A Multi-Omics Epidemiologic Study of Alzheimer's Disease

Shahzad Ahmad

A Multi-Omics Epidemiologic Study of Alzheimer's disease

SHAHZAD AHMAD

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

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

The work presented in this thesis was supported by funded by the Defining Genetic, Polygenic and Environmental Risk for Alzheimer's Disease, using multiple powerful cohorts, focused Epigenetics and Stem cell metabolomics (PERADES) Programme (Project No. 733051021), the European Union Innovative Medicine Initiative (IMI) programme: the Alzheimer's Disease Apolipoprotein Pathology for Treatment Elucidation and Development (ADAPTED, Project No. 115975) and the Alzheimer's disease sequencing project (ADSP). Other support includes Horizon 2020 programme: CoSTREAM (Project No. 667375-2) and the Gut Liver Brain Biochemical Axis in Alzheimer's Disease (Project No. 1RF1AG058942-01) and AMP-AD project (U01-AG061359 NIA).

The publication and printing of this thesis was financially supported by the department of Epidemiology, Erasmus Medical Center, Rotterdam, Erasmus University Rotterdam, and Alzheimer Nederland.

Cover design: Shahzad Ahmad and Erwin Timmerman Optima.nl, postermywall.com
Layout: Marian Sloot Printing: ProefschriftMaken
ISBN/EAN: 978-94-6380-939-9

A Multi-Omics Epidemiologic Study of Alzheimer's disease

Een multi-omics epidedemiologische studie van de ziekte van Alzheimer

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof. dr. R.C.M.E. Engels

and in accordance with the decision of the Doctorate Board.
The public defence shall be held on

Thursday 1 October 2020 at 11:30 hours

by

Shahzad Ahmad
born in Kasur, Pakistan

Erasmus University Rotterdam

Doctoral Committee:

Promotor: Prof. dr.ir. C.M. Van Duijn

Other members: Prof. dr. T. Hankemeier
Prof. dr. M.A. Ikram
Prof. dr. F. Rivadeneira

Copromotor: Dr. N. Amin

Paranymphs: Hata Karamujić-Čomić
Ivana Prokić

قَالُوا۟ سُبْحَٰنَكَ لَا عِلْمَ لَنَآ إِلَّا مَا عَلَّمْتَنَآ ۖ إِنَّكَ أَنتَ ٱلْعَلِيمُ ٱلْحَكِيمُ ﴿٣٢﴾

Dedicated to my parents, Ramzano Noor, and Noor Muhammad;

to my wife, Noreen, and my daughter, Minha;

to my teachers and mentors

Table of Contents

Publications and manuscripts in this thesis

Chapter 2

Shahzad Ahmad, Rafael A. Nafikov, Elizabeth Blue, Patrick A. Navas, Hata Karamujić-Čomić, Dina Vojinovic, Sven Van der Lee, Ashley van der Spek, Bowen Wang, Andrea R. Horimoto, Alejandro Q. Nato Jr, Harkirat K. Sohi, Mohamad Saad, Michael O. Dorschner, Debby Tsuang, Li-San Wang, CHARGE, ADSP Data Flow Work Group, ADSP Family Analysis Work Group, Lindsay A. Farrer, Jonathan L. Haines, Margaret Pericak-Vance, Gerard Schellenberg, Eric Boerwinkle, Richard P. Mayeux, M. Arfan Ikram, Anita DeStefano, Sudha Seshadri, Najaf Amin, Ellen M. Wijsman*, Cornelia M. Van Duijn* from the Alzheimer's disease sequencing project (ADSP). Variation in cadherin genes in the 5p14.3 genomic region is implicated in Alzheimer's disease in an extended ADSP pedigree. (Submitted)

*Senior authors contributed equally to the work

Chapter 3.1

Shahzad Ahmad, Christian Bannister, Sven J. van der Lee, Dina Vojinovic, Hieab H.H. Adams, Alfredo Ramirez, Valentina Escott-Price, Rebecca Sims, Emily Baker, Julie Williams, Peter Holmans, Meike W. Vernooij, M. Arfan Ikram, Najaf Amin, Cornelia M. van Duijn. Disentangling the biological pathways involved in early features of Alzheimer's disease in the Rotterdam Study. The Alzheimer's and dementia: The Journal of the Alzheimer's Association. 2018 July;14(7): 848–857

Chapter 3.2

Silvan Licher, Shahzad Ahmad, Hata Karamujić-Čomić, Trudy Voortman, Maarten J.G. Leening, M. Arfan Ikram and M. Kamran Ikram. Genetic predisposition, modifiable risk factor profile and long-term dementia risk in the general population. Nature Medicine. 2019 August; 25:1364–1369

Chapter 4.1

Shahzad Ahmad, Marta del Campo Milan, Oskar Hansson, Ayse Demirkan, Ruiz Agustin, Maria E Sáez, Nikolaos Giagtzoglou, Alfredo Cabrera-Socorro, Margot H.M. Bakker, Alfredo Ramirez, Thomas Hankemeier, Erik Stomrud, Niklas Mattsson-Carlgren, Philip Scheltens, Wiesje van der Flier, M. Arfan Ikram, Anders Malarstig, Charlotte E Teunissen, Najaf Amin, Cornelia M. van Duijn. CDH6 and HAGH protein levels in plasma associate with Alzheimer's disease in *APOE* ε4 carriers. Scientific Reports. 2020 May; 10:8233

