small. The correlation between the mFSRS and genetic risk score was low suggesting the genetic variants, which are constant over the lifetime, can add to the information provided by a single assessment of variable risk factors such as blood pressure and glucose levels. Supporting evidence of this improvement is reflected in the improved reclassification statistics. Although reclassification is seen as clinically relevant, this finding should be interpreted with caution as the value of the IDI has been questioned[31], and both the IDI and NRI may be inflated[32].

One of the most notable findings is that the GRS worked similarly in 3 of the 4 different cohorts, which, although all population-based, had different age distributions. The absence of an improvement in prediction in the FHS beyond the mFSRS may well reflect overfitting of the mFSRS in this cohort, as the original FSRS was developed on the FHS source population. Another important finding is that AUCs and AUC improvements are higher and p-values are lower for ischemic stroke compared to all stroke, despite a smaller number of cases. This difference may reflect the fact IS is an etiologically more homogenous phenotype, and the risk factors we selected are better associated to ischemic stroke than to all stroke. This is consistent with GWAS discoveries which thus far have been limited to defined subtypes[5,6]. It will be valuable to have both better clinical prediction models and improved genetic prediction models targeting ischemic and hemorrhagic stroke as two distinct clinical entities, resulting in an improvement to the FSRS, which has a low AUC. The combined clinical-genetic-and epidemiological risk model may become a valuable tool for clinicians, and such differential risk models for ischemic versus hemorrhagic stroke may even help guide treatment decisions.

As array-based genotyping becomes more and more affordable and widely available, the possibility of multi-marker genetic risk profiling as part of daily medical practice becomes more realistic. As we show in this article, clinical risk profiling alone is still superior in predictive power to genetic risk profiling alone in the setting of a population-based cohort study of middle-aged and elderly subjects. In current clinical practice, some risk factors like

diabetes, hypertension, dyslipidemia and atrial fibrillation often come to light only after a stroke has occurred. Paroxysmal atrial fibrillation may be missed during regular ECG registration or even during a 24-hour Holter monitoring. To act pre-emptively on these risk factors, a subject would have to be screened on a regular basis and at least every few years depending on their age and perhaps other factors, we suggest that their genetic profile might help select persons at higher stroke risk for more frequent or thorough screening. A genetic risk score could be estimated at a single time point early in life, and persons with a higher genetic risk be targeted for more rigorous lifestyle counseling before risk factors emerge, more stringent clinical follow-up for control of risk factors, and even preventive medication such as platelet aggregation inhibitors, statins or antihypertensive medications in persons with borderline levels of these risk factors. . We anticipate that with rapid developments in unraveling the genetic origin of various stroke risk factors, the genetic prediction of stroke risk will improve in the near future improving the efficacy of early genetic risk profiling and targeted preventive interventions.

This study in a large prospectively followed population based epidemiological cohort yields a proof of principle that genetic variants associated with risk factors for stroke combined into a risk score improves discrimination of at-risk patients. These results are based on stroke-free individuals living in the community who may be examined in first line health services, and are replicated in individuals who have had a stroke and are identified in the hospital setting. However the small improvement we found is unlikely to be of clinical significance;. and only brings a small improvement over scores based on clinical information. In the future, however, as our understanding of the genetic architecture of stroke -its sub-types and risk factors - improves, GRS could become powerful additions to clinically measured risk factors.

**ACKNOWLEDGMENTS**
The Atherosclerosis Risk in Communities Study: The authors thank the staff and participants of the ARIC study for their important contributions.
Rotterdam Study: We thank Pascal Arp BSc, Mila Jhamai BSc, Marijn Verkerk, Lizbeth Herrera MPH and Marjolein Peters MSc (Department of Internal Medicine, Erasmus University Medical Center) for their help in creating the GWAS database, and Karol Estrada (PhD, Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands, and Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA) and Maksim V. Struchalin (PhD, Department of Epidemiology, Erasmus University Medical Center) for their support in creation and analysis of imputed data. The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists.

**SOURCES OF FUNDING**
The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL70825, R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.
This CHS research was supported by National Heart, Lung, and Blood Institute contracts N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85084, N01HC85085, N01HC85086; N01HC35129, N01HC15103, N01HC55222, N01HC75150, N01HC45133, N01HC85239, and by HHSN268201200036C and NHLBI grants HL080295, HL087652, HL105756 with additional contribution from NINDS. Additional support was provided through AG023629, AG15928, AG20098, and AG027058 from the NIA. See also http://www.chs-nhlbi.org/pi.htm. DNA handling and genotyping at Cedars-Sinai Medical Center was supported in part by the National Center for Research Resources, grant UL1RR033176, and is now at the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124; in addition to the National Institute of Diabetes and Digestive and Kidney Disease grant DK063491 to the Southern California Diabetes Endocrinology Research Center.
This work was supported by the National Heart, Lung and Blood Institute's Framingham Heart Study (Contract No. N01-HC-25195) and its contract with Affymetrix, Inc. for genotyping services (Contract No. N02-HL-6-4278) and grants (U01 HL096917 and R01 HL093029). A portion of this research utilized the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. Analyses reflect intellectual input and resource development from the Framingham Heart Study

investigators participating in the SNP Health Association Resource (SHARe) project. This study was also supported by grants from the National Institute of Neurological Disorders and Stroke (NS17950) and the National Institute of Aging (AG033193, AG081220, AG16495). The content is solely the responsibility of the authors and does not necessarily represent the official views of NINDS, NHLBI, NIA, NIH or AHA.
The generation and management of GWAS genotype data for the Rotterdam Study are supported by the Netherlands Organisation of Scientific Research (NWO) Investments (nr. 175.010.2005.011, 911-03-012). This study is funded by the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) Netherlands Consortium for Healthy Ageing (NGI/NWO-NCHA; project nr. 050-060-810). 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. Bruce M. Psaty serves on a DSMB for a clinical trial of a device funded by Zoll LifeCor and on the Steering Committee of the Yale Open Data Access Project funded by Medtronic. James B. Meigs is supported by NIH grant K24 DK080140. Lenore J. Launer and Mike Nalls' participation was supported entirely by the Intramural Research Program of the NIH, National Institute on Aging (Z01 AG000954-06) and portions of Mike Nalls' contribution utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, Md. (http://biowulf.nih.gov). Abbas Dehghan is supported by Netherlands Organisation for Scientific Research (NOW) grant (veni, 916.12.154) and the Erasmus University Rotterdam (EUR) Fellowship. Stephanie Debette is a recipient of a "Chaire d'Excellence grant from the Agence National de la Recherche". The study sponsors played no role in the design and conduct of the study, collection, management, analysis, or interpretation of the data, or preparation, review, or approval of the manuscript.

**REFERENCES**


1. Mukherjee, D. & Patil, C.G. Epidemiology and the global burden of stroke. *World Neurosurg* **76**, S85-90 (2011).
2. Strong, K., Mathers, C. & Bonita, R. Preventing stroke: saving lives around the world. *Lancet Neurol* **6**, 182-7 (2007).
3. Bak, S., Gaist, D., Sindrup, S.H., Skytthe, A. & Christensen, K. Genetic liability in stroke: a long-term follow-up study of Danish twins. *Stroke* **33**, 769-74 (2002).
4. Bevan, S. et al. Genetic heritability of ischemic stroke and the contribution of previously reported candidate gene and genomewide associations. *Stroke* **43**, 3161-7 (2012).
5. International Stroke Genetics, C. et al. Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. *Nat Genet* **44**, 328-33 (2012).
6. Traylor, M. et al. Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies. *Lancet Neurol* **11**, 951-62 (2012).
7. Ikram, M.A. et al. Genomewide association studies of stroke. *N Engl J Med* **360**, 1718-28 (2009).
8. Wolf, P.A., D'Agostino, R.B., Belanger, A.J. & Kannel, W.B. Probability of stroke: a risk profile from the Framingham Study. *Stroke* **22**, 312-8 (1991).
9. D'Agostino, R.B., Wolf, P.A., Belanger, A.J. & Kannel, W.B. Stroke risk profile: adjustment for antihypertensive medication. The Framingham Study. *Stroke* **25**, 40-3 (1994).
10. Wassertheil-Smoller, S. et al. Improvement in stroke risk prediction: role of C-reactive protein and lipoprotein-associated phospholipase A(2) in the women's health initiative. in *Int J Stroke* 2012/10/24 edn (2012).
11. Kathiresan, S. et al. Polymorphisms associated with cholesterol and risk of cardiovascular events. *N Engl J Med* **358**, 1240-9 (2008).
12. Psaty, B.M. 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* **2**, 73-80 (2009).
13. ARIC-investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol* **129**, 687-702 (1989).
14. Fried, L.P. et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol* **1**, 263-76 (1991).
15. Dawber, T.R. & Kannel, W.B. The Framingham study. An epidemiological approach to coronary heart disease. *Circulation* **34**, 553-5 (1966).
16. Feinleib, M., Kannel, W.B., Garrison, R.J., McNamara, P.M. & Castelli, W.P. The Framingham Offspring Study. Design and preliminary data. *Prev Med* **4**, 518-25 (1975).

17. Hofman, A. et al. The Rotterdam Study: 2010 objectives and design update. *Eur J Epidemiol* **24**, 553-72 (2009).
18. Marroni, F. et al. A genome-wide association scan of RR and QT interval duration in 3 European genetically isolated populations: the EUROSPAN project. *Circ Cardiovasc Genet* **2**, 322-8 (2009).
19. Dehghan, A. et al. Association of novel genetic Loci with circulating fibrinogen levels: a genome-wide association study in 6 population-based cohorts. *Circ Cardiovasc Genet* **2**, 125-33 (2009).
20. Hanley, J.A. & McNeil, B.J. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* **148**, 839-43 (1983).
21. Laliberté, E. metacor: Meta-analysis of correlation coefficients. R package version 1.0-2. 21-03-2011 edn (2011).
22. Pencina, M.J., D'Agostino, R.B., Sr. & Steyerberg, E.W. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med* **30**, 11-21 (2011).
23. Pencina, M.J., D'Agostino, R.B., Sr. & Demler, O.V. Novel metrics for evaluating improvement in discrimination: net reclassification and integrated discrimination improvement for normal variables and nested models. *Stat Med* **31**, 101-13 (2012).
24. Ehret, G.B. et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103-9 (2011).
25. Wain, L.V. et al. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat Genet* **43**, 1005-11 (2011).
26. Paynter, N.P. et al. Association between a literature-based genetic risk score and cardiovascular events in women. *JAMA* **303**, 631-7 (2010).
27. Brautbar, A. et al. A genetic risk score based on direct associations with coronary heart disease improves coronary heart disease risk prediction in the Atherosclerosis Risk in Communities (ARIC), but not in the Rotterdam and Framingham Offspring, Studies. *Atherosclerosis* **223**, 421-6 (2012).
28. Hernesniemi, J.A. et al. Genetic profiling using genome-wide significant coronary artery disease risk variants does not improve the prediction of subclinical atherosclerosis: the Cardiovascular Risk in Young Finns Study, the Bogalusa Heart Study and the Health 2000 Survey--a meta-analysis of three independent studies. *PLoS One* **7**, e28931 (2012).
29. Yiannakouris, N. et al. Genetic predisposition to coronary heart disease and stroke using an additive genetic risk score: a population-based study in Greece. *Atherosclerosis* **222**, 175-9 (2012).
30. Havulinna, A.S. et al. A blood pressure genetic risk score is a significant predictor of incident cardiovascular events in 32 669 individuals. *Hypertension* **61**, 987-94 (2013).

31. Kerr, K.F., McClelland, R.L., Brown, E.R. & Lumley, T. Evaluating the incremental value of new biomarkers with integrated discrimination improvement. *Am J Epidemiol* **174**, 364-74 (2011).
32. Hilden, J. & Gerds, T.A. A note on the evaluation of novel biomarkers: do not rely on integrated discrimination improvement and net reclassification index. in *Stat Med* 2013/04/05 edn (2013).

# Chapter 5 : General discussion

Findings of this thesis and directions for future research

**FINDINGS OF THIS THESIS**

This thesis describes a comprehensive search for genetic variants influencing the risk of Alzheimer's disease (AD) in the general population. To discover common and rare genetic variation that confer small to large effects on AD and its endophenotypes I employed various statistical techniques including linkage and association covering a diverse variety of datasets ranging from candidate genes, genome-wide single nucleotide polymorphism (SNP) genotyping to exome-sequencing. In this chapter I summarize the findings of my thesis and discuss the implication of these findings towards the understanding of the pathogenesis of AD while commenting on the prospective research .

*Genetics of AD*

In Chapter 2, we sought for common genetic variants that are associated with AD. For this, we participated in a large collaboration, the International Genomics of Alzheimer's Project (IGAP). In this largest population study on AD to date, the meta-analysis of initially analyzed and replication datasets, comprising 25,580 cases and 48,466 controls, yielded 11 new genome-wide significant loci. The *SORL1* gene, sortilin-related receptor 1, which is located within one of these loci, is involved in APP trafficking. This gene harbors mutations involved in early-onset autosomal dominant AD, as well as in late-onset familial and sporadic AD, whereas findings in patients with sporadic AD have been ambiguous.[1-3] This is the first time a common variant in a known Mendelian gene for AD shows association in a GWAS. Besides several variants in genes involved in APP and Tau processing (*SORL1*, *CASS4*, *FERMT2*), other gene variants suggest involvement of other pathways including the immune response and inflammation (*HLA-DRB5/DRB1*, *INPP5D*, *MEF2C*), cell migration (*PTK2B*), lipid transport (*SORL1*), hippocampal synaptic function (*MEF2C, PTK2B*), cytoskeletal function (*CELF1, NME8, CASS4*), as well as in gene expression and post-translational modification of proteins and myeloid cell function (*INPP5D*). The yield of this collaborative effort, adding 11 new risk loci to an existing list of 10, proves that increasing the numbers of subjects and genetic variants helps to discover new associations. It was the first collaborative GWAS project on AD using an early 1000 genomes imputation panel instead of the HAPMAP panel. Using many more reference genomes than previously, to fill in the gaps between the single nucleotide variants (SNVs), the number of common variants available for analysis has been doubled.[4] The August 2010 release of the 1000 Genomes used in Chapter 2 provided us with ~11 million imputed SNVs, compared to ~2.5 million in the old standard, HapMap2.

Although this is the largest dataset analysed to date, there is no evidence that GWAS of AD has reached its limits, and new GWAS may bring to surface even more genetic variants. The analyses in Chapter 2 use the August 2010 release, a beta version of the 1000 Genomes, which has now been superseded by the phase 1 version 3. This latest version increased the number of SNVs available for imputation to ~38 million, and includes additional information on insertions and deletions. Beside the newer imputation, more power will be achieved by increasing the number of cases and controls. Since the last inclusion in 2011 (based on the

latest available datasets by then), the longitudinal cohorts have seen many former non-cases convert into AD cases and in these aging cohorts, the dementia free subjects have become much older and therefore become more informative. Further, many case-control studies have now imputed their GWAS data to the latest panel or expanded their datasets. Such increasing number of subjects and more detailed imputations will allow us to find association of rarer variants and therefore IGAP is currently performing a new round of GWAS based on the latest data based on the most recent 1000genomes version.

*Genetics of cognition*

In Chapter 3, we focused on cognitive function as an endophenotype of AD. Chapter 3.1 describes our search for rare variants linked to cognitive function in the ERF study. Our genome-wide linkage analysis of poor (lowest 10 percent of the population) performance on tests of memory, visuospatial ability, executive function and general intelligence yielded six significant linkage regions (1p13.1, 12q24.33, 19q13.43, 20p13, 21q22.13 and 21q22.3), all of which were implied in several traits. Three of these regions were described before: Chromosome 1p13.1 was previously linked to AD[5], 12q24.33 was earlier seen in schizophrenia and memory,[6,7] and 21q22 was seen in a linkage study of verbal IQ in a large linkage meta-analysis.[8] In this study, we were able to search directly for the causal variants underlying the linkage peaks using exome sequence and exome array data of 2430 to 2487 subjects (depending on the trait), using the scores on the cognitive tests as outcomes. We identified four rare to low frequency variants, all with predicted deleterious effects on the corresponding proteins. The genes harboring these variants are all implicated in developmental or degenerative processes in the brain. One variant was identified in *TRIB3,* a known gene for cognitive function,[9] on chromosome 20 (composite executive function, rs41281850 leading to a serine to asparagine substitution on codon 146, *p-value* = 0.0056). *TRIB3*, Tribbles pseudokinase 3, is a specific inducer of neuronal apoptosis activated by endoplasmatic reticulum stress or withdrawal of nerve growth factor.[10-12] It is also important in glucose homeostasis, and variants in *TRIB3* are associated to metabolic syndrome and atherosclerosis.[13] For the first time, cognitive function in healthy individuals has been found to be associated with a non-synonymous deleterious variant. Two variants were found for chromosome 19, one in *TRIM28* (associated to recognition, rs138661227, *p-value* = $3.7*10^{-5}$), and one *MZF1* (associated to the Dutch Adult Reading Test, *p-value* = $1.3*10^{-3}$). Both genes are candidate genes for cognitive function. *TRIM28*, tripartite motif containing 28, also known as *KAP1* (kinesin–II-associated protein) or *KRIP-1* (KRAB-interacting protein 1), is an essential cofactor of the KRAB-zinc finger proteins and plays a role in gene expression through regulation of gene acetylation [14,1515]. It is expressed throughout the brain and especially in the hippocampus and cerebellum in mice. Deleting *TRIM28* in the forebrain in early postnatal life of mice affects expression levels in the dentate gyrus and CA1 regions in the hippocampus mainly, resulting in increased anxiety behavior and impaired learning when exposed to subchronic stress . The identified deleterious variant causes an arginine to cysteine substitution at codon 795. The other identified deleterious

variant substituted threonine with arginine at codon 635 in *MZF1*, a myeloid zinc finger protein 1 gene (alias: Zinc Finger Protein 42, *ZNF42*) regulating DNA transcription. The gene influences the expression of cathepsin B,[16] a β-secretase involved in the production of the neurotoxic β-amyloid peptides that constitute the amyloid plaques in AD. Knockout mice show improved memory deficits and β-amyloid related biomarker levels.[17] Moreover, *MZF1* has been identified as a possible central protein in inflammatory response and cholesterol metabolism pathways in AD.[18] A question to answer is whether these two variants in the same regions are both causally related to AD or whether only one of the two is pathophsiologically relevant. They are not in linkage disequilibrium and are associated to different outcomes (*TRIM28* with recognition and *MZF1* with general intelligence), and therefore it cannot be excluded that both may have independent effects.
Finally we found a deleterious variant in*PDE4DIP* changing a proline into a serine at codon 30, associated with the composite score for executive function (rs76199660, *p-value* = $3.1*10^{-4}$). *PDE4DIP*, phosphodiesterase 4D interacting protein, is an orthologue of the centrosomin gene. In Drosophila, variants in this gene lead to a reduced cell number throughout the nervous system and death in the late embryonal or early post-hatching period. *PDE4DIP* anchors the phosphodiesterase 4D (PDE4D), the primary cAMP hydrolyzing enzyme in cells, with highly variable expression during embryogenesis,[19] to the Golgi/centrosome region of the cell, and thus plays a role in the cell division cycle.[20] PDE4D is under study as a drug target for, among other diseases, AD and schizophrenia, as inhibition of PDE4D has a beneficial effect on memory, making this finding of clinical interest.[21,22] As mutations are relatively frequent at MAFs between 0.5% and 1.9% in dbGAP populations, carriers may form an interesting subgroup to target in the evaluation in future drug trials. A *PDE4DIP* paralogue, *CDK5RAP2*, harbors causal mutations for autosomal recessive primary microcephaly and the protein also interacts with the prion protein gene (*PRNP*), suggesting that *PDE4DIP* may be involved in a wide range of physiological processes ranging from early development to late age.[23] Genetic variations in *PDE4D* might be associated with susceptibility to stroke, although findings in the Icelandic population have not been consistently replicated.[24-30] In our previous studies, The T allele of SNP39 increased the risk of small-vessel infarction 6.3 times (95% CI 1.4 to 28.7) compared with controls (*p-value* = 0.02).[31] No associations were found for large-vessel stroke. We did not find association with stroke but carriers of one of the Icelandic risk variants had an increased risk of small-vessel infarction. Our findings of *PDE4DIP* may link cognitive function to AD but also to large or small vessel pathology in the brain. It is of note that no homozygotes for the mutation in *PDE4DIP* were identified in ERF, the Rotterdam Study or in the reference genome datasets available online despite the relatively high frequency of this variant. This might indicate that the mutation is developmentally lethal in homozygous individuals, as has been described for mutations in the Drosophila ortholog of the gene. This observation asks for further functional studies, as do other identified rare mutations.

Apart from rare mutations with large effects, common variants with smaller effect may be involved in cognitive functioning. In Chapter 3.2 and 3.3, we describe our search for these mutations in genome-wide association studies in the Cohorts for Heart and Ageing (CHARGE) consortium for executive functioning and processing speed (Chapter 3.2) and memory (Chapter 3.3) respectively. We identified a variant in an alternative first intron of the cell adhesion molecule 2 (*CADM2* also known as *SYNCAM2*) gene on chromosome 3 reaching genome-wide significance for processing speed as measured by the letter-digit or digit-symbol substitution tasks in a sample of 34730 subjects (rs17518584, *p-value* = $3.91*10^{-9}$). This variant influences *CADM2* expression, especially in the cingulum (www.broadinstitute.org/gtex[32]). The protein encoded by *CADM2* is involved in glutamate signaling (*p-value* = $7.22x10^{-15}$, www.genenetwork.nl), gamma-aminobutyric acid (GABA) transport (*p-value* = $1.36x10^{-11}$) and neuron cell-cell adhesion (*p-value* = $1.48x10^{-13}$). Moreover, it has been studied as a candidate for autism and showed suggestive association to personality traits,[33,34] making it a good candidate for cognitive function as well as psychiatric disease and dementia. In the GWAS of memory (Chapter 3.3), we identified three genome-wide significant signals: rs4420638, near *APOE*, was associated with poorer delayed recall performance in discovery (*p-value* = $5.57x10^{-10}$) and replication cohorts (*p-value* = $5.65x10^{-8}$). This association was stronger in the oldest persons. Two suggestive associations (*p-value* $< 5x10^{-6}$) replicated at *p-value* $< 0.05$, reaching genome-wide significance in combined analyses of discovery and replication (rs11074779 near *HS3ST4, p-value* = $3.11x10^{-8}$, and rs6813517 near *SPOCK3*, *p-value* = $2.58x10^{-8}$). Heparan sulfate (glucosamine) 3-O-sulfotransferase 4 (*HS3ST4*) is strongly expressed in the hippocampus and is thought to play a role in Herpes Simplex Virus (HSV)-1 pathogenesis.[35] As AD-related plaques and tangles contain many HSV-1 binding proteins, the *HS3ST4* gene may mediate the relationship between HSV-1 and AD.[36] Sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 3 (*SPOCK3*) encodes a member of a novel family of calcium-binding proteoglycan proteins, which is strongly expressed in cerebral cortex and hippocampus (www.genenetwork.nl). As there is no 1:1 relation of endophenotypes to the disease of interest, the question is to what extent genes involved in cognitieve function are relevant for AD. Of note is that the CADM2 gene identified for processing speed seems to relate to cognitive processes and neuropsychiatric disease in a broad sense, while the *APOE* locus and two new candidate genes highly expressed in the hippocampus suggest that the findings in memory may be more extendable to AD. Yet there is no evidence in the IGAP analyses (Chapter 2) for the association to *HS3ST4* (*p-value* = 0.97) and *SPOCK3* (*p-value* = 0.32) to AD, implying that findings of a single endophenotype could not automatically be extended to the associated end-stage disease.[37]

*Vascular and Alzheimer pathology*

The relationship between vascular and Alzheimer pathology has been one of major research topics at the Erasmus MC Rotterdam. In the hypothesis-free genome based approaches, there has been little evidence for a genetic basis of the overlap in pathology between the

two diseases. Although none of the GWAS genes points directly into this direction as discussed earlier, our findings of *PDE4DIP* are compatible with a joint genetic pathway underlying AD, large and small vessel pathology. In Chapter 4.1, we searched for putative AD (AD) risk genes in middle-aged hypertensive subjects in the Erasmus Rucphen Family (ERF) study, using plasma amyloid β levels as an endophenotype. We performed a genome-wide linkage screen on 125 subjects from a single large pedigree, and identified two regions that were suggestively linked to plasma Aβ40 levels, on chromosomes 1q41 (LOD = 2.07) and 11q14.3 (LOD = 2.97). We subsequently screened these regions for single nucleotide polymorphisms (SNPs) associated to the Aβ40 levels in the study subjects and in 320 subjects from the population-based Rotterdam Study meeting the same inclusion criteria. We found association with SNPs in and around the Presenilin 2 (*PSEN2*) gene in the 1q41 region (*p-value* = $2.58*10^{-4}$ for rs6703170). *PSEN2* harbours mutations leading to early onset, autosomal dominant AD through a disruption of normal cleavage of APP, probably causing an increased production of the toxic amyloid β isoform Aβ42.[38]Our findings indicate for the first time that relatively common variants in the *PSEN2* gene to an AD endophenotype. Moreover, a very recent analysis of exome variants determined by the Illumina Exome Array in the Cohorts for Heart and Ageing Research in Genomic Epidemiology (CHARGE) consortium, a missense variant that encodes a *Met*174*Val* amino acid was associated to AD. The variant is listed in the AD & Frontotemporal Dementia Mutation Database (http://www.molgen.ua.ac.be/ADMutations/) as pathogenic[39,40], but it is predicted to be benign by the bio-informatics tools (Polyphen2 score 0.01).[25] This variant was not found in the exome sequence or exome chip data analysis of ERF.

Analyzing the chromosome 11 region linked to Aβ40 levels in more detail, strong associations were not found, which may be explained by the small sample size. The SNPs showing some association in ERF (top hit rs2514299, *p-value* = $3.2*10^{-3}$) did not replicate in the Rotterdam Study. The two SNPs most strongly associated with Aβ40 levels emerging from a meta-analysis of Rotterdam and ERF study were rs947937 (*p-value* = $7.3*10^{-5}$) and rs947935 (*p-value* = $9.6^{41}*10^{-5}$). However, expression quantitative trait locus (eQTL) analysis in this region showed no convincing results. The region reaching its maximum LOD at chromosome 11q14.3, with a nearly significant genome-wide linkage has been associated with various neuropsychiatric diseases, including autism and schizophrenia. Interestingly, this region is close to the region 11q25 previously found associated with depressive disorder, autism and late-onset AD.[41-46]

Our observations could not be confirmed in GWAS of plasma amyloid β levels in population cohorts. (Chouraki et al, Molecular Psychiatry, in press). The most significant finding from the GWAS, rs11241936, (p-value for Aβ42 = $2.3*10^{-7}$) is approximately 6 Kb away from the Cortexin 3 gene. Overexpression of this gene in vitro lowered Aβ42 secretion. Hypertension was not an inclusion criterium in this study, but genes involved in blood pressure were overrepresented in the nominally significant results in pathway analyses. In Chapter 4.1, we

aimed to study AD pathophysiological factors related to blood pressure, as hypertension is a risk factor for both AD and stroke. Understanding the genetics of stroke is of interest in itself given the frequency and impact of this disease. In Chapter 4.2, we took the step beyond GWAS and linkage as such, investigating whether we could calculate a risk/predcition score based on the SNPs associated to clinical risk factors for the clinical risk factors for stroke such as diabetes, hypertension or cardiac arrythmiacould be of use in stoke prediction. The identified SNPs in GWAS are not individually useful as predictors, as their individual effects on complex disease such as stroke or dementia are usually very small. However, their combination may be informative, and given the high frequency of most of these SNPs, any subject will carry several of these risk alleles. Such combined risk scores of SNPs collected for a variety of traits does indeed have statistically significant added power in discerning future stroke cases from stroke-free survivors (All stroke: $\Delta$joint AUC =0.016, *p-value* = $2.3*10^{-6}$; IS: $\Delta$ joint AUC =0.021, *p-value* = $3.7*10^{-7}$).

The absolute improvement in stroke prediction reached by the genetic risk score is small and clinically not useful at this moment. Still, we have delivered a proof of principle for the use of composite genetic risk scores in disease prediction, and with further gene discovery, they may gain their place in clinical practice in the future. The International Stroke Genetics Consortium (IGSC) replicated our findings and in parallel ran similar experiments, with a different risk score construction and in a case-control dataset (Malik et al, Stroke, in press). A more important conclusion from a methodological perspective is the finding that the variants implicated in the *risk factors* for stroke can be used to predict stroke. The question remains whether this approach may also be useful in AD – and other endophenotypes. As the genetic research of AD has yielded both rare and common variants, identification of variants associated to cognitive function goes slowly, and genes identified for related diseases are therefore of particular interest. Verhaaren et al have shown that in nondemented people, there is only a marginal joint effect of AD genes on memory function independent from *APOE*.[47], Also, these variants hardly improve prediction of AD. A GWAS of educational attainment in a discovery sample of 101,069 individuals and a replication sample of 25,490 identified and replicated three independent genome-wide significant variants (rs9320913, rs11584700, rs4851266).[48] Estimated effects sizes were small (coefficient of determination $R^2 \approx 0.02\%$). Therefore these candidate SNPs are promising for follow-up studies of cognitive function. As a proof of principle we used the variants identified for educational attainment [48] to predict cognitive function. This weighted risk score[49]of the three significant findings from the education GWAS was associated to the letter-digit substitution task (LDST) meta-analysis when correcting for age and sex ($\beta$ = 0.11, *p-value* = 0.01), suggesting that there is indeed predictive value in variants identified through GWAS.

## FUTURE RESEARCH

*Rare variants and next generation sequencing*

The search for rare variants using next generation sequencing (whole exome or genome) or arrays for rare variants (e.g. the exome array) is another avenue.[50] This approach has already yielded two rare variants for AD: *APP* and *TREM2.*[51-53] The *APP* variant (A673T) concerns a protective variant, which was identified after sequencing of 1795 persons in Iceland and subsequent chip-typing or imputation of the sequences in 3,048AD patients and 79,248 controls. This rare variant is only seen in North Europeans and a selective group of North Americans of European, possibly of Scandinavian or Icelandic ancestry, but not among any other groups of Europeans or Asians.[54] We did not find this variant among 9622 subjects from the Rotterdam and ERF studies. The combination of whole genome sequencing (30x) in a selected series of Icelandic individuals (including some patients with AD) and subsequent imputations turned out to be a successful approach in the Icelandic population. A similar approach of genome-wide sequencing was followed by targeted sequencing and imputations in the identification of the *TREM2* (R47H) variant in Iceland and simultaneously in the UK.[51,53] The association of this variant with AD has also been found elsewhere in Europe including in the Rotterdam Study in the Netherlands and the UK and France. So far the variant has not been seen in Asians.[55] Mutations in *TREM2*, including R47H, have also been associated to FTLD.[56] Thus, AD and FTLD may share common variants.
As shown in Chapter 3.1, next generation sequencing provides a powerful method to pick up possible causative genetic variants. For the near future, the expectations for discoveries of disease-associated genetic variants through next generation sequencing methods are high. The CHARGE consortium has performed joint calling of exome array data in all the core cohorts including the Rotterdam study.[50] The AD CHARGE analyses shows interesting results like the *PSEN2* mutation described above. As IGAP is now also performing a collaborative effort on AD exome chip analyses, and other cohorts will also obtain exome chip data, larger meta-analyses like for GWAS are not far away, and we will be able to demonstrate rare variants with a relatively large effect in association to AD as well as many endophenotypes. The main yield is to be expected in the relatively rare range of variants (minor allele frequencies between 0.5 and 5%) with relatively large effects.

As for exome sequence data, CHARGE (including the Rotterdam Study) and the AD Genetics Consortium (ADGC) are collaborating in The Alzheimer's Disease Sequencing Project (ADSP), genome-wide sequencing data of 582 cases of familial AD from 111 families, and exome sequencing of 5000 sporadic AD cases and 5000 controls. (https://www.niagads.org/sites/all/public_files/ADSPdocs/ADSP-SUMMARY-PLAN-41513.pdf - accessed December 23, 2013) The familial AD dataset includes a number of extended families from isolated population including those from Genomic Research in Isolated Populations (GRIP) AD cohort.[46,48,57,58] These data may yield some novel family-specific, possibly autosomal dominant, mutations, and it is expected that we will see the same mutations or other deleterious variants in the same genes in independent families.

Besides discovering new gene-disease associations, these sequencing data may allow us to finally pinpoint the exact perpetrators behind our linkage signals. The case-control dataset has been designed to find risk and protective alleles for AD at a minor allele frequency of down to 0.1% and targets specifically the exome. Both approaches may reveal new pathways in the AD pathogenesis. Collapsing variant analyses, like the sequence kernel association test (SKAT), can provide evidence for association of genes in which multiple rare mutations are found.[59]

The Decode study in Iceland has proven the value of imputation of rare variants under condition of a reliable haplotype around tha pathogenic variant, facilitating the association of rare variants to a disease or endophenotype using GWAS. This approach may also be applied to the Dutch population. In the Netherlands, the Genome of the Netherlands (GoNL) is a reference set based on the sequence of 250 trios from across the Netherlands.[60] This reference set will lead to more accurate imputations for people from Dutch or related ancestry. Using GoNL, we will be better able to pinpoint relatively rare variants that have become enriched within the Netherlands, but on the other hand we may also reliably impute non-specific variants better in the Netherlands.

*Common variants: opportunities and challenges*

Despite the successes in GWAS in the past decade, there are many questions that remain to be answered. Many of the strongest GWAS signals were found in noncoding parts of the genome, either in introns or in regions between genes. Expression data and RNA research have shown that these regions are not "junk", as has been stated in the past, but have vital roles on the regulatory level.[61,62] Also of interest is the recent finding revealing that transcription factors bind within protein- coding regions (in addition to nearby noncoding regions) in a large number of human genes.[63] How these finding translate into the interpretation of the GWAS remains to be worked out.

We have proven in the IGAP consortium that we have not reached the limit of GWAS for Alzheimer research yet. Better reference sets for imputations and more subjects are still likely to yield more information. The IGAP dataset is a rich source for additional subgroup and interaction analyses. Within IGAP we have also explored the evidence for gene-interactions, starting with a genome wide analysis of the interaction of *APOE*ε4 with other variants. The *PTK2B* locus described in Chapter 2, close to but independent of *CLU*, was first identified in non-carriers of *APOE*ε4 allele (non-carriers; OR = 1.11, *p-value* = $5.77 \times 10^{-6}$, carriers; OR = 1.04, *p-value* = 0.18**).** Within IGAP, the *APOE*ε4 interaction analyses are still ongoing. In the discovery phase, for none of the variants the SNP* *APOE*ε4 interaction term reached genome wide significance. Of interest is the interaction between a cluster of SNPs near the *TMEM106B* gene and *APOE*ε4 (*p-value* = $1.5 \times 10^{-7}$). *TMEM106B* is a gene that has been implicated in frontotemporal lobe degeneration (FTLD).[64] The top SNP in this cluster is 95 kb away from the top hit in the FTLD GWAS, and 62 kb upstream from the start of the *TMEM106B* gene. The negative regression coefficient for the interaction term suggests that

the *TMEM106B* gene plays a role in AD in the absence of an *APOE*ε4 allele only. The inverse direction of effect between strata for the SNP with lowest p-value (non-carriers: OR = 1.09, *p-value* = $4.5*10^{-4}$; carriers: OR = 0,92, *p-value* = $4.2*10^{-3}$) could point towards a very interesting mechanism, but also raises suspicion of a false positive finding. We are currently trying to confirm this finding through replication analyses. Interaction studies for other candidate genes in the Epistasis project has yielded some interesting evidence for SNP*SNP interactions, where the combination of two variants, but not the individual variants, were a risk factor for AD: variants in Interleukin-10 (*IL10*) and aromatase (*CYP19A1*),[54] *GSTM3* and the *HHEX/IDE/KIF11* gene cluster,[54] *PPARA* and several interleukins (*IL1A, IL1B, IL10*),[65] and between insulin (*INS*) and the peroxisome proliferator-activated receptor alpha (*PPARA*) genes showed evidence of such epistasis.[65] However, these findings have not been replicated yet. Although they point towards the same pathways, these genes do not overlap with those identified in the IGAP GWAS.[53] The major problem in gene interaction research is statistical power: the sample sizes needed to find interactions are more than four times that of finding main effects. The current sample sizes hardly meet the requirements for finding main effects and therefore hamper the power of simple interactions (*APOE*ε4 versus all other SNPs) and genome-wide hypothesis-free screens for SNP*SNP interactions. Apart from an interaction effect, another approach is to investigate the additive effects of cumulative risk alleles as we used in our study of stroke in Chapter 4.2. The polygenic nature of AD may imply that effects of single genes are small and joint gene effects unified in a polygenic risk score is clinically and biologically relevant.[66-69] Environmental factors known to influence AD risk, such as smoking, moderate to severe head injury, the use of antihypertensives, non-steroidal anti-inflammatory drugs, statins or estrogens, and vitamin E intake can also be included in these risk analyses.[70-75] However, unless genes have substantial effect by themselves, they will individually not be important interactants of environmental factors. For AD, gene-environment effects have been described only for *APOE*ε4 and autosomal dominant variants in *APP* and *PSEN1*.[19,66,71,76,77] From a biological perspective one may argue that it is unlikely that a genetic variant will solely interact with an environmental factor – more likely the environment interacts with a pathway of genes. As the effect of risk scores are larger than that of single variants, this implies that research of gene-environment interaction may benefit from a genetic risk score approach. Risk scores may thus be the basis for successful studies of gene interactions. However, interactions with environmental factors may be specific to distinct pathways, e.g., smoking may interact with genes involved in a certain pathway (e.g., immunity) but not with those involved in others (e.g., ubiquitination). Thus, the challenge for the near future will be to take such heterogeneity into account. The distinct pathways of AD identified in the bioinformatics/statistical analyses in IGAP include: endocytosis (*p-value* = $2.4*10^{-12}$), cholesterol transport (*p-value* = $6.1*10^{-10}$), immunity (*p-value* = $1.7*10^{-10}$) and ubiquitination (*p-value* = $6 \times 10^{-4}$) (Escott-Price et al, submitted). One of the most important challenges will be to develop pathway-specific polygenic risk scores to be used for gene-environment interaction studies.

*From genes to translation*

A major challenge will be to validate and further unravel the pathways suggested by bioinformatics tools through molecular data. The progress in understanding the molecular etiology of AD has been limited by the paucity of AD tissue at early stages and costs of constructing humanized AD animal models representing its polygenic origin involving many genes with small effects. Reprogrammed induced pluripotent stem cells (iPSC) derived from somatic cells offer a new opportunity to overcome these problems. iPSC derived from participants of cohorts who are selected based on genetic risk profiles and disease status can be used to validate pathways and discover new ones.[78] Metabolomics, including profiling of lipids (lipidomics), is especially suited to find the molecular link between genes and complex diseases such as AD, which results from the interplay of many genetic and environmental factors.[79,80] The metabolome, i.e. all low-molecular weight molecules in a system, cell or tissue, reflects the interaction of the environment with the biological system involving the (epi)genome, transcriptome and proteome. The Genetic Epidemiology Unit of Erasmus MC, along with many collaborators from within and outside the Erasmus MC, is currently setting up these metabolomics experiments on iPSCs. Beyond the validation of the pathways and gene interactions, these studies may yield new pathways, which are targets for drug and biomarker discovery.

Another, yet the ultimate step will be to use the genomic findings in the development of medication. Although the risk genes now known for AD point to many different pathways (see also Chapter 2 of this thesis), and therapeutics have focused on many of these, the translation of genetic information into personalized medicine develops slowly and is restricted to attempts to stratify the analyses of experimental trials by *APOE*ε4 status.[81,82] However, also here we may argue that studies of interaction of genetic risk scores consisting of genes in a single pathway may be a more powerful and biologically relevant approach to develop more targeted and personalized treatments. For patients for example developing AD primarily due to genetic defects in the immune system, it may be worthwhile to try out preventive strategies that primarily target the immune system.[83] Again, such approaches would require research in which one pathway is dominant (e.g. only immune) or multiple pathways are compromised (e.g. immune, ubiquination, lipids, oxidative stress, amyloid metabolism…).

Ultimately, such risk profiling may be applied long before the clinical onset of AD, so that the disease mechanisms may be timely averted or delayed beyond the human lifetime, and AD will become a preventable and treatable disease.

*Conclusion*

In this thesis, I have aimed to add to the understanding of the complex genetics of AD. The highlights comprise the discovery of 11 new genes for AD in the IGAP GWAS (Chapter 2), the discovery of four rare variants in ERF (Chapter 3.1) and four common variants in the

CHARGE consortium (Chapters 3.2 and 3.3) for cognitive function as an endophenotype for AD, a link between plasma amyloid beta levels in hypertensive subjects from the ERF cohort and a mutation in the *PSEN2* gene known for familial AD, and the demonstration that genetic variants for the clinical risk factors for a complex vascular disease (stroke) can contribute to disease prediction – a finding most likely also true for AD. Yet, a large part of genetic factors in AD remains unraveled. Exciting new developments, which may yield many new interesting findings increasing our understanding of AD pathogenesis, include genome-wide and exome-wide sequencing analyses, further GWAS studies using improved worldwide reference sets as well as population-specific reference sets, gene-environment and gene-gene interactions, and metabolomics in pluripotent cell lines. In the longer term, the knowledge gained from genetics on disease mechanisms and pathways may become clinically applicable through targets for novel drug development and through improved risk profiling, allowing for targeted disease prevention strategies, helping in the fight against this devastating neurological disease on the individual and the population level.

**REFERENCES**


1. Rogaeva, E. et al. The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. *Nat Genet* **39**, 168-77 (2007).
2. Pottier, C. et al. High frequency of potentially pathogenic SORL1 mutations in autosomal dominant early-onset Alzheimer disease. *Mol Psychiatry* **17**, 875-9 (2012).
3. Liu, F. et al. A study of the SORL1 gene in Alzheimer's disease and cognitive function. *J Alzheimers Dis* **18**, 51-64 (2009).
4. Genomes Project, C. et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).
5. Butler, A.W. et al. Meta-analysis of linkage studies for Alzheimer's disease--a web resource. *Neurobiol Aging* **30**, 1037-47 (2009).
6. Singer, J.J., Falchi, M., Macgregor, A.J., Cherkas, L.F. & Spector, T.D. Genome-wide scan for prospective memory suggests linkage to chromosome 12q22. *Behav Genet* **36**, 18-28 (2006).
7. Lien, Y.-J. A genome-wide quantitative trait loci scan of neurocognitive performances in families with schizophrenia. *Genes Brain Behav* **9**, 695-702 (2010).
8. Luciano, M. et al. Genome-wide scan of IQ finds significant linkage to a quantitative trait locus on 2q. *Behav Genet* **36**, 45-55 (2006).
9. Luciano, M. et al. Whole genome association scan for genetic polymorphisms influencing information processing speed. *Biol Psychol* **86**, 193-202 (2011).
10. Kristiansen, M., Menghi, F., Hughes, R., Hubank, M. & Ham, J. Global analysis of gene expression in NGF-deprived sympathetic neurons identifies molecular pathways associated with cell death. *BMC Genomics* **12**, 551 (2011).
11. Zou, C.G. et al. The molecular mechanism of endoplasmic reticulum stress-induced apoptosis in PC-12 neuronal cells: the protective effect of insulin-like growth factor I. *Endocrinology* **150**, 277-85 (2009).
12. Zareen, N., Biswas, S.C. & Greene, L.A. A feed-forward loop involving Trib3, Akt and FoxO mediates death of NGF-deprived neurons. *Cell Death Differ* **20**, 1719-30 (2013).
13. Prudente, S. et al. The mammalian tribbles homolog TRIB3, glucose homeostasis, and cardiovascular diseases. *Endocr Rev* **33**, 526-46 (2012).
14. Groner, A.C. et al. KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. *PLoS Genet* **6**, e1000869 (2010).
15. Jakobsson, J. et al. KAP1-mediated epigenetic repression in the forebrain modulates behavioral vulnerability to stress. *Neuron* **60**, 818-31 (2008).
16. Rafn, B. et al. ErbB2-driven breast cancer cell invasion depends on a complex signaling network activating myeloid zinc finger-1-dependent cathepsin B expression. *Mol Cell* **45**, 764-76 (2012).
17. Hook, V. et al. Cysteine Cathepsins in the secretory vesicle produce active peptides: Cathepsin L generates peptide neurotransmitters and cathepsin B produces beta-amyloid of Alzheimer's disease. *Biochim Biophys Acta* **1824**, 89-104 (2012).

18. Porcellini, E. et al. Haplotype of single nucleotide polymorphisms in exon 6 of the MZF-1 gene and Alzheimer's disease. *J Alzheimers Dis* **34**, 439-47 (2013).
19. Huang, Z. et al. Dynamic expression pattern of Pde4d and its relationship with CpG methylation in the promoter during mouse embryo development. *Biochem Biophys Res Commun* (2013).
20. Verde, I. et al. Myomegalin is a novel protein of the golgi/centrosome that interacts with a cyclic nucleotide phosphodiesterase. *J Biol Chem* **276**, 11189-98 (2001).
21. Burgin, A.B. et al. Design of phosphodiesterase 4D (PDE4D) allosteric modulators for enhancing cognition with improved safety. *Nat Biotechnol* **28**, 63-70 (2010).
22. Aznar, S. & Knudsen, G.M. Depression and Alzheimer's disease: is stress the initiating factor in a common neuropathological cascade? *J Alzheimers Dis* **23**, 177-93 (2011).
23. Bond, J. & Woods, C.G. Cytoskeletal genes regulating brain size. *Curr Opin Cell Biol* **18**, 95-101 (2006).
24. Gretarsdottir, S. et al. The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. *Nat Genet* **35**, 131-8 (2003).
25. Adzhubei, I.A. et al. A method and server for predicting damaging missense mutations. *Nat Methods* **7**, 248-9 (2010).
26. Kostulas, K. et al. PDE4D and ALOX5AP genetic variants and risk for Ischemic Cerebrovascular Disease in Sweden. *J Neurol Sci* **263**, 113-7 (2007).
27. Bevan, S. et al. Variation in the PDE4D gene and ischemic stroke risk: a systematic review and meta-analysis on 5200 cases and 6600 controls. *Stroke* **39**, 1966-71 (2008).
28. Quarta, G. et al. Phosphodiesterase 4D and 5-lipoxygenase activating protein genes and risk of ischemic stroke in Sardinians. *Eur J Hum Genet* **17**, 1448-53 (2009).
29. Kuhlenbaumer, G. et al. Evaluation of single nucleotide polymorphisms in the phosphodiesterase 4D gene (PDE4D) and their association with ischaemic stroke in a large German cohort. *J Neurol Neurosurg Psychiatry* **77**, 521-4 (2006).
30. Lohmussaar, E. et al. ALOX5AP gene and the PDE4D gene in a central European population of stroke patients. *Stroke* **36**, 731-6 (2005).
31. van Rijn, M.J. et al. Familial aggregation, the PDE4D gene, and ischemic stroke in a genetically isolated population. *Neurology* **65**, 1203-9 (2005).
32. Consortium, G.T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-5 (2013).
33. Casey, J.P. et al. A novel approach of homozygous haplotype sharing identifies candidate genes in autism spectrum disorder. *Hum Genet* **131**, 565-79 (2012).
34. Service, S.K. et al. A genome-wide meta-analysis of association studies of Cloninger's Temperament Scales. *Transl Psychiatry* **2**, e116 (2012).
35. Tiwari, V., O'Donnell, C.D., Oh, M.J., Valyi-Nagy, T. & Shukla, D. A role for 3-O-sulfotransferase isoform-4 in assisting HSV-1 entry and spread. *Biochem Biophys Res Commun* **338**, 930-7 (2005).

36. Carter, C.J. Alzheimer's disease plaques and tangles: cemeteries of a pyrrhic victory of the immune defence network against herpes simplex infection at the expense of complement and inflammation-mediated neuronal destruction. *Neurochem Int* **58**, 301-20 (2011).
37. Freimer, N. & Sabatti, C. The human phenome project. *Nat Genet* **34**, 15-21 (2003).
38. Bettens, K., Sleegers, K. & Van Broeckhoven, C. Genetic insights in Alzheimer's disease. *Lancet Neurol* **12**, 92-104 (2013).
39. Cruts, M., Theuns, J. & Van Broeckhoven, C. Locus-specific mutation databases for neurodegenerative brain diseases. *Hum Mutat* **33**, 1340-4 (2012).
40. Cruchaga, C. et al. Rare variants in APP, PSEN1 and PSEN2 increase risk for AD in late-onset Alzheimer's disease families. *PLoS One* **7**, e31039 (2012).
41. Petit, J., Boisseau, P., Taine, L., Gauthier, B. & Arveiler, B. A YAC contig encompassing the 11q14.3 breakpoint of a translocation associated with schizophrenia, and including the tyrosinase gene. *Mamm Genome* **10**, 649-52 (1999).
42. Anderson, B.M. et al. Examination of association of genes in the serotonin system to autism. *Neurogenetics* **10**, 209-16 (2009).
43. Gurling, H.M. et al. Genomewide genetic linkage analysis confirms the presence of susceptibility loci for schizophrenia, on chromosomes 1q32.2, 5q33.2, and 8p21-22 and provides support for linkage to schizophrenia, on chromosomes 11q23.3-24 and 20q12.1-11.23. *Am J Hum Genet* **68**, 661-73 (2001).
44. Schol-Gelok, S. et al. A genome-wide screen for depression in two independent Dutch populations. *Biol Psychiatry* **68**, 187-96 (2010).
45. Vorstman, J.A. et al. Identification of novel autism candidate regions through analysis of reported cytogenetic abnormalities associated with autism. *Mol Psychiatry* **11**, 1, 18-28 (2006).
46. 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 (2007).
47. Verhaaren, B.F. et al. Alzheimer's disease genes and cognition in the nondemented general population. *Biol Psychiatry* **73**, 429-34 (2013).
48. Rietveld, C.A. et al. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **340**, 1467-71 (2013).
49. International Consortium for Blood Pressure Genome-Wide Association, S. et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103-9 (2011).
50. Grove, M.L. et al. Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. *PLoS One* **8**, e68095 (2013).
51. Jonsson, T. et al. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* **488**, 96-9 (2012).
52. Jonsson, T. et al. Variant of TREM2 associated with the risk of Alzheimer's disease. *N Engl J Med* **368**, 107-16 (2013).

53. Guerreiro, R. et al. TREM2 variants in Alzheimer's disease. *N Engl J Med* **368**, 117-27 (2013).
54. Bullock, J.M. et al. Discovery by the Epistasis Project of an epistatic interaction between the GSTM3 gene and the HHEX/IDE/KIF11 locus in the risk of Alzheimer's disease. *Neurobiol Aging* **34**, 1309 e1-7 (2013).
55. Ting, S.K. et al. Absence of A673T amyloid-beta precursor protein variant in Alzheimer's disease and other neurological diseases. *Neurobiol Aging* **34**, 2441 e7-8 (2013).
56. Cuyvers, E. et al. Investigating the role of rare heterozygous TREM2 variants in Alzheimer's disease and frontotemporal dementia. *Neurobiol Aging* **35**, 726 e11-9 (2014).
57. Pericak-Vance, M.A. et al. Linkage studies in familial Alzheimer disease: evidence for chromosome 19 linkage. *Am J Hum Genet* **48**, 1034-50 (1991).
58. Lee, J.H. et al. Analyses of the National Institute on Aging Late-Onset Alzheimer's Disease Family Study: implication of additional loci. *Arch Neurol* **65**, 1518-26 (2008).
59. Wu, M.C. et al. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* **89**, 82-93 (2011).
60. Boomsma, D.I. et al. The Genome of the Netherlands: design, and project goals. *Eur J Hum Genet* (2013).
61. Willingham, A.T. & Gingeras, T.R. TUF love for "junk" DNA. *Cell* **125**, 1215-20 (2006).
62. Bonasio, R., Tu, S. & Reinberg, D. Molecular signals of epigenetic states. *Science* **330**, 612-6 (2010).
63. Weatheritt, R.J. & Babu, M.M. Evolution. The hidden codes that shape protein evolution. *Science* **342**, 1325-6 (2013).
64. Van Deerlin, V.M. et al. Common variants at 7p21 are associated with frontotemporal lobar degeneration with TDP-43 inclusions. *Nat Genet* **42**, 234-9 (2010).
65. Heun, R. et al. Interactions between PPAR-alpha and inflammation-related cytokine genes on the development of Alzheimer's disease, observed by the Epistasis Project. *Int J Mol Epidemiol Genet* **3**, 39-47 (2012).
66. Lambert, J.C. et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* **45**, 1452-8 (2013).
67. Manolio, T.A. et al. Finding the missing heritability of complex diseases. *Nature* **461**, 747-53 (2009).
68. Maher, B. Personal genomes: The case of the missing heritability. *Nature* **456**, 18-21 (2008).
69. International Schizophrenia, C. et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748-52 (2009).
70. Plassman, B.L. et al. Documented head injury in early adulthood and risk of Alzheimer's disease and other dementias. *Neurology* **55**, 1158-66 (2000).

71. Reitz, C., den Heijer, T., van Duijn, C., Hofman, A. & Breteler, M.M. Relation between smoking and risk of dementia and Alzheimer disease: the Rotterdam Study. *Neurology* **69**, 998-1005 (2007).
72. Haag, M.D., Hofman, A., Koudstaal, P.J., Breteler, M.M. & Stricker, B.H. Duration of antihypertensive drug use and risk of dementia: A prospective cohort study. *Neurology* **72**, 1727-34 (2009).
73. Devore, E.E. et al. Dietary antioxidants and long-term risk of dementia. *Arch Neurol* **67**, 819-25 (2010).
74. Slooter, A.J. et al. Estrogen use and early onset Alzheimer's disease: a population-based study. *J Neurol Neurosurg Psychiatry* **67**, 779-81 (1999).
75. Andersen, K. et al. Do nonsteroidal anti-inflammatory drugs decrease the risk for Alzheimer's disease? The Rotterdam Study. *Neurology* **45**, 1441-5 (1995).
76. Vermeiren, A.P. et al. The association between APOE epsilon4 and Alzheimer-type dementia among memory clinic patients is confined to those with a higher education. The DESCRIPA Study. *J Alzheimers Dis* **35**, 241-6 (2013).
77. Luck, T. et al. Apolipoprotein E epsilon 4 genotype and a physically active lifestyle in late life: analysis of gene-environment interaction for the risk of dementia and Alzheimer's disease dementia. *Psychol Med*, 1-11 (2013).
78. Israel, M.A. et al. Probing sporadic and familial Alzheimer's disease using induced pluripotent stem cells. *Nature* **482**, 216-20 (2012).
79. Wood, P.L. Lipidomics of Alzheimer's disease: current status. *Alzheimers Res Ther* **4**, 5 (2012).
80. Mayr, M. Metabolomics: ready for the prime time? *Circ Cardiovasc Genet* **1**, 58-65 (2008).
81. Liu, C.C., Kanekiyo, T., Xu, H. & Bu, G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nat Rev Neurol* **9**, 106-18 (2013).
82. Huang, Y. & Mucke, L. Alzheimer mechanisms and therapeutic strategies. *Cell* **148**, 1204-22 (2012).
83. Vasto, S. et al. Alzheimer's disease and genetics of inflammation: a pharmacogenomic vision. *Pharmacogenomics* **8**, 1735-45 (2007).

# Chapter 6: Summary/Samenvatting

### *6.1. Summary*

Dementia is a chronic, progressive syndrome of cognitive deterioration, more than can be expected based on normal ageing. Alzheimer's disease is the most common form of dementia of older age, and it affects one in three women and one in six men, and therefore has a great impact on healthcare systems. The main symptoms comprise dysfunction in short term memory, language and loss of visuospatial ability. There is an important association with accumulation of amyloid β and Tau proteins in the brain, and a central role of these accumulations in the pathogenesis of the disease, among others through inflammatory processes and apoptosis, has been assumed. Loss of function and loss of tissue are found in the hippocampus and temporoparietal regions. The heritability of Alzheimer's disease is estimated to be around 60-74%, but the genetic variations driving this heritability are still largely unknown. Rare mutations are known in a few genes (*APP, PSEN1, PSEN2*) which cause a familial form of Alzheimer's disease, but only variants in the *APOE* gene have in themselves, as far as we know, a large influence on the incidence of AD in the general population.

In this thesis we have tried to find new genetic factors which contribute to the susceptibility for Alzheimer's disease in the general population. As Alzheimer's disease occurs only late in life, and there is not always agreement between the cognitive functioning and measurements of biochemical and pathological processes considered to be the cause of Alzheimer's disease, we also used endophenotypes. These quantitative endophenotypes can already be measured in early adulthood to middle age, in all participants of a cohort study. All projects described in this thesis are initially based on hypothesis free, genome wide analyses: genome wide association studies (GWAS) and linkage analyses. Through follow-up analyses in independent datasets, targeted genotyping of specific variants or of the coding variants in a given region (exome array, exome sequencing) we further analysed the initially identified regions. Moreover, we developed a method to predict disease, in this case stroke, through genetic variants associated with clinical risk factors for that disease.

In Chapter 2 we present the results of a large GWAS for Alzheimer's disease. This project was conducted by the International Genomics of Alzheimer's Project (IGAP) consortium. In this project, four consortia conducting research into the genetics of Alzheimer's disease, both in case-control and in cohort studies based in the aged general population, have combined their results in meta-analysis. In the first phase of this study, we used the data of 17.008 patients with Alzheimer's disease and 37,154 controls. There were over 7 million single nucleotide polymorphisms (SNPs) which could be analysed in all cohorts. In this analysis, 11,632 SNPs outside the *APOE* locus emerged, which were possibly associated with Alzheimer's disease. These were genotyped and further investigated in an additional 8,572 patients and 11,312 controls. This yielded a total of 19 regions with genome wide significant results (p-value $< 5x10^{-8}$), 11 of which were not identified in GWAS before. The pathways these variants may play a role in, include amyloid precursor protein (APP) and Tau processing (*SORL1, CASS4, FERMT2*), the immune response and inflammation (*HLA-*

*DRB5/DRB1*, *INPP5D*, *MEF2C*), cell migration (*PTK2B*), lipid transport (*SORL1*), hippocampal synaptic function (*MEF2C, PTK2B*), and cytoskeletal function (*CELF1, NME8, CASS4*). The large number of new loci yielded by this combined analysis suggests that larger numbers, both in number of subjects and number of SNPs, help in finding new associations.

In Chapter 3, we turned towards genetic variants that are associated to normal variation in cognitive functioning in non-demented subjects. Loss of cognitive function is the main hallmark of all dementias, and a decline in cognitive functioning is a predictor for developing dementia. In Chapter 3.1 we used a genome-wide linkage approach to search for genetic regions that are associated to bad performance on neuropsychological tests in the Erasmus Rucphen Family (ERF) study. In this cohort, in which all participants are part of a large and complex family, 2881 participants have performed a series of cognitive tests. We found six regions associated with a performance in the lowest 10% (corrected for age, sex and education) on different tests. (1p13.1, 12q24.33, 19q13.43, 20p13, 21q22.13 and 21q22.3). The region on chromosome 1 was known as a linkage region for Alzheimer's disease, the region on chromosome 12 was identified earlier in association with memory and schizophrenia, and 21q22 has earlier been linked to verbal IQ. To further explore these regions, we subsequently sought for variants that may explain this linkage. For this, we performed a quantitative analysis on the cognitive test results and exome array and exome sequencing data of 2430 to 2487 participants (numbers varying per cognitive test). As in this analysis all coding variants are used, in contrast to GWAS, a directly causal relationship with the SNPs identified can be suspected. The four SNPs emerging significantly within the linkage regions all have a predicted deleterious effect on the protein they code for, and are all involved in the development and/or degeneration of the nervous system. The first association is between a combined score for execttive functioning and rs41281840 (p-value = 0.0056), located in the *TRIB3* gene involved in neuronal apoptosis. *TRIB3* was earlier identified in relation to cognitive functioning. On chromosome 19 we found two associations. The first is between the score for word recognition and rs13866122 (p-value – 3.7*10-5), located in the *TRIM28* gene involved in the regulation of gene expression and extensively expressed in, among others, the hippocampus. The second association is between the Dutch Adult Reading Test, a measure for general intelligence, and rs1506030080 in the *MZF1* gene (p-value = 1.3*10-3). This gene is also involved in DNA transcription, and influences among others the expression of Cathepsin B, which is involved in amyloid synthesis. Although these two variants lie closely together, they are not in linkage disequilibrium, and are associated to different outcomes, and therefore seem to be independent of each other. Lastly there is the variant associated with executive functioning (rs76199660, p-value – 3.1*10-4) in the *PDE4DIP* gene. This gene influences the expression of the *PDE4D* gene, which is involved in the development of the nervous system, but also seems to play a role in memory at older age. The related gene in drosophila harbours mutations that are developmentally lethal in homozygosity, and also in humans there are no recorded homozygotes for the variant we identified.

Chapter 3.2 and 3.3 describe the search for more common genetic variants associated with respectively executive function and processing speed (Chapter 3.2) and short term memory (Chapter 3.3). These two studies were performed within the Cohorts for Heart and Ageing Research in Genomic Epidemiology (CHARGE) consortium. Each of the many participating cohorts have performed GWAS for the available phenotypes, and the results were combined in meta-analyses. For processing speed we identified a genome-wide significant association: in a total of 34370 subjects, rs17518584 showed significant with letter-digit or digit-symbol substitution tasks. (p-value 3.19*10-8). This variant lies within the *CADM2* gene and influences its expression in various areas of the brain, including the cingulum and the frontal cortex. The protein *CADM2* codes for, is involved in neurotransmission through the gamma aminobutyric acid (GABA) and glutamate systems, and possibly in autism and personality. In Chapter 3.3, the GWAS for memory, the main finding is that several SNPs in and around the *APOE* gene are involved in memory in nondemented subjects (rs442-638, p-value – $5.57*10^{-10}$). Two other associations were found for rs11074779 near *HS3ST4, p-value* = $3.11x10^{-8}$, and rs6813517 near *SPOCK3*, *p-value* = $2.58x10^{-8}$. Both genes are strongly expressed in the hippocampus, the main brain structure involved in short tem memory and the formation of new memories, which is also most affected in Alzheimer's disease.

In Chapter 4, we searched for overlapping genetic risk factors between cerebrovascular pathology and dementia. In epidemiological studies, these two are clearly related, but the genes and processes responsible for this have not yet been found. In Chapter 4.1, we used a subgroup of the ERF study, including 125 participants of middle age with hypertension, to search for genetic regions which show linkage to amyloid β concentrations in the blood. Amyloid depositions in the brain form are one of the hallmarks of Alzheimer's disease, and amyloid β concentrations in cerebrospinal fluid and plasma are associated to future occurrence of Alzheimer's disease. Two regions show suggestive linkage with isoform Aβ40: their peaks are at 1q41 (LOD = 2.07) and 11q14.3 (LOD = 2.97). In the selected ERF participants and in a subgroup of 320 participants of the Rotterdam study, in the same age range and with hypertension but not not related to each other, we sought for SNPs associated to Aβ40. In the region on chromosome 1, several SNPs in and around the *PSEN2* gene are associated to these concentrations (*p-value* = $2.58*10^{-4}$ for rs6703170). *PSEN2* is one of the genes that harbour mutations for autosomal dominant forms of Alzheimer's disease. In the region on chromosome 11 we identified no significant SNPs, but the region has earlier been associated to among others schizophrenia and autism.

In Chapter 4.2, we experimented with a risk score for stroke, based upon the risk SNPs for the clinical risk factors stroke, such as obesity, hypertension, diabetes, cardiac arrhythmias and prior cardiovascular damage. We have collected over 300 such SNPs, and for each risk factor we created a risk score. We have studied the predictive value of these risk scores for future stroke in four large cohort studies within CHARGE. The main outcome measure was the added predictive value of these scores compared to a risk model based upon clinical risk

factors. The combination of genetic risk scores for the different clinical risk factors can indeed slightly improve the prediction of a future first stroke. ($\Delta$ AUC =0.021, *p-value* = $3.7*10^{-7}$). Although these results are not yet such that they have an added predictive value in clinic, they do offer perspective: with the further discovery of SNPs associated with various risk factors for brain infarctions, it may be possible to develop a clinically useful screening instrument. It needs to be further investigated whether a similar approach also works for the prediction of dementia.

In this thesis I describe how we have sought for risk genes for Alzheimer's disease through different projects. The highlights include the finding of 11 new risk loci for Alzheimer's disease (Chapter 2), four rare (Chapter 3.1) and four common loci (Chapter 3.2 and 3.3) associated to cognitive functioning in the nondemented population, mutations in the *PSEN2* gene which are associated to plasma amyloid β concentrations (Chapter 4.1) and the proof that genetic risk factors for clinical risk factors can contribute to the prediction of complex disease (Chapter 4.2). Although each project has led to interesting and plausible results, their contribution to the heritability of dementia is most likely still very small. Large developments are currently going on in the genetic empidemiology, from which large breakthroughs can be expected in the coming years, such as the combined analyses of exome chip data of dementias and endophenotypes within CHARGE and IGAP, and the collaboration between CHARGE and the AD Genetics Consortium in the Alzheimer's Disease Sequencing project, in which large scale sequencing is performed of familial and sporadic cases of Alzheimer's disease as well as control subjects. Also the development of population-specific reference genomes such as the Genomes of the Netherlands may lead to new developments, as do new worldwide GWAS based upon universal reference genomes. The further increase in size of the datasets leads to more statistical power for many gene-gene and gene-environment interactions. For breakthroughs in the translation of genetic epidemiological research results into clinic, there may be a role for stem cell research and metabolomics. These new developments may lead to the discovery of new drug targets, and to personal risk profiles indicating the most relevant disease pathways in an individual patient, and therefore allow for targeted prevention and treatment. Eventually, genetic research might thus help in the fight against the devastating effects of dementia.

## *6.2. Samenvatting*

Dementie is een chronisch, progressief syndroom van cognitieve achteruitgang, meer dan op basis van normale veroudering verwacht mag worden. De ziekte van Alzheimer is de meest voorkomende vorm van dementie op latere leeftijd, en treft tot 1 op 3 vrouwen en 1 op 6 mannen, en heeft daarmee een grote invloed op de gezondheidszorg. De belangrijkste symptomen bestaan uit stoornissen in het korte termijn geheugen, taalproblemen en verlies van visuospatiële functies. Er is een belangrijke associatie met de ophoping van amyloid β en tau eiwit in de hersenen, en een centrale rol van deze ophopingen in het ontwikkelen van de ziekte, door onder meer ontstekingsprocessen en apoptose, wordt verondersteld. Verminderde functie en weefselverlies wordt gevonden in de hippocampus en temporopariëtale regio's. De erfelijkheid van de ziekte van Alzheimer wordt geschat op 60-74%, maar de genetische variaties die deze erfelijkheid bepalen zijn nog grotendeels onbekend. Er zijn zeldzame mutaties in enkele genen (*APP, PSEN1, PSEN2*) bekend die een familiaire vorm van de ziekte van Alzheimer veroorzaken, maar alleen variaties in het *APOE* gen hebben voorzover bekend in zichzelf een grote invloed op het voorkomen van de ziekte in de algemene bevolking.

In dit proefschrift hebben wij geprobeerd nieuwe genetische factoren te vinden die bijdragen aan de vatbaarheid voor de ziekte van Alzheimer in de algemene bevolking. Omdat de ziekte van Alzheimer pas laat in het leven optreedt, en er niet altijd overeenstemming is tussen het cognitief functioneren en de metingen betreft biochemische en pathologische processen die als oorzaak voor de ziekte van Alzheimer worden gezien, hebben wij ook gebruik gemaakt van endofenotypes. Deze kwantitatieve endofenotypes kunnen al op jongvolwassen of middelbare leeftijd gemeten worden in alle deelnemers in een cohortstudie. Alle projecten beschreven in dit proefschrift zijn in beginsel gebaseerd op hypothesevrije, genoomwijde analyses: genoomwijde associatiestudies (GWAS) en linkage analyses. Met follow-up analyses in onafhankelijke datasets, gerichte genotyperingen van specifieke variaties of van coderende variaties in een bepaalde regio (exome array, exome sequencing) hebben wij de initieel gevonden regio's verder geanalyseerd. Bovendien hebben wij een methode onderzocht om ziekte, in dit geval een herseninfarct, te voorspellen door middel van genvariaties geassocieerd met klinische risicofactoren voor die ziekte.

In hoofdstuk 2 presenteren wij de resultaten van een grote GWAS voor de ziekte van Alzheimer. Dit project is uitgevoerd door het International Genomics of Alzheimer's Project (IGAP). In dit project hebben vier consortia die onderzoek doen naar de genetica van Alzheimer, in zowel klinische case-control series als in cohortstudies gebaseerd op de algemene oudere populatie, hun resultaten gecombineerd in een meta-analyse. In de eerste fase van deze studie gebruikten wij de gegevens van 17,008 patiënten met de ziekte van Alzheimer en 37,154 controles. Er waren ruim 7 miljoen single nucleotide polymorphisms (SNPs) die in alle cohorten geanalyseerd konden worden. Uit deze analyse kwamen 11,632 SNPs buiten het APOE locus naar voren die mogelijk geassocieerd waren met de ziekte van

Alzheimer, en gegenotypeerd en nader onderzocht werden in nog eens 8,572 patienten en 11,312 controles. Dit leverde in totaal 19 regio's met genoomwijd significante resultaten op (*p-waarde* < $5x10^{-8}$), waarvan 11 regio's nog niet eerder in een GWAS gevonden waren. De processen waar deze variaties mogelijk een rol in spelen, betreffen de verwerking van het amyloid precursor eiwit (APP) en Tau, (*SORL1*, *CASS4*, *FERMT2*), immuun- en ontstekingsreacties,(*HLA-DRB5/DRB1*, *INPP5D*, *MEF2C*), celmigratie (*PTK2B*), lipidentransport (*SORL1*), synapsfunctie in de hippocampus (*MEF2C, PTK2B*), en functioneren van het cytoskelet (*CELF1, NME8, CASS4*). Het grote aantal nieuwe loci dat deze gecombineerde analyse opleverde, suggereert dat schaalvergroting, zowel wat betreft aantallen patienten en controles als aantallen onderzochte SNPs, helpt om nieuwe associaties te vinden.

In hoofdstuk 3 richtten wij ons op genvariaties die geassocieerd zijn met de normale variatie in cognitief functioneren in niet-demente proefpersonen. Verlies van cognitief functioneren is het belangrijkste kenmerk van alle dementieen, en een verminderd cognitief functioneren is een voorspeller voor het ontwikkelen van dementie. In hoofdstuk 3.1 gebruikten wij een genoomwijde linkage benadering om te zoeken naar genetische gebieden die geassocieerd zijn met slechte prestaties op neuropsychologische tests in de Erasmus Rucphen Familie (ERF) studie. In dit cohort, waarvan de deelnemers allen deel uitmaken van een grote en complexe familie, hebben 2881 deelnemers een serie cognitieve testen gemaakt. Wij vonden zes regio's, die geassocieerd waren met een prestatie in de laagste 10% (gecorrigeerd voor geslacht, leeftijd en opleiding) op verschillende testen (1p13.1, 12q24.33, 19q13.43, 20p13, 21q22.13 en 21q22.3). De regio op chromosoom 1 was bekend als linkage-regio voor de ziekte van Alzheimer, de regio op chromosoom 12 is eerder gevonden in associatie met geheugen en schizofrenie, en 21q22 is eerder gelinkt aan verbaal IQ. Om deze regio's verder te onderzoeken, hebben wij vervolgens gezocht naar de variaties die deze linkage mogelijk verklaren. Hiervoor hebben wij een kwantitatieve analyse gedaan op de cognitieve testresultaten en exoom array en exoom sequencing data van 2430 tot 2487 (variatie per test) deelnemers. Omdat in deze analyse alle coderende variaties worden gebruikt, in tegenstelling tot bij GWAS, kan een direct causale relatie met de gevonden SNPs worden vermoed. De vier SNPs die significant naar voren kwamen binnen de linkage regio's, hebben alle een verondersteld zeer schadelijk effect op het eiwit waarvoor ze coderen, en zijn alle betrokken bij ontwikkeling en/of degeneratie van het zenuwstelsel. De eerste associatie is tussen een gecombineerde score voor executief functioneren en rs41281850 (p-waarde = 0.0056), gelegen in het *TRIB3* gen dat is betrokken bij neuronale apoptose. *TRIB3* was eerder gevonden in relatie met cognitief functioneren. Op chromosoom 19 vonden wij twee associaties. De eerste is tussen de score voor woordherkenning en rs13866122 (*p-waarde* = $3.7*10^{-5}$), gelegen in het *TRIM28* gen betrokken bij de regulatie van genexpressie en uitgebreid tot uiting komend in onder meer de hippocampus. De tweede associatie is tussen de Dutch Adult Reading Test, een maat voor algemene intelligentie, en rs150630080 in het *MZF1* gen *p-waarde* = $1.3*10^{-3}$). Ook dit gen is betrokken bij DNA-

transcriptie, en beinvloedt onder meer de expressie van cathepsin B, dat is betrokken bij de synthese van amyloid. Hoewel deze twee variaties dicht bij elkaar liggen, zijn ze niet in "linkage disequilibrium" en zijn ze geassocieerd met verschillende uitkomsten, en lijken daarmee onafhankelijk van elkaar. Tenslotte is er de variant geassocieerd met executief functioneren (rs76199660, *p-waarde* = $3.1*10^{-4}$) in het *PDE4DIP* gen. Dit gen beinvloedt de expressie van het *PDE4D* gen, dat is betrokken bij de aanleg van het zenuwstelsel, maar ook op latere leeftijd een rol lijkt te spelen in het geheugen. Het verwante gen in fruitvliegjes kent mutaties die in homozygotie dodelijk zijn in een vroege fase, en ook in de mens zijn er geen homozygoten bekend van de door ons gevonden variant.

Hoofdstuk 3.2 en 3.3 beschrijven de zoektocht naar meer voorkomende genvariaties geassocieerd met respectievelijk executief functioneren en snelheid van informatieverwerking (hoofdstuk 3.2) en met korte termijn geheugen (hoofdstuk 3.3). Deze beide studies zijn uitgevoerd binnen het Cohorts for Heart and Ageing Research in Genomic Epidemiology (CHARGE) consortium. Elk van de vele deelnemende cohorten heeft GWAS uitgevoerd voor de beschikbare fenotypen, en de resultaten zijn gecombineerd in meta-analyses. Voor snelheid van informatieverwerking hebben we een genoomwijd significant resultaat gevonden: in een totaal van 34730 proefpersonen bleek rs17518584 geassocieerd met de letter-cijfer of cijfer-symbool vervangingstaak (*p-waarde* = $3.91*10^{-9}$). Deze variant ligt in het *CADM2* gen en beinvloedt de expressie in diverse hersengebieden, onder meer het cingulum en de frontale cortex. Het eiwit waarvoor *CADM2* codeert, is betrokken in neurotransmissie via het gamma-aminoboterzuur (GABA) en glutamaat systeem, en mogelijk in autisme en persoonlijkheid. In hoofdstuk 3.3, de GWAS voor geheugen, is de belangrijkste bevinding dat diverse SNPs in en rondom het *APOE* gen betrokken zijn bij het geheugen in niet-demente proefpersonen (rs4420638, *p-waarde* = $5.57x10^{-10}$). Twee andere associaties worden gevonden voor rs11074779 nabij *HS3ST4, p-waarde* = $3.11x10^{-8}$, en rs6813517 nabij *SPOCK3*, *p-waarde* = $2.58x10^{-8}$. Beide genen komen sterk tot uiting in de hippocampus, de belangrijkste hersenstructuur in het korte termijn geheugen en het opslaan van nieuwe herinneringen, die ook het meest is aangedaan in de ziekte van Alzheimer.

In hoofdstuk 4 zochten wij naar overlappende genetische risicofactoren tussen cerebrovasculair lijden en dementie. In epidemiologische studies zijn deze twee duidelijk gerelateerd, maar de hiervoor verantwoordelijke genen en processen zijn nog niet gevonden. In hoofdstuk 4.1 gebruikten we een subgroep van de ERF studie, van 125 deelnemers van middelbare leeftijd met een hoge bloeddruk, om te zoeken naar genetische regio's die linkage laten zien met amyloid β concentraties in het bloed. Amyloid deposities in de hersenen vormen een van de kenmerken van de ziekte van Alzheimer, en concentraties van amyloid β in het hersenvocht en plasma zijn geassocieerd met het later optreden van de ziekte van Alzheimer. Twee regio's tonen suggestieve linkage met isoform Aβ40: de pieken hiervan liggen bij 1q41 (LOD = 2.07) en 11q14.3 (LOD = 2.97). In de

geselecteerde ERF deelnemers en een subgroep van 320 deelnemers van de Rotterdam studie, in dezelfde leeftijdsgroep en met hoge bloeddruk, maar niet aan elkaar verwant, hebben wij gezocht naar SNPs geassocieerd met Aβ40. In de region op chromosoom 1 zijn diverse SNPs in en rondom het *PSEN2* gen geassocieerd met deze concentraties (*p-waarde* = $2.58*10^{-4}$ voor rs6703170). *PSEN2* is een van de genen waarin mutaties voor autosomal dominante vormen van de ziekte van Alzheimer liggen. In de regio op chromosoom 11 vonden we geen significante SNPs, maar de regio is eerder geassocieerd met onder meer schizofrenie en autisme.

In hoofdstuk 4.2 hebben wij geëxperimenteerd met een risicoscore voor herseninfarcten, gebaseerd op de risicoSNPs voor de klinische risicofactoren voor een herseninfarct, zoals overgewicht, hoge bloeddruk, suikerziekte, hartritmestoornissen en eerdere cardiovasculaire schade. Wij hebben ruim 300 van dergelijke SNPs verzameld, en voor elke risicofactor een risicoscore gemaakt. Wij hebben de voorspellende waarde van deze risicoscores voor het later optreden van een beroerte bestudeerd in vier grote cohortstudies binnen CHARGE. De belangrijkste uitkomstmaat was de extra voorspellende waarde van deze scores in vergelijking met een risicomodel gebaseerd op klinische factoren. De combinatie van genetische risicoscores voor de verschillende klinische risicofactoren is inderdaad enigszins in staat om de voorspelling van het optreden van een eerste beroerte te verbeteren. ($\Delta$ AUC =0.021, *p-waarde* = $3.7*10^{-7}$). Hoewel deze resultaten nog niet zodanig zijn dat ze in de kliniek een toegevoegde waarde hebben, bieden ze wel perspectief: met de verdere ontdekking van SNPs geassocieerd met diverse risicofactoren voor herseninfarct kan mogelijk een klinisch bruikbaar screeningsinstrument ontwikkeld worden. Het zal moeten blijken of een dergelijke aanpak ook werkt voor dementie.

Ik beschrijf in dit proefschrift hoe wij door middel van verschillende projecten hebben gezocht naar risicogenen voor de ziekte van Alzheimer. Hoogtepunten betreffen onder meer het vinden van 11 nieuwe risico loci voor de ziekte van Alzheimer (hoofdstuk 2), vier zeldzame (hoofdstuk 3.1) en vier veel voorkomende loci (hoofdstuk 3.2 en 3.3) geassocieerd met cognitief functioneren in de niet-demente bevolking, mutaties in het *PSEN2* gen die zijn geassocieerd met plasma amyloid β concentraties (hoofdstuk 4.1), en het bewijs dat genetische risicofactoren voor de klinische risicofactoren van een complexe ziekte kunnen bijdragen aan het voorspellen van ziekte (hoofdstuk 4.2). Hoewel elk project heeft geleid tot interessante en plausibele resultaten, is hun bijdrage aan de erfelijkheid van dementie vermoedelijk nog altijd heel klein. Er zijn momenteel grote ontwikkelingen gaande binnen de genetische epidemiologie, waarvan in de komende jaren mogelijk grote doorbraken te verwachten zijn, zoals de gecombineerde analyses van exome chip data van dementieen en endofenotypes binnen CHARGE en IGAP, en de samenwerking tussen CHARGE en het AD Genetics Consortium in het Alzheimer's disease Sequencing Project waarin op grote schaal sequencing wordt verricht van familiaire en sporadische gevallen van de ziekte van Alzheimer en controles. Ook de ontwikkeling van populatie-specifieke referentie-genomen

zoals het Genoom van Nederland kan leiden tot nieuwe ontdekkingen, net als nieuwe wereldwijde GWAS gebaseerd op universele referentie-genomen. Het steeds verder vergroten van de datasets leidt tot meer statistische kracht voor diverse gen-gen en gen-omgeving interactie analyses. Voor doorbraken in de translatie van genetisch epidemiologisch onderzoek naar de kliniek is mogelijk een rol weggelegd voor stamcelonderzoek en metabolomics. Deze nieuwe ontwikkelingen leiden mogelijk tot de ontdekking van nieuwe aanknopingspunten voor medicatie, en tot persoonlijke risicoprofielen die de voor de individuele patient meest relevante pathways aanwijzen, en zo gerichte preventie en behandeling mogelijk maken. Uiteindelijk kan het genetisch onderzoek mogelijk op deze manier helpen in de strijd tegen de verwoestende effecten van dementie.

# Chapter 7: Appendices (Supplemental Material)

## *7.1. Supplemental material to Chapter 3.1*

*Genome-wide linkage and fine-mapping with exonic variants identify rare deleterious mutations affecting cognitive functioning*

**Table S1: Characteristics of the study population: linkage analysis (cases)**

| | N (cases) | Females (%) | Mean age (SD) | Only primary education (%) |
|---|---|---|---|---|
| Immediate recall | 257 | 146 (56.8) | 45.2 (16.8) | 56 (21.8) |
| Learning | 260 | 148 (56.9) | 47.4 (16.8) | 84 (32.3) |
| Delayed recall | 247 | 145 (58.7) | 47.3 (16.4) | 73 (29.6) |
| Recognition | 250 | 133 (53.2) | 52.8 (17.5) | 113 (45.2) |
| Stroop | 245 | 132 (53.9) | 55.2 (16.4) | 108 (44.1) |
| Trails | 243 | 138 (56.8) | 51.4 (18.9) | 116 (48.1) |
| Fluency | 243 | 141 (58.0) | 50.2 (17.5) | 71 (29.2) |
| Visuospatial Ability | 234 | 97 (41.5) | 49.3 (15.1) | 37 (15.8) |
| DART | 250 | 134 (53.6) | 53.2 (14.3) | 111 (44.4) |
| Zmem | 248 | 142 (57.3) | 48.9 (19.1) | 90 (36.3) |
| Zexec | 233 | 136 (58.4) | 50.9 (16.6) | 83 (35.6) |
| Zglob | 251 | 149 (59.4) | 49.2 (17.6) | 86 (34.3) |

*Cases are defined as the lowest 10 % of the residuals of the trait regressed on sex, age and education.. zmem: composite score for memory; Stroop: stroop-ratio; TMT: trail making test ratio; WF: verbal fluency; DART: Dutch Adult Reading Test; zexec: composite score for executive function;; zglob: composite score for global cognition.*

**Table S2. Characteristics of the study population, pooled exome chip and exome sequence data**

| | N | Females (%) | Mean age (SD) | Only primary education (%) |
|---|---|---|---|---|
| Immediate recall | 2485 | 1407 (56.6) | 48.7 (14.5) | 745 (30.0) |
| Learning | 2485 | 1407 (56.6) | 48.7 (14.5) | 745 (30.0) |
| Delayed recall | 2486 | 1407 (56.6) | 48.7 (14.5) | 747 (30.0) |
| Recognition | 2485 | 1406 (56.6) | 48.7 (14.5) | 746 (30.0) |
| Fluency | 2487 | 1406 (56.5) | 48.7 (14.5) | 745 (30.0) |
| Stroop | 2452 | 1403 (57.2) | 48.6 (14.5) | 723 (29.5) |
| Trails | 2475 | 1403 (56.7) | 48.7 (14.5) | 737 (29.8) |
| Visuospatial Ability | 2430 | 1376 (56.6) | 48.4 (14.3) | 724 (29.8) |
| DART | 2466 | 1399 (56.7) | 48.7 (14.5) | 732 (29.7) |
| zmem | 2483 | 1405 (56.6) | 48.7 (14.5) | 744 (30.0) |
| zexec | 2431 | 1392 (57.3) | 48.5 (14.4) | 711 (29.2) |
| zglob | 2366 | 1355 (57.3) | 48.1 (14.2) | 685 (29.0) |

*zmem: composite score for memory; Stroop: stroop-ratio; TMT: trail making test ratio; WF: verbal fluency; zexec: composite score for executive function; zglob: composite score for global cognition. DART: Dutch Adult Reading Test*

**Table S3. SNPs and genes available under the linkage peaks for each analysis**

| | ExomeSequence | | | ExomeChip | | | Pooled |
|---|---|---|---|---|---|---|---|
| **Region** | **All variants** | **Selected variants** | **Genes** | **All variants** | **Selected variants** | **Genes** | **Selected variants** |
| 1p13.1 | 3111 | 88 | 102 | 328 | 76 | 102 | 73 |
| 12q24.33 | 1229 | 21 | 22 | 145 | 28 | 22 | 27 |
| 19q13.43 | 1202 | 34 | 55 | 202 | 27 | 55 | 26 |
| 20p13 | 506 | 11 | 21 | 71 | 11 | 21 | 10 |
| 21q22.13 | 974 | 21 | 30 | 96 | 15 | 30 | 15 |
| 21q22.3 | 1600 | 24 | 19 | 190 | 22 | 19 | 22 |
| **Total** | **8622** | **199** | **249** | **1032** | **179** | **249** | **171** |

**Table S4. Association results (p-value <0.05) for pooled analysis, suggestively linked traits for each region and composite scores**

| trait | Marker | Allele | Effect | LOD | *P-value* | chromosome |
|---|---|---|---|---|---|---|
| recognition | rs13866122 | C | 1.658 | 3.7 | 0.000037 | 19q13.43 |
| zexec | rs76199660 | G | 0.134 | 2.83 | 0.00031 | 1p13.1 |
| DART | rs15063008 | G | 0.282 | 2.26 | 0.00127 | 19q13.43 |
| zmem | rs13866122 | C | 0.445 | 2.01 | 0.0024 | 19q13.43 |
| zmem | rs13888473 | T | -0.758 | 1.84 | 0.0036 | 21q22.3 |
| zglob | rs13888473 | T | -0.568 | 1.83 | 0.0037 | 21q22.3 |
| zmem | rs14729425 | G | -0.662 | 1.74 | 0.0047 | 1p13.1 |
| zexec | rs41281850 | G | 0.436 | 1.66 | 0.0056 | 20p13 |
| fluency | rs15009807 | C | -22.787 | 1.64 | 0.006 | 21q22.3 |
| zglob | rs15033543 | C | -0.258 | 1.63 | 0.0062 | 19q13.43 |
| zmem | rs15033543 | C | -0.357 | 1.61 | 0.0065 | 19q13.43 |
| zglob | rs15009807 | C | -0.627 | 1.55 | 0.0076 | 21q22.3 |
| stroop | rs61730014 | T | -0.177 | 1.54 | 0.0078 | 1p13.1 |
| recognition | rs15033543 | C | -1.135 | 1.48 | 0.0091 | 19q13.43 |
| zglob | rs76199660 | G | 0.09 | 1.47 | 0.0094 | 1p13.1 |
| imrec | rs14966744 | G | -0.609 | 1.38 | 0.0116 | 21q22.3 |
| zmem | rs20138893 | T | 0.37 | 1.37 | 0.0121 | 19q13.43 |
| stroop | rs14462817 | G | -0.092 | 1.36 | 0.0123 | 1p13.1 |
| zmem | rs61753528 | A | -0.379 | 1.35 | 0.0126 | 1p13.1 |
| zglob | rs14729425 | G | -0.279 | 1.29 | 0.0149 | 1p13.1 |
| zglob | rs13866122 | C | 0.334 | 1.23 | 0.017 | 19q13.43 |
| delrec | rs15061886 | C | -2.387 | 1.21 | 0.018 | 12q24.33 |
| zexec | rs14230361 | A | -0.287 | 1.2 | 0.019 | 21q22.13 |
| zmem | rs14803641 | G | -0.387 | 1.16 | 0.021 | 1p13.1 |
| fluency | rs61738781 | A | -2.565 | 1.16 | 0.021 | 19q13.43 |
| imrec | rs19152147 | T | -1.25 | 1.14 | 0.022 | 20p13 |
| zmem | rs35367003 | C | 0.248 | 1.13 | 0.022 | 20p13 |
| zexec | rs61730014 | T | 0.188 | 1.11 | 0.023 | 1p13.1 |
| zglob | rs41281850 | G | 0.276 | 1.13 | 0.023 | 20p13 |
| zmem | rs14966744 | G | -0.313 | 1.12 | 0.023 | 21q22.3 |
| zexec | rs15009807 | C | -0.638 | 1.07 | 0.027 | 21q22.3 |
| zmem | rs12483377 | G | -0.075 | 1.06 | 0.027 | 21q22.3 |
| zglob | rs20138893 | T | 0.292 | 1.05 | 0.028 | 19q13.43 |
| zglob | rs12483377 | G | -0.054 | 1.04 | 0.029 | 21q22.3 |
| fluency | rs13889958 | T | 11.426 | 1.03 | 0.03 | 21q22.3 |
| zmem | rs14547178 | G | -0.205 | 1.01 | 0.031 | 1p13.1 |
| stroop | rs14285111 | G | -0.1 | 1.02 | 0.031 | 19q13.43 |
| zexec | rs17293705 | G | 0.107 | 1.01 | 0.031 | 21q22.3 |

| | | | | | | |
|---|---|---|---|---|---|---|
| stroop | rs14026070 | C | -0.288 | 0.99 | 0.032 | 1p13.1 |
| fluency | rs14514427 | A | -8.581 | 1 | 0.032 | 19q13.43 |
| zmem | rs15009807 | C | -0.771 | 1 | 0.032 | 21q22.3 |
| stroop | rs17354559 | C | -0.07 | 0.99 | 0.033 | 1p13.1 |
| zglob | rs14862084 | G | -0.283 | 0.99 | 0.033 | 1p13.1 |
| DART | rs34914886 | C | 0.2 | 0.98 | 0.033 | 19q13.43 |
| zglob | rs61738781 | A | -0.107 | 0.98 | 0.033 | 19q13.43 |
| zglob | rs15063008 | G | 0.124 | 0.99 | 0.033 | 19q13.43 |
| zglob | rs14547178 | G | -0.198 | 0.97 | 0.035 | 1p13.1 |
| recognition | rs2074071 | G | 0.152 | 0.95 | 0.036 | 19q13.43 |
| zexec | rs62223645 | G | -0.092 | 0.96 | 0.036 | 21q22.13 |
| stroop | rs76199660 | G | -0.044 | 0.94 | 0.038 | 1p13.1 |
| zexec | rs61738781 | A | -0.103 | 0.92 | 0.039 | 19q13.43 |
| zglob | rs14586732 | T | -0.242 | 0.92 | 0.04 | 1p13.1 |
| block | rs61731605 | G | -0.939 | 0.91 | 0.04 | 21q22.13 |
| imrec | rs2839158 | C | 0.139 | 0.9 | 0.041 | 21q22.3 |
| stroop | rs17293705 | G | -0.063 | 0.91 | 0.041 | 21q22.3 |
| zexec | rs15063008 | G | 0.133 | 0.89 | 0.043 | 19q13.43 |
| stroop | rs45612632 | G | 0.065 | 0.88 | 0.044 | 19q13.43 |
| zexec | rs14362622 | G | -0.443 | 0.87 | 0.046 | 12q24.33 |
| learning | rs14323376 | A | -3.721 | 0.86 | 0.046 | 21q22.13 |

*zmem: composite score for memory; Stroop: stroop-ratio; DART: Dutch Adult Reading Test; TMT: trail making test ratio; WF: verbal fluency; zexec: composite score for executive function;; zglob: composite score for global cognition.*

**Figure S1. Linkage plots for chromosome 19, all traits suggestively or significantly linked to the region**

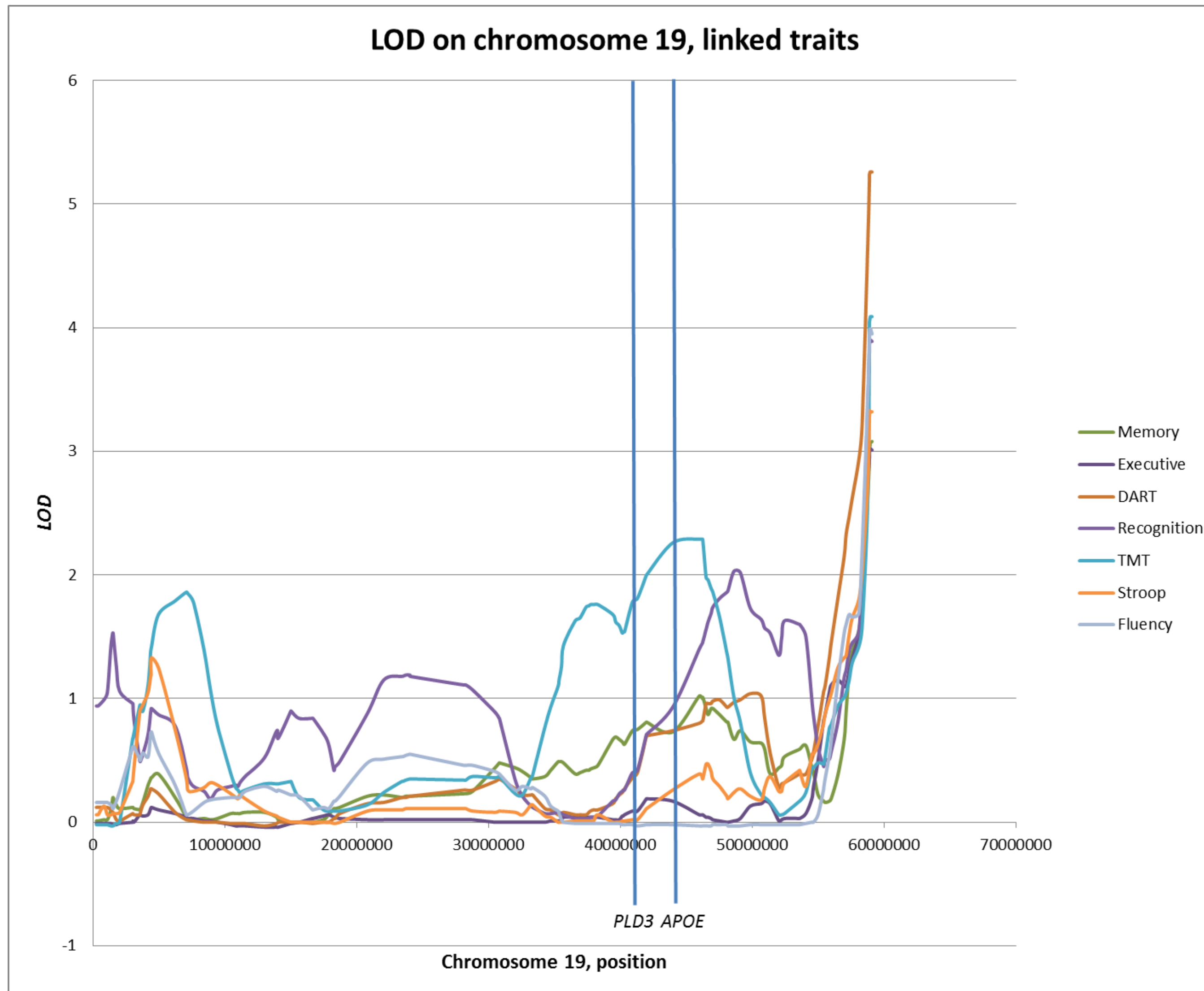
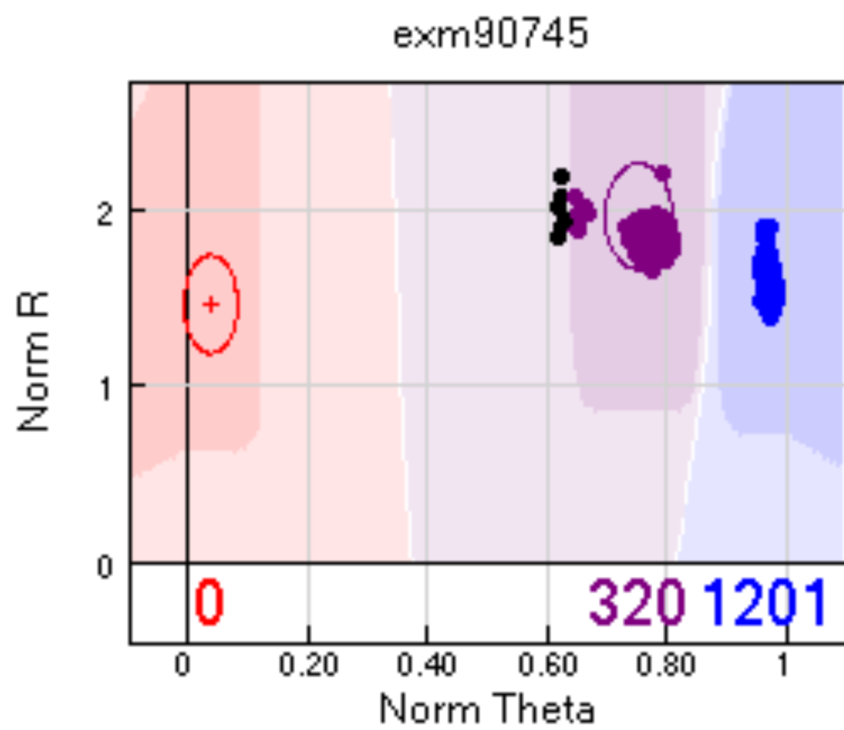


**Figure S2. Calling plot of rs76199660 on the exome chip**



## *7.2. Supplemental material to Chapter 3.2*

**GWAS for executive function and processing speed suggests involvement of the *CADM2* gene.**

**TABLE OF CONTENTS**

Supplementary notes
1. Abbreviations
2. Cohorts
3. Cognitive tests
4. Genotyping, Quality Control, and Imputation
5. Screening for Latent Population Substructure and Analysis Models
6. Additional Analyses
   - 6.1 Genotype-specific hippocampal mRNA levels (eQTLs)
   - 6.2 Gene network and functional prediction analysis

Supplementary Tables
- Supplementary Table 1: Detailed description of participating cohorts
- Supplementary Table 2: Genotyping parameters
- Supplementary Table 3: Quality control filters before imputation and methods for assessing population structure
- Supplementary Table 4: Imputation algorithms
- Supplementary Table 5: Tissue expression and functional data

Supplementary Figures
- Supplementary Figure 1: Quantile-Quantile (QQ) plots for all discovery meta-analyses
- Supplementary Figure 2: Manhattan plots for all discovery meta-analyses
- Supplementary Figure 3: CADM2 expression data from GTex

References

## 1. ABBREVIATIONS

*Test abbreviations*

LDST/DSST: Letter-digit or digit-symbol substitution task

TMT: Trail-making test

*Studies*

AAA: The Aspirin for Asymptomatic Atherosclerosis study

AGES: AGES-Reykjavik Study

ARIC: Atherosclerosis Risk in Communities Study

ASPS: Austrian Stroke Prevention Study

BLSA: Baltimore Longitudinal Study of Aging

CHS: Cardiovascular Health Study

ERF: Erasmus Rucphen Family study

FHS: Framingham Heart Study

GENOA: Genetic Epidemiology Network of Arteriopathy

GS:SFHS: Generation Scotland: Scottish Family Health Study

HBCS: Helsinki Birth Cohort Study

HCS: Hunter Community Study

LBC1921: Lothian Birth Cohort 1921

LBC1936: Lothian Birth Cohort 1936

MAP: Rush Memory and Aging Project

MAS: Sydney Memory and Aging study

NHS: Nurses' Health Study

ORCADES: Orkney Complex Disease Study

PROSPER: A Prospective Study of Pravastatin in the Elderly at Risk

ROS: Religious Orders Study

SHIP-TREND: Study of Health in Pomerania

WGHS: Women's Genome Health Study

3C: Three-City Study

## 2. COHORTS

*Discovery*

The CHARGE consortium includes large prospective community-based cohort studies that have genome-wide variation data coupled with extensive data on multiple phenotypes, as detailed previously [2]. In addition, several other community-based and family-based cohort studies have collaborated with the CHARGE consortium on this genome-wide analysis of cognitive function, both in the discovery and replication phases. Inclusion criteria for the current study were an age of 45 years or older at testing, no history of stroke or dementia and a Caucasian ethnicity; however, three African American cohorts from the same study centers as two of the core CHARGE Caucasian cohorts were included. Details on number of participants, age and sex distributions and mean scores on each test are given in Supplementary Table 1.

**Aging Gene-Environment Susceptibility - Reykjavik Study (AGES-Reykjavik)**

The AGES-Reykjavik Study is a single center prospective cohort study based on the Reykjavik Study. The Reykjavik Study was initiated in 1967 by the Icelandic Heart Association to study cardiovascular disease and risk factors. The cohort included men and women born between 1907 and 1935 who lived in Reykjavik at the 1967 baseline examination. Re-examination of surviving members of the cohort was initiated in 2002 as part of the AGES-Reykjavik Study. The AGES-Reykjavik Study is designed to investigate aging using a multifaceted comprehensive approach that includes detailed measures of brain function and structure. All cohort members were European Caucasians. Briefly, as part of a comprehensive examination, all participants answered a questionnaire, underwent a clinical examination and had blood drawn [3]. Of these, 3 660 were genotyped at the Laboratory of Neurogenetics, Intramural Research Program, NIA, Bethesda, Maryland, and 3219 participants passed QC criteria for genotyping. All consenting participants were offered to take a neuropsychological test battery [4], including the Stroop Color and Word Test [5] and the Digit Symbol Substitution Test (DSST) [6]. 603 participants were excluded for stroke or dementia. 2644 participants were available for the present analysis on Stroop, and 2 657 were available for the analysis of the DSST.

**The Atherosclerosis Risk in Communities Study (ARIC):**

The ARIC study is a prospective population-based study of atherosclerosis and clinical atherosclerotic diseases in 15 792 men and women, including 11 478 white participants, drawn from 4 United States communities (Suburban Minneapolis, Minnesota; Washington County, Maryland; Forsyth County, North Carolina; and Jackson, Mississippi). In the first 3 communities, the sample reflects the demographic composition of the community. In Jackson, only black residents were enrolled. Participants were between age 45 and 64 years at their baseline examination in 1987-1989 when blood was drawn for DNA extraction and participants consented to genetic testing [7]. A total of 15 020 participants, of which 10 898 were white, were genotyped at the Broad Institute, Boston, Massachusetts, and 9 345 of the

latter passed QC criteria for genotyping and were available for analysis after application of all exclusion criteria. Vascular risk factors and outcomes, including transient ischemic attack, stroke and dementia, were determined in a standard fashion [8]. The second clinical examination of the ARIC Study cohort in 1990–1992 included among others the DSST (revised) [9], and the Controlled Oral Word Association Test (Word Fluency Test (WFT), phonemic fluency) of the Multilingual Aphasia Examination [10,11]. Of this group, 275 subjects were excluded due to a history of stroke, leaving 9 181 subjects for DSST and 9 179 subjects for phonemic fluency. In addition, among white participants with genome-wide data, 509 have recently completed a detailed neuropsychological test battery as part of an ancillary study [12,13], including the Trail Making Test parts A and B [14], a category fluency test (animal naming), the Stroop Color and Word Test [5]. Of these, 79 participants were excluded due to stroke; 441 subjects were available for Stroop and category fluency, and 436 subjects for the trailmaking test.

**The 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 [15,16]. A total of 2 007 participants were randomly selected from the official community register stratified by gender and 5 year age groups. Individuals were excluded from the study if they had a history of neuropsychiatric disease, including previous stroke, transient ischemic attacks, and dementia, or an abnormal neurologic examination determined on the basis of a structured clinical interview and a physical and neurologic examination. During 2 study periods between September 1991 and March 1994 and between January 1999 and December 2003 an extended diagnostic work-up including neuropsychological testing was done in 1076 individuals aged 45 to 85 years randomly selected from the entire cohort: 509 from the first period and 567 from the second. The neuropsychological test battery included, among other tests [17]: part B of the Trail Making Test [14], the Stroop Color and Word Test [5] and the Letter–Digit Substitution Task (LDST) [18,19]. In 1992, blood was drawn from all study participants for DNA extraction (all were European Caucasians). Genotyping was done at the Human Genotyping Facility, Genetic Laboratory Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands, and successful in 996 participants. For executive function, there were 830 participants available for Trails B and 261 for Stroop. For processing speed, 263 individuals contributed results of the LDST.

**The Cardiovascular Health Study (CHS)**

The CHS is a population-based cohort study of risk factors for vascular disease in adults 65 years or older conducted across 4 field centers in the United States: Sacramento County, California; Washington County, Maryland; Forsyth County, North Carolina; and Pittsburgh, Allegheny County, Pennsylvania [20]. The original predominantly white cohort of 5 201 persons was recruited in 1989-1990 from a random sample of people on Medicare eligibility lists. An

additional 687 African-Americans were enrolled in 1992-1993, for a total sample of 5 888. Vascular risk factors and outcomes, including transient ischemic attack, stroke and dementia, were determined in a standard fashion [21,22]. DNA was extracted from blood samples drawn on all participants who consented to genetic testing at their baseline examination in 1989-90 or 1992-1993. In 2007-2008, genotyping was performed at the General Clinical Research Center's Phenotyping/Genotyping Laboratory at Cedars-Sinai on 3980 CHS participants who were free of cardiovascular disease at baseline and who had DNA available for genotyping. Among white participants, genotyping was attempted in 3 397 participants and was successful in 3 295 persons. Beginning in 1988/89, among other cognitive tests all participants completed the DSST [9] at their annual visits. In addition, as part of the CHS Cognition Study [22,23], in 1992-94 and again, in 1997-99, participants were invited to undergo and detailed neuropsychological assessment including among others a phonemic fluency test [24], and Trail Making Test part A and B [14], and the Stroop Color and Word Test [5]. After exclusions for stroke and dementia, there were 1 249 participants available for Trails A and B, and 2 022 individuals contributed results of the DSST.