Chapter 4.2

Shahzad Ahmad, Adelina Orellana, Isabelle Kohler, Lutz Frölich, Itziar de Rojas, Silvia Gil, Mercè Boada, Isabel Hernández, Lucrezia Hausner, Margot H.M. Bakker, Alfredo Cabrera-Socorro, Najaf Amin, Alfredo Ramírez, Agustín Ruiz, Thomas Hankemeier*, Cornelia M. Van Duijn*. Association of lysophosphatidic acids with cerebrospinal fluid biomarkers and progression to Alzheimer's disease. (Under revised submission in Alzheimer's research and Therapy)

*Senior authors contributed equally to the work

Chapter 5.1

Kwangsik Nho*, Alexandra Kueider-Paisley*, Shahzad Ahmad*, Siamak Mahmoudian Dehkordi, Matthias Arnold, Shannon L. Risacher, Gregory Louie, Colette Blach, Rebecca Baillie, Xianlin Han, Gabi Kastenmüeller, John Q. Trojanowski, Leslie M. Shaw, Michael W. Weiner, P. Murali Doraiswamy, Cornelia van Duijn, Andrew J. Saykin#, Rima Kaddurah-Daouk#, for the Alzheimer's Disease Neuroimaging Initiative and the Alzheimer Disease Metabolomics Consortium.. Association of altered liver enzymes with Alzheimer's disease diagnosis, cognition, neuroimaging measures, and cerebrospinal fluid biomarkers. JAMA Network Open. 2019 Jul 3;2(7): e197978

*First authors contributed equally to the work
#Senior authors contributed equally to the work

Chapter 5.2

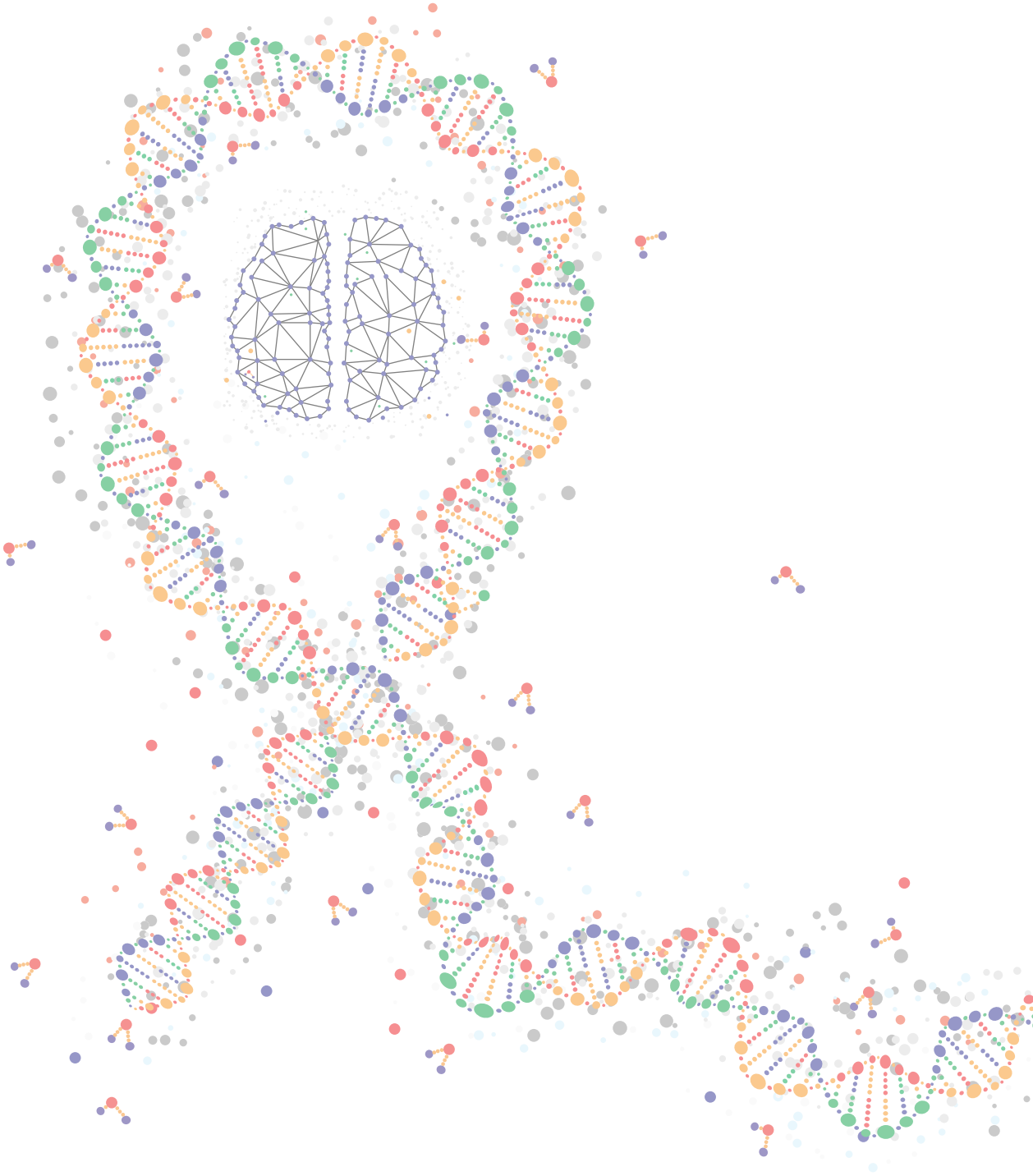
Siamak MahmoudianDehkordi*, Matthias Arnold*, Kwangsik Nho*, Shahzad Ahmad, Wei Jia, Guoxiang Xie, Gregory Louie, Alexandra Kueider-Paisley, M. Arthur Moseley, J. Will Thompson, Lisa St John Williams, Jessica D. Tenenbaum, Colette Blach, Rebecca Baillie, Xianlin Han, Sudeepa Bhattacharyya, Jon B. Toledo, Simon Schafferer, Sebastian Klein, Therese Koal, Shannon L. Risacher, Mitchel Allan Kling, Alison Motsinger-Reif, Daniel M. Rotroff, John Jack, Thomas Hankemeier, David A. Bennett, Philip L. De Jager, John Q. Trojanowski, Leslie M. Shaw, Michael W. Weiner, P. Murali Doraiswamy, Cornelia M. van Duijn, Andrew J. Saykin#, Gabi Kastenmüller#, Rima Kaddurah-Daouk#, for the Alzheimer's Disease Neuroimaging Initiative and the Alzheimer Disease Metabolomics Consortium. Altered bile acid profile associates with cognitive impairment in Alzheimer's disease – An emerging role for gut microbiome. the Alzheimer's and dementia: The Journal of the Alzheimer's Association. 2019 January;15(1): 76–92