**Framingham Heart Study (FHS)**

The FHS is a three-generation, single-site, community-based, prospective cohort study that was initiated in 1948 to investigate risk factors for cardiovascular disease including stroke. It now comprises 3 generations of participants: the original cohort followed since 1948 (Original) [25]; their offspring and spouses of the offspring, followed since 1971 (Offspring) [26]; and children from the largest offspring families enrolled in 2000 (Gen 3) [27]. The Original cohort enrolled 5 209 men and women who comprised two-thirds of the adult population then residing in Framingham, MA, USA. Survivors continue to receive biennial examinations. The Offspring cohort comprises 5 124 persons (including 3 514 biological offspring) who have been examined approximately once every 4 years. Participants in the first two generations were invited to undergo an initial neuropsychological test battery in 1999-2005 [28], including among other tests the Trail Making Test parts A and B [14] and tests for phonemic (F,A,S) and semantic fluency (animal naming). Neuropsychological testing in Gen 3 only began in 2009 and is not included in these analyses. The population of Framingham was virtually entirely whites in 1948 when the Original cohort was recruited. Vascular risk factors and outcomes, including transient ischemic attack, stroke and dementia, were identified prospectively since 1948 through an ongoing system of FHS clinic and local hospital surveillance [29,30]. Participants had DNA extracted and provided consent for genotyping in the 1990s. Genotyping was performed at Affymetrix (Santa Clara, CA) through an NHLBI funded SNP-Health Association Resource (SHARe) project and successful in 4,519 persons from the Original and Offspring cohorts. Of these 4 519 persons 4 116 were alive in 1999 when the neuropsychological study began. Of these, 2 642 participants have undergone neuropsychological testing including logical memory. After excluding 30 participants with a neurological condition that might confound the cognitive assessment (e.g. brain tumor or severe head injury), 7 participants with dementia and 52 participants with a history of

stroke, there were 2 475 participants available for the Trails tests, 1 403 for phonemic fluency and 1 421 for semantic fluency.

**Health, Aging, and Body Composition (Health ABC) Study**

The Health ABC study is a prospective cohort study designed to examine the associations between body composition, weight-related health conditions, and functional limitations in older adults aged 70-79 years at inception [31]. In 1997-1998, 3 075 participants were recruited from a random sample of white and all African-American Medicare eligible residents in the Pittsburgh, PA and Memphis, TN metropolitan areas. The DSST was administered at baseline in year 1 of the study [32]. Genotyping was performed in 1 732 white participants at the Center for Inherited Disease Research and 1663 met all QC criteria. After exclusion of 134 participants for history of stroke, 1 519 individuals with cognitive test results were available for the current analysis of processing speed. All participants provided informed consent and protocols were approved by the institutional review boards at both study sites.

**Rotterdam Study**

The Rotterdam Study (RS) 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 [33,34]. In 1990-1993, 7 983 persons participated and were re-examined every 3 to 4 years. (RS1). Additional cohorts from the same district were recruited in 2000/2001 (RS2, 3 011 subjects aged 55 years and older) and 2006-2008 (RS3, 3 932 subjects aged 45 years and older), with the exact same study protocol as RS1[35]. All participants had DNA extracted at their first visit. Genotyping for RS1 was attempted in participants with high-quality extracted DNA in 2007-2008. In total, 5 974 samples from RS1 were available with good quality genotyping data. Genotyping was done at the Human Genotyping Facility, Genetic Laboratory Department of Internal Medicine, Erasmus MC, Rotterdam, the Netherlands. Participants or RS1 underwent several neuropsychological tests at follow-up examinations [36], including the Stroop Color and Word Test [5], the LDST [18,19], a verbal fluency test in which participants had to name as many animals as possible within a 1-minute time limit [37] These tests were first assessed in the third round of examination (1997-1999). Participants are continuously monitored for major events, including dementia and stroke, by automated linkage of the general practitioners' records and hospital discharge files with the study database [38,39]. Among participants with genome-wide data, after exclusion of 124 participants with dementia and 168 participants with a history of stroke, there were 3 204 participants available for the analysis on Stroop. For processing speed, 3 250 individuals contributed results of the LDST. For animal naming there were 3 276 participants available. For RS2, genotyping data was available for 2 157 subjects. 102 subjects were excluded due to prevalent stroke, and six subjects due to prevalent dementia. Stroop (n=1 805), category fluency (n=1 842) and DSST (n=1 821) were performed at baseline.

For RS3, genotyping data was available for 2 082 participants; in this relatively young cohort, no exclusions for stroke and dementia were made. Particpants took the Stroop (n=1 923), category fluency (n=2 045) and DSST (n=2 030) tests at baseline.

**Erasmus Rucphen Family (ERF)**

The Erasmus Rucphen Family (ERF) study is a family-based cohort study in a genetically isolated population in the Netherlands [40,41], including 3 000 participants. Participants are all descendents of a limited number of founders living in the 19th century. Extensive genealogical data is available for this population. The study protocol included venous puncture for DNA isolation and chemistry, cognitive evaluation, cardiovascular examination, eye assessments and body composition measurements. All participants gave informed consent and the study was approved by the medical ethics committee at Erasmus MC University Medical Center.

Genotyping was done at the Human Genotyping Facility, Genetic Laboratory Department of Internal Medicine, Erasmus MC, Rotterdam, and at the Genotyping Center of Leiden University, The Netherlands. In total, 2 385 samples from the ERF Study were available with good quality genotyping data. Participants were invited to undergo a neuropsychological evaluation, which included among other tests [42]: verbal fluency which included a letter fluency test (DAT) and animal naming test [43], the Trail Making Test parts A and B [14] and the Stroop Color and Word Test [5]. Among participants with genome-wide data, after exclusion of 20 individuals with a history of stroke, at total of 1 238 subjects for Stroop, 1 267 for TMT and 1 255 for phonemic fluency, and 1267 subjects for semantic fluency were available.

**Lothian Birth Cohort 1921 (LBC1921) and 1936 (LBC1936)**

These samples include surviving participants from the Scottish Mental Surveys of 1932 or 1947 (SMS1932 and SMS1947), having been born in 1921 (LBC1921) and 1936 (LBC1936), respectively [44,45]. They were all Caucasian and almost all lived independently in the Lothian region (Edinburgh city and surrounding area) of Scotland. The LBC1921 cohort comprised 550 members while the LBC1936 cohort included 1 091 participants.

At mean age 79, LBC1921 participants underwent a neuropsychological examination including a phonemic fluency test [19]. In a later visit, at mean age 83, they also took the Digit Symbol coding subtest from the Wechsler Adult Intelligence Scale-III UK [46]. At age 70, LBC1936 participants took a battery of cognitive tests [44], including a phonemic fluency test [47] and the DSST.

Genotyping was performed at the Wellcome Trust Clinical Research Facility (WTCRF) Genetics Core, Western General Hospital, Edinburgh, United Kingdom, and successful in 527 participants from the LBC1921 cohort and 1 005 participants from the LBC1936 cohort. Among participants with genome-wide data, after exclusion of 4 and 0 participants for dementia and of 43 and 50 participants for history of stroke, 468 and 950 participants were available for the analysis of verbal fluency. There were 257 and 951 individuals, respectively, available for analysis of the DSST.

**Orkney Complex Disease Study (ORCADES)**

ORCADES is an ongoing, family-based, cross-sectional study that seeks to identify genetic factors influencing cardiovascular and other disease risk in the population isolate of the Orkney Isles in northern Scotland [48]. The North Isles of Orkney, the focus of this study, consist of a subgroup of ten inhabited islands with census populations varying from ~30 to ~600 people on each island. The first phase of data collection was carried out in Orkney between 2005 and 2007. Informed consent and blood samples were provided by 1 019 Orcadian volunteers who had at least one grandparent from the North Isles of Orkney. Participants were invited to take a neuropsychological test battery including the DSST [49] and verbal fluency (letter fluency) [47]. Genotyping was performed at the Helmholtz Centre in Munich on a subset of 719 participants. An additional 169 individuals were genotyped by Integragen in Paris. Among participants with genome-wide data, after exclusion of 7 participants with a history of stroke, for verbal fluency 418 participants were available. For processing speed, 417 participants over 45 years of age were available for analysis. No cases of dementia were included in the study.

**Croatian Cohorts: Split, Vis and Korčula**

The Vis and Korčula studies are part of a larger genetic epidemiology research program in Croatian island isolates, "10 001 Dalmatians." The genetic epidemiology research program in Croatian island isolates began in 1999 [50], then expanded to study human genetic variation and effects of isolation and inbreeding [51,52], and finally entered the phase of focusing on diseases and gene mapping studies [53-55]. 311 and 535 participants were included respectively in the Vis and Korčula studies. The Split study included 313 persons collected in 2009 from the general (outbred) population of Split. Split has a population of >100 000 and is second largest city in Croatia. Participants from the Vis, Korčula and Split studies were invited to undergo a neuropsychological examination including the DSST [6] and a letter fluency test[19]. Genotyping was performed at the Institute of Human Genetics, Helmholtz Zentrum München, Germany. For the analysis on letter fluency, 382 participants were available for Vis, 495 for Korčula and 306 for Split. For the processing speed analysis, 311 participants were available for Vis, 542 for Korcula, and 315 for Split.

**Religious Orders Study (ROS)**

The ROS, started in 1994, enrolled Catholic priests, nuns, and brothers, aged 53 years from about 40 groups in 12 states.[56] Since January 1994, 1 132 participants completed their baseline evaluation, of whom 1 001 were non-Hispanic white. The study was approved by the institutional review board of Rush University Medical Center. The follow-up rate of survivors exceeds 90%. Participants were free of known dementia at enrollment, agreed to annual clinical evaluations, and signed both an informed consent and an Anatomic Gift Act form donating their brains at time of death.[57] A more detailed description of the ROS has been published previously.[57] Participants were invited to take a neuropsychological test battery, including delayed recall of Story A from the logical memory subset of the Wechsler

Memory Scale-Revised,[58] and delayed word list recall from the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) battery.[59] DNA was extracted from whole blood, lymphocytes, or frozen postmortem brain tissue. Genotyping was performed at the Broad Institute's Center for Genotyping and the Translational Genomics Research Institute. Among participants with available memory tests and genome-wide genotypes, after exclusion of 58 participants with dementia and 65 participants with a history of stroke, 686 individuals were available for a GWAS of semantic fluency

**Rush Memory and Aging Project (MAP)**

The MAP, started in 1997, enrolled older men and women from assisted living facilities in the Chicago area with no evidence on dementia at baseline.[56] Since October 1997, 1 285 participants completed their baseline evaluation, of whom 1 118 were non-Hispanic white. The study was approved by the institutional review board of Rush University Medical Center. The follow-up rate of survivors exceeds 90%. Similar to the ROS, participants agreed to annual clinical evaluations and signed both an informed consent and an Anatomic Gift Act form donating their brains, spinal cords, and selected nerves and muscles to Rush investigators at the time of death.[57,60] A more detailed description of the MAP has been published previously.[57,60] Participants were invited to take a neuropsychological test battery, including delayed recall of Story A from the logical memory subset of the Wechsler Memory Scale-Revised,[58] and delayed word list recall from the CERAD battery.[59] DNA was extracted from whole blood, lymphocytes, or frozen postmortem brain tissue. Genotyping was performed at the Broad Institute's Center for Genotyping and the Translational Genomics Research Institute. Among participants with available memory tests and genome-wide genotypes, after exclusion of 51 participants with dementia and 86 participants with a history of stroke, 751 individuals were available for a GWAS of semantic fluency.

**PROspective Study of Pravastatin in the Elderly at Risk (PROSPER)**

PROSPER was a prospective multicenter randomized placebo-controlled trial to assess whether treatment with pravastatin diminishes the risk of major vascular events in elderly. Between December 1997 and May 1999, we screened and enrolled subjects in Scotland (Glasgow), Ireland (Cork), and the Netherlands (Leiden). Men and women aged 70-82 years were recruited if they had pre-existing vascular disease or increased risk of such disease because of smoking, hypertension, or diabetes. A total number of 5 804 subjects were randomly assigned to pravastatin or placebo. A large number of prospective tests were performed including Biobank tests and cognitive function measurements. A detailed description of the study has been published elsewhere [61,62] . Blood pressure was measured in all subjects during the trial every 3 months by a standardized technique using Omron M4 sphygmomanometers (Omron Healthcare Inc, Bannockburn, Illinois). A whole genome wide screening has been performed in the sequential PHASE project with the use of the Illumina 660K beadchip.

Of 5 763 subjects DNA was available for genotyping. After QC, 5 244 subjects were left for analysis. Of these 5 244 subjects,374 subjects with a history of stroke at baseline were excluded, leaving 4 870 subjects with a Stroop score at baseline. Genotyping was performed with the Illumina 660K beadchip, after QC (call rate <95%) 557,192 SNPs were left for analysis. These were imputed to 2.5 million SNPs based on the HAPMAP built 36 with MACH imputation software.

**Genetic Epidemiology Network of Arteriopathy (GENOA)**

GENOA is a study of hypertensive sibships designed to investigate the genetic underpinnings of hypertension and target organ damage [63]. In the initial phase of the GENOA study (Phase I: 1996-2001), all members of sibships containing ≥ 2 individuals with essetial hypertension clinically diagnosed before age 60 were invited to participate, including both hypertensive and normotensive siblings (1 583 non-Hispanic whites from Rochester, MN, and 1 841 African Americans from Jackson, MS). The diagnosis of essential hypertension was established based on blood pressure levels measured at the study visit (>140 mmHg average systolic BP or >90 mmHg average diastolic BP) or a prior diagnosis of hypertension and current treatment with antihypertensive medications. Exclusion criteria were secondary hypertension, alcoholism or drug abuse, pregnancy, insulin-dependent diabetes mellitus, or active malignancy. In the second phase of the GENOA study (Phase II: 2000-2004), 1 241 white and 1 482 African American participants were successfully re-recruited to measure potential target organ damage due to hypertension. The Genetics of Microangiopathic Brain Injury (GMBI) study (2001-2006) is an ancillary study of GENOA undertaken to investigate susceptibility genes for ischemic brain injury. Phase II GENOA participants that had a sibling willing and eligible to participate in the GMBI study underwent a neurocognitive testing battery to assess several domains of cognitive function including the Trails tests, the Stroop test, semantic and phonemic fluency and the DSST. Genotyping was performed at the Mayo Clinic, Rochester (MN).

*Whites – Rochester, MN*

The original sample included 967 whites from Rochester, MN that participated in the GENOA-GMBI substudy and had measurements for at least one of the ten cognitive phenotypes used in this analysis. Participants that had history of stroke (N = 22) were excluded from the analysis, leaving a total of 889 participants. After excluding participants that did not have genome-wide genotype data, 757 participants were included in the Trails tests, 762 in the semantic fluency, 760 in the phonemic fluency, 740 in the Stroop tests and 758 in the DSST.

*Replication*

**Atherosclerosis Risk in Communities Study (ARIC) - African American subcohort**

Details of the ARIC cohort in general are given in the discovery description. The 4122 black participants of the ARIC study were considered a separate group due to the differences in

population genetics between these samples; after exclusion of 118 subjects with stroke, 2 476 subjects with genotyping were available for these analyses. 413 subjects completed also the more detailed cognitive assessments in the second clinical examinations. With the same methods, tests and exclusion criteria, the ARIC African American cohort served as a replication cohort for the several tests. Respectively, 2 459 subjects for phonemic fluency, 2 457 subjects for DSST, 387 subjects for Stroop and 413 subjects for Trails were available.

**The Aspirin for Asymptomatic Atherosclerosis Study**

The aspirin for asymptomatic atherosclerosis (AAA) cohort was recruited for a double blind, placebo controlled randomized clinical trials which took place in central Scotland in 1998-2006 [64]. Participants were aged between 50 and 75 at time of recruitment. All participants gave written informed consent. Exclusions applied at recruitment were previous myocardial infarction, stroke or ongoing treatment with aspirin or warfarin. Participants underwent neuropsychological tests at three months (baseline cognitive testing) and at five years. At baseline 504 participants completed a test battery including the verbal fluency test [65], the trail making test [66] and the digit symbol test from the Wechsler Adult Intelligence Scales III [46], all participants completed this same battery at five years after baseline. For the purpose of this study the first time of testing for each individual was used thus eliminating any practice effects for those having completed the testing twice. Thirty five individuals were excluded due to stroke. Dementia diagnoses were not available on these individuals, however they had completed the MMSE and individuals with a score below 24 were excluded. There were 1 945 individuals (73.00 % females) available for these analyses, with a mean age of 66.21 (±6.66) years. For verbal fluency there were 1 908 individuals available for analysis. For the trail making test B 1 921 individuals were analysed. Processing speed was analysed in 1936 individuals using the digit symbol test.

**Cardiovascular Health Study – African American samples**

The African American subsample of CHS (see Discovery section for cohort description) contributed to the replication of TMTA. Of these 687 subjects, after exclusions for stroke and dementia, 327 subjects were available.

**Generation Scotland: Scottish Family Health Study**

The Generation Scotland: Scottish Family Health Study (GS:SFHS) is a large, family-based intensively-phenotyped cohort recruited from the general population across Scotland[67]. Participants were aged 35 – 55 years at time of recruitment and all available first degree relatives over 18 years of age were also recruited. Demographic, clinical and biological data were collected along with blood samples from which DNA was extracted. The participants were invited to take a battery of neuropsychological tests which included among others the digit symbol test from the Wechsler Adult Intelligence Scales III[3] and the verbal fluency test[4] using letters C, F, and L, with one minute for each. For the purpose of this study a sub-group

of this cohort will be used; this comprises unrelated individuals over the age of 45. There were 86 individuals removed due to a history of stroke and one individual removed due to dementia. There were 3 401 individuals (58.72 % females) available for these analyses, with a mean age of 60.31 (±8.51) years. For the verbal fluency analyses there were 3 386 individuals available. The digital symbol test analyses included 3 369 individuals.

**Baltimore Longitudinal Study of Aging (BLSA)**

The BLSA is a multidisciplinary observational study of the physiological and psychological aspects of human aging and diseases and conditions that increase with age.[68] Participants were invited to take a neuropsychological test battery, including, among others, the Trails tests, semantic and phonemic fluency and the DSST. DNA was genotyped at the Laboratory of Neurogenetics, NIA genotyping platform, USA. Among the 925 participants with available genome-wide data and cognitive test performance, after exclusion of 37 participants with a history of stroke, 422 subjects were available for the trails and fluency tests, and 438 subjects contributed to the DSST.

**3CS**

The 3C Study is a French prospective population-based study including 9294 non-institutionalized participants aged 65 years or more, living in the cities of Bordeaux, Montpellier and Dijon.[69] GWAS data were available for 6 440 subjects. At baseline and at each follows-up examination participants were invited to take a neuropsychological test battery, including among other tests the Trails test and semantic fluency tasks.[70] DNA was genotyped at the Centre National de Génotypage, Evry, France.[71] 112 subjects with dementia and 291 subjects with stroke were excluded from analysis, leaving 5 536 subjects for the Trails test and 5 643 subjects for semantic fluency.

**Genetic Epidemiology Network of Arteriopathy (GENOA)**

*African Americans – Jackson, MS*

General information on the GENOA cohort is given above. The original sample included 1 010 African Americans from Jackson, MS that participated in the GENOA-GMBI substudy and had measurements for at least one of the cognitive phenotypes used in this analysis. Participants that had missing age data (N=2) or had history of stroke (N = 51) were excluded from the analysis, leaving a total of 934 participants. Also excluded from this analysis were GENOA participants that also participated in the cognitive portion of the ARIC study (N=118) After excluding participants that did not have genome-wide genotype data, 704 participants were included in the Trails tests, 716 in the semantic fluency, 687 in the phonemic fluency, 648 in the Stroop tests and 697 in the DSST.

**Helsinki Birth Cohorts**

The source cohort for the HBCS comprised 4130 women and 4630 men born as singletons at Helsinki University Central Hospital during 1934-44, who had birth and child welfare records

and were living in Finland in 1971.[72] To approach an intended sample size of n=2 000, a random subsample of 2 902 subjects was invited to participate in the study; 2 003 of them (1 075 women and 928 men) were finally included.[73] The study was approved by the Institutional Review Board of the National Public Health Institute, and informed consent was obtained from all participants. Participants who could come to the examination center were invited to take a neuropsychological test battery, including the semantic fluency test.[59] DNA was extracted from 1 728 randomly selected participants of the HBCS. Genotyping was conducted at the Wellcome Trust Sanger Institute, Cambridge, UK. Among participants with available cognitive tests and genome-wide genotypes, after exclusion of 22 participants with a history of stroke, 889 individuals were available for a GWAS on semantic fluency.

**Hunter Community Study (HCS)**

The HCS is a community-based longitudinal investigation that was commenced in Australia in 2004–2005. The study aims to investigate retired and near-retired persons by sampling older Australians aged 55–85, randomly selected from electoral rolls in a regional area on the heavily populated east coast (New South Wales).[74] All participants were invited to take the Audio Recorded Cognitive Screen (ARCS), an instrument that uses an audio device to administer selected neuropsychological tests to unsupervised individuals,[75] including phonemic and semantic fluency tasks. Genotyping was conducted at the Hunter Medical Research Institute, Newcastle Australia. Genotype data were available for 1 230 individuals of European ancestry. Among 856 participants with available test results and genome-wide genotypes, after exclusion of 34 participants with a history of stroke, 822 subjects were available for semantic fluency and 821 subjects contributed to phonemic fluency.

**Invecchiare in Chianti (InCHIANTI)**

The InCHIANTI study is a prospective population-based study of 1 270 randomly selected persons aged 65 and older, involving subjects randomly selected from population registries from the towns of Greve in Chianti and Bagno a Ripoli, Tuscany, Italy, in 1998 [76]. As part of the baseline examination, the participants performed, among others, the Trailmaking test part A and B. 359 subjects were excluded due to prevalent stroke or dementia, leaving 845 subjects available with results for Trails A, and 758 subjects for Trails B.

**SHIP-TREND**

A separate stratified random sample of 10 000 adults aged 20 to 79 years was drawn for SHIP-TREND from the German general population (Völzke, Alte et al. 2011). The target sample size was chosen to achieve a final sample size similar to that of SHIP-0. In total 4 422 subjects participated. SHIP-TREND examinations started in September 2008 and were completed in October 2012. From the first 986 subjects who underwent whole-body MRI scanning and oral glucose tolerance testing GWAS data were available (Illumina HumanOmni2.5-Quad®). All participants performed the The Stroop Color and Word Test. All

participants gave written informed consent. After exclusions, 597, 561and 558 individuals were available for GWAS on Stroop color and word and interference test, respectively.

**Sydney Memory and Ageing Study (MAS)**

The Sydney MAS was initiated in 2005 to examine the clinical characteristics and prevalence of mild cognitive impairment and related syndromes, and to determine the rate of change in cognitive function over time.[77] Participants were invited to take a neuropsychological test battery, including among other tests semantic and phonemic fluency tests, Trailmaking tests and the DSST. DNA was genotyped at the Ramaciotti Centre, UNSW, Australia. Among the 925 participants with available genome-wide data and cognitive test performance, after exclusion of 37 participants with a history of stroke, 884 subjects were available for the GWAS on semantic fluency, 883 for phonemic fluency, 872 for the Trailmaking tests, and 873 for the DSST.

**NHS**

The Nurses' Health Study (NHS) began in 1976, when 121 700 female registered nurse women, aged 30 to 55 years, living in 11 USA states completed a mailed questionnaire on lifestyle and health [78]. Every 2 years, follow-up questionnaires have been mailed to participants to update their information, and >90% follow-up of the total possible person-time has been maintained [79].For the study of cognitive function, participants aged 70 years and older free of diagnosed stroke were selected. From 1995 to 2001, 21 085 eligible women were contacted for a baseline telephone cognitive assessment, including a task on semantic fluency (number of animals named in one minute). The study was approved by the Institutional Review Board of the Brigham and Women's Hospital.[79] Genotyping was conducted at the Broad Institute using a nested case-control design for previous research on breast cancer, type 2 diabetes, and other health conditions. Subjects with stroke were excluded from cognitive testing, and dementia was not assessed. 2 067 individuals were available for a GWAS on semantic fluency.

**WGHS**

The primary aim of the WGHS was to create a comprehensive, fully searchable genome-wide database of over 360 000 single nucleotide polymorphisms among at least 25 000 initially healthy American women participating in the ongoing NIH-funded Women's Health Study (WHS).[80] By design, participants included in the WGHS were free from dementia and stroke at baseline. Women above 70 years of age were contacted for a baseline telephone cognitive assessment, including a measure of semantic fluency (number of animals named in one minute). Genotyping was conducted at Amgen. Among participants with available tests and genome-wide genotypes, 3 907 individuals were available for a GWAS of semantic fluency.

**3. COGNITIVE TESTS**

In all of the studies, the cognitive tests were performed in eligible participants in a standardized fashion and interpreted without knowledge of genetic information. Four types of executive function tests and two tests of processing speed were performed to assess delayed recall: Trail making Test A, Trail Making Test B, Trail Making Test B minus A, Phonemic and Category Fluency Tests, Stroop Color and Word (interference and card 3 only), and the Letter-Digit Substitution Test or Digit Symbol Substitution Test.

***Trail Making Test parts A and B, used by ARIC, CHS, ERF, FHS, InCHIANTI, GENOA, MAS, 3CS and BLSA, as well as Trail Making Test part B only for ASPS and AAA***

The TMT is a time-demanding task in which participants have to connect letters and numbers as quickly as possible. In TMT-A only numbers have to be connected, in TMT-B numbers and letters have to be connected alternately (from 1 to A, to 2 to B etc). The score was defined as the time in seconds to complete the task. Participants who passed the maximum test time for TMT-B were given the maximum time of 300 seconds. There were no participants who timed out on TMT-A. For analysis, we used time in seconds on TMT-A and TMT-B. Additionally, we used the time difference between TMT-B and -A (TMTBminusA). The times were transformed by taking the natural logarithm.

***Phonemic fluency, used by ARIC, ERF, LBC1921, LBC1936, Korcula, ORCADES, SPLIT, VIS, AAA, BLSA, GENOA, GS:SFHS, FHS, Hunter Community Study, and MAS***

Verbal fluency was assessed with a phonemic task. Participants had to name as many words starting with the same letter in one minute ("F, A and S" for ARIC, FHS, BLSA and MAS, "C, F and L" for LBC1921 and LBC 36 and AAA and Generation Scotland, and "D, A and T" for non-English-speaking cohorts. The score was defined as the total number of correct words. Hunter Community Study offered only one one-minute trial, randomly with the letter "H" or "W". Results were tripled to compare with the other cohorts.

***Category fluency test, used by ARIC, ERF, FHS, RS, BLSA, GENOA, HBCS, Hunter, MAS, RUSH, WGHS, NHS, 3CS***

Category fluency was assessed by a task in which participants had to name as many items in a certain category (animals) within a certain time limit. Score was defined as the total number of unique correctly named animals within one minute.

***Stroop Color and Word Test, used by AGES, ARIC, ASPS, CHS, ERF, RS, GENOA, SHIP-TREND, PROSPER and RUSH***

The Stroop Color and Word Test is a time demanding task consisting of 3 cards. In card I, participants have to name the right words as quickly as possible, in card II they have to name the right colors as quickly as possible and in card III they have to name the colors in which the words are printed as quickly as possible. On this card the meaning of the word is different than the color the word is printed in (e.g. blue is written whereas the word is printed in red). ARIC,

CHS, GENOA and RUSH recorded the number of correctly read items in a certain time limit; AGES, ASPS, ERF, RS, SHIP-TREND and PROSPER recorded time in seconds needed to complete the card. For analysis, we used the score on card 3 alone, and the difference between the scores on card III and II (Stroop interference).

***Digit Symbol Substitution Test (DSST)[49], used by ARIC, AGES, CHS, Health ABC, Korcula, LBC1921, LBC1936, ORCADES, Split, Vis, AAA, BLSA, GENOA, GS:SFHS, MAS, Rush, PROSPER***

The Digit Symbol Substitution Subtest of the Wechsler Adult Intelligence Scale-Revised (WAIS-R) is a paper-and-pencil task requiring timed translation of numbers (1-9) to symbols using a key given at the top of the test page. The test was scored as the number of correct translations completed within 90 seconds (range 0 – 93). MAS used the test from the WAIS-III which was scored as the number correct within 120 seconds (range 0-133).

***Letter-Digit Substitution Task (LDST), used by ASPS and RS***

In the Letter-Digit Substitution Task, each number from 1-9 is paired with a different letter of the alphabet using a key given at the top of the test page. The participant is required to assign numbers to as many of a series of randomized letters as possible within 60 seconds. The score was defined as the total number of correct combinations completed.

**4. GENOTYPING, QUALITY CONTROL, AND IMPUTATION**

The consortium was formed after the individual studies had finalized their GWAS platforms, and the studies included used different platforms (**Supplementary Table 2**).

As detailed previously,[81] participant-specific quality controls included filters for call rate, heterozygosity, and number of Mendelian errors per individual. SNP-specific quality controls included filters for call rate, minor allele frequency, Hardy-Weinberg equilibrium, and differential missingness by outcome or genotype (mishap test in PLINK, http://pngu.mgh.harvard.edu/purcell/plink/).

The set of genotyped input SNPs used for imputation in each study was selected based on their highest quality GWA data (see **Supplementary Table 3** for quality control filter applied prior to imputation). We used either the Markov Chain Haplotyping (MaCH) package (http://www.sph.umich.edu/csg/abecasis/MACH) or BIM-BAM15 programs for imputation, as detailed in **Supplementary Table 4**. For each imputed SNP, imputation quality was estimated using the R-square metric or the ratio of the empirically observed dosage variance to the expected binomial dosage variance.

Of note, for imputation in African-Americans from the ARIC study, given the admixed nature of the study population, SNPs were not removed if they only deviated from the Hardy-Weinberg equilibrium criterion. A combined CEU + YRI reference panel was generated prior to imputation that included approximately $2.74 \times 10^6$ SNPs that were either in both the CEU

and YRI panels, or that were present in one panel and were monomorphic and non-missing in the other.

The GS:SFHS and AAA studies did not provide GWAS data, but performed genotyping of SNPs emerging in the discovery phase in the ABI Openarray custom array; failed SNPs on the array were genotyped using Taqman.

**5. SCREENING FOR LATENT POPULATION SUBSTRUCTURE AND ANALYSIS MODELS**

Studies were screened for latent population substructure, including cryptic relatedness, using suitable programs (**Supplementary Table 3**).[82-84] When appropriate, components related to the phenotype under study were included as covariates in the linear regression. ERF, CROATIA-Vis, CROATIA-Korčula, GENOA and FHS included related individuals and used the following methods to adjust for relatedness of the population: ERF, CROATIA-Vis and CROATIA-Korčula performed association analyses in GenABEL (R-library) [85], using the mmscore method, with a kinship matrix that was estimated from the genotype data[86]; GENOA performed association analyses using linear mixed effects modeling with family as a random intercept; FHS used a linear mixed effects model accounting for familial relatedness.[87]

We studied quantile-quantile (Q-Q) plots to ensure that the p-value distributions in each of the cohorts conformed to a null distribution at all but the extreme tail. We also calculated the genomic inflation factor lambda, which measures over-dispersion of test-statistics from association tests indicating population stratification and can be used to apply genomic control.[88]

For discovery, only SNPs present in all studies for the respective trait were used. For meta-analyses combining studies using exactly the same tests, we used an inverse-variance weighted meta-analysis as our primary method after applying genomic control within each individual study. Beta estimates were weighted by their inverse variance and a combined estimate was obtained by summing the weighted betas and dividing by the summed weights. Hence results for SNPs imputed with low certainty were down-weighted because the low quality of imputation ensures a large variance. In contrast, studies with large sample sizes and with directly genotyped or well-imputed SNPs had a greater effect on the meta-analyses p-value because of small variances.
For meta-analyses combining studies using similar but not identical measurements (Stroop), we used as a primary method an effective sample size weighted meta-analysis technique after applying genomic control within each individual study. For each SNP the z-statistic was weighted by the effective sample size (product of the sample size and the ratio of the empirically observed dosage variance to the expected binomial dosage variance for imputed SNPs). A combined estimate was obtained by summing the weighted z-statistics and dividing by the summed weights. Hence results for SNPs imputed with low certainty were down-

weighted. In contrast, studies with large sample sizes and with directly genotyped or well-imputed SNPs had a greater effect on the meta-analyses p-value.
We undertook the meta-analyses using METAL.

We estimated the genomic inflation factor lambda after meta-analysis. Lambdas were below 1.05 for all analyses (see details in **Supplementary Figure 1**), indicating no significant inflation of p-values. The quantile-quantile (Q-Q) plots of our meta-analysis results for the various memory traits (**Supplementary Figure 1**) show the distribution of the observed test statistic (negative log of p-values, on the *y*-axis) plotted against the distribution of test statistic expected under the null-hypothesis (on the *x*-axis). All SNPs with a p-value < $5x10^{-6}$ and a minor allele frequency > 0.05 were selected for in silico replication.

## 6. ADDITIONAL ANALYSES

### 6.1 Genotype-specific hippocampal mRNA levels (eQTLs)

To characterize the functional relevance of significant and suggestive memory risk loci, we analyzed their association with human hippocampus cell line gene expression profiles, in a unique set of 138 pre-mortem human hippocampus samples from patients undergoing surgery for treatment-resistant epilepsy, as part of the Epilepsy Surgery Program at Bonn University, Germany.[89] Hippocampal biopsy samples were obtained from patients with chronic pharmaco-resistant temporal lobe epilepsy in the Epilepsy Surgery Program at Bonn University, Germany.[89] Surgical removal was indicated to achieve seizure control after standardized pre-surgical evaluation using a combination of noninvasive and invasive procedures.[90] Procedures were carried out in accordance with the Helsinki-Declaration and approved by the local ethics committee. All patients signed a written informed consent. Fresh frozen hippocampal segments derived from the epilepsy surgery were prepared as tissue slices via cryostat-conditions. Total DNA and RNA of hippocampus samples were isolated using the AllPrep DNA/RNA Micro Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. Quality of total RNA of each individual tissue sample was checked for degradation via gel electrophoresis in a BioAnalyzer 2 100 (Agilent Technologies, Waldbronn, Germany) with RNA 6 000 nano lab chips following the instructions of the manufacturers protocol. All RNA samples showed intact 28S and 18S ribosomal RNA signals and a RNA integrity number (RIN) of >7.9. All total RNA and DNA samples were used for a systematic chip-based genome-wide association study (GWAS) and gene expression (GEX) analysis. 50 ng of total RNA was reverse transcribed into cRNA and biotin-UTP labelled using the Illumina TotalPrep 96-RNA Amplification Kit (Ambion/Applied Biosystems, Darmstadt, Germany). Labelled cRNA was hybridized to Illumina human HT-12 Expression BeadChips using standard protocols (Illumina, San Diego, USA). Human HT-12 microarrays interrogate for more than 99.99% of all known human genes (approx. 25,000 annotated RefSeq and UniGene genes) containing more than 48,000 probes. All expression profiles were extracted and average/quantile/cubic spline/rank invariant normalized using GenomeStudio software (Illumina, San Diego, USA). For genome-wide SNP-genotyping 200 ng of DNA was hybridized

to the Illumina Human660W-Quad v1 DNA Analysis BeadChip (>658 000 markers per sample) according to the Infinium® HD Assay Super manual from Illumina (Illumina, San Diego, USA). Allele calls were determined using Illumina BeadStudio (Illumina, San Diego, USA).[89]
We applied a hidden factor analysis that identifies and corrects for unknown confounding factors in the data and thus diminishes the false-positive and false-negative eQTL rate (PEER, https://github.com/PMBio/peer/wiki)[91,92]. Fifteen hidden factors were identified and used as co-variates for expression analysis. Statistical Analysis was performed using GenABEL® software.

**6.2 Gene network and functional prediction analysis**
Prediction of gene function can be conducted using a guilt-by-association approach: E.g. if there are 100 genes that are known to be involved in apoptosis, identification of a gene that is strongly co-expressed with these 100 genes suggests that that gene is likely to be involved in apoptosis as well.
As such co-expression data can be used to predict likely functions for genes. However, important to realize is that some phenomena exert very strong transcriptomic effects and therefore will overshadow more subtle effects. In order to be able to identify such subtle relationships as well, we conducted a principal component analysis on an unprecedented scale (**Fehrmann *et al*, manuscript in preparation**): We collected gene expression data for three different species (homo sapiens, mus musculus and rattus norvegicus) from the Gene Expression Omnibus. We confined analyses to four different Affymetrix expression platforms (Affymetrix Human Genome U133A Array, Affymetrix Human Genome U133 Plus 2.0 Array, Affymetrix Mouse Genome 430 2.0 Array and Affymetrix Rat Genome 230 2.0 Array). For each of these platforms we downloaded the raw CEL files (20 108, 43 278, 18 639 and 6 124 arrays, respectively), and used RMA for normalization. We could run RMA on all samples at once for the 20 108 Human Genome U133A Array, 18 639 Mouse Genome 430 2.0 Array and 6 123 Rat Genome 230 2.0 Array. For the 43,278 Human Genome U133 Plus 2.0 Array samples we ran RMA in eight batches due to its size, by randomly assigning the samples to one of these batches. We subsequently conducted quality control (QC) on the data. We first removed duplicate samples, and subsequently conducted a principal component analysis (PCA) on the sample correlation matrix. The first principal component ($PC_{qc}$) on such a matrix describes nearly always a constant pattern (dominating the data) which explains around 80-90% of the total variance. This pattern can be regarded as probe-specific variance, independent of the biological sample hybridized to the array. The correlation of each individual microarray with this $PC_{qc}$ can be used to detect outliers, as arrays of lesser quality will have a lower correlation with the $PC_{qc}$. We removed samples that had a correlation $R < 0.75$. After QC in total 77 840 different samples remained for downstream analysis (54 736 human samples, 17 081 mouse samples, 6 023 rat samples). Although this QCed dataset can be well used for the aforementioned guilt-by-association co-expression analysis, we reasoned that the presence of profound effects on many genes will make it difficult to identify the more subtle relationships that exist between genes. Therefore we conducted a

PCA on the probe correlation matrix, resulting in the identification of in total 2 206 robustly estimated principal components (377 for Human Genome U133A, 777 for Human Genome U133 Plus 2.0, 677 for Mouse Genome 430 2.0 and 375 for Rat Genome 230 2.0) by requiring a Cronbach's alpha > 0.70 for each individual principal component. Jointly these components explain between 79% and 90% of the variance in the data per Affymetrix expression platform, and many of these are well conserved across the three species. Subsequent Gene Set Enrichment Analysis (GSEA) revealed that each of these 2 206 components are significantly enriched (False discovery rate < 0.05) for at least one GO term, KEGG, BioCarta or Reactome pathway, indicating that these components are describing biologically relevant but often diverse phenomena. While per species the very first components describe profound effects on expression (i.e. many enriched pathways and GO terms), the other components are potentially equally biologically relevant, as each of the components describe certain biological phenomena. We therefore used the individual components and integrated the different platforms and species by collapsing the probe identifiers to human Ensembl genes and used orthology information from Ensembl for the mouse and rat platform, resulting in a harmonized matrix of 19,997 unique Ensembl genes x 2 206 principal components.
We subsequently predicted the most likely Gene Ontology (GO) biological process using the following strategy: We first ascertained each individual GO term and assessed per PC whether the genes that were explicitly annotated with this GO term showed a significant difference from the genes that were not annotated with this GO term using a T-Test. We converted the resulting P-Value into an 'enrichment' Z-Score (to ensure normality). We subsequently investigated *CADM2* and correlated the 2,206 PC eigenvector coefficients of *CADM2* with each GO term by taking the 2 206 'enrichment' Z-Scores as the expression profile for that GO term. A significant positive correlation means *SMIM1* has an expression profile that is comparable to the GO term. We have visualized this method at **www.genenetwork.nl/genenetwork**, click on “Method”). In order to correct for multiple testing, we permuted Ensembl gene identifiers: Using permuted data we redid the ‘enrichment’ Z-score calculation and investigated how strong *CADM2* correlated with permuted pathway. We repeated this analysis 100 times, and observed that a P-Value cut-off of 1.18 x 10-5 corresponded to a false discovery rate of 0.05. This resulted in significant prediction of 58 significant GO Biological Process functions and 2 significant KEGG pathways.

## SUPPLEMENTARY TABLES

### Supplementary Table 1. Details of the participating discovery and replication cohorts

| Cohort | | TMT-A | TMT-B | Semantic fluency | Phonemic fluency | Stroop interference | Stroop card 3 | LDST/DSST |
|---|---|---|---|---|---|---|---|---|
| **Discovery Cohorts** | | | | | | | | |
| AGES | N | - | - | - | - | 2644 | 2790 | 2657 |
| | Age | - | - | - | - | 75.9 | 75.9 | 75.9 |
| | (SD) | | | | | (5.3) | (5.3) | (5.3) |
| | Gender | - | - | - | - | 58.9 | 59.1 | 59 |
| | Test score | - | - | - | - | 41.7 | 70.0 | 29.9 |
| | (SD) | | | | | (21.5) | (24.7) | (10.6) |
| ARIC | N | 438 | 436 | 441 | 9179 | 430 | 441 | 9181 |
| | Age | 72.6 | 72.6 | 72.6 | 57.2 | 72.5 | 72.6 | 57.2 |
| | (SD) | (4.2) | (4.2) | (4.2) | (5.7) | (4.2) | (4.2) | (5.7) |
| | Gender | 59.8 | 59.6 | 59.9 | 53.2 | 60 | 59.9 | 53.2 |
| | Test score | 41.6 | 114.6 | 17.7 | 35.0 | 29.9 | 29.3 | 49.1 |
| | (SD) | (17.6) | (49.6) | | (11.8) | (9.5) | | (11.4) |
| ASPS | N | - | 830 | - | - | 261 | 256 | 263 |
| | Age | - | 65.2 | - | - | 65.1 | 65.0 | 65.1 |
| | (SD) | | (8.0) | | | (7.5) | (7.5) | (7.6) |
| | Gender | - | 56.9 | - | - | 57.1 | 57 | 56.4 |
| | Test score | - | 128.0 | - | - | 47.2 | 45.1 (14.4) | 26.6 |
| | (SD) | | (60.2) | | | (22.7) | | (6.9) |
| CHS | N | 1249 | 1249 | - | - | - | - | 2022 |
| | Age | 79.4 (3.8) | 79.4 | - | - | - | - | 71.5 |
| | (SD) | | (3.8) | | | | | (4.7) |
| | Gender | 61 | 61 | - | - | - | - | 61.9 |
| | Test score | 52.3 | 141.5 | - | - | - | - | 41.4 |
| | (SD) | (22.0) | (62.5) | | | | | (11.6) |
| ERF | N | 1,267 | 1255 | 1263 | 1255 | 1238 | 1238 | - |
| | Age | 58.7 (8.9) | 58.7 | 58.6 | 58.6 | 58.6 | 58.6 | - |
| | (SD) | | (8.8) | (8.9) | (8.8) | (8.8) | (8.8) | |
| | Gender | 54.8 | 54.7 | 54.8 | 54.5 | 55.5 | 55.5 | - |
| | Test score | 45.9 | 139.5 | 33.7 | 20.3 | 54.2 | 119.5 (41.6) | - |
| | (SD) | (20.8) | (78.0) | (9.5) | (5.7) | (33.0) | | |
| FHS | N | 2475 | 2440 | 1403 | - | - | - | - |
| | Age | 64.7 | 64.6 | 67.8 | - | - | - | - |
| | (SD) | (11.9) | (10.9) | | | | | |
| | Gender | 54.8 | 54.7 | 54 | - | - | - | - |
| | Test score | 35.7 | 92.0 | 18.1 | - | - | - | - |
| | (SD) | (20.2) | (55.6) | | | | | |
| GENOA | N | - | - | - | - | - | - | 758 |
| Rochester | Age | - | - | - | - | - | - | 61.27 |
| | (SD) | | | | | | | (8.84) |
| | Gender | - | - | - | - | - | - | 58.1 |
| | Test score | - | - | - | - | - | - | 50.18 |
| | (SD) | | | | | | | (12.43) |
| Health ABC | N | - | - | - | - | - | - | 1519 |
| | Age | - | - | - | - | - | - | 73.7 |
| | (SD) | | | | | | | (2.8) |
| | Gender | - | - | - | - | - | - | 47 |
| | Test score | - | - | - | - | - | - | 41.2 |
| | (SD) | | | | | | | (12.0) |
| KORCULA | N | - | - | - | 495 | - | - | 542 |
| | Age | - | - | - | 61.28 | - | - | 60.3 |
| | (SD) | | | | (9.3) | | | (10.0) |
| | Gender | - | - | - | 64.9 | - | - | 61 |
| | Test score | - | - | - | 27.52 | - | - | 41.3 |
| | (SD) | | | | (9.7) | | | (16.4) |
| LBC1921 | N | - | - | - | 468 | - | - | 257 |
| | Age | - | - | - | 79.1 | - | - | 83.3 |
| | (SD) | | | | (0.6) | | | (0.6) |
| | Gender | - | - | - | 59.6 | - | - | 58.6 |
| | Test score | - | - | - | 40.5 | - | - | 41.8 |
| | (SD) | | | | (12.2) | | | (12.9) |
| LBC1936 | N | - | - | - | 951 | - | - | 951 |
| | Age | - | - | - | 69.6 | - | - | 69.5 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | (SD) | | | | (0.8) | | | (0.8) |
| | Gender | - | - | - | 49.4 | - | - | 49 |
| | Test score | - | - | - | 42.4 | - | - | 56.9 |
| | (SD) | | | | (12.5) | | | (12.8) |
| ORCADES | N | - | | - | 418 | - | - | 417 |
| | Age | - | - | - | 62.78 | - | - | 62.7 |
| | (SD) | | | | (9.5) | | | (9.5) |
| | Gender | - | - | - | 58.6 | - | - | 58 |
| | Test score | - | - | - | 37.3 | - | - | 26.7 |
| | (SD) | | | | (12.5) | | | (7.1) |
| RUSH | N | - | - | - | - | - | - | 732 |
| MAP | Age | - | - | - | - | - | - | 80.5 |
| | (SD) | | | | | | | (6.5) |
| | Gender | - | - | - | - | - | - | 72.3 |
| | Test score | - | - | - | - | - | - | 37.4 |
| | (SD) | | | | | | | (10.4) |
| RUSH | N | - | - | - | - | - | - | 685 |
| ROS | Age | - | - | - | - | - | - | 75.1 |
| | (SD) | | | | | | | (7.1) |
| | Gender | - | - | - | - | - | - | 65.4 |
| | Test score | - | - | - | | - | - | 41.3 |
| | (SD) | | | | | | | (10.3) |
| RS | N | - | - | 3276 | - | 3204 | 3271 | 3250 |
| | Age | - | - | 65.6 | - | 64.4 | 65.5 | 65.4 |
| | (SD) | | | (6.7) | | (6.7) | (6.7) | (6.7) |
| | Gender | - | - | 57.5 | - | 57.9 | 57.7 | 57.6 |
| | Test score | - | - | 20.9 | - | 32.1 | 57.2 | 26.7 |
| | (SD) | | | (5.5) | | (18.9) | | (7.1) |
| PROSPER | N | - | - | - | - | - | 4870 | 4370 |
| | Age | - | - | - | - | - | 75.3 | 75.2 |
| | (SD) | | | | | | (3.3) | |
| | Gender | - | - | - | - | - | 53 | 53 |
| | Test score | - | - | - | - | - | 66.7 (27.3) | 23.2 |
| | (SD) | | | | | | | (7.8) |
| RS2 | N | - | - | - | - | - | - | 1821 |
| | Age | - | - | - | - | - | - | 64.1 |
| | (SD) | | | | | | | (7.6) |
| | Gender | - | - | - | - | - | - | 55 |
| | Test score | - | - | - | - | - | - | 27.8 |
| | (SD) | | | | | | | (6.7) |
| RS3 | N | - | - | - | - | - | - | 2030 |
| | Age | - | - | - | - | - | - | 56.0 |
| | (SD) | | | | | | | (5.7) |
| | Gender | - | - | - | - | - | - | 56 |
| | Test score | - | - | | | | - | 32.4 |
| | (SD) | | | | | | | (6.6) |
| SPLIT | N | - | - | - | 306 | - | - | 315 |
| | Age | - | - | - | 58.33 | - | - | 58.2 |
| | (SD) | | | | (8.17) | | | (8.3) |
| | Gender | - | - | - | 61.1 | - | - | 58.8 |
| | Test score | - | - | - | 24.37 | - | - | 27.6 |
| | (SD) | | | | (8.83) | | | (8.4) |
| VIS | N | - | - | - | 382 | - | - | 311 |
| | Age | - | - | - | 64.58 | - | - | 60.6 |
| | (SD) | | | | (10.62) | | | (9.2) |
| | Gender | - | - | - | 62.9 | - | - | 63 |
| | Test score | - | - | - | 20.16 | - | - | 45.2 |
| | (SD) | | | | (9.07) | | | (18.0) |
| | Total N | 5,429 | 6210 | 6383 | 13454 | 7777 | 12866 | 32080 |

| Cohort | | TMT-A | TMT-B | Semantic fluency | Phonemic fluency | Stroop interference | Stroop card 3 | LDST/DSST |
|---|---|---|---|---|---|---|---|---|
| **Replication cohorts** | | | | | | | | |
| AAA | N | - | 1910 | - | 1895 | - | - | - |
| | Age | - | 66.2 | - | 66.2 | - | - | - |
| | (SD) | | (6.6) | | (6.6) | | | |
| | Gender | - | 73 | - | 73 | - | - | - |
| | Test score | - | 107.3 | - | 37.6 | - | - | - |
| | (SD) | | (46.6) | | (12.4) | | | |
| ARIC black | N | 413 | 373 | 424 | 2459 | 387 | - | - |
| | Age | 71.7 | 71.5 | 71.8 | 56.0 | 71.7 | - | - |
| | (SD) | (4.4) | (4.3) | (4.4) | (5.7) | (4.4) | | |
| | Gender | 64.9 | 63.5 | 64.2 | 64.7 | 65.4 | - | - |
| | Test score | 80.5 | 188.8 | 14.3 | 28.5 | 27.4 | - | - |
| | (SD) | (52.1) | (61.1) | (4.4) | (12.9) | (11.4) | | |
| BLSA | N | 422 | 422 | 422 | 422 | - | - | 438 |
| | Age | 73.1 | 73.1 | 73.1 | 73.1 | - | - | 71.2 |
| | (SD) | (7.9) | (7.9) | (7.9) | (7.9) | | | (11.5) |
| | Gender | 36 | 36 | 36 | 36 | - | - | 48 |
| | Test score | 38.3 | 94.1 | 18.2 | 44.0 | - | - | 47.0 |
| | (SD) | (17.5) | (46.0) | (4.8) | (12.7) | | | (11.9) |
| 3CS | N | 5540 | 5229 | 5643 | - | - | - | - |
| | Age | 73.99 | 73.90 | 74.03 | - | - | - | - |
| | (SD) | (5.35) | (5.30) | (5.39) | | | | |
| | Gender | 60.8 | 60.7 | 61.0 | - | - | - | - |
| | Test score | 56.4 | 111.3 | 13.0 | - | - | - | - |
| | (SD) | (23.9) | (48.0) | (3.6) | | | | |
| GENOA | N | 704 | 631 | 716 | 687 | 648 | 648 | - |
| Jackson | Age | 63.29 | 63.29 | 63.29 | 63.29 | 63.29 | 63.29 (8.22) | - |
| | (SD) | (8.22) | (8.22) | (8.22) | (8.22) | (8.22) | | |
| | Gender | 72.6 | 72.6 | 72.6 | 72.6 | 72.6 | 72.6 | - |
| | Test score | 58.56 | 162.39 | 14.92 | 28.63 | 33.25 | 22.28 | - |
| | (SD) | (1.65) | (1.70) | (4.47) | (11.71) | (11.83) | (10.10) | |
| GENOA | N | 757 | 673 | 762 | 760 | 740 | 740 | - |
| Rochester | Age | 61.27 | 61.27 | 61.27 | 61.27 | 61.27 | 61.27 (8.84) | - |
| | (SD) | (8.84) | (8.84) | (8.84) | (8.84) | (8.84) | | |
| | Gender | 58.1 | 58.1 | 58.1 | 58.1 | 58.1 | 58.1 | - |
| | Test score | 30.26 | 74.44 | 19.25 | 32.29 | 32.81 | 34.54 (9.30) | - |
| | (SD) | (1.46) | (1.51) | (4.86) | (13.63) | (9.26) | | |
| GS:SFHS | N | - | - | - | 3386 | - | - | - |
| | Age | - | - | - | 60.3 | - | - | - |
| | (SD) | | | | (8.5) | | | |
| | Gender | - | - | - | 58.7 | - | - | - |
| | Test score | - | - | - | 40.5 | - | - | - |
| | (SD) | | | | (11.9) | | | |
| HBCS | N | - | - | 889 | - | - | - | - |
| | Age | - | - | 75.1 | - | - | - | - |
| | (SD) | | | (7.1) | | | | |
| | Gender | - | - | 65 | - | - | - | - |
| | Test score | - | - | 24.3 (6.0) | - | - | - | - |
| | (SD) | | | | | | | |
| Hunter | N | - | - | 855 | 855 | - | - | - |
| | Age | - | - | 64 | 64 | - | - | - |
| | (SD) | | | | | | | |
| | Gender | - | - | 45.5 | 45.5 | - | - | - |
| | Test score | - | - | 13.3 | 10.0 | - | - | - |
| | (SD) | | | (3.4) | (3.8) | | | |
| FHS | N | - | - | - | 1421 | - | - | - |
| | Age | - | - | - | 67.9 | - | - | - |
| | (SD) | | | | | | | |
| | Gender | - | - | - | 54 | - | - | - |
| | Test score | - | - | - | 37.7 | - | - | - |
| | (SD) | | | | | | | |
| InCHIANTI | N | 845 | 758 | - | - | - | - | - |
| | Age | 74.1 | 73.1 | - | - | - | - | - |
| | (SD) | (7.0) | (6.2) | | | | | |
| | Gender | 55.9 | 54.5 | - | - | - | - | - |
| | Test score | 90.0 | 179.4 | - | - | - | - | - |
| | (SD) | (1.9) | (1.6) | | | | | |
| MAS | N | 872 | 821 | 884 | 883 | - | - | 873 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Age | 78.6 | 78.6 | 78.7 | 78.7 | - | - | 78.7 |
| | (SD) | (4.8) | (4.8) | (4.4) | (4.8) | | | (4.8) |
| | Gender | 55.8 | 56.5 | 55.9 | 55.8 | - | - | 55.9 |
| | Test score | 46.1 | 119.2 | 15.8 | 36.9 | - | - | 48.3 |
| | (SD) | (16.1) | (51.7) | (4.4) | (12.6) | | | (12.2) |
| RUSH | N | - | - | 751 | - | - | 704 | - |
| MAP | Age | - | - | 80.7 | - | - | 80.9 | - |
| | (SD) | | | (6.5) | | | | |
| | Gender | - | - | 72.8 | - | - | 73.7 | - |
| | Test score | - | - | 16.4 | - | - | 18.0 | - |
| | (SD) | | | (5.2) | | | (7.6) | |
| RUSH | N | - | - | 686 | - | - | - | - |
| ROS | Age | - | - | 75.1 | - | - | - | - |
| | (SD) | | | (7.1) | | | | |
| | Gender | - | - | 65.5 | - | - | - | - |
| | Test score | - | - | 17.8 | | - | - | - |
| | (SD) | | | (5.2) | | | | |
| SHIP-TREND | N | - | - | - | - | 558 | 558 | - |
| | Age | - | - | - | - | 58.0 | 58.0 | - |
| | (SD) | | | | | (8.7) | (8.7) | |
| | Gender | - | - | - | - | 56.2 | 56.2 | - |
| | Test score | - | - | - | - | 42.8 | 18.5 | - |
| | (SD) | | | | | (7.6) | (6.4) | |
| PROSPER | N | - | - | - | - | 4870 | - | - |
| | Age | - | - | - | - | 75.3 | - | - |
| | (SD) | | | | | (3.3) | | |
| | Gender | - | - | - | - | 53 | - | - |
| | Test score | - | - | - | - | 38.9 | | - |
| | (SD) | | | | | (23.3) | | |
| RS2 | N | - | - | 1842 | - | 1805 | 1805 | - |
| | Age | - | - | 64.2 | - | 64.2 | 64.2 | - |
| | (SD) | | | (7.8) | | (7.8) | (7.8) | |
| | Gender | - | - | 55 | - | 55 | 55 | - |
| | Test score | - | - | 22.2 | - | 26.8 | 49.9 | - |
| | (SD) | | | (5.2) | | (14.3) | (16.4) | |
| RS3 | N | - | - | 2045 | - | 1927 | 1927 | - |
| | Age | - | - | 56.0 | - | 56.0 | 56.0 | - |
| | (SD) | | | (5.8) | | (5.6) | (5.6) | |
| | Gender | - | - | 56 | - | 57 | 57 | - |
| | Test score | - | - | 24.4 | - | 20.3 | 42.5 (13.6) | - |
| | (SD) | | | (6.0) | | (10.2) | | |
| NHS | N | - | - | 2067 | - | - | - | - |
| | Age | - | - | 74.1 | - | - | - | - |
| | (SD) | | | (2.2) | | | | |
| | Gender | - | - | 100 | - | - | - | - |
| | Test score | - | - | 17.4 | - | - | - | - |
| | (SD) | | | (4.7) | | | | |
| WGHS | N | - | - | 3907 | - | - | - | - |
| | Age | - | - | | - | - | - | - |
| | (SD) | | | | | | | |
| | Gender | - | - | 100 | - | - | - | - |
| | Test score | - | - | 17.7 | - | - | - | - |
| | (SD) | | | (4.9) | | | | |
| Total N | Replication | 9 553 | 10 817 | 15 826 | 26 222 | 7 203 | 2 650 | 1 311 |
| D+R | Total | 14 982 | 17,027 | 22 209 | 39 676 | 14 980 | 15 516 | 33 391 |

## Supplementary Table 2. Genotyping parameters

| Study | Genotyping platforms - SNP panel | Gentoyping center | Genotype calling |
|---|---|---|---|
| **Discovery Cohorts** | | | |
| AGES | Illumina HumanCNV370 Duo BeadChip® | NIA, NIH, USA | Illumina Bead Studio |
| ARIC-whites | Affymetrix GeneChip SNP Array 6.0® | Broad Institute, USA | Birdseed |
| ASPS | Illumina Human610-Quad BeadChip® | Erasmus MC, Rotterdam, NL | Illumina |
| CHS | Illumina HumanCNV370 Duo BeadChip® | Genotyping Laboratory at Cedars- | Illumina Bead Studio |
| Croatia- | Illumina HumanHap370-Duo® and HumanHap370-Quad | Helmholtz Centre, Munich, D | Illumina Bead Studio |
| Croatia-Split | Illumina HumanHap370-Quad BeadChip® | AROS Applied Biotechnology, Aarhus, | Illumina Bead Studio |
| Croatia-Vis | Illumina Hap300-V1 | | Illumina Bead Studio |
| ERF | Illumina HumanHap 300K array®, Illumina HumanHap | Leiden University Medical Center, | Illumina BeadStudio& |
| FHS | Affymetrix GeneChip Human Mapping 500K Array® | Affymetrix (Santa Clara), USA | Affymetrix BRLMM |
| GENOA- | Affymetrix GeneChip SNP Array 6.0® | Mayo Clinic, Rochester (MN), USA | Birdseed |
| Health ABC | Illumina Human 1M-duo | Center for Inherited Disease Research | Illumina Bead Studio |
| LBC1921 | Ilumina 610-Quadv1 | Wellcome Trust Clinical Research | Illumina Bead Studio |
| LBC1936 | Ilumina 610-Quadv1 | Wellcome Trust Clinical Research | Illumina Bead Studio |
| MAP | Affymetrix Genechip 6.0® | Broad Institute; Translational | Birdsuite, Broad |
| ORCADES | Illumina HumanCNV370-Duo and –HumanHap300K | Helmholtz Centre, Munich, D and | Illumina Bead Studio |
| ROS | Affymetrix Genechip 6.0® | Broad Institute; Translational | Illumina GenCall |
| PROSPER | Illumina 660K BeadChip | Erasmus MC, Rotterdam, NL | Illumina Bead Studio |
| Rotterdam | Illumina HumanHap550-Duo BeadChip® | Erasmus MC, Rotterdam, NL | Birdsuite, Broad |
| Rotterdam | Illumina HumanHap550 Duo BeadChip® and Illumina | Erasmus MC, Rotterdam, NL | Illumina Bead Studio |
| Rotterdam | Illumina Human 610 Quad BeadChip® | Erasmus MC, Rotterdam, NL | Illumina Genome |
| **Replication cohorts** | | | |
| ARIC-blacks | Affymetrix GeneChip SNP Array 6.0® | Broad Institute, USA | Birdseed |
| BLSA | Illumina 550K ® | Laboratory of Neurogenetics, NIA | Illumina Bead Studio |
| CHS-blacks | Illumina HumanCNV370 Duo BeadChip® | Genotyping Laboratory at Cedars- | Illumina Bead Studio |
| GENOA- | Affymetrix GeneChip SNP Array 6.0® and Illumina | Mayo Clinic, Rochester (MN), USA | Birdseed |
| HBCS | modified Illumina Infinum 610K Quad chip® | Wellcome Trust Sanger Institute, | Illumina Bead Studio |
| HCS | Illumina Human610-Quad BeadChip® | Hunter Medical Research Institute, | Genomestudio |
| InCHIANTI | Illumina 550K | | Illumina Bead Studio |
| NHS | Affymetrix Genechip 6.0®, Illumina Infinium Sentrix | Broad Institute, USA | Birdseed |
| SHIP-TREND | Affymetrix GeneChip SNP Array 6.0® | Affymetrix (Santa Clara), USA | Birdseed2 |
| MAS | Ramaciotti Centre, UNSW | Diamantina Institute and Institute of | Illumina GenCall |
| WGHS | HumanHap300 Duo ‘’+’’ chips or HumanHuman300 Duo | Amgen | Illumina Bead Studio |
| 3CS | Illumina Human 610-Quad BeadChip | Centre National de Génotypage | Illumina BeadStudio |

## Supplementary Table 3: Quality control filters before imputation and methods for assessing population structure

| Study | Sample call rate | SNP call rate | MAF | HWE p-value | Assessment of Population Stratification |
|---|---|---|---|---|---|
| **Discovery Cohorts** | | | | | |
| AGES | < 97% | < 98% | < 0.01 | < 10-6 | EIGENSTRAT |
| ARIC-whites | < 95% | < 95% | < 0.01 | < 10-5 | EIGENSTRAT |
| ASPS | < 98% | < 98% | < 0.01 | < 10-6 | IBD matrix |
| CHS | ≤ 95% | < 97% | < 0.01 | < 10-5 | PCA |
| Croatia-Korcula | < 97% | < 98% | < 0.01 | < 10-6 | PCA |
| Croatia-Split | < 97% | < 98% | < 0.01 | < 10-6 | PCA |
| Croatia-Vis | < 97% | < 98% | < 0.01 | < 10-6 | PCA |
| ERF | < 95% | < 98% | <0.005 | < 10-6 | n.a. |
| FHS | < 97% | < 97% | < 0.01 | < 10-6 | EIGENSTRAT |
| GENOA-whites | < 95% | < 95% | < 0.01 | n.a. | PCA |
| Health ABC | <99% | <97% | < 0.01 | <10 -7 | |
| LBC1921 | < 95% | < 98% | < 0.01 | < 10-3 | MDS |
| LBC1936 | < 95% | < 98% | < 0.01 | < 10-3 | MDS |
| MAP | < 95% | < 95% | < 0.01 | < 10-6 | EIGENSTRAT |
| ORCADES | < 97% | < 98% | < 0.01 | < 10-6 | MDS |
| PROSPER | <97.5% | <98% | <0.01 | <10-6 | IBD matrix |
| ROS | < 95% | < 95% | < 0.01 | < 10-6 | EIGENSTRAT |
| Rotterdam Study | < 97.5% | < 98% | < 0.01 | < 10-6 | IBD matrix |
| Rotterdam Study-II | < 97.5% | < 98% | < 0.01 | < 10-6 | IBD matrix |
| Rotterdam Study-III | < 97.5% | < 98% | < 0.01 | < 10-6 | IBD matrix |
| **Replication Cohorts** | | | | | |
| ARIC-blacks | < 95% | < 95% | < 0.01 | < 10-5 | EIGENSTRAT |
| BLSA | < 98.5% | < 99% | < 0.01 | < 10-4 | EIGENSTRAT |
| CHS - blacks | < 95% | < 97% | < 0.01 | < 10-5 | PCA |
| GENOA-blacks | < 95% | < 95% | < 0.01 | n.a. | PCA |
| HBCS | < 95% | < 95% | < 0.01 | < 10-5 | MDS |
| HCS | < 95% | < 95% | < 0.01 | < 10-6 | EIGENSTRAT |
| InCHIANTI | <97% | <99% | <0.01 | < 10-4 | |
| NHS | < 95% | < 98% | n.a. | < 10-4 | EIGENSTRAT |
| SHIP-TREND | < 92% | < 92% | n.a. | n.a. | PCA and MDS |
| MAS | < 95% | < 95% | < 0.01 | < 10-6 | EIGENSTRAT |
| WGHS | < 98% | < 90% | < 0.01 | < 10-6 | EIGENSTRAT |
| 3CS | < 95% | < 98% | < 0.01 | < 10-6 | EIGENSTRAT |

PCA: Principal Component Analysis; MDS: Multidimensional Scaling

## Supplementary Table 4: Imputation algorithms

| Study | Imputation software | Imputation reference panel |
|---|---|---|
| **Discovery Cohorts** | | |
| AGES | MACH | HapMap II CEU (build 36, release 22) |
| ARIC-whites | MACH (v1.0.16) | HapMap II CEU (build 36) |
| ASPS | MACH (v1.0.15) | HapMap II CEU (build 36, release 22) |
| CHS | BIM-BAM15 | HapMap II CEU (build 36) |
| Croatia- Korcula | MACH | HapMap II CEU (build 36) |
| Croatia-Split | MACH | HapMap II CEU (build 36) |
| Croatia-Vis | MACH | HapMap II CEU (build 36) |
| ERF | MACH | HapMap II CEU (build 36) |
| FHS | MACH | HapMap II CEU (build 36) |
| GENOA-whites | MACH | HapMap II, 60 unrelated CEU, build 36, release 22 |
| Health ABC | MACH | HapMap II CEU (build 36) |
| LBC1921 | MACH | HapMap II CEU (build 36, release 22) |
| LBC1936 | MACH | HapMap II CEU (build 36, release 22) |
| MAP | MACH (version 1.0.16a) | HapMap II CEU (build 36, release 22) |
| ORCADES | MACH | HapMap II CEU (build 36, release 22) |
| PROSPER | MACH | Hapmap II CEU (build 36, release 22) |
| ROS | MACH (version 1.0.16a) | HapMap II CEU (build 36, release 22) |
| Rotterdam Study | MACH (v1.0.15) | HapMap II CEU (build 36, release 22) |
| Rotterdam Study-II | MACH (v1.0.16) | HapMap II CEU (build 36, release 22) |
| Rotterdam Study-III | MACH (v1.0.16) | HapMap II CEU (build 36, release 22) |
| **Replication Cohorts** | | |
| ARIC-blacks | MACH (v1.0.16) | HapMap II CEU and YRI (build 35, release 21) |
| BLSA | MACH | HapMap II CEU (build 36) |
| CHS-blacks | | |
| GENOA-blacks | MACH | HapMap II, 60 unrelated YRI and 60 unrelated CEU, build 36, release 22 |
| HBCS | MACH | HapMap II CEU |
| HCS | MACH (version 1.0.16) | HapMap II CEU (build 36.1, release 24) |
| InCHIANTI | MACH | HapMap II CEU |
| NHS | MACH | HapMap II CEU (build 36, release 22) |
| SHIP-TREND | IMPUTE (v1.2.1.3) | HapMap II |
| MAS | MACH / minimac | HapMap II CEU (build 36, release 22) |
| WGHS | MACH (v1.0.15) | HapMap II CEU (build 36, release 22) |
| 3CS | IMPUTE (v2.2) | HapMap II CEU (build 36.3, release 22) |

**Supplementary Table 5: Expression and functional data for *CADM2***

| Neuronal_tissues_expressed | AUC | P_value | KEGG (<0.01) | P_value |
|---|---|---|---|---|
| Prefrontal Cortex | 1.00 | 1.29E-31 | Long-term depression | 5.38E-06 |
| Frontal Lobe | 0.99 | 5.57E-41 | Neuroactive ligand-receptor interaction | 8.37E-05 |
| Cerebral Cortex | 0.99 | 2.47E-171 | Long-term potentiation | 0.004329 |
| Hippocampus | 0.99 | 8.26E-36 | | |
| Parietal Lobe | 0.98 | 5.22E-12 | | |
| Temporal Lobe | 0.98 | 4.96E-57 | | |
| Entorhinal Cortex | 0.98 | 4.20E-52 | | |
| Substantia Nigra | 0.98 | 6.04E-15 | | |
| Mesencephalon | 0.98 | 3.04E-26 | | |
| Cerebellum | 0.98 | 3.30E-23 | | |
| Thalamus | 0.98 | 3.85E-11 | | |
| Putamen | 0.97 | 4.87E-11 | | |
| Occipital Lobe | 0.97 | 2.51E-26 | | |
| Visual Cortex | 0.97 | 2.21E-21 | | |
| Subthalamic Nucleus | 0.97 | 1.96E-08 | | |
| Cerebrum | 0.96 | 3.89E-194 | | |
| Hypothalamus | 0.94 | 3.47E-09 | | |
| Ganglia | 0.93 | 6.52E-07 | | |
| Spinal Cord | 0.93 | 1.28E-10 | | |
| Motor Neurons | 0.92 | 3.51E-07 | | |
| Neural Stem Cells | 0.88 | 1.39E-05 | | |
| Neurons | 0.71 | 8.37E-06 | | |
| Eye | 0.48 | 0.393603736 | | |
| Dendritic Cells | 0.47 | 0.056655426 | | |
| Astrocytes | 0.45 | 0.517064871 | | |

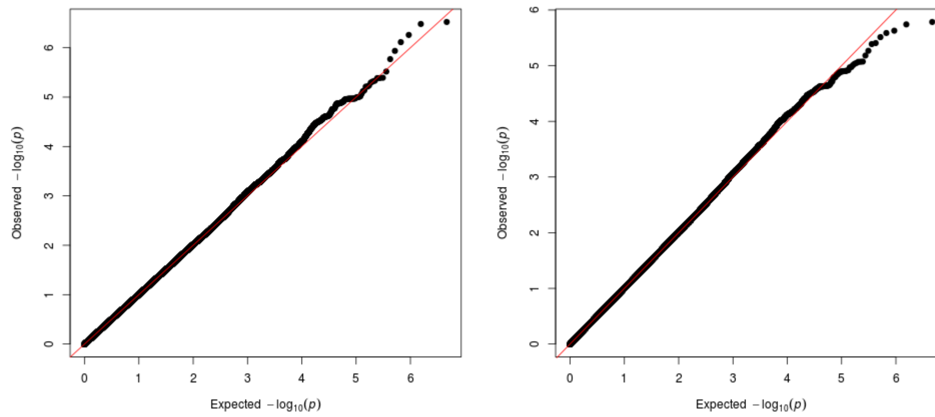
## Supplementary Table 5: (continued) Predicted biological process of *CADM2*

| GO Biological Process (<10^-6) | P-value |
|---|---|
| glutamate signaling pathway | 7.22E-15 |
| neuron cell-cell adhesion | 1.48E-13 |
| startle response | 1.36E-11 |
| gamma-aminobutyric acid transport | 6.04E-11 |
| eye photoreceptor cell development | 8.83E-10 |
| gamma-aminobutyric acid signaling pathway | 1.07E-09 |
| photoreceptor cell development | 1.25E-09 |
| L-amino acid import | 1.54E-09 |
| photoreceptor cell differentiation | 3.53E-09 |
| regulation of membrane potential | 3.82E-09 |
| sensory perception | 4.22E-09 |
| eye photoreceptor cell differentiation | 4.94E-09 |
| neuron-neuron synaptic transmission | 5.29E-09 |
| amino acid import | 7.20E-09 |
| regulation of synaptic transmission, glutamatergic | 7.95E-09 |
| positive regulation of synaptic transmission | 1.07E-08 |
| visual perception | 1.21E-08 |
| sensory perception of light stimulus | 1.60E-08 |
| acidic amino acid transport | 2.08E-08 |
| detection of abiotic stimulus | 2.29E-08 |
| adult behavior | 7.13E-08 |
| positive regulation of transmission of nerve impulse | 7.50E-08 |
| photoreceptor cell maintenance | 1.28E-07 |
| reflex | 1.33E-07 |
| regulation of synaptic transmission | 1.70E-07 |
| negative regulation of synaptic transmission | 1.90E-07 |
| positive regulation of neurological system process | 1.95E-07 |
| phototransduction | 2.84E-07 |
| receptor clustering | 3.37E-07 |
| detection of external stimulus | 5.12E-07 |
| synaptic transmission, glutamatergic | 6.21E-07 |
| regulation of transmission of nerve impulse | 7.06E-07 |
| cognition | 7.08E-07 |
| L-glutamate transport | 9.67E-07 |

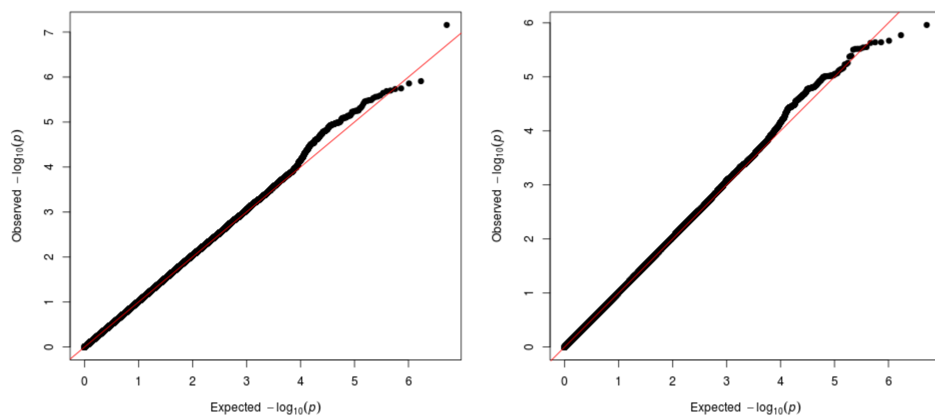
## SUPPLEMENTARY FIGURES

**Supplementary Figure 1. Quantile-Quantile (QQ) plots for all discovery analyses**

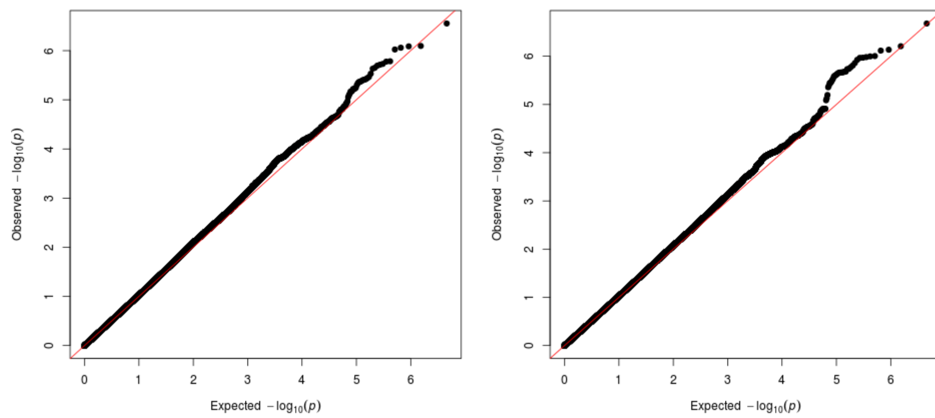
*Trailmaking A*



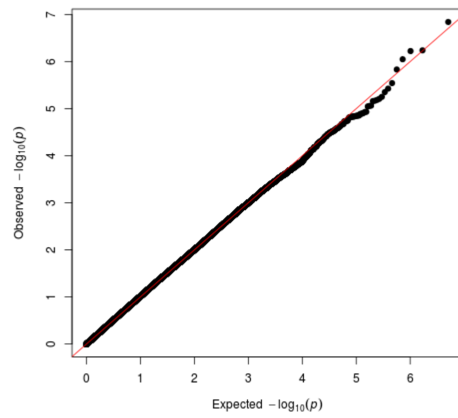

*Trailmaking B*

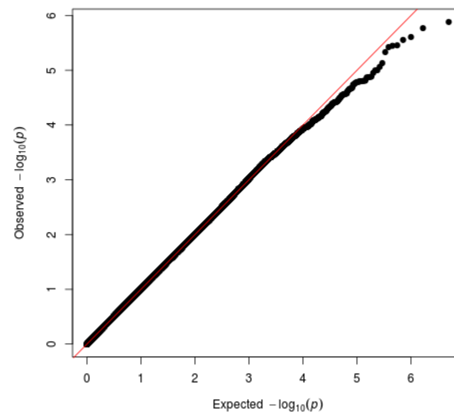



*Semantic fluency*

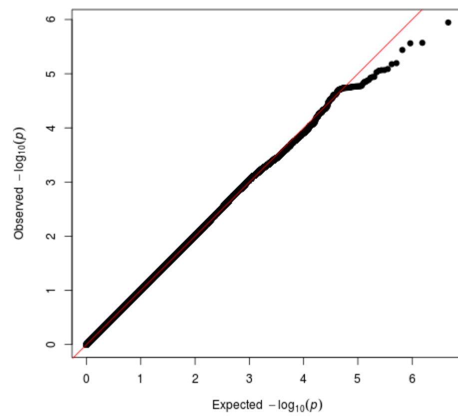


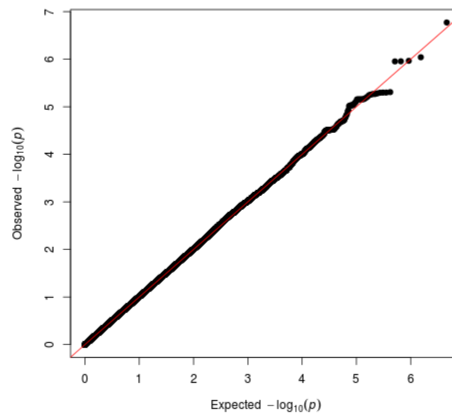
*Phonemic fluency*



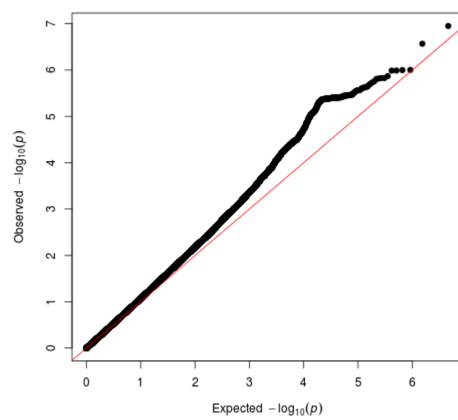



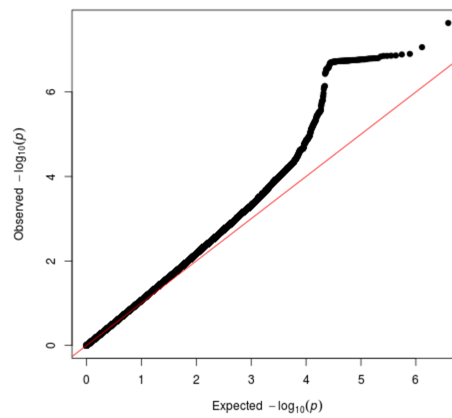

*Stroop*




*LDST/DSST*

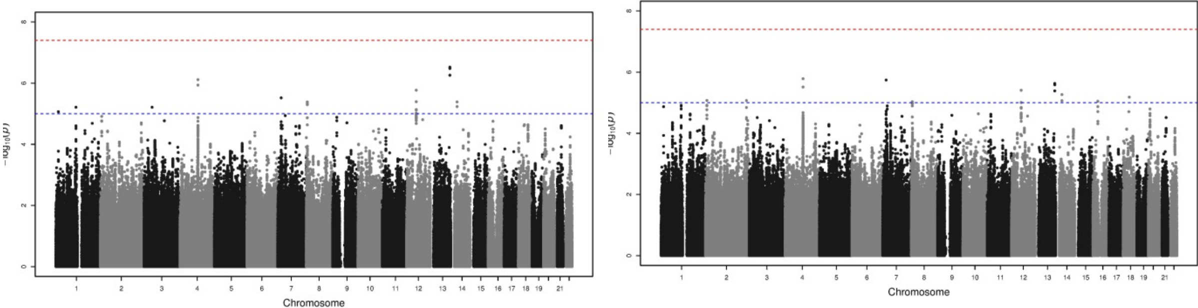


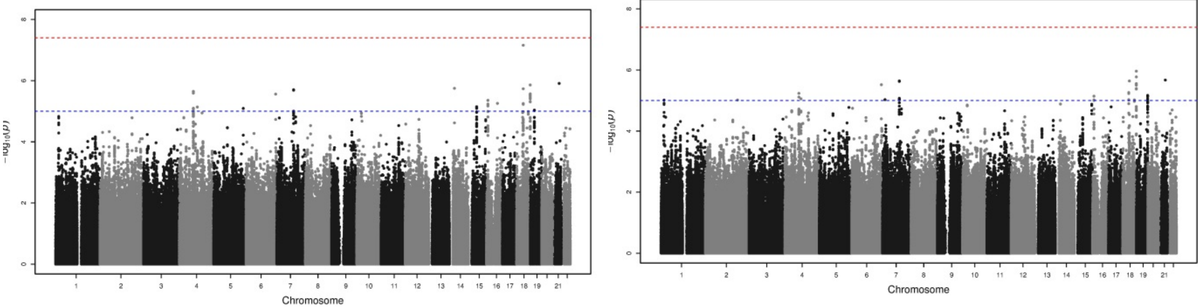

*Left panes: Model adjusted for age and sex. Right panes: adjusted for age, sex, and education*

**Supplementary Figure 2. Manhattan (position by -10log P-value) plots for all discovery analyses**

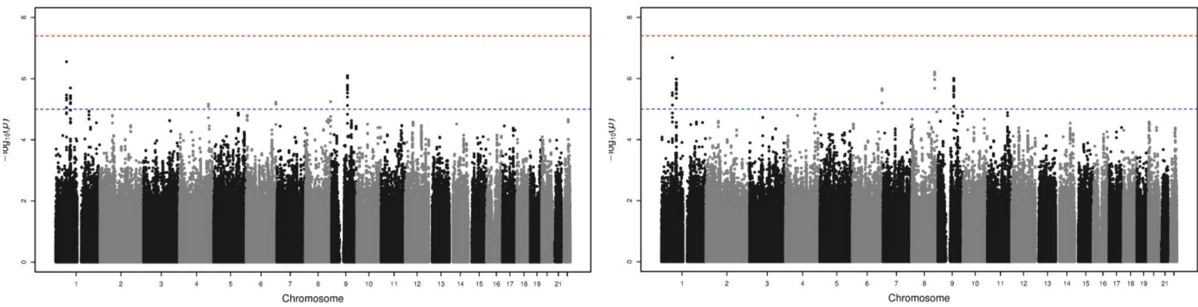
*Trailmaking A*



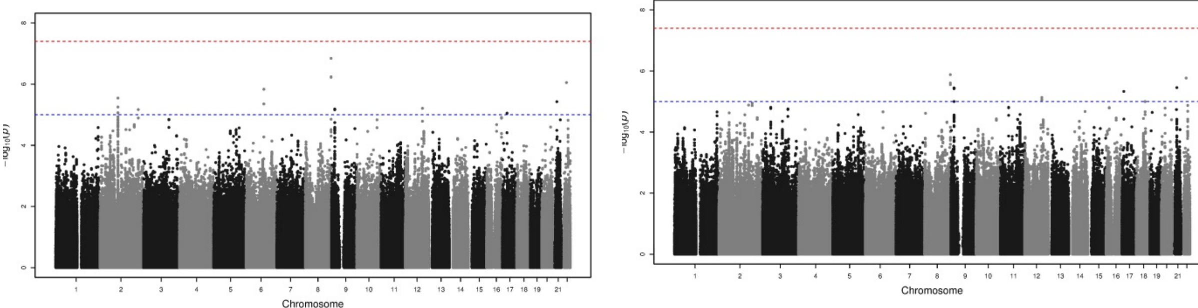

*Trailmaking B*



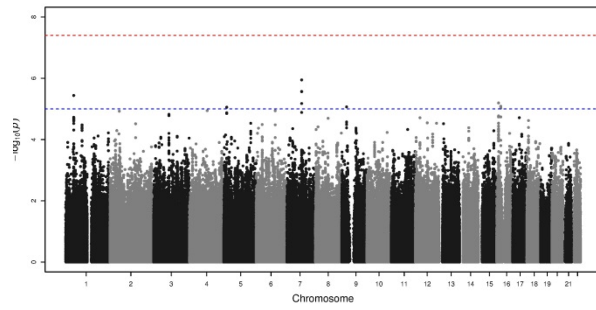

*Semantic fluency*

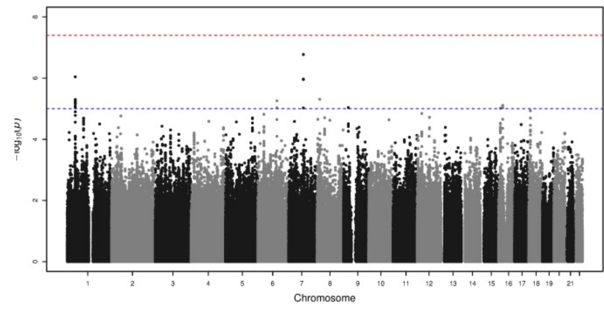



*Phonemic fluency*

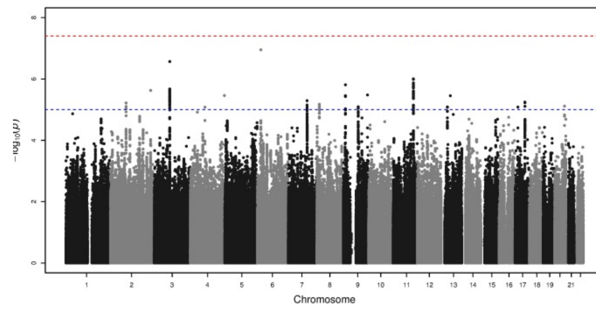


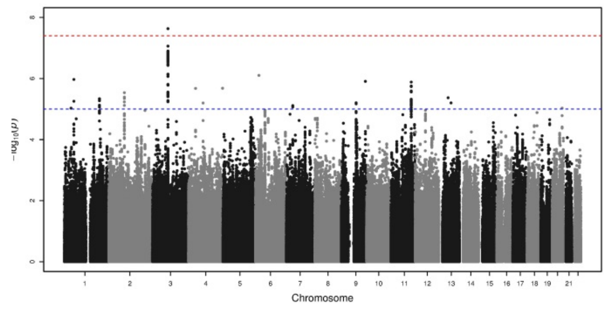
*Stroop*




*LDST/DSST*



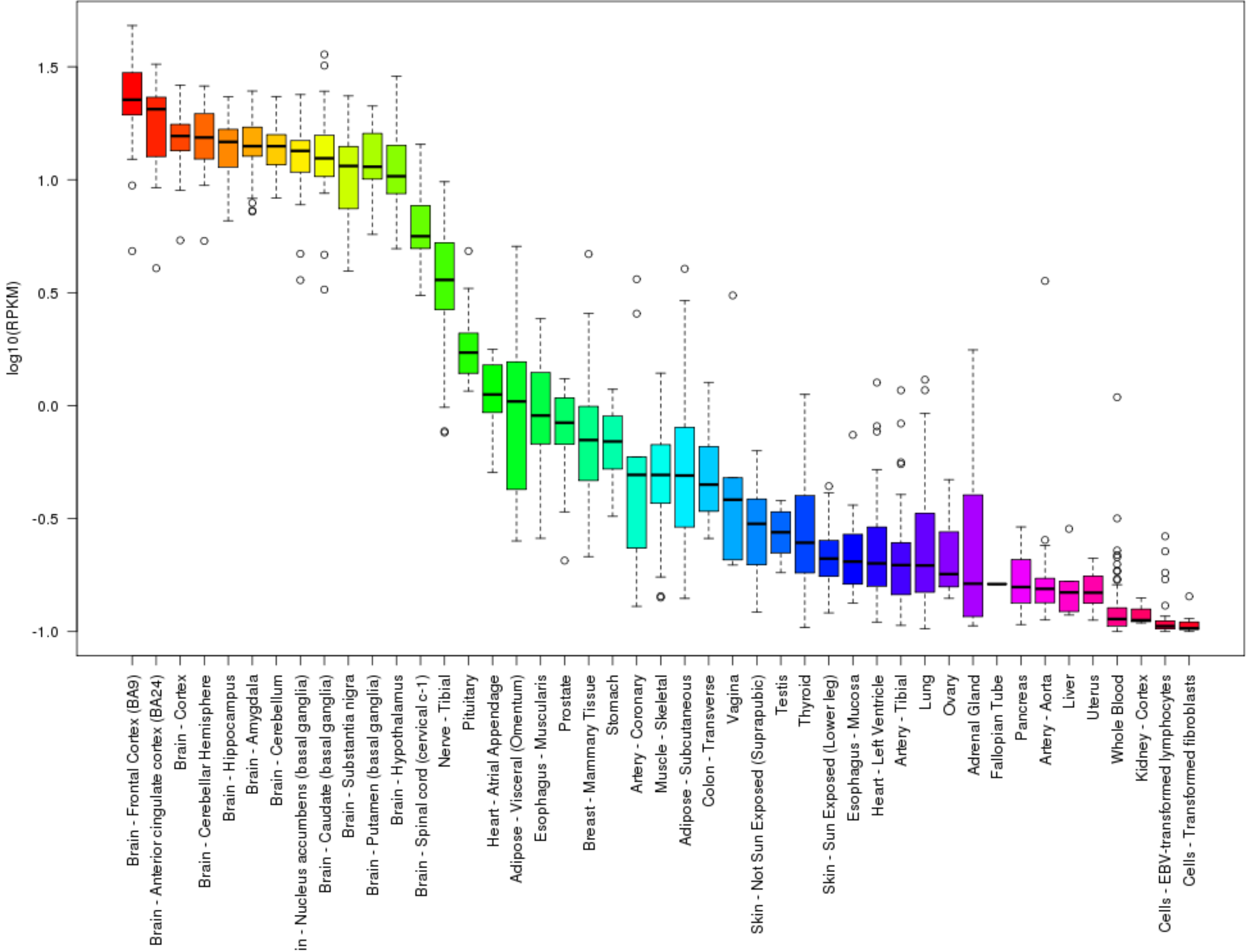




*Left panes: Model adjusted for age and sex. Right panes: adjusted for age, sex, and education*

**Supplementary Figure 3. Tissue expression data for CADM2 from GTex**

Source: http://www.broadinstitute.org/gtex/searchGenes[1]

*X-axis: tissues assessed. Y-axis: $^{10}$log n of reads*



**REFERENCES**


1. Consortium, G.T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-5 (2013).
2. Psaty, B.M. et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from five cohorts. *Circulation:Cardiovascular Genetics* **2**, 73-80 (2009).
3. Harris, T.B. et al. Age, Gene/Environment Susceptibility-Reykjavik Study: multidisciplinary applied phenomics. *Am J Epidemiol* **165**, 1076-87 (2007).
4. Palm, W.M. et al. Ventricular dilation: association with gait and cognition. *Ann Neurol* **66**, 485-93 (2009).
5. Stroop, J. Studies of interference in serial verbal reactions. *J Exp Psychol* **18**, 643–662 (1935).
6. Wechsler, D. *The measurement and appraisal of adult intelligence. Manual for the Wechsler Adult Intelligence Scale.* , (New York: Psychological Corporation, 1955).
7. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol* **129**, 687-702 (1989).
8. Rosamond, W.D. et al. Stroke incidence and survival among middle-aged adults: 9-year follow-up of the Atherosclerosis Risk in Communities (ARIC) cohort. *Stroke* **30**, 736-43 (1999).
9. Wechsler, D. *WAIS-R manual.* , (Cleveland, OH: The Psychologic Corporation, 1981).
10. Benton, A. *Multilingual aphasia examination. 2nd ed.*, (Iowa City, IA: AJA Associates, 1989).
11. Tranel, D. Neuropsychological assessment. *Psychiatr Clin North Am* **15**, 283-99 (1992).
12. Knopman, D.S., Mosley, T.H., Catellier, D.J. & Coker, L.H. Fourteen-year longitudinal study of vascular risk factors, APOE genotype, and cognition: the ARIC MRI Study. *Alzheimers Dement* **5**, 207-14 (2009).
13. Lesage, S.R. et al. Retinal microvascular abnormalities and cognitive decline: the ARIC 14-year follow-up study. *Neurology* **73**, 862-8 (2009).
14. *Army Individual Test Battery Manual of directions and scoring*, (Washington, DC: United States War Department, Adjutant General's Office, 1944).
15. Schmidt, R., Fazekas, F., Kapeller, P., Schmidt, H. & Hartung, H.P. MRI white matter hyperintensities: three-year follow-up of the Austrian Stroke Prevention Study. *Neurology* **53**, 132-9 (1999).
16. Schmidt, R. et al. Assessment of cerebrovascular risk profiles in healthy persons: definition of research goals and the Austrian Stroke Prevention Study (ASPS). *Neuroepidemiology* **13**, 308-13 (1994).
17. Schmidt, R. et al. White matter lesion progression, brain atrophy, and cognitive decline: the Austrian stroke prevention study. *Ann Neurol* **58**, 610-6 (2005).

18. Jolles, J., Houx, P., van Boxtel, M. & Ponds, R. *The Maastricht Aging Study: Determinants of cognitive aging*, (Maastricht: Neuropsych Publishers, 1995).
19. Lezak, M. *Neuropsychological assessment*, (New York: Oxford University Press, 1995).
20. Fried, L.P. et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol* **1**, 263-76 (1991).
21. Longstreth, W.T., Jr. et al. Frequency and predictors of stroke death in 5,888 participants in the Cardiovascular Health Study. *Neurology* **56**, 368-75 (2001).
22. Lopez, O.L. et al. Evaluation of dementia in the cardiovascular health cognition study. *Neuroepidemiology* **22**, 1-12 (2003).
23. Lopez, O.L. et al. Neuropsychological characteristics of mild cognitive impairment subgroups. *J Neurol Neurosurg Psychiatry* **77**, 159-65 (2006).
24. Benton, A. Differential behavioral effects in frontal lobe disease. *Neuropsychologia* **6**, 53–60 (1968).
25. Dawber, T.R. & Kannel, W.B. The Framingham study. An epidemiological approach to coronary heart disease. *Circulation* **34**, 553-5 (1966).
26. Feinleib, M., Kannel, W.B., Garrison, R.J., McNamara, P.M. & Castelli, W.P. The Framingham Offspring Study. Design and preliminary data. *Prev Med* **4**, 518-25 (1975).
27. 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**, 1328-35 (2007).
28. Au, R. et al. New norms for a new generation: cognitive performance in the framingham offspring cohort. *Exp Aging Res* **30**, 333-58 (2004).
29. Carandang, R. et al. Trends in incidence, lifetime risk, severity, and 30-day mortality of stroke over the past 50 years. *Jama* **296**, 2939-46 (2006).
30. Seshadri, S. et al. The lifetime risk of stroke: estimates from the Framingham Study. *Stroke* **37**, 345-50 (2006).
31. Pahor, M. & Kritchevsky, S. Research hypotheses on muscle wasting, aging, loss of function and disability. *J Nutr Health Aging* **2**, 97-100 (1998).
32. Yaffe, K. et al. PPAR-gamma Pro12Ala genotype and risk of cognitive decline in elders. *Neurobiol Aging* **29**, 78-83 (2008).
33. Hofman, A. et al. The Rotterdam Study: 2010 objectives and design update. *Eur J Epidemiol* **24**, 553-72 (2009).
34. Hofman, A., Grobbee, D.E., de Jong, P.T. & van den Ouweland, F.A. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. *Eur J Epidemiol* **7**, 403-22 (1991).
35. Hofman, A. et al. The Rotterdam Study: 2012 objectives and design update. *Eur J Epidemiol* **26**, 657-86 (2011).
36. Prins, N.D. et al. Cerebral small-vessel disease and decline in information processing speed, executive function and memory. *Brain* **128**, 2034-41 (2005).

37. Welsh, K.A. et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part V. A normative study of the neuropsychological battery. *Neurology* **44**, 609-14 (1994).
38. Bots, M.L. et al. Prevalence of stroke in the general population. The Rotterdam Study. *Stroke* **27**, 1499-501 (1996).
39. Hollander, M. et al. Incidence, risk, and case fatality of first ever stroke in the elderly population. The Rotterdam Study. *J Neurol Neurosurg Psychiatry* **74**, 317-21 (2003).
40. Service, S. et al. Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. *Nat Genet* **38**, 556-60 (2006).
41. Aulchenko, Y.S. et al. Linkage disequilibrium in young genetically isolated Dutch population. *Eur J Hum Genet* **12**, 527-34 (2004).
42. van Koolwijk, L.M. et al. Association of cognitive functioning with retinal nerve fiber layer thickness. *Invest Ophthalmol Vis Sci* **50**, 4576-80 (2009).
43. Luteijn, F. & Van der Ploeg, F. *Groninger Intelligentie Test Handleiding*, (Lisse, The Netherlands: Swets & Zeitlinger, 1983).
44. Deary, I.J. et al. The Lothian Birth Cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond. *BMC Geriatr* **7**, 28 (2007).
45. Deary, I.J., Whiteman, M.C., Starr, J.M., Whalley, L.J. & Fox, H.C. The impact of childhood intelligence on later life: following up the Scottish mental surveys of 1932 and 1947. *J Pers Soc Psychol* **86**, 130-47 (2004).
46. Wechsler, D. *WAIS-IIIUK administration and scoring manual* (London, UK, Psychological Corporation, 1998).
47. Lezak, M. *Neuropsychological testing*, (Oxford, UK, Oxford University Press, 2004).
48. McQuillan, R. et al. Runs of homozygosity in European populations. *Am J Hum Genet* **83**, 359-72 (2008).
49. Wechsler, D. *Wechsler adult intelligence scale–revised*, (New York: The Psychological Corporation, 1981).
50. Rudan, I., Campbell, H. & Rudan, P. Genetic epidemiological studies of eastern Adriatic Island isolates, Croatia: objective and strategies. *Coll Antropol* **23**, 531-46 (1999).
51. Rudan, I. et al. Effects of inbreeding, endogamy, genetic admixture, and outbreeding on human health: a (1001 Dalmatians) study. *Croat Med J* **47**, 601-10 (2006).
52. Campbell, H. et al. Effects of genome-wide heterozygosity on a range of biomedically relevant human quantitative traits. *Hum Mol Genet* **16**, 233-41 (2007).
53. Polasek, O. et al. Genome-wide association study of anthropometric traits in Korcula Island, Croatia. *Croat Med J* **50**, 7-16 (2009).
54. Vitart, V. et al. Heritabilities of ocular biometrical traits in two croatian isolates with extended pedigrees. *Invest Ophthalmol Vis Sci* **51**, 737-43.
55. Vitart, V. et al. SLC2A9 is a newly identified urate transporter influencing serum urate concentration, urate excretion and gout. *Nat Genet* **40**, 437-42 (2008).

56. Chibnik, L.B. et al. CR1 is associated with amyloid plaque burden and age-related cognitive decline. *Ann Neurol* **69**, 560-9.
57. Bennett, D.A. et al. Neuropathology of older persons without cognitive impairment from two community-based studies. *Neurology* **66**, 1837-44 (2006).
58. Wechsler, D. *Wechsler memory-scale-revised (manual)*, (Psychological Corporation, New York, 1987).
59. Morris, J.C. et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology* **39**, 1159-65 (1989).
60. Bennett, D.A. et al. The Rush Memory and Aging Project: study design and baseline characteristics of the study cohort. *Neuroepidemiology* **25**, 163-75 (2005).
61. Shepherd, J. et al. Pravastatin in elderly individuals at risk of vascular disease (PROSPER): a randomised controlled trial. *Lancet* **360**, 1623-30 (2002).
62. Shepherd, J. et al. The design of a prospective study of Pravastatin in the Elderly at Risk (PROSPER). PROSPER Study Group. PROspective Study of Pravastatin in the Elderly at Risk. *Am J Cardiol* **84**, 1192-7 (1999).
63. Investigators, F. Multi-center genetic study of hypertension: The Family Blood Pressure Program (FBPP). *Hypertension* **39**, 3-9 (2002).
64. Price, J.F. et al. Low dose aspirin and cognitive function in middle aged to elderly adults: randomised controlled trial. *BMJ* **337**, a1198 (2008).
65. Lezak, M.D. *Neuropsychological Assessment*, (Oxford University Press, USA, 1995).
66. Spreen, O.S., E. *A compendium of neuropsychological tests: administration, norms, and commentary*, (Oxford University Press, 1998).
67. Smith, B.H. et al. Cohort profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int J Epidemiol* (2012).
68. Shock, N. et al. *Normal human aging: The Baltimore Longitudinal Study of Aging*, (Washington (DC): U.S. Government Printing Office, November 1984 Report No. NIH Publ No. (NIH)84–2450).
69. 3C-Study-Group. Vascular factors and risk of dementia: design of the Three-City Study and baseline characteristics of the study population. *Neuroepidemiology* **22**, 316-25 (2003).
70. Auriacombe, S. et al. Validity of the free and cued selective reminding test in predicting dementia: the 3C study. *Neurology* **74**, 1760-7 (2010).
71. Lambert, J.C. et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet* **41**, 1094-9 (2009).
72. Lahti, J. et al. Glucocorticoid receptor gene haplotype predicts increased risk of hospital admission for depressive disorders in the Helsinki birth cohort study. *J Psychiatr Res* **45**, 1160-4.

73. Barker, D.J., Osmond, C., Forsen, T.J., Kajantie, E. & Eriksson, J.G. Trajectories of growth among children who have coronary events as adults. *N Engl J Med* **353**, 1802-9 (2005).
74. McEvoy, M. et al. Cohort profile: The Hunter Community Study. *Int J Epidemiol* **39**, 1452-63.
75. Schofield, P.W. et al. The Audio Recorded Cognitive Screen (ARCS): a flexible hybrid cognitive test instrument. *J Neurol Neurosurg Psychiatry* **81**, 602-7.
76. Ferrucci, L. et al. Subsystems contributing to the decline in ability to walk: bridging the gap between epidemiology and geriatric practice in the InCHIANTI study. *J Am Geriatr Soc* **48**, 1618-25 (2000).
77. Sachdev, P.S. et al. The Sydney Memory and Ageing Study (MAS): methodology and baseline medical and neuropsychiatric characteristics of an elderly epidemiological non-demented cohort of Australians aged 70-90 years. *Int Psychogeriatr* **22**, 1248-64 (2010).
78. Grodstein, F., Chen, J., Wilson, R.S., Manson, J.E. & Nurses' Health, S. Type 2 diabetes and cognitive function in community-dwelling elderly women. *Diabetes Care* **24**, 1060-5 (2001).
79. Kang, J.H. & Grodstein, F. Postmenopausal hormone therapy, timing of initiation, APOE and cognitive decline. *Neurobiol Aging* **33**, 1129-37 (2012).
80. Ridker, P.M. et al. Rationale, design, and methodology of the Women's Genome Health Study: a genome-wide association study of more than 25,000 initially healthy american women. *Clin Chem* **54**, 249-55 (2008).
81. Ikram, M.A. et al. Genomewide association studies of stroke. *N Engl J Med* **360**, 1718-28 (2009).
82. Price, A.L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-9 (2006).
83. Patterson, N., Price, A.L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet* **2**, e190 (2006).
84. Richards, J.B. et al. Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. *Lancet* **371**, 1505-12 (2008).
85. Aulchenko, Y.S., Ripke, S., Isaacs, A. & van Duijn, C.M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294-6 (2007).
86. Chen, W.M. & Abecasis, G.R. Family-based association tests for genomewide association scans. *Am J Hum Genet* **81**, 913-26 (2007).
87. Chen, M.H. & Yang, Q. GWAF: an R package for genome-wide association analyses with family data. *Bioinformatics* **26**, 580-1.
88. Bacanu, S.A., Devlin, B. & Roeder, K. The power of genomic control. *Am J Hum Genet* **66**, 1933-44 (2000).
89. Kohli, M.A. et al. The neuronal transporter gene SLC6A15 confers risk to major depression. *Neuron* **70**, 252-65 (2011).

90. Kral, T. et al. Preoperative evaluation for epilepsy surgery (Bonn Algorithm). *Zentralbl Neurochir* **63**, 106-10 (2002).
91. Stegle, O., Parts, L., Durbin, R. & Winn, J. A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. *PLoS Comput Biol* **6**, e1000770 (2010).
92. Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat Protoc* **7**, 500-7 (2012).

## *7.3. Supplemental material to Chapter 3.3*

**Genome-wide studies of verbal declarative memory in non-demented older people: the CHARGE consortium**

**TABLE OF CONTENTS**

1. Abbreviations
2. Cohorts
3. Memory Tests
   3.1. Word list delayed recall
      3.1.1. Tests with verbal presentation of the word list
      3.1.2. Tests with visual presentation of the word list
   3.2. Paragraph delayed recall
4. Genotyping, Quality Control, and Imputation
5. Screening for Latent Population Substructure and Analysis Models
6. Genotype-specific hippocampal mRNA levels (eQTLs)
7. Pathway analysis
8. AD pathology
9. Memory genetic score
10. Supplemental Tables
   10.1. Table S1: Detailed description of participating cohorts
   10.2. Table S2: Description of memory test parameters for participating cohort
   10.3. Table S3: Correlation between performance on paragraph delayed recall and word list delayed recall
   10.4. Table S4: Genotyping parameters
   10.5. Table S5: Quality control filters before imputation and methods for assessing population structure
   10.6. Table S6: Imputation algorithms
   10.7. Table S7: Results and annotation for all SNPs yielding suggestive associations with delayed recall
   10.8. Table S8: Association of top memory hits with Amyloid plaques and Neurofibrillary Tangles
   10.9. Table S9: Association of top memory hits with hippocampal mRNA levels (eQTL)
   10.10. Table S10: Hypergeometric distribution test for Ubiquitin C (UBC) interactions
   10.11. Table S11: Association with published memory risk variants
   10.12. Table S12: Association with published AD risk variants
11. Supplemental Figures
   11.1. Figure S1: QQ plots for all meta-analyses
   11.2. Figure S2: Manhattan plots for all meta-analyses
   11.3. Figure S3: Pathway analyses
12. References

## 1. ABBREVIATIONS

*Memory Tests*
ALL-dr: combined tests of verbal delayed recall
CERAD-dr: Consortium to Establish a Registry for Alzheimer's Disease delayed recall
CVLT-dr: California Verbal Learning Test delayed recall (belongs to WL-dr category)
DWRT-dr: Delayed Word Recall Test (belongs to WL-dr category)
HVLT-dr: Hopkins Verbal Learning Test delayed recall (belongs to WL-dr category)
PAR-dr: paragraph (or story) delayed recall
RAVLT-dr: Rey's Auditory Verbal Learning Test delayed recall (belongs to WL-dr category)
VPWL-dr: delayed recall for visually presented word list
WL-dr: word list delayed recall
*Studies*
AGES: AGES-Reykjavik Study
ARIC: Atherosclerosis Risk in Communities Study
BLSA: Baltimore Longitudinal Study of Aging
CHS: Cardiovascular Health Study
ERF: Erasmus Rucphen Family study
FHS: Framingham Heart Study
GENOA: Genetic Epidemiology Network of Arteriopathy
HBCS: Helsinki Birth Cohort Study
HCS: Hunter Community Study
LBC1921: Lothian Birth Cohort 1921
LBC1936: Lothian Birth Cohort 1936
MAP: Rush Memory and Aging Project
NHS: Nurses' Health Study
ORCADES: Orkney Complex Disease Study
REGARDS: REasons for Geographic and Racial Differences in Stroke
ROS: Religious Orders Study
SHIP: Study of Health in Pomerania
SMAS: Sydney Memory and Aging Study
Swiss MGS: Swiss Memory Genetics Study
TASCOG: Tasmanian Study of Cognition and Gait
WGHS: Women's Genome Health Study
3C-Bordeaux: Three-City Bordeaux Study

## 2. COHORTS

### *Discovery GWAS*

The CHARGE consortium includes large prospective community-based cohort studies that have genome-wide variation data coupled with extensive data on multiple phenotypes, as detailed previously.[1] In addition, several other community-based cohort studies have collaborated with the CHARGE consortium on this genome-wide analysis of memory. What follows are some details about each study.

**Aging Gene-Environment Susceptibility - Reykjavik Study (AGES-Reykjavik)**

The AGES-Reykjavik Study is a single center prospective cohort study based on the Reykjavik Study. The Reykjavik Study was initiated in 1967 by the Icelandic Heart Association to study cardiovascular disease and risk factors. The cohort included men and women born between 1907 and 1935 who lived in Reykjavik at the 1967 baseline examination. Re-examination of surviving members of the cohort was initiated in 2002 as part of the AGES-Reykjavik Study. The AGES-Reykjavik Study is designed to investigate aging using a multifaceted comprehensive approach that includes detailed measures of brain function and structure. All cohort members were European Caucasians. Briefly, as part of a comprehensive examination, all participants answered a questionnaire, underwent a clinical examination and had blood drawn.[2] All consenting participants were offered to take a neuropsychological test battery,[3] including the California Verbal Learning Test (CVLT),[4] to assess memory performance. Among participants with genome-wide data, after exclusion of 126 participants with dementia and 266 participants with a history of stroke, 2616 participants were available for a GWAS of delayed word list recall (CVLT).

**The Atherosclerosis Risk in Communities Study (ARIC):**

The ARIC study is a prospective population-based study of atherosclerosis and clinical atherosclerotic diseases in 15 792 men and women, including 11 478 white participants, drawn from 4 United States communities (Suburban Minneapolis, Minnesota; Washington County, Maryland; Forsyth County, North Carolina; and Jackson, Mississippi). In the first 3 communities, the sample reflects the demographic composition of the community. In Jackson, only black residents were enrolled. Participants were between age 45 and 64 years at their baseline examination in 1987-1989 when blood was drawn for DNA extraction and participants consented to genetic testing.[5] A total of 15 020 participants, of which 10 898 were white, were genotyped at the Broad Institute, Boston, Massachusetts, and 9,345 of the latter passed QC criteria for genotyping and were available for analysis after application of all exclusion criteria. Vascular risk factors and outcomes, including transient ischemic attack and stroke, were determined in a standard fashion [6]. The second clinical examination of the ARIC Study cohort in 1990–1992 included a neuropsychological evaluation,[7] with the Delayed Word Recall Test (DWRT),[8] to assess memory performance. Given the relatively young age of the cohort at the second examination (age range = 48-67 years) and low expected prevalence of dementia, a formal assessment for dementia was not performed.

Among white participants with genome-wide data, after excluding 275 participants with a history of stroke, 9 188 participants were available for a GWAS of delayed word list recall (visually presented word list). In addition, among white participants with genome-wide data, 509 have recently completed a detailed neuropsychological test battery as part of an ancillary study,[9,10] including the logical memory test from the revised version of the Wechsler Memory Scale (WMS-R).[11] After excluding 79 participants with a history of stroke, 430 ARIC White participants were available for a GWAS of delayed paragraph recall.

**The Cardiovascular Health Study (CHS)**

The CHS is a population-based cohort study of risk factors for vascular disease in adults 65 years or older conducted across 4 field centers in the United States: Sacramento County, California; Washington County, Maryland; Forsyth County, North Carolina; and Pittsburgh, Allegheny County, Pennsylvania.[12] The original predominantly white cohort of 5 201 persons was recruited in 1989-1990 from a random sample of people on Medicare eligibility lists. An additional 687 African-Americans were enrolled in 1992-1993, for a total sample of 5 888. Vascular risk factors and outcomes, including transient ischemic attack, stroke and dementia, were determined in a standard fashion.[13,14] DNA was extracted from blood samples drawn on all participants who consented to genetic testing at their baseline examination in 1989-90 or 1992-1993. In 2007-2008, genotyping was performed at the General Clinical Research Center's Phenotyping/Genotyping Laboratory at Cedars-Sinai on 3 980 CHS participants who were free of cardiovascular disease at baseline and who had DNA available for genotyping. Because most other cohorts were predominantly white, the African American participants were excluded from this analysis to limit the potential for false positive associations due to population stratification. Among white participants, genotyping was attempted in 3 397 participants and was successful in 3 295 persons. As part of the CHS Cognition Study,[14,15] in 1992-94 and again, in 1997-99, participants were invited to undergo and detailed neuropsychological assessment including the CVLT[4] and the logical memory test from the WMS-R.[11] Among participants with genome-wide data: after exclusion of 47 participants with dementia and of 26 participants with history of stroke, 334 participants were available a GWAS of delayed word list recall (CVLT); after exclusion of 4 participants with dementia, 472 participants were available for a GWAS of delayed paragraph recall.

**Croatian Cohorts: Split and Korčula**

The Croatia-Korčula study is part of a larger genetic epidemiology research program in Croatian island isolates, “10,001 Dalmatians.” The genetic epidemiology research program in Croatian island isolates began in 1999,[16] then expanded to study human genetic variation and effects of isolation and inbreeding,[17,18] and finally entered the phase of focusing on diseases and gene mapping studies.[19-21] A total of 969 participants were included in the CROATIA-Korčula study. The Croatia-Split study included 535 persons collected in 2009 from the general (outbred) population Split. Split has a population of >100 000 and is second

largest city in Croatia. Participants from the Croatia-Korčula and Croatia-Split studies were invited to undergo a neuropsychological examination including the Rey's Auditory Verbal Learning Test (RAVLT).[22] Croatia- Korčula genotyping was performed at the Institute of Human Genetics, Helmholtz Zentrum München, Germany and CROATIA-Split genotyping was performed at AROS Applied Biotechnology, Aarhus, Denmark. Genotyping was successful in 898 and 499 participants respectively for Croatia-Korčula and Croatia-Split. Among participants with genome-wide data, 25 and 10 participants with a history of stroke were excluded from the Croatia-Korčula and Croatia-Split studies respectively. Of the remaining 577 and 471 individuals, aged 20 years or more, 472 and 303 participants aged 45 years or older were available for a GWAS of delayed word list recall (RAVLT).

**Erasmus Rucphen Family (ERF)**

The Erasmus Rucphen Family (ERF) study is a family-based cohort study in a genetically isolated population in the Netherlands,[23,24] including 3 000 participants. Participants are all descendants of a limited number of founders living in the 19th century. Extensive genealogical data is available for this population. The study protocol included venous puncture for DNA isolation and chemistry, cognitive evaluation, cardiovascular examination, eye assessments and body composition measurements. All participants gave informed consent and the study was approved by the medical ethics committee at Erasmus MC University Medical Center.

Genotyping was done at the Human Genotyping Facility, Genetic Laboratory Department of Internal Medicine, Erasmus MC, Rotterdam, and at the Genotyping Center of Leiden University, The Netherlands. In total, 2 385 samples from the ERF Study were available with good quality genotyping data. Participants were invited to undergo a neuropsychological evaluation,[25] which included the Dutch version of Rey's Auditory Verbal Learning Test (RAVLT).[22,26] Among participants with genome-wide data, we excluded 20 individuals with a history of stroke. Of the remaining 2 119 participants, aged 20 years or older, 1 267 participants were aged 45 years or more and were available for performance GWAS of delayed word list recall (RAVLT).

**Framingham Heart Study (FHS)**

The FHS is a three-generation, single-site, community-based, prospective cohort study that was initiated in 1948 to investigate risk factors for cardiovascular disease including stroke. It now comprises 3 generations of participants: the original cohort followed since 1948 (Original);[27] their offspring and spouses of the offspring, followed since 1971 (Offspring);[28] and children from the largest offspring families enrolled in 2000 (Gen 3).[29] The Original cohort enrolled 5 209 men and women who comprised two-thirds of the adult population then residing in Framingham, MA, USA. Survivors continue to receive biennial examinations. The Offspring cohort comprises 5 124 persons (including 3 514 biological offspring) who have been examined approximately once every 4 years. Participants in the first two generations were invited to undergo an initial neuropsychological test battery in 1999-

2005,[30] including the logical memory test from the Original WMS.[31] Neuropsychological testing in Gen 3 only began in 2009 and is not included in these analyses. The population of Framingham was virtually entirely whites in 1948 when the Original cohort was recruited. Vascular risk factors and outcomes, including transient ischemic attack, stroke and dementia, were identified prospectively since 1948 through an ongoing system of FHS clinic and local hospital surveillance.[32,33] Participants had DNA extracted and provided consent for genotyping in the 1990s. Genotyping was performed at Affymetrix (Santa Clara, CA) through an NHLBI funded SNP-Health Association Resource (SHARe) project and successful in 4,519 persons from the Original and Offspring cohorts. Of these 4 519 persons 4 116 were alive in 1999 when the neuropsychological study began. Of these, 2 642 participants have undergone neuropsychological testing including logical memory. We excluded 30 participants with a neurological condition that might confound the cognitive assessment (e.g. brain tumor or severe head injury), 7 participants with dementia and 52 participants with a history of stroke. Of the remaining 2 553 participants, 2 493 were aged 45 years or older and were available for a GWAS of delayed paragraph recall.

**Genetic Epidemiology Network of Arteriopathy (GENOA)**

GENOA is a study of hypertensive sibships designed to investigate the genetic underpinnings of hypertension and target organ damage. In the initial phase of the GENOA study (Phase I: 1996-2001), all members of sibships containing ≥ 2 individuals with essential hypertension clinically diagnosed before age 60 were invited to participate, including both hypertensive and normotensive siblings (1 583 non-Hispanic whites from Rochester, MN, and 1 841 African Americans from Jackson, MS). The diagnosis of essential hypertension was established based on blood pressure levels measured at the study visit (>140 mmHg average systolic BP or >90 mmHg average diastolic BP) or a prior diagnosis of hypertension and current treatment with antihypertensive medications. Exclusion criteria were secondary hypertension, alcoholism or drug abuse, pregnancy, insulin-dependent diabetes mellitus, or active malignancy. In the second phase of the GENOA study (Phase II: 2000-2004), 1 241 white and 1 482 African American participants were successfully re-recruited to measure potential target organ damage due to hypertension. The Genetics of Microangiopathic Brain Injury (GMBI) study (2001-2006) is an ancillary study of GENOA undertaken to investigate susceptibility genes for ischemic brain injury. Phase II GENOA participants that had a sibling willing and eligible to participate in the GMBI study underwent a neurocognitive testing battery to assess several domains of cognitive function including Rey's Auditory Verbal Learning Test (RAVLT)[22] (967 whites and 1010 African Americans). Genotyping was performed at the Mayo Clinic, Rochester (MN). GENOA White participants who were less than 45 years of age (N=56) or had history of stroke (N=22) were excluded from the analysis, leaving a total of 889 participants. Among participants with genome-wide genotype data, 758 GENOA White participants from 378 sibships were available for a GWAS on delayed word list recall (RAVLT).

**Helsinki Birth Cohort Study (HBCS)**

The source cohort for the HBCS comprised 4 130 women and 4 630 men born as singletons at Helsinki University Central Hospital during 1934-44, who had birth and child welfare records and were living in Finland in 1971.[34] To approach an intended sample size of N= 2 000, a random subsample of 2 902 subjects was invited to participate in the study; 2 003 of them (1 075 women and 928 men) were finally included.[35] Participants who could come to the examination center were invited to take a neuropsychological test battery, including the delayed word list recall from the CERAD battery.[36] 1 063 participants attended neuropsychological testing between February 2005 and February 2011. DNA was extracted from 1 728 randomly selected participants of the HBCS. Genotyping was conducted at the Wellcome Trust Sanger Institute, Cambridge, UK. Among participants with available cognitive tests and genome-wide genotypes, after exclusion of 19 participants with a history of stroke, 888 individuals were available for a GWAS of delayed word list recall (visually presented word list). The study was approved by the Institutional Review Board of the National Public Health Institute, and informed consent was obtained from all participants.

**Lothian Birth Cohort 1921 (LBC1921) and 1936 (LBC1936)**

These samples include surviving participants from the Scottish Mental Surveys of 1932 or 1947 (SMS1932 and SMS1947), having been born, respectively in 1921 (LBC1921) and 1936 (LBC1936).[37,38] They were all Caucasian and almost all lived independently in the Lothian region (Edinburgh city and surrounding area) of Scotland. The LBC1921 cohort comprised 550 members while the LBC1936 cohort included 1 091 participants. At age 79 approximately, LBC1921 participants underwent a neuropsychological examination including the logical memory test from the WMS-R.[11] At age 70, LBC1936 participants took a battery of cognitive tests,[37] including the logical memory test from the Wechsler Memory Scale-IIIUK (WMS-IIIUK).[39] Genotyping was performed at the Wellcome Trust Clinical Research Facility (WTCRF) Genetics Core, Western General Hospital, Edinburgh. Among participants with good quality genome-wide data, after exclusion of 5 participants for dementia and 43 participants for history of stroke, 469 individuals from the LBC1921 cohort were available for a GWAS on paragraph delayed recall . Among participants with genome-wide data, after exclusion of 50 participants for history of stroke, 953 individuals from the LBC1936 cohort were available for a GWAS on delayed paragraph recall.

**Orkney Complex Disease Study (ORCADES)**

ORCADES is an ongoing, family-based, cross-sectional study that seeks to identify genetic factors influencing cardiovascular and other disease risk in the population isolate of the Orkney Isles in northern Scotland.[40] The North Isles of Orkney, the focus of this study, consist of a subgroup of ten inhabited islands with census populations varying from ~30 to ~600 people on each island. The first phase of data collection was carried out in Orkney between 2005 and 2007. Informed consent and blood samples were provided by 1019 Orcadian volunteers who had at least one grandparent from the North Isles of Orkney.

Participants were invited to take a neuropsychological test battery including the logical memory test adapted from the Original Wechsler Memory Scale.[31] Genotyping was performed at the Helmholtz Centre in Munich on a subset of 719 participants. An additional 169 individuals were genotyped by Integragen in Paris. Among participants with genome-wide data, we excluded 7 participants with a history of stroke. Of the 537 remaining individuals, aged 20 years or older, 419 participants aged 45 years or older were available for a GWAS of delayed paragraph recall.

**Rotterdam Study, Rotterdam Study-II, Rotterdam Study-III**

The Rotterdam Study is a population-based prospective 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.[41,42] In 1990-1993, 7 983 persons 55 years of age or over participated and were re-examined every 3 to 4 years. In 1999, 3 011 individuals who had become 55 years of age or moved into the study district since the start of the study were added to the cohort (Rotterdam Study-II), and in 2006 a further extension of the cohort was initiated in which 3 932 subjects aged 45–54 years and living in the same district were included (Rotterdam Study-III).[43] All participants had DNA extracted at their first visit. Genotyping was attempted in participants with high-quality extracted DNA. Genotyping was done at the Human Genotyping Facility, Genetic Laboratory Department of Internal Medicine, Erasmus MC, Rotterdam, the Netherlands. Participants underwent several neuropsychological tests at the baseline and follow-up examinations,[44] including a 15-word verbal learning test based on Rey's recall of words.[45] Participants are continuously monitored for major events, including dementia and stroke, by automated linkage of the general practitioners' records and hospital discharge files with the study database.[46,47] Among participants with good quality genome-wide data: after exclusion of participants with dementia (N=124 for the Rotterdam Study, N=56 for the Rotterdam Study-II and N=6 for the Rotterdam Study-III) and participants with a history of stroke (N=168 for the Rotterdam Study and N=102 for the Rotterdam Study-II), 2 067 participants from the Rotterdam Study, 1 533 participants from the Rotterdam Study-II and 1 935 participants from the Rotterdam Study-3 were available for a GWAS of word list delayed recall (visually presented word list).

**Religious Orders Study (ROS)**

The ROS, started in 1994, enrolled Catholic priests, nuns, and brothers, aged 53 years from about 40 groups in 12 states.[48] Since January 1994, 1 132 participants completed their baseline evaluation, of whom 1 001 were non-Hispanic white. The study was approved by the institutional review board of Rush University Medical Center. The follow-up rate of survivors exceeds 90%. Participants were free of known dementia at enrollment, agreed to annual clinical evaluations, and signed both an informed consent and an Anatomic Gift Act form donating their brains at time of death.[49] A more detailed description of the ROS has been published previously.[49] Participants were invited to take a neuropsychological test

battery, including delayed recall of Story A from the logical memory subset of the Wechsler Memory Scale-Revised,[11] and delayed word list recall from the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) battery.[36] DNA was extracted from whole blood, lymphocytes, or frozen postmortem brain tissue. Genotyping was performed at the Broad Institute's Center for Genotyping and the Translational Genomics Research Institute. Among participants with available memory tests and genome-wide genotypes, after exclusion of 58 participants with dementia and 65 participants with a history of stroke, 687 individuals were available for a GWAS of delayed paragraph recall and delayed word list recall (visually presented word list).

**Rush Memory and Aging Project (MAP)**

The MAP, started in 1997, enrolled older men and women from assisted living facilities in the Chicago area with no evidence on dementia at baseline.[48] Since October 1997, 1 285 participants completed their baseline evaluation, of whom 1 118 were non-Hispanic white. The study was approved by the institutional review board of Rush University Medical Center. The follow-up rate of survivors exceeds 90%. Similar to the ROS, participants agreed to annual clinical evaluations and signed both an informed consent and an Anatomic Gift Act form donating their brains, spinal cords, and selected nerves and muscles to Rush investigators at the time of death.[49,50] A more detailed description of the MAP has been published previously.[49,50] Participants were invited to take a neuropsychological test battery, including delayed recall of Story A from the logical memory subset of the Wechsler Memory Scale-Revised,[11] and delayed word list recall from the CERAD battery.[36] DNA was extracted from whole blood, lymphocytes, or frozen postmortem brain tissue. Genotyping was performed at the Broad Institute's Center for Genotyping and the Translational Genomics Research Institute. Among participants with available memory tests and genome-wide genotypes, after exclusion of 51 participants with dementia and 86 participants with a history of stroke, 751 individuals were available for a GWAS of delayed paragraph recall and delayed word list recall (visually presented word list).

**Study of Health in Pomerania (SHIP)**

The "Study of Health in Pomerania" is a population-based epidemiological study in the region of Western Pomerania, Germany.[51] In brief, from the total population of West Pomerania comprising 213 057 inhabitants in 1996, a two-stage stratified cluster sample of adults aged 20–79 years was drawn. The net sample (without migrated or deceased persons) comprised 6 265 eligible subjects, out of which 4 308 completed their baseline examinations. From July 2007 to October 2010 the 'Life-Events and Gene-Environment Interaction in Depression' (LEGENDE) study was carried out in the SHIP cohort. A diagnostic interview for mental disorders was performed based on Diagnostic and Statistical Manual for Mental Disorders (IV edition) diagnostic criteria.[52]

As part of the SHIP-LEGENDE project,[53] participants have been invited to take a cognitive test battery, including a German adaptation of Rey's Auditory Verbal Learning Test

(RAVLT).[54] Genotyping was performed at Affymetrix (Santa Clara, CA). The genetic data analysis workflow was created using the Software InforSense. Genetic data were stored using the database Caché (InterSystems). Among participants with available cognitive tests and genome-wide genotypes, we excluded 62 participants with dementia and 19 participants with a history of stroke. Of the remaining 2 027 individuals, aged 20 years or older, 1 474 participants aged 45 years or older were available for a GWAS of delayed word list recall (RAVLT).

**Tasmanian Study of Cognition and Gait (TASCOG)**

TASCOG is a study of cerebrovascular mechanisms underlying gait, balance and cognition in a population-based sample of Tasmanian people aged at least 60 years. Individuals aged 60–86 years (n = 395) living in Southern Tasmania, Australia, were randomly selected from the electoral roll to participate in the study. Individuals were excluded if they lived in a nursing home, had a contraindication for magnetic resonance scanning (MRI) or were unable to walk without a gait aid.[55] Participants were invited to take a neuropsychological test battery,[56] including among other tests the Hopkins Verbal Learning Test (HVLT).[57] DNA was extracted from peripheral blood samples by proteinase K digestion following cell lysis, then phenol-chloroform purification. DNA was genotyped at the Diamantina Institute and Institute of Molecular Biosciences, University of Queensland, Australia, for 370 participants, and call rates were greater than 97% for all samples. Genotypes for 22 individuals were excluded, either because they were closely related to other individuals, they were outliers in a population ancestry analysis or their sex predicted from genotypes did not match sex as recorded in the database. Among the 348 remaining participants with available genome-wide data, after exclusion of 2 participants with dementia and 15 participants with a history of stroke, 331 individuals were available for performance GWAS of delayed word list recall (HVLT).

***Replication Study***

**Aging Gene-Environment Susceptibility - Reykjavik Study (AGES-Reykjavik)**

A subset of white participants from the AGES-Reykjavik study (see above for details on the AGES-Reykjavik cohort) who had not undergone genome-wide genotyping at the time of the study, but had undergone the neuropsychological test battery [3], including the California Verbal Learning Test (CVLT)[4] to assess memory performance, were included in the replication study. SNPs were genotyped on an Illumina custom-designed chip. After exclusion of 159 participants with dementia and 235 participants with a history of stroke, 1 525 participants with genotype data were used in the present analysis.

**Baltimore Longitudinal Study of Aging (BLSA)**

The BLSA is a multidisciplinary observational study of the physiological and psychological aspects of human aging and diseases and conditions that increase with age.[58] Participants were invited to take a neuropsychological test battery, including among other tests the

CVLT.[4] DNA was genotyped at the Laboratory of Neurogenetics, NIA genotyping platform, USA. Among the 925 participants with available genome-wide data and cognitive test performance, after exclusion of 37 participants with a history of stroke, 888 individuals were available for performance GWAS of delayed word list recall (CVLT).

**Hunter Community Study (HCS)**

The HCS is a community-based longitudinal investigation that was commenced in Australia in 2004–2005. The study aims to investigate retired and near-retired persons by sampling older Australians aged 55–85, randomly selected from electoral rolls in a regional area on the heavily populated east coast (New South Wales).[59] All participants were invited to take the Audio Recorded Cognitive Screen (ARCS), an instrument that uses an audio device to administer selected neuropsychological tests to unsupervised individuals,[60] including a 12-word list learning task, modeled on the Hopkins Verbal Learning Test -revised (HVLT-R). Genotyping was conducted at the Hunter Medical Research Institute, Newcastle Australia. Among participants with available memory tests and genome-wide genotypes, after exclusion of 34 participants with a history of stroke, 821 individuals were available for a GWAS of delayed word list recall (HVLT).

**Nurses' Health Study (NHS)**

The NHS began in 1976, when 121,700 female registered nurse women, aged 30 to 55 years, living in 11 USA states completed a mailed questionnaire on lifestyle and health. Every 2 years, follow-up questionnaires have been mailed to participants to update their information, and >90% follow-up of the total possible person-time has been maintained.[61] For the study of cognitive function, participants aged 70 years and older free of diagnosed stroke were selected (dementia diagnosis was not ascertained). From 1995 to 2001, 21 085 eligible women were contacted for a baseline telephone cognitive assessment, including a measure of delayed paragraph recall using the East Boston Memory Test, and the Telephone Interview of Cognitive Status 10 –item word list; 19 415 (92%) completed the interview. The study was approved by the Institutional Review Board of the Brigham and Women's Hospital.[61] Genotyping was conducted at the Broad Institute using a nested case-control design. Among participants from 6 independent case-control studies in the NHS cohort with available memory tests and genome-wide genotypes, 2 066 / 1 773 individuals were available for a GWAS of delayed paragraph / word list recall (TICS): 756 / 655 from the diabetes case-control dataset, 358 / 307 from the coronary heart disease case-control dataset, 553 / 450 from the breast cancer case-control dataset, 179 / 166 from the glaucoma case-control dataset, 80 / 76 from the kidney stone case-control dataset, and 141 / 119 from the endometrial cancer case-control dataset.

**Sydney Memory and Ageing Study (Sydney MAS)**

The Sydney MAS was initiated in 2005 to examine the clinical characteristics and prevalence of mild cognitive impairment and related syndromes, and to determine the rate of change in

cognitive function over time.[62] Participants were invited to take a neuropsychological test battery, including among other tests the RAVLT,[22] and Story A of the Wechsler Memory Scale, 3rd edition, (WMS-III).[63] DNA was genotyped at the Ramaciotti Centre, UNSW, Australia. Among the 925 participants with available genome-wide data, after exclusion of 37 participants with a history of stroke, 880 individuals were available for performing a GWAS of delayed word list recall (RAVLT) and 887 for delayed paragraph recall (WMS-III).

**Three-City Bordeaux (3C-Bordeaux)**

The 3C-Bordeaux Study is a French prospective population-based study including 2 104 non-institutionalized participants aged 65 years or more.[64] At baseline and at each follows-up examination participants were invited to take a neuropsychological test battery, including among other tests the FCSRT.[65] DNA was genotyped at the Centre National de Génotypage, Evry, France.[66] Among 1 139 participants with available genome-wide data and cognitive test performance, after exclusion of 27 participants with dementia and 49 participants with a history of stroke, 1 063 individuals were available for performing a GWAS of delayed word list recall (FCSRT).

**Women's Genome Health Study (WGHS)**

The WGHS is a large cohort for genome-wide genetic analysis of a wide range of clinical phenotypes among >25 000 women, 45 years or older at baseline and with ongoing follow-up observation, now for approximately 18 years. [67] The population is derived from participants in the Women's Health Study who provided a blood sample at baseline. By design, participants included in the WGHS were free from dementia and stroke at baseline. Women above 70 years of age were contacted for a baseline telephone cognitive assessment, including a measure of delayed paragraph recall using the East Boston Memory Test, and the Telephone Interview of Cognitive Status 10 –item word list. Genome-wide genotyping in the WGHS was performed using the Illumina (San Diego, CA) HumanHap 300 Duo "+" platform including a total 339 596 SNPs passing quality control filters among 23 294 participants with verified European ancestry. Among participants with available memory tests and genome-wide genotypes, 3 542 individuals of verified European ancestry were available for genetic analysis of delayed word list recall (TICS) and of delayed paragraph recall.

***Extension to African-American populations***

**The Atherosclerosis Risk in Communities Study (ARIC):**

The overall design of the ARIC study is described above. African-American participants underwent the same neuropsychological evaluation,[7] including the Delayed Word Recall Test (DWRT) [8] to assess memory performance. Genotyping on an Affymetrix GeneChip SNP Array 6.0® was available on 2 594 African-American participants. After excluding 118 participants with a history of stroke, 2 476 ARIC African-American participants were available for a GWAS of delayed word list recall (visually presented word list).

**Genetic Epidemiology Network of Arteriopathy (GENOA)**

The overall design of the GENOA study is described above. GENOA African-Americans participants who were less than 45 years of age (N=23), had missing age data (N=2) or had history of stroke (N=51) were excluded from the analysis, leaving a total of 934 participants. Also excluded from this analysis were GENOA participants who also participated in the cognitive portion of the ARIC study (N=118). After excluding participants who did not have imputed genome-wide genotype data, a final sample of 708 GENOA African-American participants from 406 sibships were available for a GWAS on delayed word list recall (RAVLT).

**REasons for Geographic and Racial Differences in Stroke (REGARDS)**

The REGARDS study is a population-based longitudinal cohort study of 30 239 black and white individuals 45 years and older, recruited between 2003 and 2007 and followed up through 2011. The study oversampled blacks and residents of the southeastern “Stroke Belt” (Alabama, Arkansas, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, and Tennessee). Risk factors were assessed through a combination of telephone interview and a physical examination conducted in the participant's home. Study design details and event identification and adjudication are available elsewhere.[68-70] All REGARDS participants are assessed using the Word List Learning and Delayed Recall from the Consortium to Establish a Registry for Alzheimer’s Disease (CERAD) battery,[36] which are telephone administered in REGARDS. These measures were first incorporated into the REGARDS follow-up telephone interviews in Jan. 2006 during assessments occurring at 18-mo. intervals. The frequency of assessment was changed to 2-yr. intervals in Feb. 2008. The original sample for this analysis consisted of 2 398 black participants from a substudy examining the genetic influences on diabetes in African Americans, with all 2 398 having DNA genotyping using the Affymetrix 6.0 platform. Of the original sample, 1 149 were classified as having diabetes and 1 249 not meeting this definition served as controls; only controls were used for the present analysis. Of the 744 controls with available genome-wide data and cognitive test performance, 80 were excluded due to baseline impairment on the Six-Item Screener (defined as a score of 4 or less), 37 for stroke prior to cognitive assessment. The remaining 627 comprised the sample used in this analysis for performing a GWAS of delayed word list recall (CERAD).

***Extension to young adults***

**Swiss Memory Genetics Study (Swiss MGS):**

A total of 1 561 healthy, young Swiss university students or age-matched employees/trainees (658 from the University of Zürich and 903 from the University of Basel) were recruited for a genetic study of memory performance.[71] All were between 18 and 45 years of age. The study was approved by the ethics committee of the Canton of Zurich, Switzerland. DNA was obtained from blood or saliva. Participants were genotyped at the Biozentrum, University of Basel, Switzerland.

**3. MEMORY TESTS**

In all cohorts, cognitive tests were administered by trained personnel following standardized protocols and blind to genetic information. To assess the delayed recall component of verbal declarative memory, cohorts administered one of two, or both types of memory tests: word list delayed recall (WL-dr) and paragraph (or story) delayed recall (PAR-dr). WL-dr tests were further subdivided into tests using visually presented or verbally presented word lists, with or without semantic relatedness between the words (Supplementary Table 2). In each case participants were asked to remember as many words or story elements as possible after a specified delay interval, preceded by an immediate recall task. Specific test characteristics are detailed below.

In contrast with WL-dr, in PAR-dr the story elements to be remembered are contextually linked, so the two tests may activate different, if partly overlapping, neural networks. In studies where both PAR-dr and WL-dr were tested, the correlation between performances on PAR-dr and WL-dr was of the same magnitude as the correlation of PAR-dr or WL-dr with executive function or processing speed performance, i.e. non-memory cognitive domains (Supplementary Table 3), suggesting that these two measures of verbal declarative memory are not interchangeable. Whether the learning material is presented visually or verbally and whether words are unrelated or chosen from a limited number of semantic categories probably also has an impact on the neural circuits involved in the learning process.[72] Mechanisms are also likely to differ depending on the number of exposure trials and the delay after which the participant is asked to recall the learned material, repeated stimulations and longer delays being more likely to involve long-term memory storage with protein synthesis.[73,74]

Given these differences, we decided a priori to run global meta-analyses combining all tests, as well as meta-analyses combining similar tests only. Meta-analyses thus comprised a combination of: all measures of delayed recall (ALL-dr); PAR-dr; WL-dr; visually presented, semantically unrelated, word lists, including Consortium to Establish a Registry for Alzheimer's Disease delayed recall (CERAD-dr) and Delayed Word Recall Test (DWRT-dr); verbally presented semantically unrelated word lists (Rey Auditory Verbal Learning Test, RAVLT-dr); and verbally presented semantically related word lists, including the California Verbal Learning Test (CVLT-dr) and the Hopkins Verbal Learning Test (HVLT-dr). In the ALL-dr meta-analysis, for cohorts with both PAR-dr and WL-dr we used the test with the largest sample size; when sample size was identical, we used PAR-dr.

### 3.1 Word list delayed recall

#### 3.1.1 Tests with verbal presentation of the word list

***Rey's Auditory Verbal Learning Test (RAVLT),[22] used by ERF, Croatia-Split, Croatia-Korčula, GENOA, SHIP, AAA, Sydney MAS***

In the first part of the test a first list of 15 semantically unrelated common words (list A) is read to the participant followed by an immediate recall test where the participant is asked to say the words he/she can recall. There are 5 such exposure trials (3 trials for SHIP). After that, an interference task unrelated to the RAVLT follows in the ERF protocol. In the CROATIA-Split, CROATIA-Korčula and Sydney-MAS protocol, after the 5 exposure trials a second list (list B) with 15 different words is read to the participant, followed by an immediate recall test from list B; after that the participant is asked to recall the words from list A (6th recall of list A for the participant without prior exposure this time). After a given delay (20 minutes for ERF and SHIP, 15 minutes for CROATIA-Split and CROATIA-Korčula, 30 minutes for Sydney MAS) the participant is asked to again say all the words he/she remembers from list A. The delayed recall score corresponds to the total number of words correctly recalled in the 7th recall trial of list A for CROATIA-Split, CROATIA-Korčula, Sydney MAS and the 6th recall trial of list A for ERF.

***California Verbal Learning Test (CVLT),[4] used by AGES, CHS and the BLSA***

The CVLT consists of lists of 16 items, belonging to one of four semantic categories of "shopping list" items: for example, the first-"Monday's" list-contains four names of fruits, of herbs and spices, of articles of clothing, and of tools. The delayed recall score corresponds to the total number of words recalled after a given number of exposure trials (4 in a different order for AGES, 5 in the same order for CHS and BLSA) and following a given delay (15 minutes for AGES, 30 minutes for CHS, 15-20 minutes for BLSA). In all cases only list A is used and the exposure trials are separated from the delayed recall by interference activities.

***Hopkins Verbal Learning Test (HVLT)[57] used by TASCOG and the HCS***

The HVLT involves presenting 12 words (4 words from 3 semantic categories) verbally. All 12 words are presented for 3 trials. After each trial participants are to recall as many words from the list as they can remember. This is followed by a 20 minute delayed recall trial. The exposure trials are separated from the delayed recall by non-memory tests as interference activities. In the HCS three alternative 12 word lists have been generated and for the baseline exam used in the present analyses, one of these three lists was chosen randomly. The three lists were shown to be equivalent.[60]

***Telephone Interview of Cognitive Status (TICS),[75] used by the NHS and WGHS***

The TICS is a 10-item non-semantically related word list, which the participant is asked to recall both immediately and after a delay of about 20 minutes, filled with distractor questions. The TICS was shown to be as reliable and valid as face-to-face administration.[75]

**3.1.2 Tests with visual presentation of the word list**

***Delayed Word Recall Test (DWRT),[8] used by ARIC***

In the DWRT test, a set of 10 common semantically unrelated words is presented visually, one word at a time. The words are taken from lists B and C of the RAVLT. In response to reading each word on an index card, the subject is required to make up a sentence using the word. A second exposure to the list using the same protocol immediately follows. The delayed recall score corresponds to the total number of words recalled after a 5 minute delay, during which a non-memory test is given as an interference activity.

***Rey's recall of words,[45] used by the Rotterdam Study, Rotterdam Study-II and Rotterdam Study-III***

In this test participants are shown 15 different semantically unrelated words (from list A of the RAVLT) and then asked to recall as many as possible. This is done 3 consecutive times. After 10 minutes, during which a non-memory test is given as an interference activity, the participants are asked again to recall the 15 words. The total number of words recalled is the outcome variable for the delayed recall score.

***Consortium to Establish a Registry for Alzheimer's Disease (CERAD),[36] used by the HBCS, the ROS and MAP, and REGARDS***

The list learning test is a part of a larger set of tests from CERAD. The word list learning task consists of learning a list of 10 semantically unrelated written words, read aloud by the participant. This list is repeated 3 times (in a different order) and after each presentation, the participant recalls as many words as possible. Delayed recall task is conducted after 3 to 5 minutes during which an interference task is given.

In REGARDS, the CERAD is administered according to the standard protocol, but over telephone, with no simultaneous visual presentation of the word list. The instructions for each learning trial, including the oral presentations of the word list, are administered via a recording so that all participants are exposed to the same narration, thereby avoiding any differences in dialect, tone, gender, or volume that might affect participants' performance.

***FCSRT,[65,76] used by 3C-Bordeaux***

The FCSRT comprises 16 semantically unrelated items presented 4 at a time on successive cards. In the first presentation, aimed at inducing semantic encoding, the subject is asked to read aloud the name of the item corresponding to a specific semantic category (e.g., "what is the name of the flower?"). When all 4 items have been named, the card is removed and the participant is asked for immediate recall of the 4 items in response to their semantic cue. The procedure of encoding is repeated 3 times, until the subject can recall all 4 items, then the following card is proposed. This is followed by 3 successive trials during which a free recall of the 16 items is asked from the subject in 2 minutes, followed by a selective semantic cuing of the items not spontaneously recalled. After 20 minutes, subjects undergo a delayed free and cued recall, only the free recall was used for analyses.

***Visually presented word list delayed recall,[77] used by the Swiss MGS***

In this test, subjects view six series of five semantically unrelated nouns presented at a rate of one word per second with the instruction to learn the words for immediate free recall after each series. After 5 min, subjects undergo an unexpected delayed free-recall test of the learned words.

**3.2 Paragraph delayed recall**

All tests with paragraph recall used verbal presentation.

***Logical Memory adapted from the Original Wechsler Memory Scale (WMS),[31] used by FHS, ORCADES***

Story A (Anna Thompson) from the original WMS version is read to the participant and followed by an immediate free recall. After a 20 minute delay during which other neuropsychological tests are administered and serve as an interference activity, the participant is asked again to recall the story. The total score is the average number of ideas recalled from story A after the 20-minute delay. Story A contains 25 memory units or "ideas", hence the maximum score is 25.

***Logical Memory from the Wechsler Memory Scale Revised (WMS-R),[11] used by ARIC, CHS, LBC1921, ROS and MAP***

The examiner reads two stories, stopping after each reading for an immediate free recall. Compared to the original version of the WMS the stories have been changed to make them more contemporary in content and language. After a 30 minute delay during which a non-memory test is given as an interference activity, the participant is asked again to recall the story. The total score is the average number of ideas recalled from story A and B after the 30-minute delay. As each story contains 25 scoring units, the maximum score is 25 (25+25/2). Of note, in ROS and MAP only story A is used.

***Logical Memory from the Wechsler Memory Scale-IIIUK (WMS-IIIUK),[39] used by LBC1936 and GS:SFHS***

This version is very similar to the logical memory test from the WMS-R, but the second story is read twice.

***Logical Memory from the Wechsler Memory Scale (3rd edition, WMS-III),[63] used by Sydney MAS***

This version is very similar to the logical memory test from the WMS-R but only the first story was presented.

***East Boston Memory Test (EBMT),[78] used by NHS and WGHS***

The EBMT consists of a short story read to participants, the delay in recall is about 20 minutes, and the score is based on participant's ability to repeat 12 key items in the story, for a range in scores of 0-12.

## 4. GENOTYPING, QUALITY CONTROL, AND IMPUTATION

*Discovery cohorts*

The consortium was formed after the individual studies had finalized their GWAS platforms, and the studies included used different platforms (Supplementary Table 4).

As detailed previously,[79] participant-specific quality controls included filters for call rate, heterozygosity, and number of Mendelian errors per individual. SNP-specific quality controls included filters for call rate, minor allele frequency, Hardy-Weinberg equilibrium, and differential missingness by outcome or genotype (mishap test in PLINK, http://pngu.mgh.harvard.edu/purcell/plink/).

The set of genotyped input SNPs used for imputation in each study was selected based on their highest quality GWA data (see Supplementary Table 5 for quality control filter applied prior to imputation). We used either the Markov Chain Haplotyping (MaCH) package (http://www.sph.umich.edu/csg/abecasis/MACH) or BIM-BAM15 programs for imputation, as detailed in Supplementary Table 6. For each imputed SNP, imputation quality was estimated using the R-square metric or the ratio of the empirically observed dosage variance to the expected binomial dosage variance.

*Replication cohorts*

Apart from the AGES-Reykjavik replication sample, replication was performed in silico in all follow-up cohorts. Genotyping platforms, quality control filters and imputation methods are detailed in Supplementary Tables 4-6.

Of note, for imputation in African-Americans from the ARIC study, given the admixed nature of the study population, SNPs were not removed if they only deviated from the Hardy-Weinberg equilibrium criterion. A combined CEU + YRI reference panel was generated prior to imputation that included approximately $2.74 \times 10^6$ SNPs that were either in both the CEU and YRI panels, or that were present in one panel and were monomorphic and non-missing in the other.

For replication genotyping in the AGES-Reykjavik cohort, genotyping was performed on an Illumina Bead Xpress, following the manufacturer's instructions. The typed SNPs were included in a 96-well panel.

## 5. SCREENING FOR LATENT POPULATION SUBSTRUCTURE AND ANALYSIS MODELS

Studies were screened for latent population substructure, including cryptic relatedness, using suitable programs (Supplementary Table 5).[80-82] When appropriate, components related to the phenotype under study were included as covariates in the linear regression. ERF, CROATIA-Vis, CROATIA-Korčula, GENOA and FHS included related individuals and used the following methods to adjust for relatedness of the population: ERF, CROATIA-Vis and

CROATIA-Korčula performed association analyses in GenABEL (R-library),[83] using the mmscore method, with a kinship matrix that was estimated from the genotype data;[84] GENOA performed association analyses using linear mixed effects modeling with family as a random intercept; FHS used a linear mixed effects model accounting for familial relatedness.[85]

We studied quantile-quantile (Q-Q) plots to ensure that the p-value distributions in each of the cohorts conformed to a null distribution at all but the extreme tail. We also calculated the genomic inflation factor lambda, which measures over-dispersion of test-statistics from association tests indicating population stratification and can be used to apply genomic control.[86]

We restricted the present meta-analysis to the 2 720 831 autosomal SNPs common to all studies. For meta-analyses combining studies using exactly the same memory test (CVLT-dr, RAVLT-dr, CERAD-dr), we used an inverse-variance weighted meta-analysis as our primary method after applying genomic control within each individual study. Beta estimates were weighted by their inverse variance and a combined estimate was obtained by summing the weighted betas and dividing by the summed weights. Hence results for SNPs imputed with low certainty were down-weighted because the low quality of imputation ensures a large variance. In contrast, studies with large sample sizes and with directly genotyped or well-imputed SNPs had a greater effect on the meta-analyses p-value because of small variances. For meta-analyses combining studies using similar but not identical memory tests (ALL-dr, PAR-dr, WL-dr, VPWL-dr, CVLT-HVLT-dr), we used as a primary method an effective sample size weighted meta-analysis technique after applying genomic control within each individual study. This approach is recommended when the dependent variable is measured on different scales between cohorts and does not yield any directly comparable effect estimate. For each SNP the z-statistic was weighted by the effective sample size (product of the sample size and the ratio of the empirically observed dosage variance to the expected binomial dosage variance for imputed SNPs). A combined estimate was obtained by summing the weighted z-statistics and dividing by the summed weights. Hence results for SNPs imputed with low certainty were down-weighted. In contrast, studies with large sample sizes and with directly genotyped or well-imputed SNPs had a greater effect on the meta-analyses p-value.
We undertook the meta-analyses using METAL.[87]

We estimated the genomic inflation factor lambda after meta-analysis. The estimates of lambda ranged between 0.999 and 1.049 (see details in Supplementary Figure 1), indicating no significant inflation of p-values. The quantile-quantile (Q-Q) plots of our meta-analysis results for the various memory traits (Supplementary Figure 1) show the distribution of the observed test statistic (negative log of p-values, on the *y*-axis) plotted against the distribution of test statistic expected under the null-hypothesis (on the *x*-axis).

All SNPs with a p-value < $5 \times 10^{-6}$ and a minor allele frequency > 0.05 were selected for in silico replication. In the AGES-repli study, replication had to be performed by de novo genotyping, thus a more limited number of SNPs were selected for replication, arising from the CVLT-dr and CVLT-HVLT-dr meta-analyses.

## 6. GENOTYPE-SPECIFIC HIPPOCAMPAL MRNA LEVELS (EQTLS)

To characterize the functional relevance of significant and suggestive memory risk loci, we analyzed their association with human hippocampus cell line gene expression profiles, in a unique set of 138 pre-mortem human hippocampus samples from patients undergoing surgery for treatment-resistant epilepsy, as part of the Epilepsy Surgery Program at Bonn University, Germany.[88]

*Hippocampal sample preparation*

Hippocampal biopsy samples were obtained from patients with chronic pharmaco-resistant temporal lobe epilepsy in the Epilepsy Surgery Program at Bonn University, Germany.[88] Surgical removal was indicated to achieve seizure control after standardized pre-surgical evaluation using a combination of noninvasive and invasive procedures.[89] Procedures were carried out in accordance with the Helsinki-Declaration and approved by the local ethics committee. All patients signed a written informed consent. Fresh frozen hippocampal segments derived from the epilepsy surgery were prepared as tissue slices via cryostat-conditions.

*Whole genome expression profiling and genome-wide genotyping*

Total DNA and RNA of hippocampus samples were isolated using the AllPrep DNA/RNA Micro Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. Quality of total RNA of each individual tissue sample was checked for degradation via gel electrophoresis in a BioAnalyzer 2100 (Agilent Technologies, Waldbronn, Germany) with RNA 6000 nano lab chips following the instructions of the manufacturers protocol. All RNA samples showed intact 28S and 18S ribosomal RNA signals and a RNA integrity number (RIN) of >7·9. All total RNA and DNA samples were used for a systematic chip-based genome-wide association study (GWAS) and gene expression (GEX) analysis. 50 ng of total RNA was reverse transcribed into cRNA and biotin-UTP labelled using the Illumina TotalPrep 96-RNA Amplification Kit (Ambion/Applied Biosystems, Darmstadt, Germany). Labelled cRNA was hybridized to Illumina human HT-12 Expression BeadChips using standard protocols (Illumina, San Diego, USA). Human HT-12 microarrays interrogate for more than 99.99% of all known human genes (approx. 25 000 annotated RefSeq and UniGene genes) containing more than 48 000 probes. All expression profiles were extracted and average/quantile/cubic spline/rank invariant normalized using GenomeStudio software (Illumina, San Diego, USA). For genome-wide SNP-genotyping 200 ng of DNA was hybridized to the Illumina Human660W-Quad v1 DNA Analysis BeadChip (> 658 000 markers per sample) according to

the Infinium® HD Assay Super manual from Illumina (Illumina, San Diego, USA). Allele calls were determined using Illumina BeadStudio (Illumina, San Diego, USA).[88]

*Imputation*

Prephasing was performed using the Shapeit software. Imputation was performed using the Impute2 software.[90] As reference the Phase I integrated variant set release (v3) (in NCBI build 37 (hg19) coordinates) was used (download and details can be found here: http://mathgen.stats.ox.ac.uk/impute/data_download_1000G_phase1_integrated.html). This dataset comprises for chromosomes 1-22 in total: 36 648 992 SNPs, 1 380 736 INDELs and 13 805 SVs. Only SNPs with an imputation Info-score > 0.8 were used for analysis.

*Probe annotation*

The nucleotide sequences of all probes embedded on the HumanHT-12 v3 Expression BeadChip were re-aligned to the UCSC version 18 of the human genome (hg18). The alignment was conducted in two steps using BWA,[91] and allowing only perfect matches. During the first step all probes were aligned to the mRNA sequences of the RefSeq database,[92] and the physical position was obtained using the UCSC "refGene" table for hg18. Probes that did not align to the mRNA sequences were aligned to the whole genome, here annotations were based on both the UCSC known genes and RefSeq genes. The physical position of Probes that could not be aligned in either step was obtained from the UCSC database ("illuminaProbes" table for hg18). Subsequently, probes containing either intrinsic polymorphisms as listed in HapMap-CEU (for hg18) or matched to multiple positions in the human genome were excluded from the downstream analysis. Furthermore, the physical position and gene annotations of all probes were based on the described realignment procedure. The remaining probes were pre-analysed using the GenomeStudioV2010.1 software from Illumina to merge data of replicated samples of the same individuals as well as to identify probes with a detection P-value < 0.01 in at least 5% of all samples. Data not matching these criteria were considered as non-expressed transcripts and dropped from further analysis. The remaining 15 426 probes (accounting for 13 842 distinct transcripts) were then normalized to consider background noise using the vsn2 option implemented in the package Variance stabilization and calibration for microarray data (VSN) for R.

*Analysis*

We applied a hidden factor analysis that identifies and corrects for unknown confounding factors in the data and thus diminishes the false-positive and false-negative eQTL rate (PEER).[93,94] Fifteen hidden factors were identified and used as co-variates for expression analysis. Statistical Analysis was performed using GenABEL® software.

*Data extraction for suggestive memory loci*

We used exploratory p-value thresholds of 0.05 for *cis*-eQTL (eQTL with distance between SNP and probe was $\leq$2 Mb) and $10^{-4}$ for *trans*-eQTLs, for a total number of 172 SNPs tested

(rs10924730 and rs13052909 could not be reliably imputed), representing 56 independent loci ($r^2$<0.25).

**7. PATHWAY ANALYSIS**

Pathway analyses were carried out using the core analysis function of the Ingenuity Pathway Analysis software (IPA©, Ingenuity Systems, www.ingenuity.com).

*Selection of focus molecules*

We performed gene-based tests for association based on the results from the PAR-dr and WL-dr GWAS (model adjusted for age and gender), using the Versatile Gene-based Association Study (VEGAS) software.[95] The full list of genes and gene-based p-values generated by VEGAS was uploaded into IPA for use as a reference set. From this list p-value cut-offs of 0.01 or 0.05 were used to identify IPA focus molecules. IPA will not accept a molecule if there is any ambiguity in the identifier or if there are no annotations for that molecule in the IPA database; therefore not all of the molecules in the input list were able to be “mapped” to IPA database objects. In the WL-dr dataset, of 17 898 genes, 16 965 were available for analysis. Applying a p-value cut-off of 0.01 left 175 focus molecules to be included in the analysis, while a cut-off of 0.05 resulted in 886 molecules. In the WL-dr dataset, of 17 874 genes, 16 953 were available for analysis. Applying a p-value cut-off of 0.01 left 183 focus molecules to be included in the analysis, while a cut-off of 0.5 resulted in 895 molecules.

*Generating IPA networks*

Networks generated by IPA provide insight into the molecular interactions of the focus molecules, independent of any predictions of biological function. For this analysis only “direct” interactions were used, that is, where there is physical contact between the two molecules. This could be a protein-protein interaction or a transcriptional regulation event. Chemical modifications are also included where it is known that one molecule acts on another without an intermediate agent. The relationships indicated between the nodes are based on one or more findings in the Ingenuity knowledge base.
The size of the network can be determined by the user, here the default IPA network size of 35 nodes was used. The program will incorporate as many focus molecules as possible into the network, and then add non-focus molecules from its internal database to make up the required size. Although the pool of focus molecules is defined using a p-value cut-off, the network building algorithm does not prioritize focus molecules based on p-value.
The network building algorithm ranks focus molecules by interconnectedness (the number of triangular connections with other pairs of genes). The top molecule is taken as a seed gene and additional focus molecules are added, prioritizing those that have most overlap with the existing network. The degree of overlap is calculated by considering the “neighborhood” of the potential new genes (all molecules one edge away) and dividing the number that are already in the network by the total number to give a specific overlap score.

Focus molecules are added until the target network size is reached. If no further focus molecules can be added and the target network size has not been reached, the algorithm will attempt to merge smaller networks by linking them through a single gene that connects to both networks. If the networks have still not reached the target size after this process, additional molecules will be added from the IPA global molecular network. When all focus molecules have been added to a network or the target number of networks has been reached, the networks will be displayed ranked by network score.
The score for a network represents the chance of finding that number of focus molecules in a network of that size randomly selected from the global molecular network. It is the –log(p-value) calculated using Fisher's exact test. The network score is not an indicator of biological significance and it is possible to get high scores using random molecule lists, due to the high interconnectedness of the global molecular network.

*Description of IPA networks obtained*

The networks reaching the highest score for each of the steps described above are shown in Supplementary Figure 3.
In the PAR-dr pathway analysis using a gene-based p-value cut-off of 0.01, the top 3 networks and networks 5 and 6 (ordered by decreasing scores, respectively of 33, 28, 26, 25, 24, 21) all include UBC (Ubiquitin C) as a central hub. UBC was also a central hub in networks 3, 4, 5, and 6 (respective scores 27, 26, 23 and 23) of the WL-dr pathway analysis using a gene-based p-value cut-off of 0.01. When using a gene-based p-value cut-off of 0.05 several PAR-dr and WL-dr networks also included UBC as a central hub. As UBC has a large number (N=8332) of directly related molecules, we calculated whether there are more UBC interactions in the whole focus molecule set than would be expected by chance, given the number of UBC interactions in the reference set, using the hypergeometric distribution test.

**8. AD PATHOLOGY**

We examined whether the memory risk alleles were associated with AD pathology, consisting of intracellular neurofibrillary tangles and extracellular amyloid plaques, in the Religious Orders Study and the Rush Memory and Aging Project (N=725).[96] Autopsy procedures have been described in detail previously.[49,97,98] Briefly, persons were classified as having pathologic AD based on intermediate or high likelihood of AD by National Institute on Aging–Reagan criteria using Consortium to Establish a Registry for Alzheimer's Disease estimates of neuritic plaque density and Braak staging of neurofibrillary pathology.[49,99-101]
We standardized the raw counts of neuritic plaques and neurofibrillary tangles by dividing each person's count by the standard deviation for that particular measure. The square-root of these standardized counts was used as dependent variables for the genetic association analyses.

**9. MEMORY GENETIC SCORE**

The memory genetic score comprised all independent SNPs ($r^2<0.25$[102]) associated with delayed recall at $p<5x10^{-6}$ (identified as lead SNPs in Supplementary Table 7). Each participant was assigned a score value, determined by summing up the number of copies (or imputed dosage) of each memory risk allele (i.e. the allele associated with weaker verbal declarative memory performance).[103] As most meta-analyses used effective sample size weighting, which does not yield any effect estimate, we have weighted each SNP in the score by the corresponding Z-score from the memory GWAS meta-analysis. Two versions of the score were computed, one with the *APOE* locus (58 SNPs) and one without (57 SNPs). One-sided p-values <0.05 were considered statistically significant.

To test the association of the memory genetic score with incident AD directly in the CHARGE meta-analysis of AD GWAS, without requiring additional study-specific analyses, we used the method implemented by the International Consortium for Blood Pressure GWAS:[104] Briefly, the beta coefficients of the independent SNPs on AD, corresponding to the effect of the memory risk allele, were summed up to examine if the cumulative effect of these SNPs on AD is different from zero. The corresponding Z-score from the memory GWAS meta-analysis were used as weight in the summation. Of note, for this method complete independence between SNPs is required, thus rs17053512 was removed from the score, as it is in weak LD with another more significant memory risk SNP in the list, rs6813517 ($r^2$ = 0.23).

## 10. SUPPLEMENTAL TABLES

## 10.1 Table S1: Demographic characteristics of participating cohorts

## Table S1a: Cohorts with paragraph delayed recall

| ***Discovery*** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Cohort** | **FHS** | **ORCADES** | **ROS** | **MAP** | **ARIC** | **CHS** | **LBC1921** | **LBC1936** |
| Exclusions for prevalent dementia (N) | 7 | 0 | 58 | 51 | 0 | 4 | 5 | 0 |
| Exclusions for prevalent stroke | 52 | 6 | 65 | 86 | 32 | 0 | 43 | 50 |
| N in analysis | 2493 | 419 | 687 | 751 | 430 | 472 | 469 | 953 |
| Age (years), mean±SD | 64.9±11.1 | 62.9±9.6 | 75.1±7.1 | 80.7±6.5 | 72.6±4.3 | 73.4±4.4 | 79.0±0.6 | 69.5±0.8 |
| Women, N (%) | 1373 (55.0) | 200 (57.1) | 450 (65.5) | 547 (72.8) | 258 (60.0) | 292 (61.9) | 279 (59.4) | 471 (49.4) |
| Education level: | | | | | | | | |
| < than high school [0 - 11 years] | 123 (4.9) | 228 (65.1) | 40 (5.3) | 17 (2.5) | 40 (9.3) | 75 (15.9) | 331(70.6) | 705 (74.0) |
| high school / GED degree [12 years] | 828 (33.2) | 38 (10.9) | 178 (23.7) | 35 (5.1) | 155 (36.0) | 137 (29.0) | 44 (9.4) | 141 (14.8) |
| some college [> 12 yrs], no college degree | 644 (25.8) | 30 (8.6) | 192 (25.6) | 33 (4.8) | 37 (8.6) | 98 (20.8) | 53 (11.3) | 54 (5.7) |
| college degree and above | 898 (36.3) | 54 (15.4) | 341 (45.4) | 602 (87.6) | 198 (46.0) | 162 (34.3) | 41 (8.7) | 53 (5.6) |

| ***Replication*** | | | | ***Extension to African-Americans*** |
|---|---|---|---|---|
| **Cohort** | **NHS** | **WGHS** | **Sydney MAS** | **ARIC-blacks** |
| Exclusions for prevalent dementia (N) | N/A | N/A | 0 | 0 |
| Exclusions for prevalent stroke | 0 | 0 | 37 | 65 |
| N in analysis | 2066 | 3542 | 887 | 416 |
| Age (years), mean±SD | 74.1±2.2 | 71.9±4.1 | 78.7±4.8 | 71.8 (4.4) |
| Women, N (%) | 2066 (100) | 3542 (100) | 495 (55.8) | 266 (63.9) |
| Education level: | | | | |
| < than high school [0 - 11 years] | 0 | 0 | 372 (41.9) | 138 (33.2) |
| high school / GED degree [12 years] | 0 | 0 | 172(19.4) | 76 (18.3) |
| some college [> 12 yrs], no college degree | 1577 (76.3) | 2359 (66.6) | 79 (8.9) | 28 ( 6.7) |
| college degree and above | 490 (23.7) | 1183 (33.4) | 264 (29.8) | 174 (41.8) |

## Table S1b: Cohorts with word list delayed recall

| *Discovery* | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Test** | **CVLT** | | **RAVLT** | | | | | **HVLT** | **DWRT** | **Rey** | | | **CERAD** | | |
| **Cohort** | **AGES** | **CHS** | **ERF** | **Croatia-Korcula** | **Croatia-Split** | **GENOA-whites** | **SHIP** | **TASCOG** | **ARIC** | **RS** | **RS-II** | **RS-III** | **HBCS** | **ROS** | **MAP** |
| Exclusions for prevalent dementia (N) | 126 | 47 | 0 | 0 | 0 | 0 | 62 | 2 | 0 | 124 | 56 | 6 | 0 | 58 | 51 |
| Exclusions for prevalent stroke | 266 | 26 | 21 | 25 | 10 | 22 | 19 | 15 | 275 | 168 | 102 | 0 | 19 | 65 | 86 |
| N in analysis | 2616 | 334 | 1267 | 472 | 303 | 758 | 1474 | 331 | 9188 | 2067 | 1533 | 1935 | 888 | 687 | 751 |
| Age (years), mean±SD | 75.9 (5.4) | 79.9 (4.0) | 58.8±8.9 | 61.1±9.1 | 58.3±8.1 | 61.3±8.8 | 60.8±10.1 | 71.8±6.9 | 57.2±5.7 | 63.9±5.8 | 67.8±7.0 | 56.0±5.7 | 68.2±2.9 | 75.1±7.1 | 80.7±6.5 |
| Women, N (%) | 1545 | 213 | 692 | 311 (65.9) | 185 (61.1) | 443 (58.1) | 753 (51.1) | 144 | 4889 (53.2) | 1193 | 827 | 1091 | 537 | 450 | 547 (72.8) |
| Education level: | | | | | | | | | | | | | | | |
| < than high school [0 - 11 years] | 589 (22.2) | 48 (14.4) | 542 | 173 (36.6) | 33 (10.9) | 40 (5.2) | 1063 | 127 (38) | 1449 (15.8) | 269 (13.0) | 304 | 192 (10.0) | 394 | 40 (5.3) | 17 (2.5) |
| high school / GED degree [12 years] | 1306 | 103 | 542 | 196 (41.5) | 126 (41.6) | 329 (43.2) | 44 (3.0) | 51 (16) | 3365 (36.6) | 628 (30.4) | 551 | 665 (34.7) | 69 (7.8) | 178 | 35 (5.1) |
| some college [>12yrs], no college | 422 (16.1) | 66 (19.8) | 148 | 50 (10.6) | 58 (19.1) | 246 (32.3) | 140 (9.5) | 109 (33) | 839 ( 9.1) | 912 (44.1) | 411 | 549 (28.6) | 163 | 192 | 33 (4.8) |
| college degree and above | 303 (11.6) | 117 (35.0) | 35 (2.8) | 53 (11.2) | 86 (28.4) | 147 (19.3) | 227 (15.4) | 44 (13) | 3535 (38.5) | 242 (11.7) | 247 (16.3) | 513 (26.7) | 174 (19.6) | 341 (45.4) | 602 (87.6) |

| *Replication* | | | | | | | | *Extension to African-Americans* | | | *Extension to young adults* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Test** | **CVLT** | | **RAVLT** | **HVLT** | **FCSRT** | **TICS** | | **RAVLT** | **DWRT** | **CERAD** | **Swiss test** |
| **Cohort** | **AGES-Repli*** | **BLSA** | **Sydney MAS** | **HCS** | **3C-Bordeaux** | **NHS** | **WGHS** | **GENOA-blacks** | **ARIC-blacks** | **REGARDS-blacks** | **Swiss MGS** |
| Exclusions for prevalent dementia (N) | 159 | 0 | 0 | 0 | 27 | N/A | N/A | 0 | 0 | 80 | 0 |
| Exclusions for prevalent stroke | 235 | 10 | 37 | 34 | 49 | 0 | 0 | 51 | 118 | 37 | 0 |
| N in analysis | 1525 | 712 | 880 | 821 | 1063 | 1773 | 3542 | 708 | 2476 | 627 | 1561 |
| Age (years), mean±SD | 76.2±5.4 | 65.3±12.2 | 78.7±4.8 | 64.0±7.4 | 75.1±4.8 | 74.2±2.3 | 71.9±4.1 | 63.3±8.2 | 56.0±5.7 | 62.7 (7.78) | 22.3±3.4 |
| Women, N (%) | 865 (56.7) | 313 (44.0) | 493 (56.0) | 389 (47.4) | 636 (60%) | 1773 (100) | 3542 (100) | 523 (72.6) | 1596 (64.5) | 413 (65.8) | 1078 (69.1%) |
| Education level: | | | | | | | | | | | |
| < than high school [0 - 11 years] | 430 (23) | 7 (1) | 370 (42.0) | 178 (21.7) | 610 (57.0) | 0 | 0 | 204 (28.3) | 930 (37.6) | 82 (13.1) | n.a. |
| high school / GED degree [12 years] | 930 (50) | 100 (14) | 172 (19.5) | 205 (24.9) | 115 (11.0) | 0 | 0 | 205 (28.5) | 537 (21.7) | 191 (30.4) | n.a. |
| some college [>12yrs], no college degree | 288 (19) | 64 (9) | 78 (8.9) | 218 (26.5) | 338 (32.0) | 1358 (76.6) | 2359 (66.6) | 127 (17.6) | 164 ( 6.6) | 164 (26.1) | n.a. |
| college degree and above | 196 (11) | 541 (76) | 260 (29.5) | 215 (26.2)† | | 415 (23.4) | 1183 (33.4) | 184 (25.6) | 845 (34.1) | 191 (30.4) | n.a. |

* *information on education is available only in 1 497 participants;* † *39 individuals (4.7%) are in "other" education category; n.a.: not available*

## 10.2 Table S2: Main characteristics of the delayed memory tests across cohorts

### Table S2a: Cohorts with paragraph delayed recall

| Cohort | Test name | Type of presentation | # Stories | Delay (minutes) | Maximum score | Mean score ± SD |
|---|---|---|---|---|---|---|
| ***Discovery*** | | | | | | |
| FHS | WMS | verbal | 1 | 20 | 24 | 10.2±3.8 |
| ORCADES | WMS | verbal | 1 | 20 | 25 | 13.0±5.0 |
| ROS | WMS-R | verbal | 1 | 30 | 25 | 10.6±4.3 |
| MAP | WMS-R | verbal | 1 | 30 | 25 | 8.9±4.5 |
| ARIC | WMS-R | verbal | 2 | 30 | 50 | 16.0±7.9 |
| CHS | WMS-R | verbal | 2 | 30 | 25 | 18.7±7.5 |
| LBC1921 | WMS-R | verbal | 2 | 30 | 25 | 12.7±7.0 |
| LBC1936 | WMSIII-UK | verbal | 2 | 30 | 50 | 27.2±8.1 |
| ***Replication*** | | | | | | |
| NHS | EBMT | verbal | 1 | 20 | 12 | 9.2±1.9 |
| WGHS | EBMT | verbal | 1 | 20 | 12 | 9.4±2.0 |
| Sydney MAS | WMSIII | verbal | 1 | 30 | 25 | 9.3±4.1 |
| ***Extension African-Americans*** | | | | | | |
| ARIC-blacks | WMS-R | verbal | 2 | 30 | 50 | 11.34±6.9 |

## Table S2b: Cohorts with word list delayed recall

| Cohort | Test name | Type of presentation | # Words in list | # Exposure trials | Delay (minutes) | Mean score ± SD |
|---|---|---|---|---|---|---|
| ***Discovery*** | | | | | | |
| AGES | CVLT | verbal | 16 | 4 | 15 | 6.2±3.1 |
| CHS | CVLT | verbal | 16 | 5 | 30 | 7.9±3.3 |
| TASCOG | HVLT | verbal | 12 | 3 | 20 | 7.7±2.9 |
| ERF | RAVLT | verbal | 15 | 5 | 20 | 6.5±2.8 |
| Croatia-Korčula | RAVLT | verbal | 15 | 5 | 15 | 9.5±2.8 |
| Croatia-Split | RAVLT | verbal | 15 | 5 | 15 | 11.2±2.5 |
| GENOA-whites | RAVLT | verbal | 15 | 5 | 30 | 9.1±3.3 |
| SHIP | RAVLT | verbal | 15 | 3 | 20 | 7.5±3.0 |
| ARIC | DWRT | visual (VPWL) | 10 | 2 | 5 | 6.8±1.4 |
| Rotterdam Study | Rey's recall of words | visual (VPWL) | 15 | 3 | 10 | 6.3±2.7 |
| Rotterdam Study-II | Rey's recall of words | visual (VPWL) | 15 | 3 | 10 | 6.3±2.7 |
| Rotterdam Study-III | Rey's recall of words | visual (VPWL) | 15 | 3 | 10 | 6.3±2.7 |
| HBCS | CERAD-list | visual (VPWL) | 10 | 3 | 5 | 7.4±1.7 |
| ROS | CERAD-list | visual (VPWL) | 10 | 3 | 3 | 5.8±2.1 |
| MAP | CERAD-list | visual (VPWL) | 10 | 3 | 3 | 5.1±2.4 |
| ***Replication*** | | | | | | |
| AGES-plus | CVLT | verbal | 16 | 4 | 15 | 6.2±3.1 |
| BLSA | CVLT | verbal | 16 | 5 | 15-20 | 11.2±3.4 |
| NHS | TICS | verbal | 10 | 1 | 20 | 2.4±2.0 |
| WGHS | TICS | verbal | 10 | 1 | 20 | 3.4±2.3 |
| HCS | HVLT | verbal | 12 | 3 | 20 | 8.8±2.5 |
| Sydney MAS | RAVLT | verbal | 15 | 5 | 30 | 7.5±3.5 |
| 3C-Bordeaux | FCSRT | visual | 16 | 3 | 20 | 9.3±3.0 |
| ***Extension African-Americans*** | | | | | | |
| GENOA-blacks | RAVLT | verbal | 15 | 5 | 30 | 6.8±3.4 |
| ARIC-blacks | DWRT | visual | 10 | 2 | 5 | 6.2±1.6 |
| REGARDS-blacks | CERAD-list | visual (VPWL) | 10 | 3 | 5 | 6.3±2.1 |
| ***Extension Young*** | | | | | | |
| Swiss MGS | Swiss test | visual | 30 | 1 | 5 | 8.0±3.1 |

## 10.3. Table S3: Correlation between performance on paragraph delayed recall and word list delayed recall

### Table S3a: Correlation between PAR-dr and WL-dr in ARIC

| | **WL-dr (DWRT)** | **PAR-dr (WMS)** | **Exe/PS DSST** | **Exe/PS TMT-A** | **Exe/PS TMT-B** | **Exe/PS Stroop** | **Exe/PS WFT** | **Exe/PS An-Naming** |
|---|---|---|---|---|---|---|---|---|
| WL-dr (DWRT) | 1<br><br>507 | | | | | | | |
| PAR-dr (WMS) | 0.35<br><0.0001<br>503 | 1<br><br>504 | | | | | | |
| Exe/PS DSST | 0.31<br><0.0001<br>504 | 0.24<br><0.0001<br>502 | 1<br><br>505 | | | | | |
| Exe/PS TMT-A | -0.23<br><0.0001<br>504 | -0.13<br>0.0026<br>502 | -0.53<br><0.0001<br>504 | 1<br><br>505 | | | | |
| Exe/PS TMT-B | -0.26<br><0.0001<br>504 | -0.32<br><0.0001<br>502 | -0.61<br><0.0001<br>504 | 0.55<br><0.0001<br>504 | 1<br><br>505 | | | |
| Exe/PS Stroop | -0.12<br>0.0073<br>498 | -0.02<br>0.64<br>496 | -0.17<br>0.0002<br>498 | 0.13<br>0.0032<br>498 | 0.09<br>0.044<br>498 | 1<br><br>499 | | |
| Exe/PS WFT | 0.25<br><0.0001<br>507 | 0.16<br>0.0003<br>504 | 0.41<br><0.0001<br>505 | -0.31<br><0.0001<br>505 | -0.40<br><0.0001<br>505 | -0.0037<br>0.93<br>499 | 1<br><br>508 | |
| Exe/PS An-Naming | 0.21<br><0.0001<br>507 | 0.25<br><0.0001<br>504 | 0.36<br><0.0001<br>505 | -0.29<br><0.0001<br>505 | -0.38<br><0.0001<br>505 | -0.0012<br>0.98<br>499 | 0.46<br><0.0001<br>508 | 1<br><br>508 |

*Numbers from top to bottom correspond to Pearson correlation coefficient, p-value, and number of observations; An-Naming: Animal Naming; DWRT, Delayed Word Recall Test; DSST, Digit Symbol Substitution Test; Exe/PS: Executive Function / Processing Speed; PAR-dr: Paragraph delayed recall; Stroop: Stroop Interference Test; TMT-A, Trails Making Test A; TMT-B, Trails Making Test B; WFT, Word Fluency Test; WL-dr: Word List delayed recall*

## Table S3b: Correlation between PAR-dr and WL-dr in CHS

| | WL-dr (CVLT) | PAR-dr (WMS) | Exe/PS DSST | Exe/PS TMT-A | Exe/PS TMT-B | Exe/PS TMT-B-A |
|---|---|---|---|---|---|---|
| WL-dr (CVLT) | 1<br><br>334 | | | | | |
| PAR-dr (WMS) | 0.32<br><0.0001<br>196 | 1<br><br>472 | | | | |
| Exe/PS DSST | 0.34<br><0.0001<br>322 | 0.32<br><0.0001<br>452 | 1<br><br>2022 | | | |
| Exe/PS TMT-A | -0.22<br><0.0001<br>307 | -0.13<br>0.022<br>327 | -0.25<br><0.0001<br>1239 | 1<br><br>1297 | | |
| Exe/PS TMT-B | -0.30<br><0.0001<br>296 | -0.24<br><0.0001<br>327 | -0.45<br><0.0001<br>1239 | 0.47<br><0.0001<br>1267 | 1<br><br>1267 | |
| Exe/PS TMT-B-A | -0.25<br><0.0001<br>296 | -0.21<br><0.0001<br>323 | -0.39<br><0.0001<br>1239 | 0.007<br>0.80<br>1267 | 0.88<br><0.0001<br>1267 | 1<br><br>1267 |

*Numbers from top to bottom correspond to Pearson correlation coefficient, p-value, and number of observations; CVLT: California Verbal Learning Test; DSST, Digit Symbol Substitution Test; Exe/PS: Executive Function / Processing Speed; PAR-dr: Paragraph delayed recall; TMT-A, Trails Making Test A; TMT-B, Trails Making Test B; TMT-B-A, Trails Making Test B-A; WFT, Word Fluency Test; WL-dr: Word List delayed recall*

**Table S3c: Correlation between PAR-dr and WL-dr in Sydney MAS**

| | WL-dr (RAVLT) | PAR-dr (WMS) | Exe/PS DSST | Exe/PS TMT-A | Exe/PS TMT-B | Exe/PS TMT-B-A | Exe/PS WFT | Exe/PS An-Naming |
|---|---|---|---|---|---|---|---|---|
| WL-dr (RAVLT) | 1<br><br>880 | | | | | | | |
| PAR-dr (WMS) | 0.46<br><0.0001<br>879 | 1<br><br>887 | | | | | | |
| Exe/PS DSST | 0.23<br><0.0001<br>865 | 0.21<br><0.0001<br>872 | 1<br><br>873 | | | | | |
| Exe/PS TMT-A | -0.23<br><0.0001<br>864 | -0.20<br><0.0001<br>871 | -0.56<br><0.0001<br>872 | 1<br><br>872 | | | | |
| Exe/PS TMT-B | -0.23<br><0.0001<br>815 | -0.23<br><0.0001<br>820 | -0.61<br><0.0001<br>821 | 0.58<br><0.0001<br>820 | 1<br><br>821 | | | |
| Exe/PS TMT-B-A | -0.18<br><0.0001<br>814 | -0.20<br><0.0001<br>819 | -0.49<br><0.0001<br>820 | 0.26<br><0.0001<br>820 | 0.92<br><0.0001<br>820 | 1<br><br>820 | | |
| Exe/PS WFT | 0.22<br><0.0001<br>876 | 0.169<br><0.0001<br>882 | 0.40<br><0.0001<br>868 | -0.25<br><0.0001<br>867 | -0.35<br><0.0001<br>817 | -0.33<br><0.0001<br>816 | 1<br><br>883 | |
| Exe/PS An-Naming | 0.26<br><0.0001<br>877 | 0.32<br><0.0001<br>883 | 0.42<br><0.0001<br>869 | -0.33<br><0.0001<br>868 | -0.37<br><0.0001<br>818 | -0.31<br><0.0001<br>817 | 0.38<br><0.0001<br>883 | 1<br><br>884 |

*Numbers from top to bottom correspond to Pearson correlation coefficient, p-value, and number of observations; DSST, Digit Symbol Substitution Test; Exe/PS: Executive Function / Processing Speed; PAR-dr: Paragraph delayed recall; RAVLT: Rey's Auditory Verbal Learning Test; TMT-A, Trails Making Test A; TMT-B, Trails Making Test B; TMT-B-A, Trails Making Test B-A; WFT, Word Fluency Test; WL-dr: Word List delayed recall*

## 10.4. Table S4: Genotyping parameters

| Study | Genotyping platforms - SNP panel | Gentoyping center | Genotype calling |
|---|---|---|---|
| AGES | Illumina HumanCNV370 Duo BeadChip® | NIA, NIH, USA | Illumina Bead Studio |
| ARIC-blacks | Affymetrix GeneChip SNP Array 6.0® | Broad Institute, USA | Birdseed |
| ARIC-whites | Affymetrix GeneChip SNP Array 6.0® | Broad Institute, USA | Birdseed |
| BLSA | Illumina 550K genotype calling algorithm | Laboratory of Neurogenetics, NIA genotyping platform | Illumina Bead Studio |
| CHS | Illumina HumanCNV370 Duo BeadChip® | Genotyping Laboratory at Cedars-Sinai, USA | Illumina Bead Studio |
| Croatia- Korcula | Illumina HumanHap370-Duo® and HumanHap370-Quad BeadChip® | Helmholtz Centre, Munich, D | Illumina Bead Studio |
| Croatia-Split | Illumina HumanHap370-Quad BeadChip® | AROS Applied Biotechnology, Aarhus, DK | Illumina Bead Studio |
| ERF | Illumina HumanHap 300K array®, Illumina HumanHap 6k Beadchip®, Illumina Human 370K-Duo SNP array®, Affymetrix GeneChip® Human Mapping 250K Nsp Array | Leiden University Medical Center, Leiden Erasmus MC, Rotterdam, NL | Illumina BeadStudio& Affymetrix BRLMM |
| FHS | Affymetrix GeneChip Human Mapping 500K Array® +50K Human Gene Focused Panel® | Affymetrix (Santa Clara), USA | Affymetrix BRLMM |
| GENOA-blacks | Affymetrix GeneChip SNP Array 6.0® and Illumina Human1M-Duo BeadChip® (N=118) | Mayo Clinic, Rochester (MN), USA | Birdseed |
| GENOA-whites | Affymetrix GeneChip SNP Array 6.0® | Mayo Clinic, Rochester (MN), USA | Birdseed |
| HBCS | modified Illumina Infinum 610K Quad chip® | Wellcome Trust Sanger Institute, Cambridge, UK | Illumina Bead Studio |
| HCS | Illumina Human610-Quad BeadChip® | Hunter Medical Research Institute, Newcastle Australia | Genomestudio |
| LBC1921 | Ilumina 610-Quadv1 | Wellcome Trust Clinical Research Facility (WTCRF) Genetics | Illumina Bead Studio |
| LBC1936 | Ilumina 610-Quadv1 | Core, Western General Hospital, Edinburgh | Illumina Bead Studio |
| MAP | Affymetrix Genechip 6.0® | Broad Institute; Translational Genomics Research Institute, USA | Birdsuite, Broad Institute |
| NHS | Affymetrix Genechip 6.0®, Illumina Infinium Sentrix HumanHap550, Illumina 610Q, Illumina 660, Omni Express | Broad Institute, USA | Birdseed |
| ORCADES | Illumina HumanCNV370-Duo and –HumanHap300K | Helmholtz Centre, Munich, D and Integragen, Paris, F | Illumina GenCall |
| REGARDS | Affymetrix Genechip 6.0® | University of Virginia | Affymetrics Power Tools, apt-probeset-genotype |
| ROS | Affymetrix Genechip 6.0® | Broad Institute; Translational Genomics Research Institute, USA | Birdsuite, Broad Institute |
| Rotterdam Study | Illumina HumanHap550-Duo BeadChip® | Erasmus MC, Rotterdam, NL | Illumina Bead Studio |
| Rotterdam Study-II | Illumina HumanHap550 Duo BeadChip® and Illumina Human 610 Quad BeadChip® | Erasmus MC, Rotterdam, NL | Illumina Genome Studio |
| Rotterdam Study-III | Illumina Human 610 Quad BeadChip® | Erasmus MC, Rotterdam, NL | Illumina Bead Studio |
| SHIP | Affymetrix GeneChip SNP Array 6.0® | Affymetrix (Santa Clara), USA | Birdseed2 |
| Sydney MAS | Affymetrix GeneChip SNP Array 6.0® | Ramaciotti Centre, University of New South Wales, Sydney Australia | CRLMM (v1.10.0) in R |
| Swiss MGS | Affymetrix GeneChip SNP Array 6.0® | Biozentrum, University of Basel, Switzerland | Birdseed |
| TASCOG | Illumina HumanCNV370-Duo BeadChip® | Diamantina Institute and Institute of Molecular Biosciences, University of Queensland, AU | Illumina GenCall |
| WGHS | HumanHap300 Duo ‘‘+’’ chips or HumanHuman300 Duo + iSelect chips | Amgen | Illumina Bead Studio |
| 3C-Bordeaux | Illumina Human 610-Quad BeadChip | Centre National de Génotypage | Illumina BeadStudio |

## 10.5. Table S5: Quality control filters before imputation and methods for assessing population structure

| Study | Sample call rate | SNP call rate | MAF | HWE p-value | Assessment of Population Stratification |
|---|---|---|---|---|---|
| AGES | < 97% | < 98% | < 0.01 | < 10-6 | EIGENSTRAT |
| ARIC-blacks | < 95% | < 95% | < 0.01 | < 10-5 | EIGENSTRAT |
| ARIC-whites | < 95% | < 95% | < 0.01 | < 10-5 | EIGENSTRAT |
| BLSA | < 98.5% | < 99% | < 0.01 | < 10-4 | EIGENSTRAT |
| CHS | ≤ 95% | < 97% | < 0.01 | < 10-5 | PCA |
| Croatia-Korcula | < 97% | < 98% | < 0.01 | < 10-6 | PCA |
| Croatia-Split | < 97% | < 98% | < 0.01 | < 10-6 | PCA |
| ERF | < 95% | < 98% | < 0.005 | < 10-6 | n.a. |
| FHS | < 97% | < 97% | < 0.01 | < 10-6 | EIGENSTRAT |
| GENOA-blacks | < 95% | < 95% | < 0.01 | n.a. | PCA |
| GENOA-whites | < 95% | < 95% | < 0.01 | n.a. | PCA |
| HBCS | < 95% | < 95% | < 0.01 | < 10-5 | MDS |
| HCS | < 95% | < 95% | < 0.01 | < 10-6 | EIGENSTRAT |
| LBC1921 | < 95% | < 98% | < 0.01 | < 10-3 | MDS |
| LBC1936 | < 95% | < 98% | < 0.01 | < 10-3 | MDS |
| MAP | < 95% | < 95% | < 0.01 | < 10-6 | EIGENSTRAT |
| NHS | < 95% | < 98% | n.a. | < 10-4 | EIGENSTRAT |
| ORCADES | < 97% | < 98% | < 0.01 | < 10-6 | MDS |
| REGARDS | < 90% | < 97% | < 0.01 | < 10-6 | EIGENSTRAT 2.0 |
| ROS | < 95% | < 95% | < 0.01 | < 10-6 | EIGENSTRAT |
| Rotterdam Study | < 97.5% | < 98% | < 0.01 | < 10-6 | IBD matrix |
| Rotterdam Study-II | < 97.5% | < 98% | < 0.01 | < 10-6 | IBD matrix |
| Rotterdam Study-III | < 97.5% | < 98% | < 0.01 | < 10-6 | IBD matrix |
| SHIP | < 92% | < 92% | n.a. | n.a. | PCA and MDS |
| Sydney MAS | < 95% | < 95% | < 0.01 | < 10-6 | EIGENSTRAT |
| Swiss MGS | < 95% | < 95% | < 0.05 | 0.05 | EIGENSTRAT |
| TASCOG | < 97% | < 97% | < 0.005 | < 10-7 | EIGENSTRAT |
| WGHS | < 98% | < 90% | < 0.01 | < 10-6 | EIGENSTRAT |
| 3C-Bordeaux | < 95% | < 98% | < 0.01 | < 10-6 | EIGENSTRAT |

*PCA: Principal Component Analysis; MDS: Multidimensional Scaling; HWE: Hardy-Weinberg Equilibrium*

## 10.6. Table S6: Imputation algorithms

| Study | Imputation software | Imputation reference panel |
|---|---|---|
| AGES | MACH | HapMap II CEU (build 36, release 22) |
| ARIC-blacks | MACH (v1.0.16) | HapMap II CEU and YRI (build 35, release 21) |
| ARIC-whites | MACH (v1.0.16) | HapMap II CEU (build 36) |
| BLSA | MACH | HapMap II CEU (build 36) |
| CHS | BIM-BAM15 | HapMap II CEU (build 36) |
| Croatia- Korcula | MACH | HapMap II CEU (build 36) |
| Croatia-Split | MACH | HapMap II CEU (build 36) |
| ERF | MACH | HapMap II CEU (build 36) |
| FHS | MACH | HapMap II CEU (build 36) |
| GENOA-blacks | MACH | HapMap II, 60 unrelated YRI and 60 unrelated CEU, build 36, release 22 |
| GENOA-whites | MACH | HapMap II, 60 unrelated CEU, build 36, release 22 |
| HBCS | MACH | HapMap II CEU |
| HCS | MACH (version 1.0.16) | HapMap II CEU (build 36.1, release 24) |
| LBC1921 | MACH | HapMap II CEU |
| LBC1936 | MACH | HapMap II CEU |
| MAP | MACH (version 1.0.16a) | HapMap II CEU (build 36, release 22) |
| NHS | MACH | HapMap II CEU (build 36, release 22) |
| ORCADES | MACH | HapMap II CEU (build 36, release 22) |
| REGARDS | MACH | HapMap III |
| ROS | MACH (version 1.0.16a) | HapMap II CEU (build 36, release 22) |
| Rotterdam Study | MACH (v1.0.15) | HapMap II CEU (build 36, release 22) |
| Rotterdam Study-II | MACH (v1.0.16) | HapMap II CEU (build 36, release 22) |
| Rotterdam Study-III | MACH (v1.0.16) | HapMap II CEU (build 36, release 22) |
| SHIP | IMPUTE (v0.5.0) | HapMap II |
| Sydney MAS | MACH / minimac | HapMap II CEU (build 36, release 22) |
| Swiss MGS | IMPUTE | 1000G, March 2012 release, CEU (build 37) |
| TASCOG | MACH (v1.0.16) | HapMap II CEU (build 36) |
| WGHS | MACH (v1.0.15) | HapMap II CEU (build 36, release 22) |
| 3C-Bordeaux | IMPUTE (v2.2) | HapMap II CEU (build 36.3, release 22) |

### 10.7. Table S7: Results and annotation for all SNPs yielding suggestive associations with delayed recall

Lead-SNP: SNP with lowest p-value in locus (500kb window), or in weak LD ($r^2$<0·25) with the most significant SNP in the locus (if several SNPs in weak LD with top SNP in locus were in LD with each other, the most significant one was retained) – 58 lead SNPs were identified and used to compute the memory genetic score (Appendix, Section 9); Gene1: closest gene; Dist1: distance of the closest gene; Gene2: second closest gene; Dist2: distance of the second closest gene; A1: coded allele; A2: non-coded allele; Beta_A1: effect estimate for A1; SE_A1: standard error of effect estimate for A1; Risk_All: risk allele (i.e. allele associated with poorer memory performance; RAF: risk allele frequency; Beta_RA: effect estimate for RA; SE_RA: standard error of effect estimate for RA; Pval_GC: p-value after genomic control; Direction_meta: direction of effects in each study included in the meta-analysis (listed in alphabetical order); HetPval: p-value for test of heterogeneity

### 10.8. Table S8: Association of top memory hits with Amyloid plaques and Neurofibrillary Tangles

Lead-SNP: SNP with lowest p-value in locus (500kb window), or in weak LD ($r^2$<0·25) with the most significant SNP in the locus (if several SNPs in weak LD with top SNP in locus were in LD with each other, the most significant one was retained) – 58 lead SNPs were identified and used to compute the memory genetic score (Appendix, Section 9)

### 10.9. Table S9: Association of top memory hits with hippocampal mRNA levels (eQTL)

Distance: distance between SNP and probe start; Info: imputation quality metric from IMPUTE2

*NOTE: the tables mentioned on this page are too large to print, and are digitally available upon request – c.ibrahim-verbaas@erasmusmc.nl*

## 10.10. Table S10: Hypergeometric distribution test for Ubiquitin C (UBC) interactions

| Dataset | UBC Associated Networks | Non UBC Associated Networks | Total Networks | Hypergeometric Distribution Test* | |
|---|---|---|---|---|---|
| | **N** | **N** | **N** | **Exact *P-value*** | **Sum *P-value*** |
| WL-dr p<0.01 | 86 | 89 | 175 | 0.038 | 0.184 |
| WL-dr p<0.05 | 434 | 457 | 891 | 0.004 | 0.026 |
| WL-dr reference set | 7771 | 9312 | 17083 | - | - |
| PAR-dr p<0.01 | 97 | 87 | 184 | 0.008 | 0.040 |
| PAR-dr p<0.05 | 424 | 437 | 861 | 0.002 | 0.015 |
| PAR-dr reference set | 7767 | 9305 | 17072 | - | - |

* *Test assessing the probability for UBC interactions in the focus molecule set to occur more often than by chance, given the number of UBC interactions in the reference set: Three out of four datasets (WL-dr with p<0.05, PAR-dr with p<0.01 and p<0.05) appear to be enriched for UBC interaction at sum p-value p<0.05*

## 10.11. Table S11: Association with published verbal memory risk variants

| Gene | SNP | Risk Allele | Proxy | Papers | Combined meta-analyses | | | | | | Meta-analyses for main subtests of WL-dr | | | | | | Gene-based analyses | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | ALL-dr | | PAR-dr | | WL-dr | | RAVLT-dr | | CVLT-dr | | CERAD-dr | | PAR-dr | WL-dr |
| | | | | | Dir | Pval | Dir | Pval | Dir | Pval | Dir | Pval | Dir | Pval | Dir | Pval | gene-pval | gene-pval |
| **AKT1** | rs1130214 | A | | Pietiläinen [105], 2009 | - | 0.48 | - | 0.83 | - | 0.47 | + | 0.16 | - | 0.06 | - | 0.75 | 0.25 | 0.45 |
| **BDNF** | rs6265 | G | | Egan [106], 2003; Miyajima [107], 2008 | + | 0.25 | + | 0.65 | + | 0.22 | + | 0.93 | + | 0.17 | + | 0.57 | 0.51 | 0.53 |
| **CAMTA1** | rs4908449 | T | | Huentelman [108], 2007 | - | 0.45 | + | 0.24 | - | 0.27 | - | 0.64 | + | 0.22 | + | 0.97 | 0.67 | 0.31 |
| **CTNNBL1** | rs16986890 | G | | Papassotiropoulos [71], 2011 | + | 0.08 | + | 0.30 | + | 0.18 | + | 0.07 | - | 0.69 | + | 0.72 | 0.78 | 0.72 |
| **DTNBP1** | rs2619522 | G | | Luciano [109], 2009 | + | 0.09 | + | 0.98 | + | 0.08 | + | 0.27 | + | 0.017 | - | 0.16 | 0.74 | 0.48 |
| **ERBB4** | rs2272024 | G | | Nicodemus [110], 2006 | + | 0.53 | + | 0.75 | + | 0.81 | + | 0.33 | + | 0.71 | + | 0.83 | | |
| | rs3791709* | T | | | - | 0.45 | - | 0.52 | - | 0.70 | + | 0.36 | - | 0.98 | + | 0.047 | 0.68 | 0.72 |
| | rs4673628* | T | | | - | 0.72 | - | 0.32 | + | 0.87 | + | 0.74 | - | 0.07 | + | 0.06 | | |
| **HTR2A** | rs6314 | T | | De Quervain [77], 2003; Papassotiropoulos [111], 2005; Reynolds [112], 2006 | - | 0.23 | - | 0.69 | - | 0.27 | + | 0.29 | - | 0.82 | - | 0.81 | 0.08 | 0.31 |
| | rs6311† | A | | | - | 0.19 | - | 0.85 | - | 0.10 | + | 0.83 | + | 0.11 | + | 0.35 | | |
| **Il1B** | rs16944 | G | | Baune [113], 2008 | + | 0.83 | - | 0.58 | + | 0.41 | - | 0.10 | + | 0.34 | - | 0.78 | 0.48 | 0.82 |
| **KIBRA** | rs17070145 | C | | Papassotiropoulos [114], 2006; Schaper [115], 2008; Almeida [116], 2008; Yasuda [117], 2010; Preuschhof [118], 2010 ; Milnik [119], 2012 | + | 0.34 | + | 0.0067 | - | 0.72 | - | 0.96 | - | 0.59 | - | 0.78 | 0.40 | 0.91 |
| **KLK8** | rs1612902 | G | | Izumi [120], 2008 | + | 0.40 | + | 0.80 | + | 0.33 | + | 0.53 | + | 0.64 | + | 0.26 | 0.87 | 0.29 |
| **PDYN** | rs1997794 | T | | Kolsch [121], 2009 | - | 0.55 | + | 0.28 | - | 0.23 | - | 0.61 | + | 0.46 | - | 0.95 | 0.39 | 0.59 |
| | rs910080‡ | A | | | + | 0.59 | + | 0.028 | - | 0.78 | - | 0.28 | + | 0.08 | + | 0.57 | | |
| **WRN** | rs2725335 | A | | Houlihan [122], 2009 | - | 0.49 | - | 0.024 | + | 0.42 | + | 0.63 | n.a. | n.a. | + | 0.45 | 0.62 | 0.80 |
| **CPEB3** | rs11186856 | C | | Vogler [123], 2009 | | | | | | | | | | | | | 0.31 | 0.75 |
| | | | rs6583802‖ | | - | 0.13 | - | 0.57 | - | 0.34 | + | 0.62 | - | 0.42 | - | 0.85 | | |
| **PTPRO** | rs17222089 | A | | LeBlanc [124], 2012 | - | 0.18 | + | 0.62 | - | 0.12 | + | 0.12 | - | 0.007 | - | 0.81 | 0.97 | 0.34 |
| | rs11056571§ | G | | | - | 0.17 | + | 0.64 | - | 0.12 | + | 0.12 | - | 0.0072 | - | 0.40 | | |
| | rs2300290§ | G | | | - | 0.15 | + | 0.69 | - | 0.11 | + | 0.11 | - | 0.0085 | - | 0.91 | | |

*p-values <0.0029 and gene-based pvalues < 0.0036 were considered statistically significant after Bonferroni correction for the number of independent SNPs / genes tested; +: the reported risk allele is associated with better delayed recall performance in the present study; -: the reported risk allele is associated with worse delayed recall performance in the present study; ALL: combined analysis of all studies; PAR: paragraph delayed recall; WL: word list delayed recall; CVLT: California Verbal Learning Test; CVLT_HVLT: California Verbal Learning Test or Hopkins Verbal Learning Test (verbally presented, semantically related word list); CERAD: Consortium to Establish a Registry for Alzheimer's Disease word list (visually presented, semantically unrelated word list); VPWL: combined analysis of all studies using visually presented, semantically unrelated word lists); n.a.: not available; * $r^2$ with rs2272024 is 0.072 for rs3791709 and 0.006 for rs4673628; † $r^2$ with rs2272024 is 0.004 for rs6311; ‡ $r^2$ with rs1997794 is 0.423 for rs910080; § $r^2$ with rs17222089 is 1.00 for rs11056571 and 0.951 for rs2300290; ‖ $r^2$ = 1 with rs11186856*

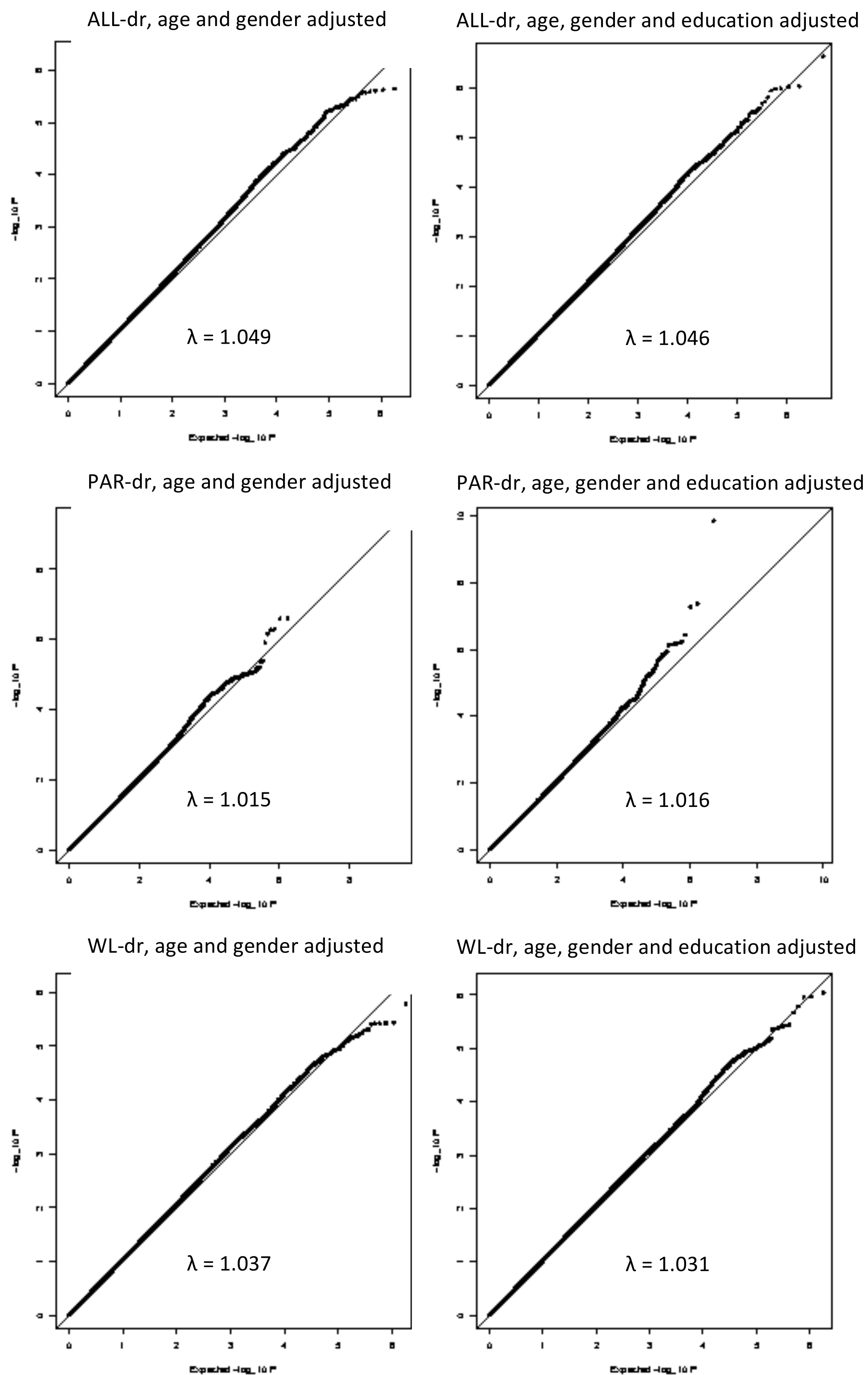
## 10.12. Table S12: Association with published AD risk variants

| | | AD | Combined meta-analyses | | | | | | Meta-analyses for main subtests of WL-dr | | | | | | Gene-based analyses | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Risk | ALL-dr | | PAR-dr | | WL-dr | | RAVLT-dr | | CVLT-dr | | CERAD-dr | | PAR-dr | WL-dr |
| **Gene** | **SNP** | **Allele** | **Dir** | **Pval*** | **Dir** | **Pval*** | **Dir** | **Pval*** | **Dir** | **Pval*** | **Dir** | **Pval*** | **Dir** | **Pval*** | **gene-pval** | **gene-pval** |
| ***Adjusted for age and gender*** | | | | | | | | | | | | | | | | |
| **APOC1/APOE** | rs4420638 | **G** | - | $9.31x10^{-7}$ | - | $5.571x0^{-10}$ | - | $3.31x10^{-4}$ | - | 0.22 | - | 0.023 | - | $7.61x10^{-4}$ | 0.0012 / 0.0015 | 0.017 / 0.018 |
| **PVRL2** | rs6857 | **T** | - | $2.21x10^{-5}$ | - | $3.06x10^{-7}$ | - | $1.52x10^{-3}$ | - | 0.73 | - | 0.18 | - | $7.97x10^{-3}$ | $1.27x10^{-4}$ | 0.06 |
| **TOMM40** | rs2075650 | **G** | - | $3.37x10^{-5}$ | - | $3.15x10^{-7}$ | - | 0.0036 | - | 0.64 | - | 0.29 | - | 0.0026 | $2.24x10^{-4}$ | 0.017 |
| **CLU** | rs11136000 | **C** | - | 0.0090 | - | 0.045 | - | 0.020 | - | 0.21 | - | 0.33 | - | 0.78 | 0.28 | 0.36 |
| **EPHA1** | rs11767557 | **T** | - | 0.32 | - | 0.019 | + | 0.78 | + | 0.061 | + | 0.70 | - | 0.79 | 0.53 | 0.16 |
| **CD2AP** | rs9349407 | **C** | - | 0.24 | + | 0.46 | - | 0.17 | - | 0.17 | - | 0.07 | + | 0.017 | 0.19 | 0.23 |
| **PICALM** | rs3851179 | **C** | - | 0.25 | - | 0.38 | - | 0.16 | - | 0.14 | - | 0.95 | - | 0.83 | 0.41 | 0.07 |
| **MS4A6A** | rs610932 | **G** | - | 0.19 | - | 0.72 | - | 0.52 | - | 0.26 | + | 0.93 | + | 0.39 | 0.19 | 0.68 |
| **BIN1** | rs744373 | **G** | - | 0.28 | + | 0.82 | - | 0.21 | - | 0.05 | - | 0.53 | - | 0.37 | 0.96 | 0.39 |
| **CD33** | rs3865444 | **C** | - | 0.38 | - | 0.64 | - | 0.91 | + | 0.96 | + | 0.84 | + | 0.3 | 0.51 | 0.22 |
| **CR1** | rs3818361 | **A** | - | 0.63 | - | 0.09 | - | 0.95 | - | 0.37 | + | 0.72 | - | 0.27 | 0.21 | 0.92 |
| **ABCA7** | rs3764650 | **G** | - | 0.93 | - | 0.20 | + | 0.65 | + | 0.20 | - | 0.19 | + | 0.39 | 0.045 | 0.30 |
| ***Adjusted for age, gender and education*** | | | | | | | | | | | | | | | | |
| **APOC1** | rs4420638 | **G** | - | $4.28x10^{-7}$ | - | $1.94x10^{-10}$ | - | $2.26x10^{-4}$ | - | 0.32 | - | $5.98x10^{-3}$ | - | $1.30x10^{-3}$ | $6.5x10^{-4}$ | $9.2x10^{-3}$ |
| **PVRL2** | rs6857 | **T** | - | $5.12x10^{-6}$ | - | $5.36x10^{-8}$ | - | $7.20x10^{-4}$ | - | 0.73 | - | 0.06 | - | 0.017 | $6.3x10^{-5}$ | 0.05 |
| **TOMM40** | rs2075650 | **G** | - | $1.66x10^{-5}$ | - | $6.77x10^{-8}$ | - | 0.0030 | - | 0.64 | - | 0.14 | - | 0.0071 | $9.7x10^{-5}$ | 0.01 |
| **CLU** | rs11136000 | **C** | - | 0.0054 | - | 0.072 | - | 0.011 | - | 0.24 | - | 0.29 | - | 0.84 | 0.37 | 0.24 |
| **EPHA1** | rs11767557 | **T** | - | 0.25 | - | 0.0040 | + | 0.77 | + | 0.053 | + | 0.64 | - | 0.61 | 0.3 | 0.22 |
| **CD2AP** | rs9349407 | **C** | - | 0.22 | + | 0.50 | - | 0.15 | - | 0.21 | - | 0.042 | + | 0.022 | 0.29 | 0.07 |
| **PICALM** | rs3851179 | **C** | - | 0.21 | - | 0.42 | - | 0.12 | - | 0.09 | - | 0.98 | - | 0.68 | 0.4 | 0.08 |
| **MS4A6A** | rs610932 | **G** | - | 0.21 | + | 0.92 | - | 0.47 | - | 0.27 | - | 0.99 | + | 0.41 | 0.25 | 0.64 |
| **BIN1** | rs744373 | **G** | - | 0.40 | + | 0.66 | - | 0.28 | - | 0.07 | - | 0.53 | - | 0.24 | 0.88 | 0.43 |
| **CD33** | rs3865444 | **C** | - | 0.44 | - | 0.68 | - | 0.97 | - | 0.80 | + | 0.51 | + | 0.30 | 0.6 | 0.26 |
| **CR1** | rs3818361 | **A** | - | 0.71 | - | 0.18 | - | 0.89 | - | 0.41 | + | 0.95 | - | 0.25 | 0.34 | 0.89 |
| **ABCA7** | rs3764650 | **G** | - | 0.77 | - | 0.22 | + | 0.83 | + | 0.23 | - | 0.14 | + | 0.43 | 0.21 | 0.26 |

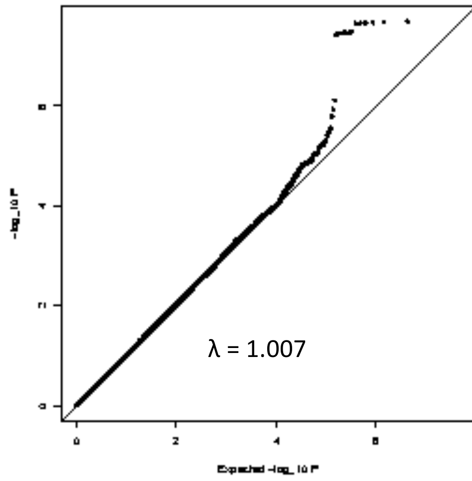
** p-values <0·0056 were considered statistically significant after Bonferroni correction for the number of independent SNPs tested*

## 11. SUPPLEMENTAL FIGURES

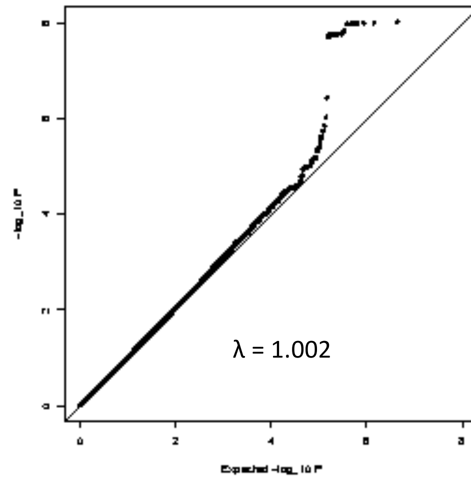
### 11.1. Figure S1: QQ plots for GWAS meta-analyses



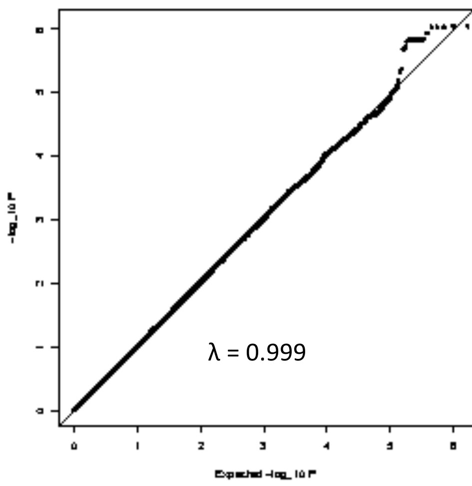
CVLT-dr, age and gender adjusted



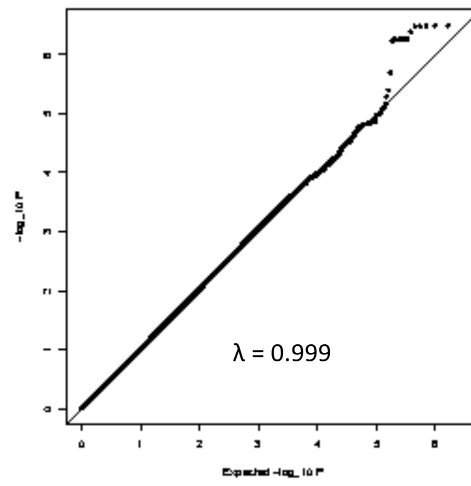

CVLT-dr, age, gender and education adjusted



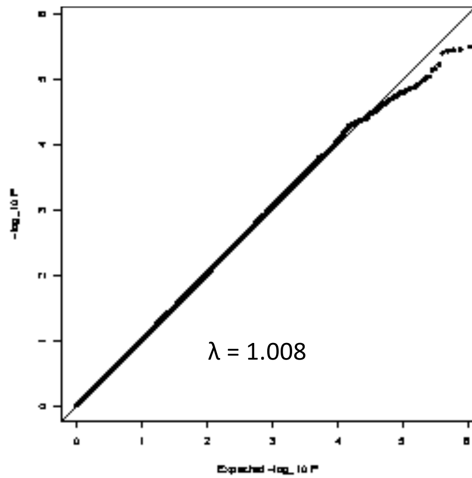

CVLT_HVLT-dr, age and gender adjusted



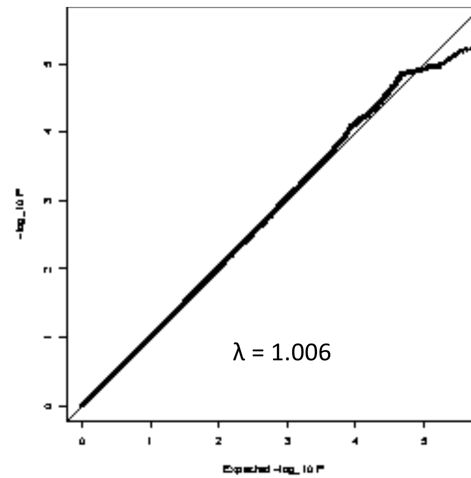

CVLT_HVLT-dr, age, gender and education adjusted



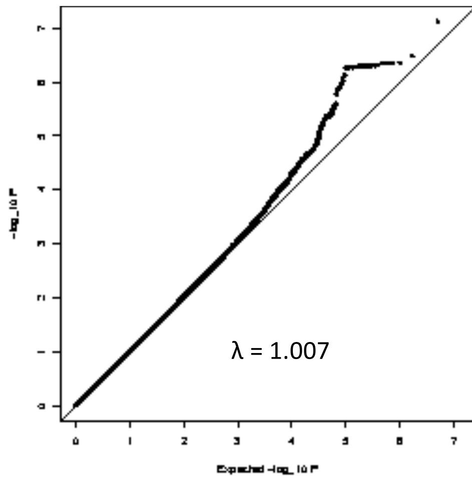

RAVLT-dr, age and gender adjusted



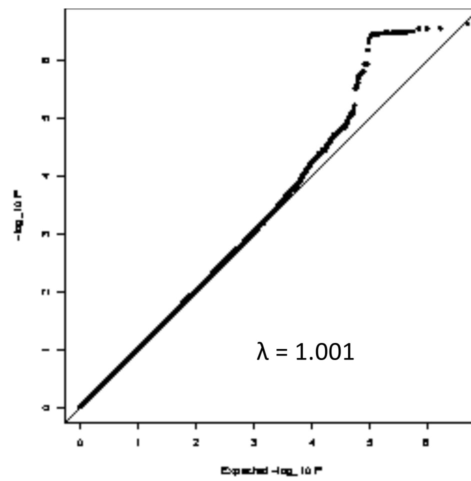

RAVLT-dr, age, gender and education adjusted



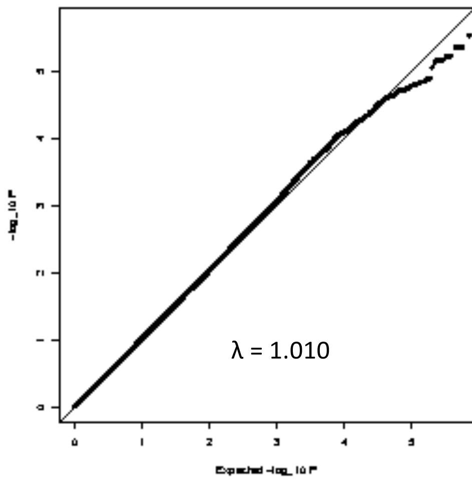
CERAD-dr, age and gender adjusted



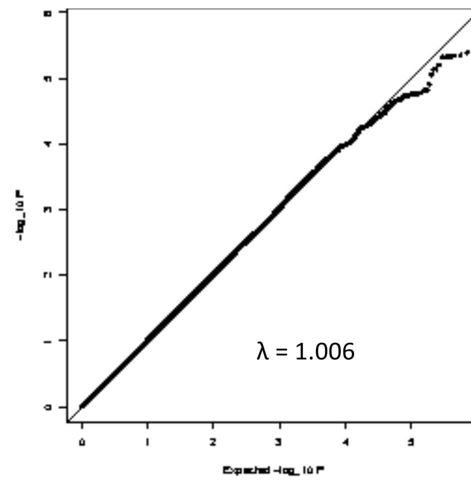

CERAD-dr, age, gender and education adjusted



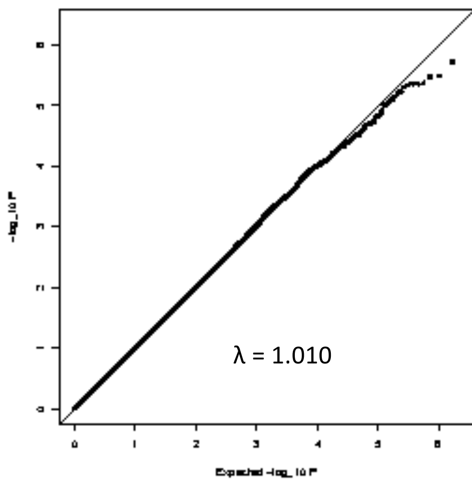

VPWL-dr, age and gender adjusted



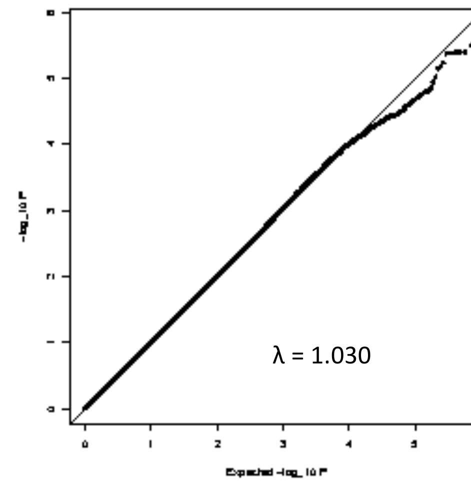

VPWL-dr, age, gender and education adjusted



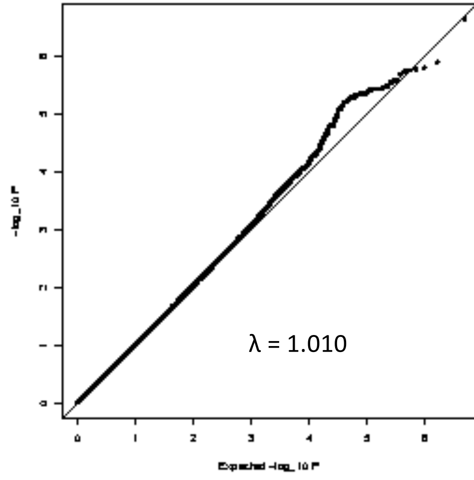

DWRT-dr, age and gender adjusted



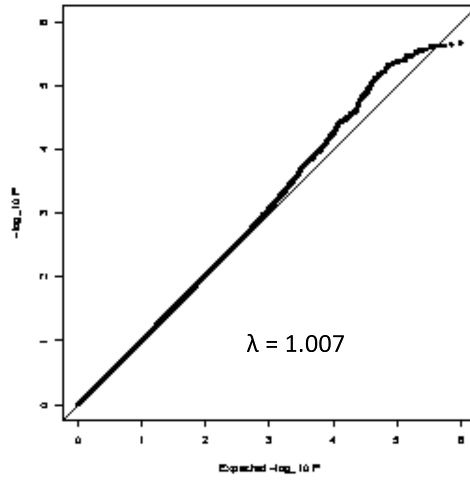

DWRT-dr, age, gender and education adjusted



HVLT-dr, age and gender adjusted

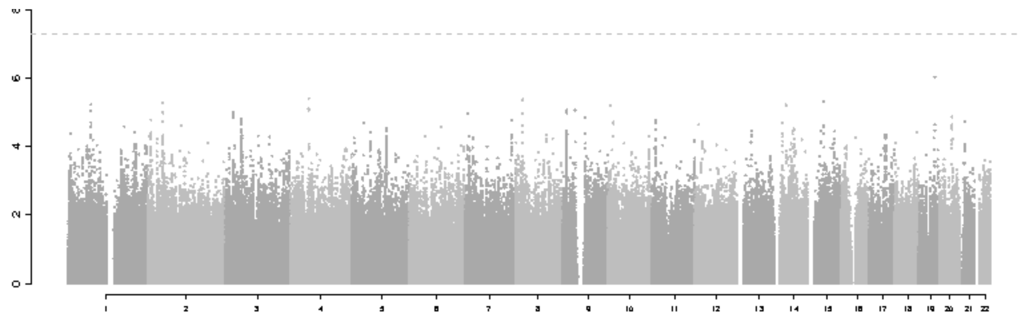



HVLT-dr, age, gender and education adjusted

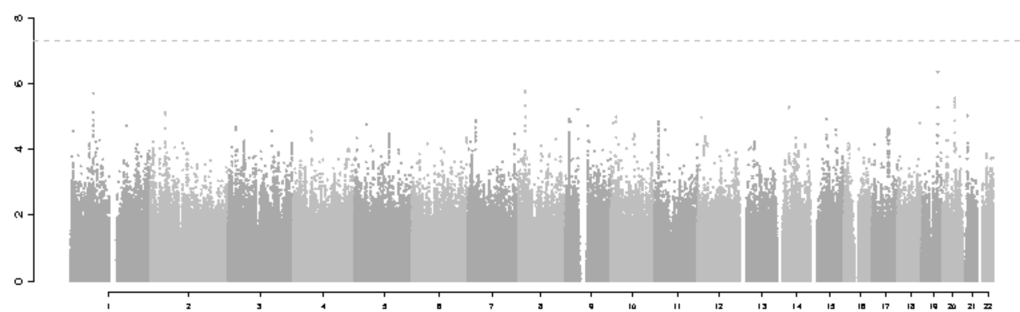


## 11.2. Figure S2: Manhattan plots for all meta-analyses

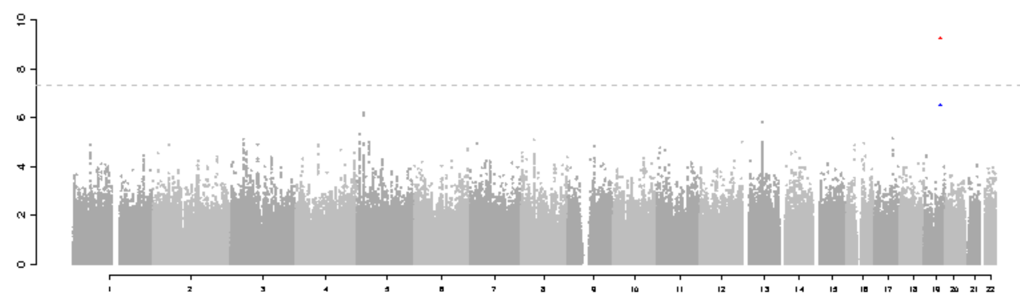
ALL-dr, age and gender adjusted



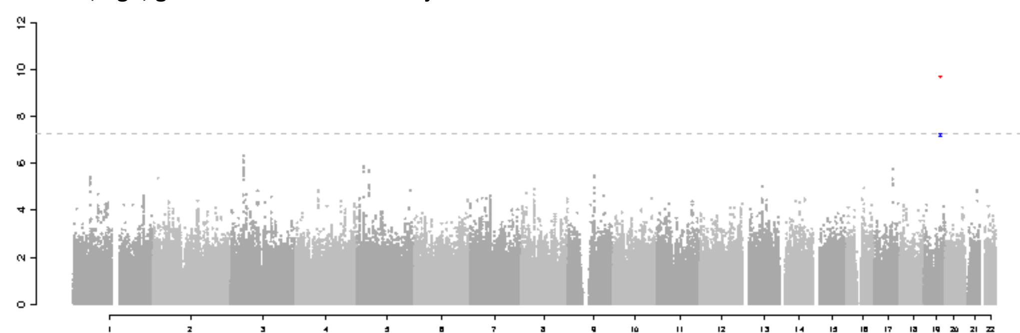
ALL-dr, age, gender and education adjusted



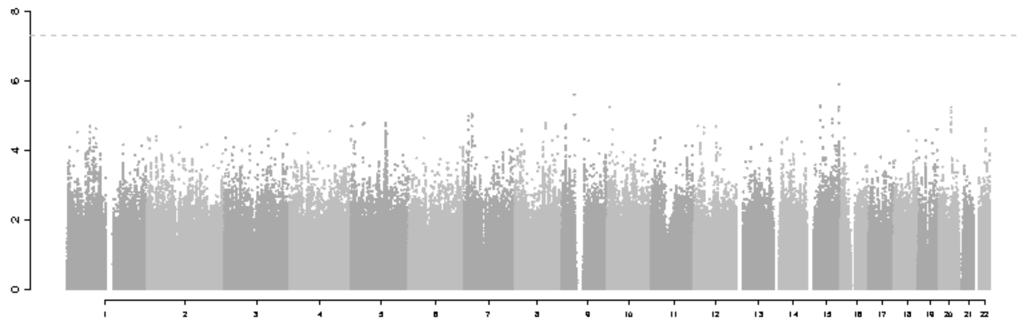
PAR-dr, age and gender adjusted



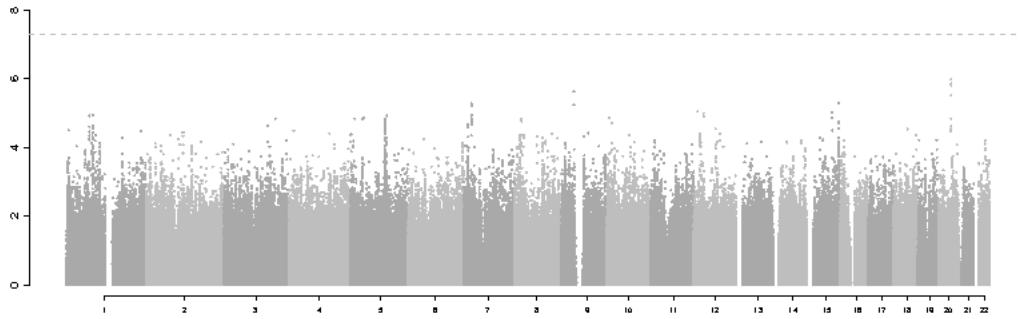
PAR-dr, age, gender and education adjusted



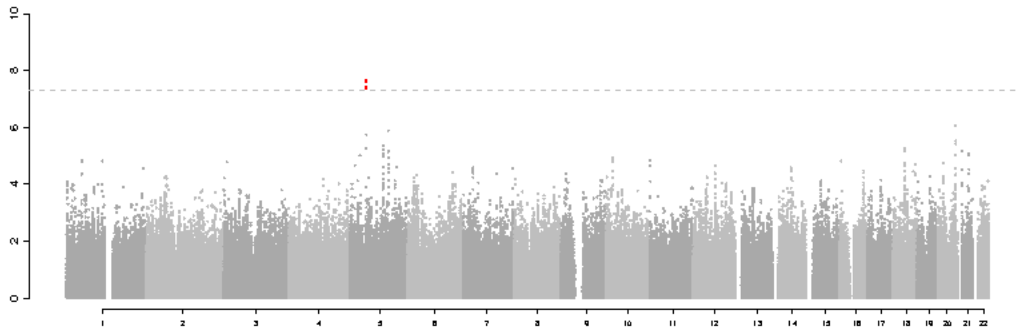
WL-dr, age and gender adjusted



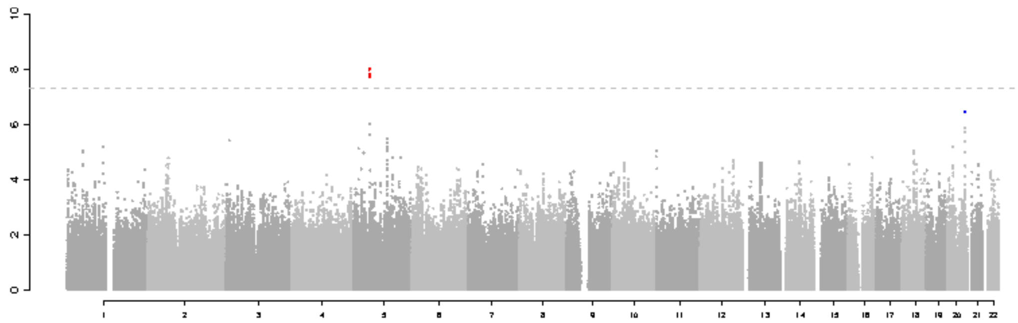
WL-dr, age, gender and education adjusted



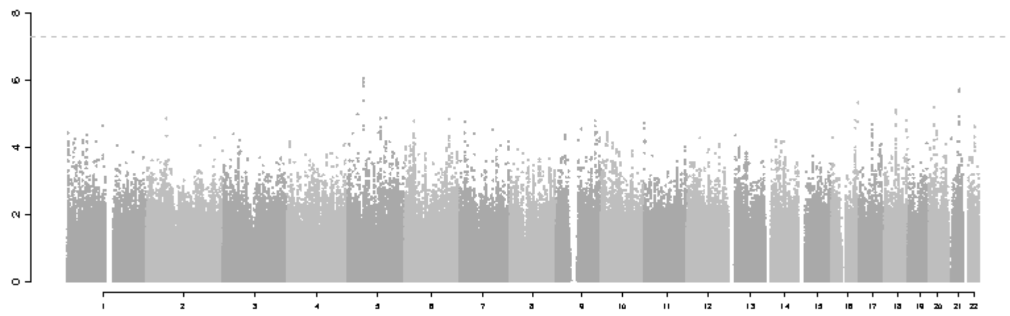
CVLT-dr, age and gender adjusted



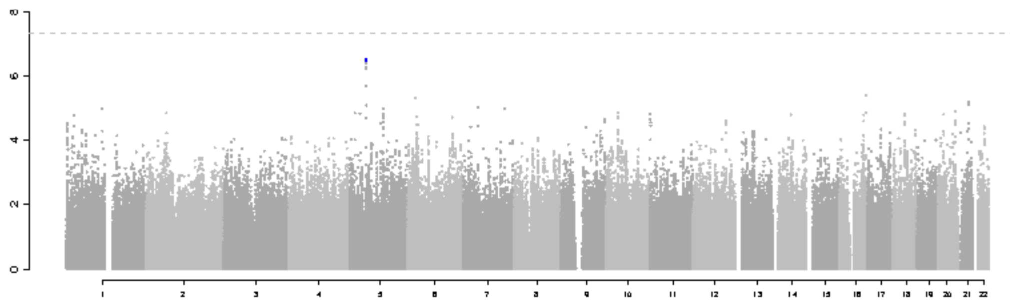
CVLT-dr, age, gender and education adjusted



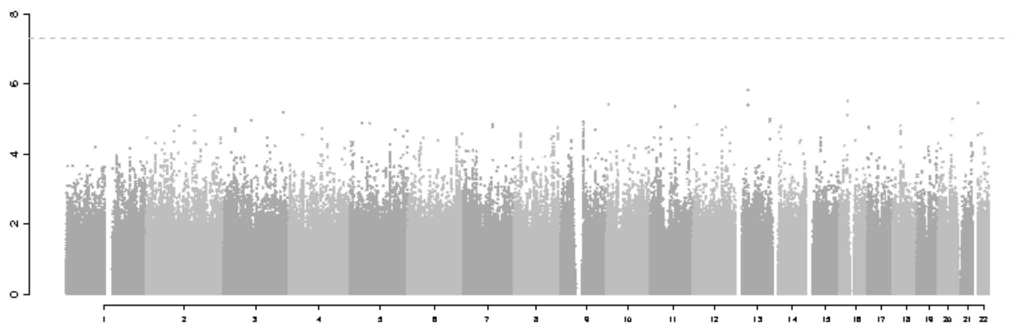
CVLT_HVLT-dr, age and gender adjusted



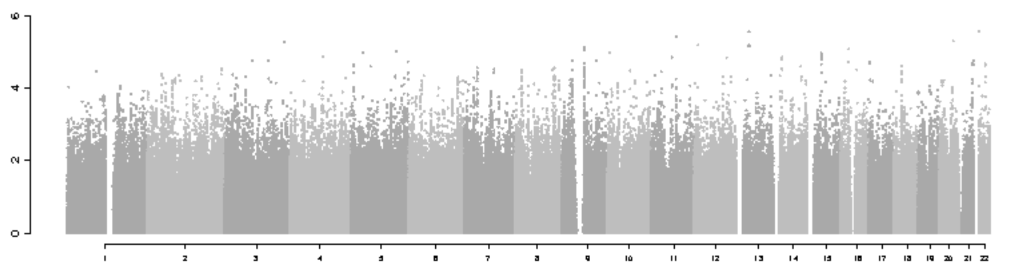
CVLT_HVLT-dr, age, gender and education adjusted



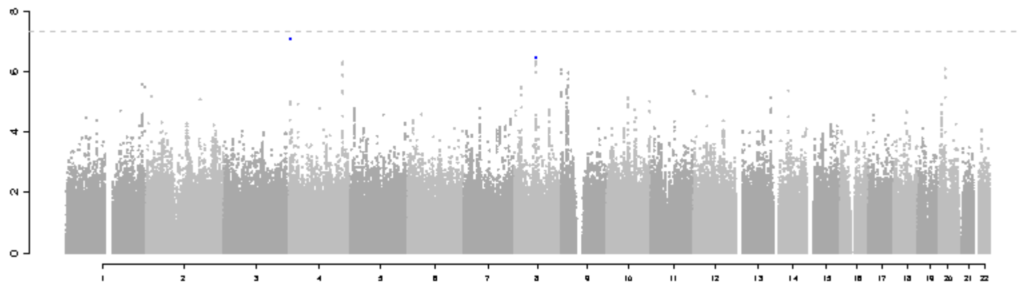
RAVLT-dr, age and gender adjusted



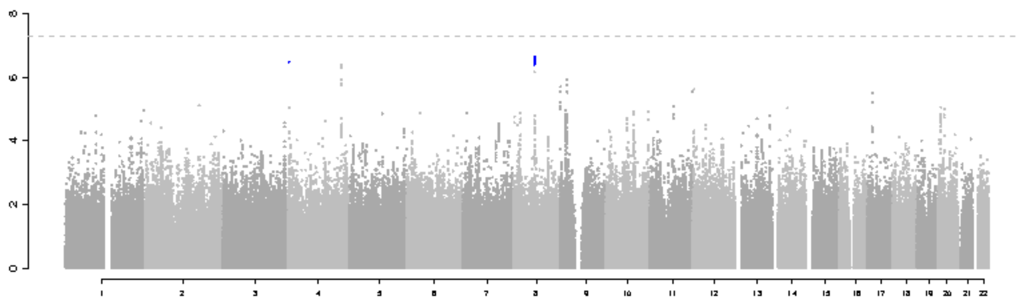
RAVLT-dr, age, gender and education adjusted



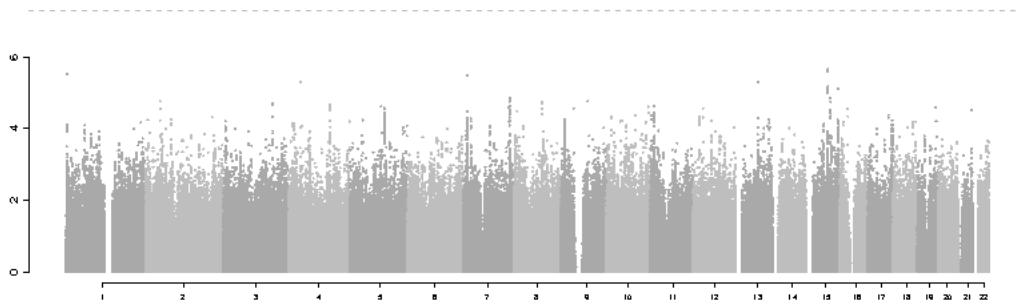
CERAD-dr, age and gender adjusted



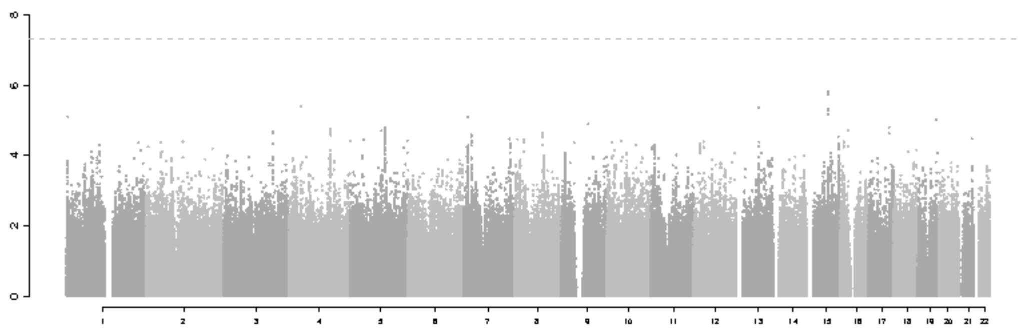
CERAD-dr, age, gender and education adjusted



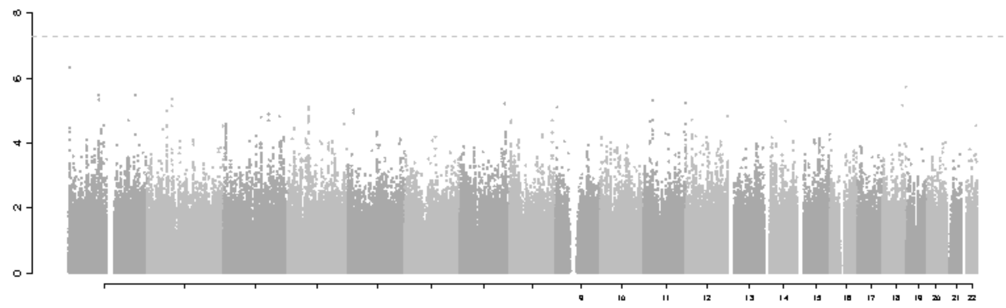
VPWL-dr, age and gender adjusted



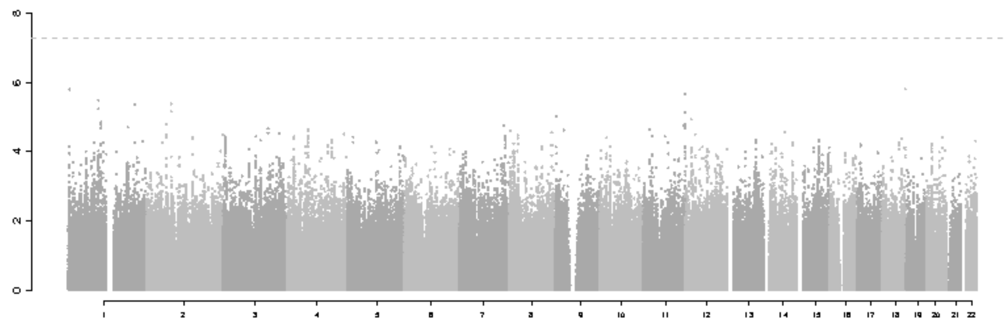
VPWL-dr, age, gender and education adjusted



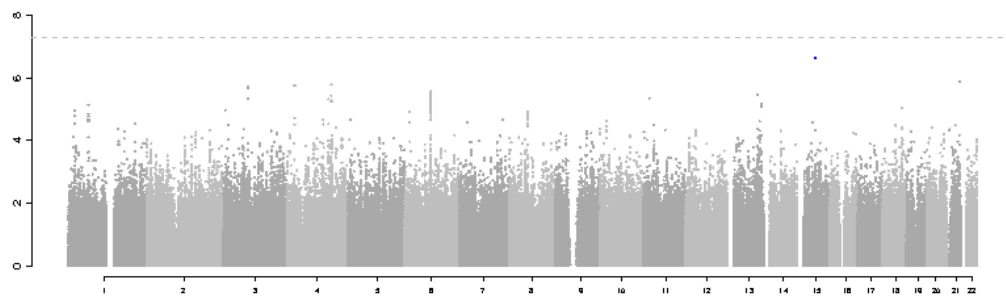
DWRT-dr, age and gender adjusted



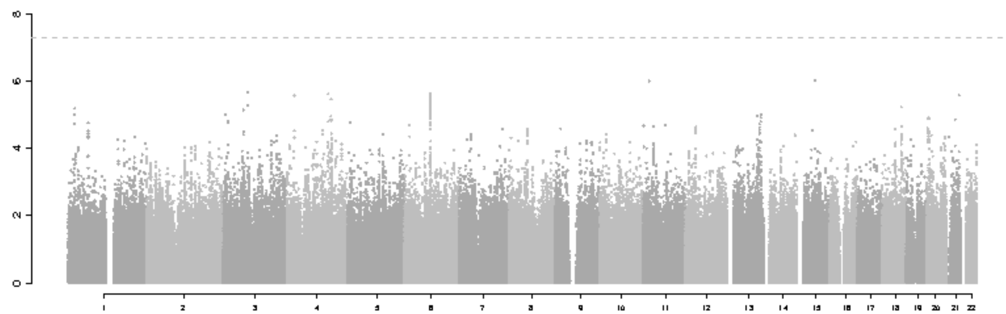
DWRT-dr, age, gender and education adjusted



HVLT-dr, age and gender adjusted

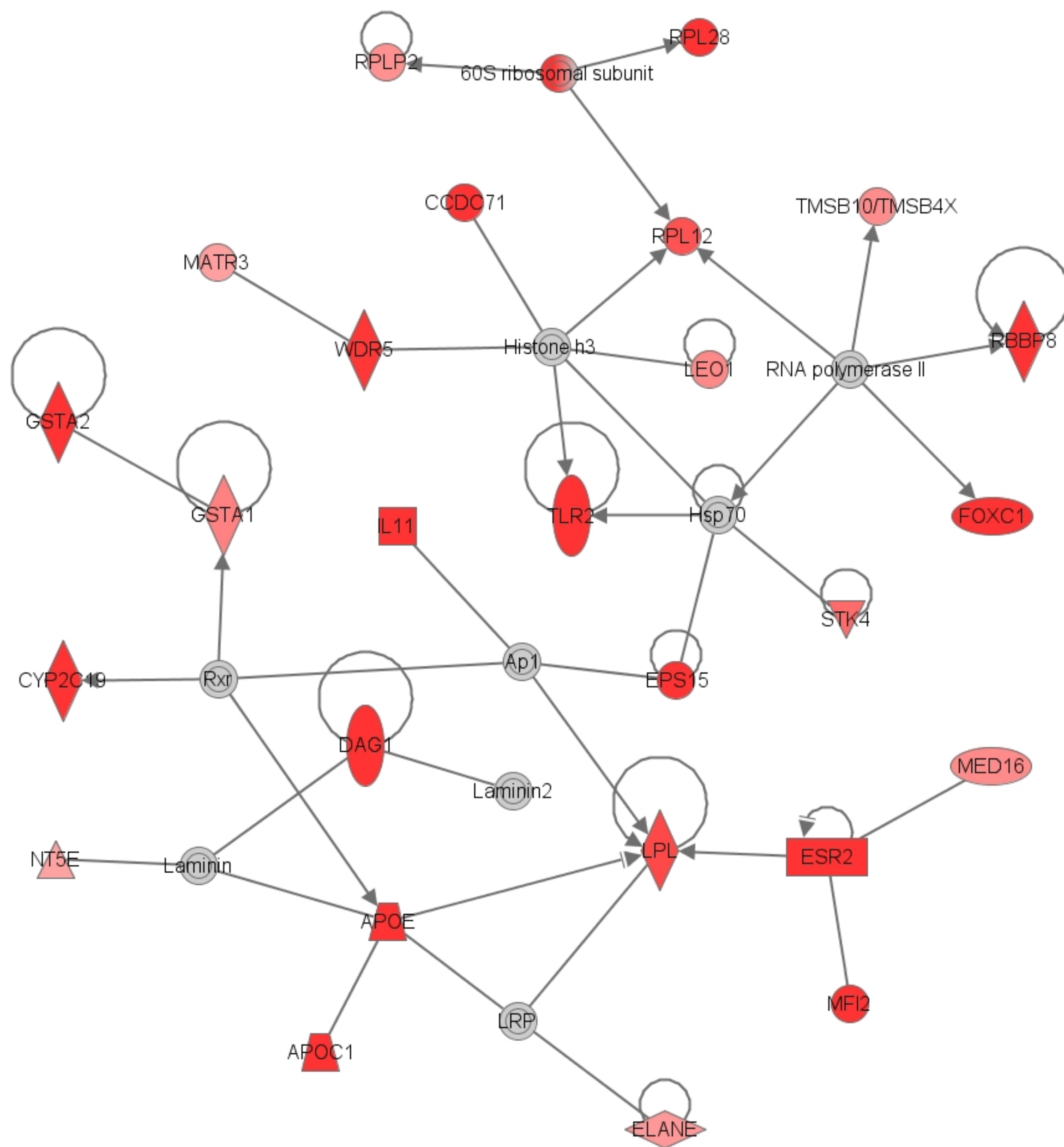


HVLT-dr, age, gender and education adjusted



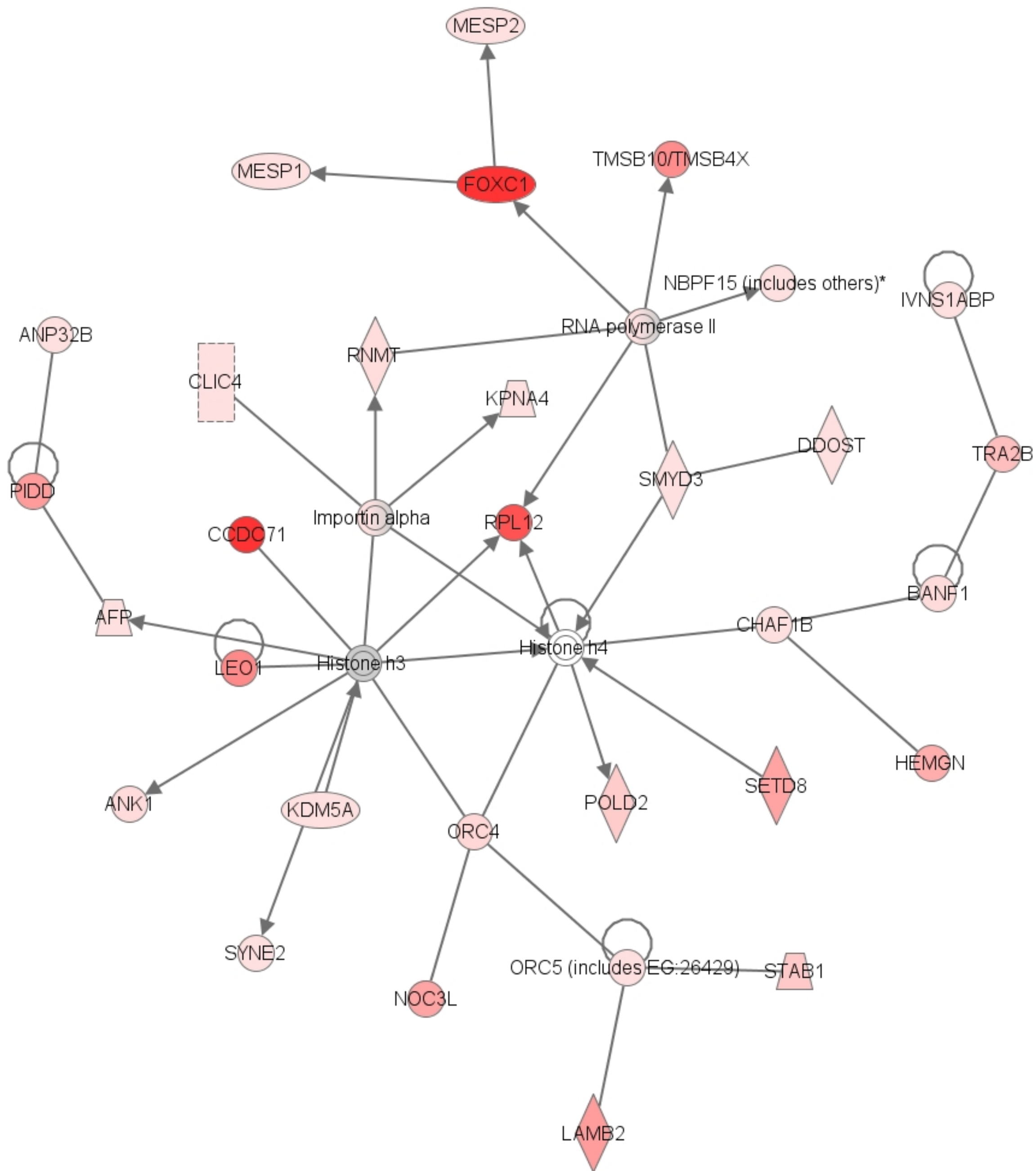
## 11.3. Figure S3: Pathway analysis

### Figure S3a: IPA analysis of genes closest to SNPs with p<0·01 from PAR-dr GWAS



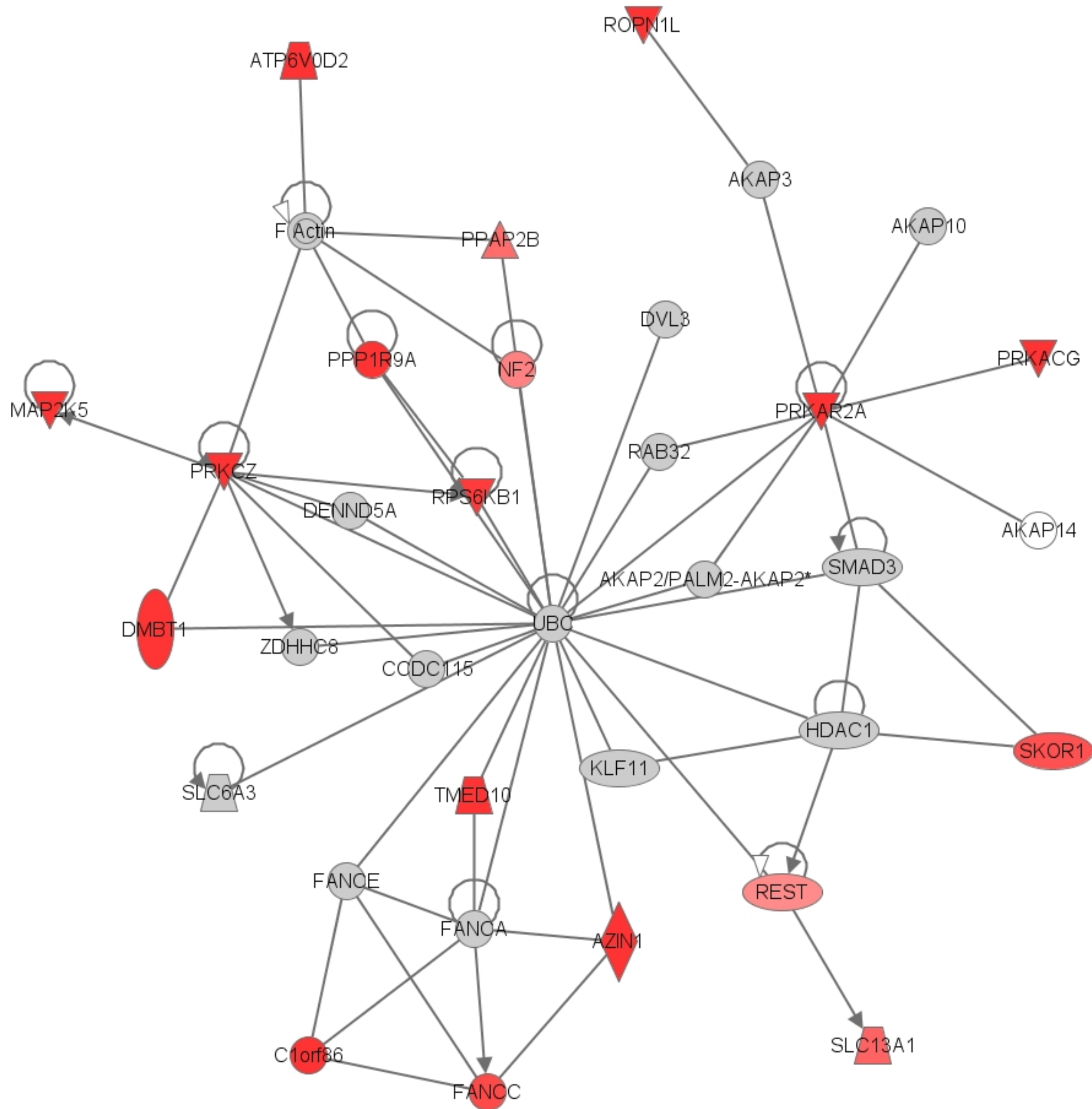

*Focus molecule nodes (p<0·01 from PAR-dr GWAS) are colored by the VEGAS[95] p-value for the gene (intense red = low p-value). Nodes that were in the data set but did not meet the cut off are grey and nodes that were not in the GWAS at all are white. Network score: 54*

**Figure S3b: IPA analysis of genes closest to SNPs with p<0·05 from PAR-dr GWAS**



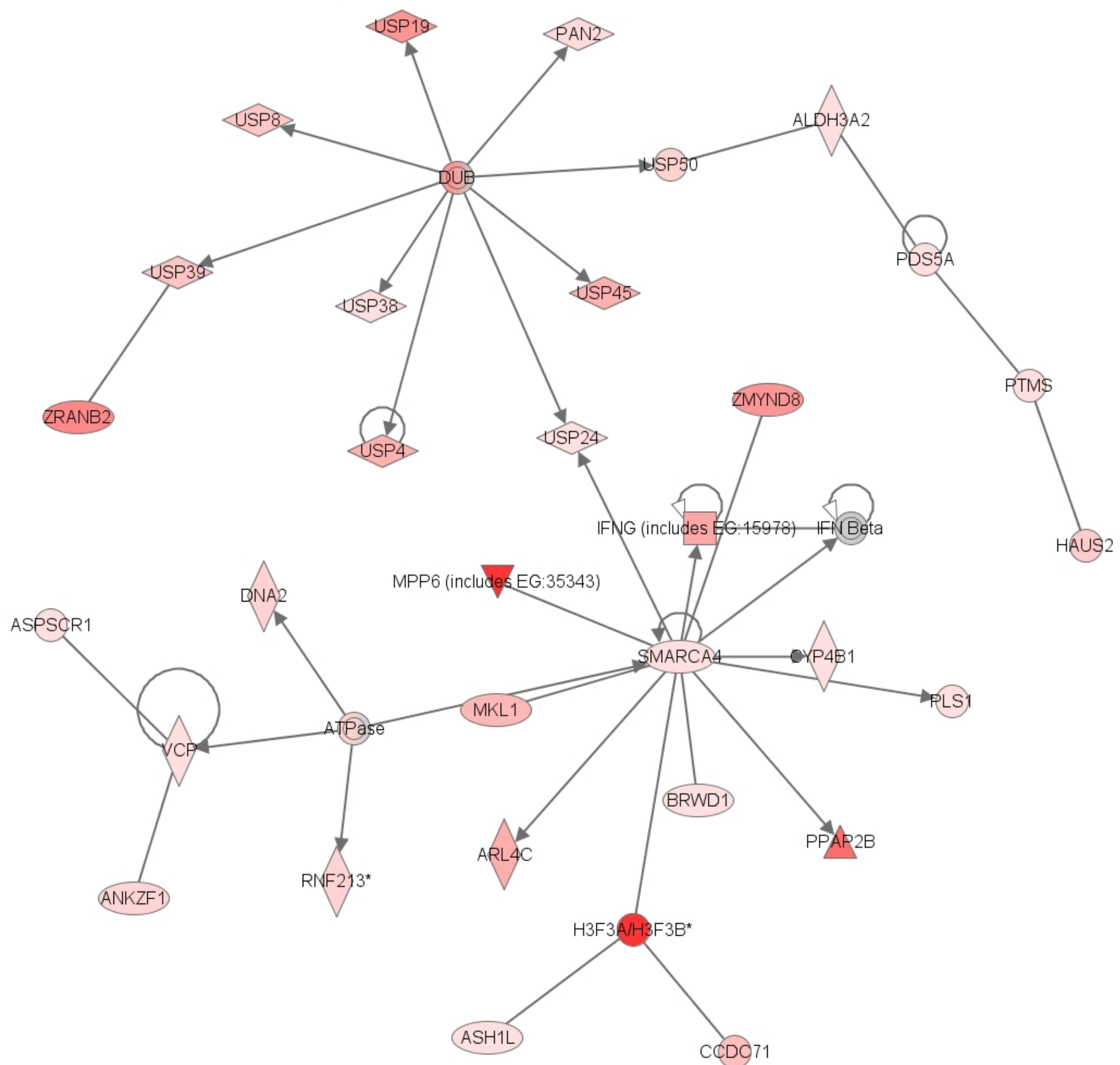

*Focus molecule nodes (p<0·05 from PAR-dr GWAS) are colored by the VEGAS[95] p-value for the gene (intense red = low p-value). Nodes that were in the data set but did not meet the cut off are grey and nodes that were not in the GWAS at all are white. Network score: 47*

**Figure S3c: IPA analysis of genes closest to SNPs with p<0·01 from WL-dr GWAS**




*Focus molecule nodes (p<0·01 from WL-dr GWAS) are colored by the VEGAS[95] p-value for the gene (intense red = low p-value). Nodes that were in the data set but did not meet the cut off are grey and nodes that were not in the GWAS at all are white. Network score: 33*

**Figure S3d: IPA analysis of genes closest to SNPs with p<0·05 from WL-dr GWAS**




*Focus molecule nodes (p<0·05 from WL-dr GWAS) are colored by the VEGAS[95] p-value for the gene (intense red = low p-value). Nodes that were in the data set but did not meet the cut off are grey and nodes that were not in the GWAS at all are white. Network score: 49*

**12. REFERENCES**


1. Psaty, B.M. et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from five cohorts. *Circulation:Cardiovascular Genetics* **2**, 73-80 (2009).
2. Harris, T.B. et al. Age, Gene/Environment Susceptibility-Reykjavik Study: multidisciplinary applied phenomics. *Am J Epidemiol* **165**, 1076-87 (2007).
3. Palm, W.M. et al. Ventricular dilation: association with gait and cognition. *Ann Neurol* **66**, 485-93 (2009).
4. Delis, D., Kramer, J., Kaplan, E. & al., e. *The California Verbal Learning Test—Research Edition*, (New York, NY: Psychological Corporation, 1987).
5. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol* **129**, 687-702 (1989).
6. Rosamond, W.D. et al. Stroke incidence and survival among middle-aged adults: 9-year follow-up of the Atherosclerosis Risk in Communities (ARIC) cohort. *Stroke* **30**, 736-43 (1999).
7. Mosley, T.H., Jr. et al. Cerebral MRI findings and cognitive functioning: the Atherosclerosis Risk in Communities study. *Neurology* **64**, 2056-62 (2005).
8. Knopman, D.S. & Ryberg, S. A verbal memory test with high predictive accuracy for dementia of the Alzheimer type. *Arch Neurol* **46**, 141-5 (1989).
9. Knopman, D.S., Mosley, T.H., Catellier, D.J. & Coker, L.H. Fourteen-year longitudinal study of vascular risk factors, APOE genotype, and cognition: the ARIC MRI Study. *Alzheimers Dement* **5**, 207-14 (2009).
10. Lesage, S.R. et al. Retinal microvascular abnormalities and cognitive decline: the ARIC 14-year follow-up study. *Neurology* **73**, 862-8 (2009).
11. Wechsler, D. *Wechsler memory-scale-revised (manual)*, (Psychological Corporation, New York, 1987).
12. Fried, L.P. et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol* **1**, 263-76 (1991).
13. Longstreth, W.T., Jr. et al. Frequency and predictors of stroke death in 5,888 participants in the Cardiovascular Health Study. *Neurology* **56**, 368-75 (2001).
14. Lopez, O.L. et al. Evaluation of dementia in the cardiovascular health cognition study. *Neuroepidemiology* **22**, 1-12 (2003).
15. Lopez, O.L. et al. Neuropsychological characteristics of mild cognitive impairment subgroups. *J Neurol Neurosurg Psychiatry* **77**, 159-65 (2006).
16. Rudan, I., Campbell, H. & Rudan, P. Genetic epidemiological studies of eastern Adriatic Island isolates, Croatia: objective and strategies. *Coll Antropol* **23**, 531-46 (1999).
17. Rudan, I. et al. Effects of inbreeding, endogamy, genetic admixture, and outbreeding on human health: a (1001 Dalmatians) study. *Croat Med J* **47**, 601-10 (2006).

18. Campbell, H. et al. Effects of genome-wide heterozygosity on a range of biomedically relevant human quantitative traits. *Hum Mol Genet* **16**, 233-41 (2007).
19. Polasek, O. et al. Genome-wide association study of anthropometric traits in Korcula Island, Croatia. *Croat Med J* **50**, 7-16 (2009).
20. Vitart, V. et al. Heritabilities of ocular biometrical traits in two croatian isolates with extended pedigrees. *Invest Ophthalmol Vis Sci* **51**, 737-43.
21. Vitart, V. et al. SLC2A9 is a newly identified urate transporter influencing serum urate concentration, urate excretion and gout. *Nat Genet* **40**, 437-42 (2008).
22. Rey, A. *L'Examen Clinique en Psychologie*, (Paris: Presses Universitaires de France, 1964).
23. Service, S. et al. Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. *Nat Genet* **38**, 556-60 (2006).
24. Aulchenko, Y.S. et al. Linkage disequilibrium in young genetically isolated Dutch population. *Eur J Hum Genet* **12**, 527-34 (2004).
25. van Koolwijk, L.M. et al. Association of cognitive functioning with retinal nerve fiber layer thickness. *Invest Ophthalmol Vis Sci* **50**, 4576-80 (2009).
26. Saan, R. & Deelman, B. De 15-woordentest A en B: Een Voorlopige Handleiding [Internal Report]. (Groningen, The Netherlands: University Hospital Groningen, Department of Neuropsychology, 1986).
27. Dawber, T.R. & Kannel, W.B. The Framingham study. An epidemiological approach to coronary heart disease. *Circulation* **34**, 553-5 (1966).
28. Feinleib, M., Kannel, W.B., Garrison, R.J., McNamara, P.M. & Castelli, W.P. The Framingham Offspring Study. Design and preliminary data. *Prev Med* **4**, 518-25 (1975).
29. 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**, 1328-35 (2007).
30. Au, R. et al. New norms for a new generation: cognitive performance in the framingham offspring cohort. *Exp Aging Res* **30**, 333-58 (2004).
31. Wechsler, D. *Wechsler memory scale*, (San Antonio, TX, US: Psychological Corporation, 1945).
32. Carandang, R. et al. Trends in incidence, lifetime risk, severity, and 30-day mortality of stroke over the past 50 years. *Jama* **296**, 2939-46 (2006).
33. Seshadri, S. et al. The lifetime risk of stroke: estimates from the Framingham Study. *Stroke* **37**, 345-50 (2006).
34. Lahti, J. et al. Glucocorticoid receptor gene haplotype predicts increased risk of hospital admission for depressive disorders in the Helsinki birth cohort study. *J Psychiatr Res* **45**, 1160-4.

35. Barker, D.J., Osmond, C., Forsen, T.J., Kajantie, E. & Eriksson, J.G. Trajectories of growth among children who have coronary events as adults. *N Engl J Med* **353**, 1802-9 (2005).
36. Morris, J.C. et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology* **39**, 1159-65 (1989).
37. Deary, I.J. et al. The Lothian Birth Cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond. *BMC Geriatr* **7**, 28 (2007).
38. Deary, I.J., Whiteman, M.C., Starr, J.M., Whalley, L.J. & Fox, H.C. The impact of childhood intelligence on later life: following up the Scottish mental surveys of 1932 and 1947. *J Pers Soc Psychol* **86**, 130-47 (2004).
39. Wechsler, D. *WMS-IIIUK administration and scoring manual*, (London, UK, Psychological Corporation, 1998).
40. McQuillan, R. et al. Runs of homozygosity in European populations. *Am J Hum Genet* **83**, 359-72 (2008).
41. Hofman, A. et al. The Rotterdam Study: 2010 objectives and design update. *Eur J Epidemiol* **24**, 553-72 (2009).
42. Hofman, A., Grobbee, D.E., de Jong, P.T. & van den Ouweland, F.A. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. *Eur J Epidemiol* **7**, 403-22 (1991).
43. Hofman, A. Recent trends in cardiovascular epidemiology. *Eur J Epidemiol* **24**, 721-3 (2009).
44. Prins, N.D. et al. Cerebral small-vessel disease and decline in information processing speed, executive function and memory. *Brain* **128**, 2034-41 (2005).
45. Brand, N. & Jolles, J. Learning and retrieval rate of words presented auditorily and visually. *J Gen Psychol* **112**, 201-10 (1985).
46. Bots, M.L. et al. Prevalence of stroke in the general population. The Rotterdam Study. *Stroke* **27**, 1499-501 (1996).
47. Hollander, M. et al. Incidence, risk, and case fatality of first ever stroke in the elderly population. The Rotterdam Study. *J Neurol Neurosurg Psychiatry* **74**, 317-21 (2003).
48. Chibnik, L.B. et al. CR1 is associated with amyloid plaque burden and age-related cognitive decline. *Ann Neurol* **69**, 560-9.
49. Bennett, D.A. et al. Neuropathology of older persons without cognitive impairment from two community-based studies. *Neurology* **66**, 1837-44 (2006).
50. Bennett, D.A. et al. The Rush Memory and Aging Project: study design and baseline characteristics of the study cohort. *Neuroepidemiology* **25**, 163-75 (2005).
51. Volzke, H. et al. Cohort profile: the study of health in Pomerania. *Int J Epidemiol* **40**, 294-307.
52. Wittchen, H.U., Lachner, G., Wunderlich, U. & Pfister, H. Test-retest reliability of the computerized DSM-IV version of the Munich-Composite International Diagnostic Interview (M-CIDI). *Soc Psychiatry Psychiatr Epidemiol* **33**, 568-78 (1998).

53. Appel, K. et al. Moderation of adult depression by a polymorphism in the FKBP5 gene and childhood physical abuse in the general population. *Neuropsychopharmacology* **36**, 1982-91 (2011).
54. Oswald, W.D. & Fleischmann, U.M. Nürnberger-Alters-Inventar NAI. Testkasten und Kurzmanual. (Nürnberg: Universität Erlangen-Nürnberg. 1986).
55. Callisaya, M.L. et al. A population-based study of sensorimotor factors affecting gait in older people. *Age Ageing* **38**, 290-5 (2009).
56. Martin, K. et al. Visuospatial ability and memory are associated with falls risk in older people: a population-based study. *Dement Geriatr Cogn Disord* **27**, 451-7 (2009).
57. Lezak, M. *Neuropsychological testing*, (Oxford, UK, Oxford University Press, 2004).
58. Shock, N. et al. *Normal human aging: The Baltimore Longitudinal Study of Aging*, (Washington (DC): U.S. Government Printing Office, November 1984 Report No. NIH Publ No. (NIH)84–2450).
59. McEvoy, M. et al. Cohort profile: The Hunter Community Study. *Int J Epidemiol* **39**, 1452-63.
60. Schofield, P.W. et al. The Audio Recorded Cognitive Screen (ARCS): a flexible hybrid cognitive test instrument. *J Neurol Neurosurg Psychiatry* **81**, 602-7.
61. Kang, J.H. & Grodstein, F. Postmenopausal hormone therapy, timing of initiation, APOE and cognitive decline. *Neurobiol Aging*.
62. Sachdev, P.S. et al. The Sydney Memory and Ageing Study (MAS): methodology and baseline medical and neuropsychiatric characteristics of an elderly epidemiological non-demented cohort of Australians aged 70-90 years. *Int Psychogeriatr* **22**, 1248-64 (2010).
63. Wechsler, D. *Wechsler Memory Scale - Third edition*, (San Antonio TX: The Psychological Corporation, 1997).
64. 3C-Study-Group. Vascular factors and risk of dementia: design of the Three-City Study and baseline characteristics of the study population. *Neuroepidemiology* **22**, 316-25 (2003).
65. Auriacombe, S. et al. Validity of the free and cued selective reminding test in predicting dementia: the 3C study. *Neurology* **74**, 1760-7 (2010).
66. Lambert, J.C. et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet* **41**, 1094-9 (2009).
67. Ridker, P.M. et al. Rationale, design, and methodology of the Women's Genome Health Study: a genome-wide association study of more than 25,000 initially healthy american women. *Clin Chem* **54**, 249-55 (2008).
68. Howard, V.J. et al. The reasons for geographic and racial differences in stroke study: objectives and design. *Neuroepidemiology* **25**, 135-43 (2005).
69. Howard, G. & Howard, V.J. Ethnic disparities in stroke: the scope of the problem. *Ethn Dis* **11**, 761-8 (2001).
70. Howard, V.J. et al. Disparities in stroke incidence contributing to disparities in stroke mortality. *Ann Neurol* **69**, 619-27 (2011).

71. Papassotiropoulos, A. et al. A genome-wide survey and functional brain imaging study identify CTNNBL1 as a memory-related gene. *Mol Psychiatry* (2011).
72. Vigneau, M. et al. What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? Insights from a meta-analysis. *Neuroimage* **54**, 577-93 (2011).
73. Kandel, E.R. The molecular biology of memory storage: a dialogue between genes and synapses. *Science* **294**, 1030-8 (2001).
74. Bacskai, B.J. et al. Spatially resolved dynamics of cAMP and protein kinase A subunits in Aplysia sensory neurons. *Science* **260**, 222-6 (1993).
75. Brandt, J., Spencer, M. & Folstein, M. The telephone interview for cognitive status. *Neuropsychiatry Neuropsychol Behav Neurol* **1**, 111-117 (1988).
76. Van der Linden, M.
77. de Quervain, D.J. et al. A functional genetic variation of the 5-HT2a receptor affects human memory. *Nat Neurosci* **6**, 1141-2 (2003).
78. Albert, M. et al. Use of brief cognitive tests to identify individuals in the community with clinically diagnosed Alzheimer's disease. *Int J Neurosci* **57**, 167-78 (1991).
79. Ikram, M.A. et al. Genomewide association studies of stroke. *N Engl J Med* **360**, 1718-28 (2009).
80. Price, A.L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-9 (2006).
81. Patterson, N., Price, A.L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet* **2**, e190 (2006).
82. Richards, J.B. et al. Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. *Lancet* **371**, 1505-12 (2008).
83. Aulchenko, Y.S., Ripke, S., Isaacs, A. & van Duijn, C.M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294-6 (2007).
84. Chen, W.M. & Abecasis, G.R. Family-based association tests for genomewide association scans. *Am J Hum Genet* **81**, 913-26 (2007).
85. Chen, M.H. & Yang, Q. GWAF: an R package for genome-wide association analyses with family data. *Bioinformatics* **26**, 580-1 (2010).
86. Bacanu, S.A., Devlin, B. & Roeder, K. The power of genomic control. *Am J Hum Genet* **66**, 1933-44 (2000).
87. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
88. Kohli, M.A. et al. The neuronal transporter gene SLC6A15 confers risk to major depression. *Neuron* **70**, 252-65 (2011).
89. Kral, T. et al. Preoperative evaluation for epilepsy surgery (Bonn Algorithm). *Zentralbl Neurochir* **63**, 106-10 (2002).
90. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529 (2009).

91. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60 (2009).
92. Pruitt, K.D., Tatusova, T. & Maglott, D.R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **35**, D61-5 (2007).
93. Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat Protoc* **7**, 500-7 (2012).
94. Stegle, O., Parts, L., Durbin, R. & Winn, J. A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. *PLoS Comput Biol* **6**, e1000770 (2010).
95. Liu, J.Z. et al. A versatile gene-based test for genome-wide association studies. *Am J Hum Genet* **87**, 139-45 (2010).
96. Chibnik, L.B. et al. CR1 is associated with amyloid plaque burden and age-related cognitive decline. *Ann Neurol* **69**, 560-9 (2011).
97. Bennett, D.A., Schneider, J.A., Bienias, J.L., Evans, D.A. & Wilson, R.S. Mild cognitive impairment is related to Alzheimer disease pathology and cerebral infarctions. *Neurology* **64**, 834-41 (2005).
98. Schneider, J.A., Wilson, R.S., Bienias, J.L., Evans, D.A. & Bennett, D.A. Cerebral infarctions and the likelihood of dementia from Alzheimer disease pathology. *Neurology* **62**, 1148-55 (2004).
99. Braak, H. & Braak, E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol* **82**, 239-59 (1991).
100. Mirra, S.S. et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology* **41**, 479-86 (1991).
101. Bennett, D.A., Schneider, J.A., Tang, Y., Arnold, S.E. & Wilson, R.S. The effect of social networks on the relation between Alzheimer's disease pathology and level of cognitive function in old people: a longitudinal cohort study. *Lancet Neurol* **5**, 406-12 (2006).
102. International Schizophrenia, C. et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748-52 (2009).
103. Kathiresan, S. et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet* **41**, 334-41 (2009).
104. Ehret, G.B. et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103-9 (2011).
105. Pietilainen, O.P. et al. Association of AKT1 with verbal learning, verbal memory, and regional cortical gray matter density in twins. *Am J Med Genet B Neuropsychiatr Genet* **150B**, 683-92 (2009).

106. Egan, M.F. et al. The BDNF val66met polymorphism affects activity-dependent secretion of BDNF and human memory and hippocampal function. *Cell* **112**, 257-69 (2003).
107. Miyajima, F. et al. Brain-derived neurotrophic factor polymorphism Val66Met influences cognitive abilities in the elderly. *Genes Brain Behav* **7**, 411-7 (2008).
108. Huentelman, M.J. et al. Calmodulin-binding transcription activator 1 (CAMTA1) alleles predispose human episodic memory performance. *Hum Mol Genet* **16**, 1469-77 (2007).
109. Luciano, M. et al. Variation in the dysbindin gene and normal cognitive function in three independent population samples. *Genes Brain Behav* **8**, 218-27 (2009).
110. Nicodemus, K.K. et al. Further evidence for association between ErbB4 and schizophrenia and influence on cognitive intermediate phenotypes in healthy controls. *Mol Psychiatry* **11**, 1062-5 (2006).
111. Papassotiropoulos, A. et al. Age-dependent effects of the 5-hydroxytryptamine-2a-receptor polymorphism (His452Tyr) on human memory. *Neuroreport* **16**, 839-42 (2005).
112. Reynolds, C.A., Jansson, M., Gatz, M. & Pedersen, N.L. Longitudinal change in memory performance associated with HTR2A polymorphism. *Neurobiol Aging* **27**, 150-4 (2006).
113. Baune, B.T. et al. Association between genetic variants of IL-1beta, IL-6 and TNF-alpha cytokines and cognitive performance in the elderly general population of the MEMO-study. *Psychoneuroendocrinology* **33**, 68-76 (2008).
114. Papassotiropoulos, A. et al. Common Kibra alleles are associated with human memory performance. *Science* **314**, 475-8 (2006).
115. Schaper, K., Kolsch, H., Popp, J., Wagner, M. & Jessen, F. KIBRA gene variants are associated with episodic memory in healthy elderly. *Neurobiol Aging* **29**, 1123-5 (2008).
116. Almeida, O.P. et al. KIBRA genetic polymorphism influences episodic memory in later life, but does not increase the risk of mild cognitive impairment. *J Cell Mol Med* **12**, 1672-6 (2008).
117. Yasuda, Y. et al. Association study of KIBRA gene with memory performance in a Japanese population. *World J Biol Psychiatry* **11**, 852-7 (2010).
118. Preuschhof, C. et al. KIBRA and CLSTN2 polymorphisms exert interactive effects on human episodic memory. *Neuropsychologia* **48**, 402-8 (2010).
119. Milnik, A. et al. Association of KIBRA with episodic and working memory: a meta-analysis. *Am J Med Genet B Neuropsychiatr Genet* **159B**, 958-69 (2012).
120. Izumi, A. et al. Genetic variations of human neuropsin gene and psychiatric disorders: polymorphism screening and possible association with bipolar disorder and cognitive functions. *Neuropsychopharmacology* **33**, 3237-45 (2008).
121. Kolsch, H. et al. Gene polymorphisms in prodynorphin (PDYN) are associated with episodic memory in the elderly. *J Neural Transm* **116**, 897-903 (2009).

122. Houlihan, L.M. et al. Replication study of candidate genes for cognitive abilities: the Lothian Birth Cohort 1936. *Genes Brain Behav* **8**, 238-47 (2009).
123. Vogler, C. et al. CPEB3 is associated with human episodic memory. *Front Behav Neurosci* **3**, 4 (2009).
124. LeBlanc, M. et al. Genome-wide study identifies PTPRO and WDR72 and FOXQ1-SUMO1P1 interaction associated with neurocognitive function. *J Psychiatr Res* **46**, 271-8 (2012).

## Chapter 8: Epilogue

### *8.1. Dankwoord/Acknowledgments*

Klaar. Bijna. Tijd voor de kleine en grote bedankjes, met hier en daar een dike papieren knuffel. Het was intensief, fantastisch, soms even doorbijten, maar altijd leerzaam, en dat heb ik aan vele mensen te danken.

De eerste plaats is voor degene die de afgelopen jaren de lijnen uitzette, de touwtjes strak hield en af en toe alles radicaal omgooide: mijn promotor Prof. Dr. C.M. van Duijn, hoofd van de Genetische Epidemiologie groep. Cock, ik denk dat ik niet vaak meer in zo'n korte tijd zoveel kansen krijg om nieuwe dingen te doen, mensen te ontmoeten, en elke dag bij te leren. Je bent een inspirator en geweldig leermeester door je kennis, creativiteit, humor en nuchterheid. Bedankt voor alles, en ik hoop dat alle nieuwe dingen die nu gaande zijn tot mooie ontdekkingen zullen leiden.

Prof. Dr. J.C. van Swieten, mijn andere promotor. John, de volgende plek is aan jou. Je hebt me getipt en enthousiast gemaakt over deze promotieplaats, en je hebt niets teveel gezegd. Terwijl mijn onderzoek zich ontspon in methodologische richting en consortiumwerk, en onze samenwerking niet altijd even intensief was, was je goede raad er als ik hem nodig had. Ook in de kliniek kan ik nog veel van je leren, en niet alleen medisch-inhoudelijk, ik verheug me erop.

My very learned copromotor, dr. N.Amin. This one is for you Najaf. If this thesis were to have a starred second author, it would be you. Critical as ever, demanding nothing less than perfection in layout and statistics, but also there like an older sister, whether it was about work, home, or the combination of those.  Your help throughout, but also in the last phase of my PhD, getting this thesis into shape, has been indispensible. A big paper hug for you.

Ook gaat mijn dank uit naar de kleine commissie. Dr. M.A. Ikram, beste Arfan, we hebben intensief samengewerkt aan de Rotterdam data, en binnen CHARGE. Ik heb veel van je geleerd, en heb respect voor hoe je in korte tijd je plaats hebt gevonden als hoofd van de neuroepidemiologie groep in de Rotterdam studie. Bedankt dat je ook in deze laatste fase een grote rol wilt spelen als lid van de leescommissie. Dr. L.J. Launer, dear Lenore, I am so glad you are coming all the way to Rotterdam for my defense, and that you are willing to be in the reading committee. Inviting me to Bethesda on the CHARGE fellowship, you gave me a lifetime experience. Can one be homesick to somewhere she lived only for three months? Prof.Dr.R.M.W. Hofstra, dank voor het plaatsnemen in de commissie en het zo snel beoordelen van mijn proefschrift.

Ook de grote commissie wil ik graag hartelijk danken. Prof.Dr. B.A. Oostra: Ben, een van de grondleggers van de ERF studie, alle proefschriften van de afgelopen jaren binnen de genetische epidemiologie loven je vaderlijke betrokkenheid. Ik kan me daar alleen maar bij aansluiten. Dat je twee jaar na je pensioen nog terugkomt om plaats te nemen in de commissie, waardeer ik enorm. Dr. W.M. van der Flier, bedankt dat je wilt plaatsnemen in de

commissie, je bent een voorbeeld voor elke jonge onderzoeker in de dementie. Prof.Dr. S.A. Kushner, basic science and clinical psychiatry seem to be so far apart, yet you combine them. I am really curious what question you will come up with, thank you for being in my committee.

Een ereplaats in dit dankwoord is ook voor de duizenden deelnemers aan de ERGO studie (Rotterdam studie) en ERF studie. Ik weet voor het grootste deel niet wie jullie zijn, maar doordat jullie vrijwillig je tijd hebben gegeven en je geduldig door het bloedprikken, meten, wegen, de echo's, de ECG's, de foto's en de hele cognitieve testbatterij worstelden, is dit onderzoek mogelijk geweest. Prof.Dr. A. Hofman, afdelingshoofd van de epidemiologie en grondlegger van de Rotterdam studie, en Dr. J. Heeringa, studiecoordinator van de Rotterdam Studie, dankzij jullie leiding helpen de inmiddels hoogbejaarde deelnemers nog altijd om meer en meer data te krijgen, en krijgen jonge onderzoekers de kans om mee te doen op het hoogste niveau binnen de genetische epidemiologie.

Het onderzoek binnen onze afdeling zou ook niet mogelijk zijn zonder de mensen die de gegevens verzamelen en de hele bak aan data minutieus archiveren. Dames van het ERGO centrum, het was altijd gezellig en prettig samenwerken. Jeannette, Andy, Andrea en Bernadette, wat ik ook aan gegevens of samples van jaren terug nodig had, jullie wisten het te vinden, en het was nooit te laat om terug te komen met CJD bloedjes. Petra, je schept orde in de chaos. Elza, bedankt voor je hulp in talloze situaties. Erica, Hetty, Jacqueline: bedankt voor jullie praktische hulp en het aannemen van vele CJD-telefoontjes.

Most of the projects described in this thesis originate from intensive international collaborations. First, there is the CHARGE consortium. The people to thank here would fill the entire acknowledgment section so I can't name them all, but some people I would like to thank personally. Prof.Dr. Bruce Psaty, so many thanks to you and the RSC for creating the great climate for young researchers to work in, with so many opportunities. Your support for the stroke risk score project and the CHARGE fellowship has given me a unique experience. I would also like to name my colleagues from the NeuroCHARGE group: fellow data crunchers in many projects - Josh, Anita, Jan, Seung, Stephanie, Albert, Gail, Jari and all the others: I learnt a lot from you guys, and it was great exchanging ideas. PIs Sudha, Tom, Annette, Myriam, Lenore, Ian and others: we made long and winding roads with our projects, but seem to end up in beautiful places. Thanks for the collaboration.

With the neuroCHARGE group, we also stepped into the IGAP project. Here, too, I unfortunately have to limit myself to the key persons who made this collaboration so great to work in. Fellow data crunchers Gyungah, Denise, Jean-Charles, Benjamin and Celine, and PIs Gerard, Lindsay, Philippe and Julie, it has been an honor to work with you.

The linkage studies described in this thesis would not have been possible without the help of our Russian colleagues at the University of Novosibirsk. Irina and Tatiana, apart from running analyses, with enough explanations for a doctor to understand them, you also gave indispensible advice on our data and methods. Thank you!

Hoewel het in dit proefschrift niet aan bod komt, heeft het werk voor de Landelijke Registratie Prionziekten een belangrijk onderdeel uitgemaakt van mijn promotietijd. Pathologen Annemieke Rozemuller en Wim Spliet van het prionlab UMC Utrecht, met ondersteuning van Will Hermsen, samen konden we de kliniek en pathologie koppelen en duiden. Ik heb veel van jullie geleerd. Corien Swaan van het RIVM en GGD-artsen, bedankt voor de vele malen overleg over casus, het volksgezondheid-denken voegde een extra dimensie toe aan mijn werk. Prof. Dr. Pim van Gool en dr. Pascual Sanchez Juan, bedankt voor de hulp in de moeilijkste puzzels van casuistiek. Wan Zheng, Sara en Eline, bedankt voor de back-up als ik weer eens afwezig was, en Sven voor het opvolgen. My EUROCJD colleagues, led by prof.dr. Bob Will from the University of Edinburgh, I learnt a lot about this strange disease from working with you. En om niet te vergeten, veel dank aan alle patiënten en families die op de moeilijkste momenten van hun leven de tijd en moeite namen om deel te nemen aan het langlopende onderzoek.

Als volgende mijn collega's van de genetische epidemiologie unit, met wie ik bij vlagen meer tijd doorbracht dan thuis. En toch was weer een spontaan afdelingsdiner (sushi!!!) altijd een feest! Sara, we zijn bijna tegelijk begonnen, en hebben vele hoogtepunten en frustraties gedeeld, van een cursus aan het strand in Egmond aan Zee tot grote GWASes onder tijdsdruk met programmatuur die niet meewerkte, en alles wat zich thuis afspeelde. Jij bent de volgende die gaat promoveren, hopelijk heb je een super tijd in Engeland! Linda, je hebt me wegwijs gemaakt op de afdeling als kamergenootje van het eerste uur. Je bent een lief en authentiek persoon, veel succes in de groep van Uitterlinden! Claudia, together with Sara you would keep me alive with tea and cookies when I was too busy to think of food. Looking forward to seeing you again at your own defense! Elisa, het was stil op de kamer als jij er niet was. Heel veel geluk met kleine Charlotte! Adriana, you are one of the sweetest persons I know, and so patient with your zebrafish. Maybe they will tell you their genetic secrets if you kindly ask them! Dina, you compete with Adriana in sweetness. In this hypothesis-free era, you focused on one gene and one trait for your first project, and made it work. I hope you will make yourself more visible, you're worth it! Yurii, you translate the most complex maths into something a non-statistician can work with, both in teaching and in software. Lennart, je vindt taalkundige discussies bijna even leuk als scripten, en weet simpele oplossingen voor complexe problemen. Veel succes aan jullie beiden in jullie eigen bedrijf! Maarten, de man die alles kan, van databeheer tot servermanagement tot programmeren. Veel success met alles! Aaron, thanks for all the good advice and I enjoyed our parenting talk, Noah is lucky to have such a proud father! Maksim, Sophia, Catharina, Revanius, it was nice to have you as a colleague. Good luck with the rest of your careers! En last but not least,

mijn opvolger Sven. Zijn we zeker weten geen familie? Onze voorliefde voor droge rijstwafels (niet de bovenste van de open rol pakken), het uit ons hoofd kennen van kinderliedjes van ver voor onze tijd, de vreemde gelijkenissen zijn soms uncanny. Je hebt niet alleen met verve opgepakt waar ik niet aan toe kwam, maar stort je vol overgave op de nieuwste data en methoden en stevent af op een van de mooiste boekjes binnen ons vakgebied, met nog zo'n twee jaar te gaan. Zie ik je straks ook terug als collega-AIOS?

Ook de collega's van de epidemiologie in bredere zin, bedankt voor jullie collegialiteit en samenwerking! In het bijzonder wil ik Hieab, Ben, Renske, Renee en Elisabeth van de neurogroep, de oogmeisjes Henriette, Gabrielle en Virginie, en Ling en Abbas noemen. De uitwisseling van ideeen in informele gesprekken levert soms nog het meeste op.

Van de afdeling neurologie, prof. Peter Koudstaal bedankt voor je hulp bij de stroke paper, en je altijd persoonlijke inslag. Prof. Peter Sillevis Smitt, bedankt voor het gestelde vertrouwen, gestalte gegeven door de opleidingsplaats neurologie, en je flexibiliteit waardoor ik kliniek en onderzoek kon combineren. Maaike, jij hebt veel van de onderzoeken in dit proefschrift opgestart. Hopelijk ben je tevreden over wat ik er verder mee gedaan heb. Minah, Susanne en collega's, super bedankt voor het mede mogelijk maken van soms last minute ingeplande stukjes schrijfvrij. Mede-AIOS en bazen: we gaan er mooie jaren van maken!

My dear paranimfen. Ayse, you too are one of my colleagues from day 1, and you've always been a special colleague. Your dedication to the work may have taken a self-destructive turn at times, but it made you one of the core persons in the department. When we were no longer roommates, I could always come to you for a talk, often with some exotic tea, or some practical issue. For all this, I'm very glad to have you at my side as my paranimf today. Jennifer, je bent uitgegroeid tot een van mijn beste vriendinnen. We waren samen zwanger, en onze meisjes worden ook al echte vriendinnetjes. Hopelijk kunnen we straks niet alleen de hele dag appen, maar ook nog vaker afspreken, met of zonder de mannen! Met jou als paranimf op mijn promotie voegen we nog een herinnering toe.

Papa en mama (2x), jullie geestelijke steun en belangstelling voor mijn onderzoek, en af en toe jullie misschien niet geheel onterechte bezorgdheid, is minstens even belangrijk geweest als alle keren dat jullie ons uit de brand hielpen door op het laatste moment op te passen.

Lieve Mehdi, mijn voorstelling van zaken toen ik je ruim vier jaar geleden vertelde dat ik wilde gaan promoveren, was misschien niet altijd de werkelijkheid. Ondanks de strubbelingen die we hebben gehad door het werk, en de afspraken die ik soms niet kon nakomen (kan je nog een keer sorry horen?), heb je me tot het einde gesteund. De vele keren dat je me hebt vergezeld op zakenreizen, hebben ons mooie herinneringen gebracht. Ik hou van je.

Lieve Rihana, je bent de beste en meest blijvende herinnering aan mijn promotietijd, sinds je verwekt bent op een paar honderd meter van het Witte Huis tijdens mijn stage bij de NIH. Er is geen dag dat je me niet ontroert, versteld doet staan of aan het lachen maakt! Met zijn drietjes, en straks met zijn viertjes, gaan we heel gelukkig zijn!

## *8.2 About the author*

Carla Antoinette Ibrahim-Verbaas was born in The Hague, The Netherlands, on June 21$^{st}$, 1984. She completed her gymnasium (pre-scientific) education at the Krimpenerwaard college in Krimpen aan den IJssel, The Netherlands, in 2001, in which year she also started her Medicine studies at Leiden University, Leiden, The Netherlands. After completing her university (drs.) degree in Medicine in 2006 and her medical training (artsexamen) in 2008, she started working as a resident not in training (ANIOS) in the Department of Neurology at the Groene Hart Ziekenhuis in Gouda, The Netherlands. In 2009 she came to the Erasmus University Medical Center (Erasmus MC) in Rotterdam, The Netherlands, where she first worked as an ANIOS in the Department of Neurology. In 2010, she started the work on this PhD thesis under supervision of Prof. Dr. Cornelia van Duijn at the Department of Epidemiology, Genetic Epidemiology Unit, and Prof. Dr. John van Swieten at the Department of Neurology at Erasmus MC. During her PhD training, she continued to perform clinical duties at the Department of Neurology. From September to November 2011, she worked at the National Institute on Ageing, Laboratory of Epidemiology and Population Sciences, Bethesda, MD, USA, under the supervision of Dr. L.J. Launer, as part of her PhD training; this research fellowship was sponsored by a grant from the Cohorts for Heart and Ageing Research in Genomic Epidemiology (CHARGE). She obtained her Master of Science degree in Genetic Epidemiology at the Netherlands Institute of Health Sciences (NIHES), Rotterdam, The Netherlands, in 2013. She started her specialty training in Neurology at Erasmus MC, under supervision of Prof. Dr. P.A.E. Sillevis Smitt, in August 2013. She plans to continue her specialty training in the coming years. As part of her specialty training, she started her peripheral residency at the Sint Franciscus Gasthuis in Rotterdam, The Netherlands, supervised by Dr. S.L.M. Bakker, on April 1, 2014.

## *8.3. List of publications*

1. Bis, J.C., DeCarli, C., Smith, A.V., van der Lijn, F., Crivello, F., Fornage, M., Debette, S., Shulman, J.M., Schmidt, H., Srikanth, V., Schuur, M., Yu, L., Choi, S.H., Sigurdsson, S., Verhaaren, B.F., DeStefano, A.L., Lambert, J.C., Jack, C.R., Jr., Struchalin, M., Stankovich, J., **Ibrahim-Verbaas, C.A.,** Fleischman, D., Zijdenbos, A., den Heijer, T., Mazoyer, B., Coker, L.H., Enzinger, C., Danoy, P., Amin, N., Arfanakis, K., van Buchem, M.A., de Bruijn, R.F., Beiser, A., Dufouil, C., Huang, J., Cavalieri, M., Thomson, R., Niessen, W.J., Chibnik, L.B., Gislason, G.K., Hofman, A., Pikula, A., Amouyel, P., Freeman, K.B., Phan, T.G., Oostra, B.A., Stein, J.L., Medland, S.E., Vasquez, A.A., Hibar, D.P., Wright, M.J., Franke, B., Martin, N.G., Thompson, P.M., Enhancing Neuro Imaging Genetics through Meta-Analysis, C., Nalls, M.A., Uitterlinden, A.G., Au, R., Elbaz, A., Beare, R.J., van Swieten, J.C., Lopez, O.L., Harris, T.B., Chouraki, V., Breteler, M.M., De Jager, P.L., Becker, J.T., Vernooij, M.W., Knopman, D., Fazekas, F., Wolf, P.A., van der Lugt, A., Gudnason, V., Longstreth, W.T., Jr., Brown, M.A., Bennett, D.A., van Duijn, C.M., Mosley, T.H., Schmidt, R., Tzourio, C., Launer, L.J., Ikram, M.A., Seshadri, S., Cohorts for, H. & Aging Research in Genomic Epidemiology, C. Common variants at 12q14 and 12q24 are associated with hippocampal volume. *Nat Genet* **44**, 545-51 (2012).
2. Heun, R., Kolsch, H., **Ibrahim-Verbaas, C.A.,** Combarros, O., Aulchenko, Y.S., Breteler, M., Schuur, M., van Duijn, C.M., Hammond, N., Belbin, O., Cortina-Borja, M., Wilcock, G.K., Brown, K., Barber, R., Kehoe, P.G., Coto, E., Alvarez, V., Lehmann, M.G., Deloukas, P., Mateo, I., Morgan, K., Warden, D.R., Smith, A.D. & Lehmann, D.J. Interactions between PPAR-alpha and inflammation-related cytokine genes on the development of Alzheimer's disease, observed by the Epistasis Project. *Int J Mol Epidemiol Genet* **3**, 39-47 (2012).
3. **Ibrahim-Verbaas, C.A.,** Zorkoltseva, I.V., Amin, N., Schuur, M., Coppus, A.M., Isaacs, A., Aulchenko, Y.S., Breteler, M.M., Ikram, M.A., Axenovich, T.I., Verbeek, M.M., van Swieten, J.C., Oostra, B.A. & van Duijn, C.M. Linkage analysis for plasma amyloid beta levels in persons with hypertension implicates Abeta-40 levels to presenilin 2. *Hum Genet* **131**, 1869-76 (2012).
4. Jansen, C., Parchi, P., Capellari, S., **Ibrahim-Verbaas, C.A.,** Schuur, M., Strammiello, R., Corrado, P., Bishop, M.T., van Gool, W.A., Verbeek, M.M., Baas, F., van Saane, W., Spliet, W.G., Jansen, G.H., van Duijn, C.M. & Rozemuller, A.J. Human prion diseases in the Netherlands (1998-2009): clinical, genetic and molecular aspects. *PLoS One* **7**, e36333 (2012).
5. Kolsch, H., Lehmann, D.J., **Ibrahim-Verbaas, C.A.,** Combarros, O., van Duijn, C.M., Hammond, N., Belbin, O., Cortina-Borja, M., Lehmann, M.G., Aulchenko, Y.S., Schuur, M., Breteler, M., Wilcock, G.K., Brown, K., Kehoe, P.G., Barber, R., Coto, E., Alvarez, V., Deloukas, P., Mateo, I., Maier, W., Morgan, K., Warden, D.R., Smith, A.D. & Heun, R. Interaction of insulin and PPAR-alpha genes in Alzheimer's disease: the Epistasis Project. *J Neural Transm* **119**, 473-9 (2012).

6. Bullock, J.M., Medway, C., Cortina-Borja, M., Turton, J.C., Prince, J.A., **Ibrahim-Verbaas, C.A.,** Schuur, M., Breteler, M.M., van Duijn, C.M., Kehoe, P.G., Barber, R., Coto, E., Alvarez, V., Deloukas, P., Hammond, N., Combarros, O., Mateo, I., Warden, D.R., Lehmann, M.G., Belbin, O., Brown, K., Wilcock, G.K., Heun, R., Kolsch, H., Smith, A.D., Lehmann, D.J. & Morgan, K. Discovery by the Epistasis Project of an epistatic interaction between the GSTM3 gene and the HHEX/IDE/KIF11 locus in the risk of Alzheimer's disease. *Neurobiol Aging* **34**, 1309 e1-7 (2013).
7. Jonsson, T., Stefansson, H., Steinberg, S., Jonsdottir, I., Jonsson, P.V., Snaedal, J., Bjornsson, S., Huttenlocher, J., Levey, A.I., Lah, J.J., Rujescu, D., Hampel, H., Giegling, I., Andreassen, O.A., Engedal, K., Ulstein, I., Djurovic, S., **Ibrahim-Verbaas, C.,** Hofman, A., Ikram, M.A., van Duijn, C.M., Thorsteinsdottir, U., Kong, A. & Stefansson, K. Variant of TREM2 associated with the risk of Alzheimer's disease. *N Engl J Med* **368**, 107-16 (2013).
8. Klug, G.M., Wand, H., Simpson, M., Boyd, A., Law, M., Masters, C.L., Matej, R., Howley, R., Farrell, M., Breithaupt, M., Zerr, I., van Duijn, C., **Ibrahim-Verbaas, C.,** Mackenzie, J., Will, R.G., Brandel, J.P., Alperovitch, A., Budka, H., Kovacs, G.G., Jansen, G.H., Coulthard, M. & Collins, S.J. Intensity of human prion disease surveillance predicts observed disease incidence. *J Neurol Neurosurg Psychiatry* **84**, 1372-7 (2013).
9. Lambert, J.-C., **Ibrahim-Verbaas, C.A.,** Harold, D., Naj, A.C., Sims, R., Bellenguez, C., Jun, G., DeStefano, A.L., Bis, J.C., Beecham, G.W., Grenier-Boley, B., Russo, G., Thornton-Wells, T.A., Jones, N., Smith, A.V., Chouraki, V., Thomas, C., Ikram, M.A., Zelenika, D., Vardarajan, B.N., Kamatani, Y., Lin, C.-F., Gerrish, A., Schmidt, H., Kunkle, B., Dunstan, M.L., Ruiz, A., Bihoreau, M.-T., Choi, S.-H., Reitz, C., Pasquier, F., Hollingworth, P., Ramirez, A., Hanon, O., Fitzpatrick, A.L., Buxbaum, J.D., Campion, D., Crane, P.K., Baldwin, C., Becker, T., Gudnason, V., Cruchaga, C., Craig, D., Amin, N., Berr, C., Lopez, O.L., De Jager, P.L., Deramecourt, V., Johnston, J.A., Evans, D., Lovestone, S., Letenneur, L., Moron, F.J., Rubinsztein, D.C., Eiriksdottir, G., Sleegers, K., Goate, A.M., Fievet, N., Huentelman, M.J., Gill, M., Brown, K., Kamboh, M.I., Keller, L., Barberger-Gateau, P., McGuinness, B., Larson, E.B., Green, R., Myers, A.J., Dufouil, C., Todd, S., Wallon, D., Love, S., Rogaeva, E., Gallacher, J., St George-Hyslop, P., Clarimon, J., Lleo, A., Bayer, A., Tsuang, D.W., Yu, L., Tsolaki, M., Bossu, P., Spalletta, G., Proitsi, P., Collinge, J., Sorbi, S., Sanchez-Garcia, F., Fox, N.C., Hardy, J., Naranjo, M.C.D., Bosco, P., Clarke, R., Brayne, C., Galimberti, D., Mancuso, M., Matthews, F., European Alzheimer's Disease, I., Genetic, Environmental Risk in Alzheimer's, D., Alzheimer's Disease Genetic, C., Cohorts for, H., Aging Research in Genomic, E., Moebus, S., Mecocci, P., Del Zompo, M., Maier, W., Hampel, H., Pilotto, A., Bullido, M., Panza, F., Caffarra, P., Nacmias, B., Gilbert, J.R., Mayhaus, M., Lannfelt, L., Hakonarson, H., Pichler, S., Carrasquillo, M.M., Ingelsson, M., Beekly, D., Alvarez, V., Zou, F., Valladares, O., Younkin, S.G., Coto, E., Hamilton-Nelson, K.L., Gu, W., Razquin, C., Pastor, P., Mateo, I., Owen, M.J., Faber, K.M., Jonsson, P.V.,

Combarros, O., O'Donovan, M.C., Cantwell, L.B., Soininen, H., Blacker, D., Mead, S., Mosley Jr, T.H., Bennett, D.A., Harris, T.B., Fratiglioni, L., Holmes, C., de Bruijn, R.F.A.G., Passmore, P., Montine, T.J., Bettens, K., Rotter, J.I., Brice, A., Morgan, K., Foroud, T.M., Kukull, W.A., Hannequin, D., Powell, J.F., Nalls, M.A., Ritchie, K., Lunetta, K.L., Kauwe, J.S.K., Boerwinkle, E., Riemenschneider, M., Boada, M., Hiltunen, M., Martin, E.R., Schmidt, R., Rujescu, D., Wang, L.-S., Dartigues, J.-F., Mayeux, R., Tzourio, C., Hofman, A., Nothen, M.M., Graff, C., Psaty, B.M., Jones, L., Haines, J.L., Holmans, P.A., Lathrop, M., Pericak-Vance, M.A., Launer, L.J., Farrer, L.A., van Duijn, C.M., Van Broeckhoven, C., Moskvina, V., Seshadri, S., Williams, J., Schellenberg, G.D. & Amouyel, P. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* **45**, 1452-1458 (2013).

10. Rietveld, C.A., Medland, S.E., Derringer, J., Yang, J., Esko, T., Martin, N.W., Westra, H.J., Shakhbazov, K., Abdellaoui, A., Agrawal, A., Albrecht, E., Alizadeh, B.Z., Amin, N., Barnard, J., Baumeister, S.E., Benke, K.S., Bielak, L.F., Boatman, J.A., Boyle, P.A., Davies, G., de Leeuw, C., Eklund, N., Evans, D.S., Ferhmann, R., Fischer, K., Gieger, C., Gjessing, H.K., Hagg, S., Harris, J.R., Hayward, C., Holzapfel, C., **Ibrahim-Verbaas, C.A.,** Ingelsson, E., Jacobsson, B., Joshi, P.K., Jugessur, A., Kaakinen, M., Kanoni, S., Karjalainen, J., Kolcic, I., Kristiansson, K., Kutalik, Z., Lahti, J., Lee, S.H., Lin, P., Lind, P.A., Liu, Y., Lohman, K., Loitfelder, M., McMahon, G., Vidal, P.M., Meirelles, O., Milani, L., Myhre, R., Nuotio, M.L., Oldmeadow, C.J., Petrovic, K.E., Peyrot, W.J., Polasek, O., Quaye, L., Reinmaa, E., Rice, J.P., Rizzi, T.S., Schmidt, H., Schmidt, R., Smith, A.V., Smith, J.A., Tanaka, T., Terracciano, A., van der Loos, M.J., Vitart, V., Volzke, H., Wellmann, J., Yu, L., Zhao, W., Allik, J., Attia, J.R., Bandinelli, S., Bastardot, F., Beauchamp, J., Bennett, D.A., Berger, K., Bierut, L.J., Boomsma, D.I., Bultmann, U., Campbell, H., Chabris, C.F., Cherkas, L., Chung, M.K., Cucca, F., de Andrade, M., De Jager, P.L., De Neve, J.E., Deary, I.J., Dedoussis, G.V., Deloukas, P., Dimitriou, M., Eiriksdottir, G., Elderson, M.F., Eriksson, J.G., Evans, D.M., Faul, J.D., Ferrucci, L., Garcia, M.E., Gronberg, H., Guethnason, V., Hall, P., Harris, J.M., Harris, T.B., Hastie, N.D., Heath, A.C., Hernandez, D.G., Hoffmann, W., Hofman, A., Holle, R., Holliday, E.G., Hottenga, J.J., Iacono, W.G., Illig, T., Jarvelin, M.R., Kahonen, M., Kaprio, J., Kirkpatrick, R.M., Kowgier, M., Latvala, A., Launer, L.J., Lawlor, D.A., Lehtimaki, T., Li, J., Lichtenstein, P., Lichtner, P., Liewald, D.C., Madden, P.A., Magnusson, P.K., Makinen, T.E., Masala, M., McGue, M., Metspalu, A., Mielck, A., Miller, M.B., Montgomery, G.W., Mukherjee, S., Nyholt, D.R., Oostra, B.A., Palmer, L.J., Palotie, A., Penninx, B.W., Perola, M., Peyser, P.A., Preisig, M., Raikkonen, K., Raitakari, O.T., Realo, A., Ring, S.M., Ripatti, S., Rivadeneira, F., Rudan, I., Rustichini, A., Salomaa, V., Sarin, A.P., Schlessinger, D., Scott, R.J., Snieder, H., St Pourcain, B., Starr, J.M., Sul, J.H., Surakka, I., Svento, R., Teumer, A., LifeLines Cohort, S., Tiemeier, H., van Rooij, F.J., Van Wagoner, D.R., Vartiainen, E., Viikari, J., Vollenweider, P., Vonk, J.M., Waeber, G., Weir, D.R., Wichmann, H.E., Widen, E., Willemsen, G., Wilson, J.F., Wright, A.F., Conley, D., Davey-Smith, G., Franke, L., Groenen, P.J., Hofman, A.,

Johannesson, M., Kardia, S.L., Krueger, R.F., Laibson, D., Martin, N.G., Meyer, M.N., Posthuma, D., Thurik, A.R., Timpson, N.J., Uitterlinden, A.G., van Duijn, C.M., Visscher, P.M., Benjamin, D.J., Cesarini, D. & Koellinger, P.D. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **340**, 1467-71 (2013).

11. Chouraki, V., De Bruijn, R.F., Chapuis, J., Bis, J.C., Reitz, C., Schraen, S., **Ibrahim-Verbaas, C.A.,** Grenier-Boley, B., Delay, C., Rogers, R., Demiautte, F., Mounier, A., Fitzpatrick, A.L., The Alzheimer's Disease Neuroimaging, I., Berr, C., Dartigues, J.F., Uitterlinden, A.G., Hofman, A., Breteler, M., Becker, J.T., Lathrop, M., Schupf, N., Alperovitch, A., Mayeux, R., van Duijn, C.M., Buee, L., Amouyel, P., Lopez, O.L., Ikram, M.A., Tzourio, C. & Lambert, J.C. A genome-wide association meta-analysis of plasma Abeta peptides concentrations in the elderly. *Mol Psychiatry* (2014) (epub ahead of print, doi:10.1038/mp.2013.185)

12. **Ibrahim-Verbaas, C.A.,** Fornage, M., Bis, J.C., Choi, S.H., Psaty, B.M., Meigs, J.B., Rao, M., Nalls, M., Fontes, J.D., O'Donnell, C.J., Kathiresan, S., Ehret, G.B., Fox, C.S., Malik, R., Dichgans, M., Schmidt, H., Lahti, J., Heckbert, S.R., Lumley, T., Rice, K., Rotter, J.I., Taylor, K.D., Folsom, A.R., Boerwinkle, E., Rosamond, W.D., Shahar, E., Gottesman, R.F., Koudstaal, P.J., Amin, N., Wieberdink, R.G., Dehghan, A., Hofman, A., Uitterlinden, A.G., Destefano, A.L., Debette, S., Xue, L., Beiser, A., Wolf, P.A., Decarli, C., Ikram, M.A., Seshadri, S., Mosley, T.H., Jr., Longstreth, W.T., Jr., van Duijn, C.M. & Launer, L.J. Predicting stroke through genetic risk functions: the CHARGE Risk Score Project. *Stroke* **45**, 403-12 (2014).

13. Malik, R., Bevan, S., Nalls, M.A., Holliday, E.G., Devan, W.J., Cheng, Y.C., **Ibrahim-Verbaas, C.A.,** Verhaaren, B.F., Bis, J.C., Joon, A.Y., de Stefano, A.L., Fornage, M., Psaty, B.M., Ikram, M.A., Launer, L.J., van Duijn, C.M., Sharma, P., Mitchell, B.D., Rosand, J., Meschia, J.F., Levi, C., Rothwell, P.M., Sudlow, C., Markus, H.S., Seshadri, S., Dichgans, M. & Wellcome Trust Case Control, C. Multilocus genetic risk score associates with ischemic stroke in case-control and prospective cohort studies. *Stroke* **45**, 394-402 (2014).

14. Medway, C., Combarros, O., Cortina-Borja, M., Butler, H.T., **Ibrahim-Verbaas, C.A.**, de Bruijn, R.F., Koudstaal, P.J., van Duijn, C.M., Ikram, M.A., Mateo, I., Sanchez-Juan, P., Lehmann, M.G., Heun, R., Kolsch, H., Deloukas, P., Hammond, N., Coto, E., Alvarez, V., Kehoe, P.G., Barber, R., Wilcock, G.K., Brown, K., Belbin, O., Warden, D.R., Smith, A.D., Morgan, K. & Lehmann, D.J. The sex-specific associations of the aromatase gene with Alzheimer's disease and its interaction with IL10 in the Epistasis Project. *Eur J Hum Genet* **22**, 216-20 (2014).

### *8.4. PhD portfolio*

| 1. PhD training | | |
|---|---|---|
| | **Year** | **Workload (Hours/ECTS)** |
| **In-depth courses (e.g. Research school, Medical Training)** | | |
| **Master of Science in Genetic Epidemiology** | **2010-2013** | |
| Erasmus Summer Programme 2010 | 2010 | 4.5 |
| Erasmus Summer Programme 2011 | 2011 | 2.8 |
| CC01 Study Design | 2012 | 4.3 |
| CC02 Classical Methods for Data-analysis | 2010 | 5.7 |
| EP03 Modern Statistical Methods | 2010 | 4.3 |
| GE02 Genetic-Epidemiologic Research Methods | 2012 | 5.7 |
| GE08 SNPs and human diseases | 2010 | 1.4 |
| EP17 Courses for the Quantitative Researcher | 2010 | 1.4 |
| EWP11 Introduction to Clinical and Public Health | 2011 | 1.4 |
| Genomics | 2011 | 1.4 |
| GE03 Advances in Genome-Wide Association Studies | 2012 | 1.4 |
| GE05 Family-based Genetic Analyses | 2011 | 0.9 |
| GE10 Mendelian Randomization | 2012 | 0.6 |
| SC09 Working with Linux | | |
| **Presentations** | | |
| ***Oral*** | | |
| **Project proposal: risk score models**<br>CHARGE meeting February 2011, Boston, MA, USA | 2011 | 1 |
| **IGAP APOE analyses: initial results**<br>IGAP meeting at AAIC 2011, Paris, France | 2011 | 1 |
| **Genome-wide association study of memory performance**<br>AAIC 2011, Paris, France | 2011 | 1 |
| **Update on the stroke risk score project**<br>CHARGE meeting October 2011, Redondo Beach, CA, USA | 2011 | 1 |
| **The CHARGE stroke risk score project**<br>ISGC meeting November 2012, Krakow, Poland | 2012 | 1 |
| **IGAP APOE working group update**<br>Paris, France | 2013 | 1 |
| **APOE-stratified genome-wide association analysis reveals differential effects by APOE status and potential new risk loci for Alzheimer's disease**<br>Alzheimer's Association International Conference 2013, Boston, MA, USA | 2013 | 3 |

| | | |
|---|---|---|
| ***Poster*** | | |
| **Genome-wide Linkage and Association Analyses in Nondemented Hypertensive Individuals Suggest Association of PSEN2 with Plasma Amyloid Beta**<br>Alzheimer's and Parkinson's Disease Congress 2011, Barcelona, Spain | 2011 | 0.5 |
| **Genome-wide Linkage and Association Analyses in Nondemented Hypertensive Individuals Suggest Association of PSEN2 with Plasma Amyloid Beta**<br>European Society of Human Genetics conference 2011,<br>Amsterdam, The Netherlands | 2011 | 0.5 |
| **Genome-wide association study of executive function**<br>Alzheimer's Association International Conference 2011, Paris, France | 2011 | 0.5 |
| **CJD with numerous Aβ plaques in a 58-year old patient 28 years after dura mater grafting**<br>PRION 2012, Amsterdam, The Netherlands | 2012 | 0.5 |
| **A Combined Linkage-sequencing Approach Yields Candidate Variants for Cognitive Function**<br>Alzheimer's and Parkinson's Disease Congress 2013, Florence, Italy | 2013 | 0.5 |
| ***International conferences*** | | |
| **PRION 2010**<br>Salzburg, Austria | 2010 | 1 |
| **Alzheimer's and Parkinson's Disease Congress 2011**<br>Barcelona, Spain | 2011 | 1.5 |
| **European Society of Human Genetics conference 2011**<br>Amsterdam, The Netherlands | 2011 | 1 |
| **Alzheimer's Association International Conference 2011**<br>Paris, France | 2011 | 1.5 |
| **PRION 2012**<br>Amsterdam, The Netherlands | 2012 | 1 |
| **Alzheimer's and Parkinson's Disease Congress 2013**<br>Florence, Italy | 2013 | 1.5 |
| **Alzheimer's Association International Conference 2013**<br>Boston, MA, USA | 2013 | 1.5 |

| *Consortium meetings* | | |
|---|---|---|
| **Cohorts for Heart and Ageing Research in Genomic Epidemiology (CHARGE) meeting 2010**<br>Houston, TX, USA | 2010 | 1 |
| **EUROCJD meeting**<br>Rotterdam, The Netherlands | 2010 | 0.5 |
| **CHARGE meeting February 2011**<br>Boston, MA, USA | 2011 | 1 |
| **EUROCJD meeting**<br>Prague, Czech Republic | 2011 | 0.5 |
| **IGAP meeting at AAIC 2011**<br>Paris, France | 2011 | 0.3 |
| **CHARGE meeting October 2011**<br>Redondo Beach, CA, USA | 2011 | 1 |
| **CHARGE meeting May 2012**<br>Reykjavik, Iceland | 2012 | 1 |
| **ISGC meeting November 2012**<br>Krakow, Poland | 2012 | 0.3 |
| **International Genomics of Alzheimer's (IGAP) meeting**<br>Paris, France | 2013 | 0.5 |
| **CHARGE meeting June 2013**<br>Rotterdam, The Netherlands | 2013 | 0.5 |
| **IGAP meeting at AAIC 2013**<br>Boston, MA, USA | 2013 | 0.5 |
| **Seminars and workshops** | | |
| **Debating Techniques**<br>Pre-conference workshop, PRION 2010, Salzburg, Austria | 2010 | 1 |
| **Weekly seminars at the Department of Epidemiology**<br>Erasmus MC, Rotterdam, The Netherlands | 2010-2013 | 2 |

**2. Teaching activities**

| | **Year** | **Workload (Hours/ECTS)** |
|---|---|---|
| **Lecturing** | | |
| **Presentation "De ziekte van Alzheimer"**<br>Introductory lesson about Alzheimer's disease and genetics for first and second class of secondary school<br>RSG Hoeksche Waard, Oud-Beijerland | 2010 | 1.5 |
| **Presentation "Wetenschap!"**<br>Introductory lesson about my work as a researcher | 2012 | 1 |

| | | |
|---|---|---|
| and developing a hypothesis in the “Familiealbum” project<br>Basisschool “Buiten de Grenzen”, Hellevoetsluis | | |
| **Supervising Master’s theses** | | |
| **Eline Nijskens, medical student (master phase)**<br>Exome-sequencing, in a genetically isolated population that is not selected for any disease, reveals mutations with effects on cognitive performance | 2012 | 3 |
| **Other** | | |
| **Creutzfeldt Jakob disease registry**<br>Maintenance of the Dutch CJD registration, patient visits and inclusion in biobank, information service for clinicians, participation in research | 2010-2012 | 25 |
| **Department of Epidemiology Seminar organization**<br>Speaker invitations and contact, hosting weekly seminar sessions | 2010-2013 | 5 |
| **Reviewer of various papers for international journals** | 2010-2013 | 2 |