*First authors contributed equally to the work
#Senior authors contributed equally to the work

Chapter 5.3

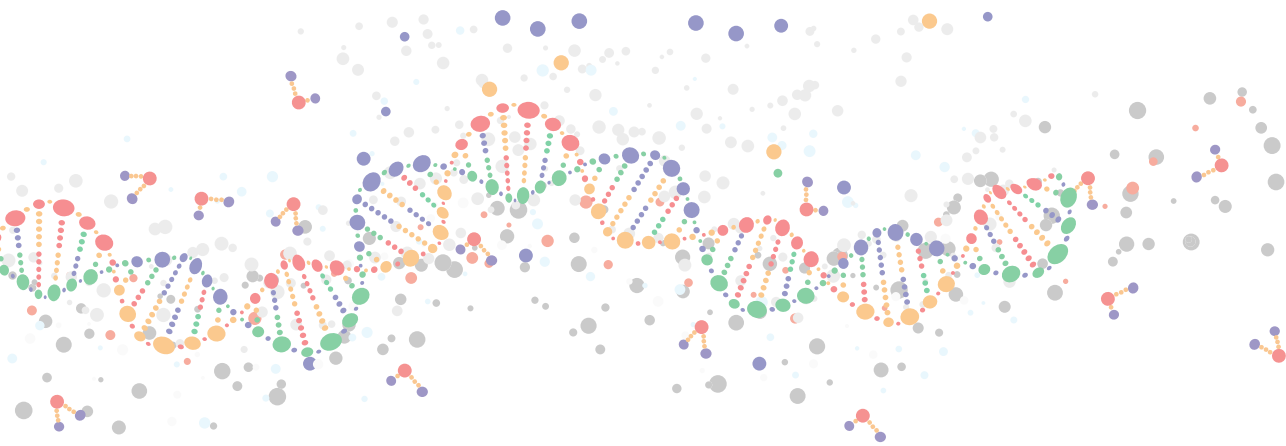
Shahzad Ahmad*, Hata Karamujić-Čomić*, Djawad Radjabzadeh*, Bruno Bonnechere, Dina Vojinovic, Thomas Hankemier, M. Arfan Ikram, Andre G. Uitterlinden, Rima Kaddurah-Daouk, Robert Kraaij, Najaf Amin, Cornelia M. van Duijn. The Alzheimer's disease genetic risk variant in *ABCA7* is associated to the gut microbiome. (In preparation)

*First authors contributed equally to the work



Chapter 1

General Introduction



Alzheimer's disease (AD) is a neurodegenerative disorder that accounts for 50-70% of the worldwide dementia cases in elderly people[1,2]. Almost 47 million individuals are reported to suffer from dementia worldwide, and this number is expected to double every 20 years[1,3]. AD is characterized by progressive loss of memory, problems in executive functioning and daily life activities[4]. The prevalence of AD is rising sharply in aging populations[5]. The increasing disease burden is posing serious health and economic challenge in modern society.

Pathologically, AD is characterized by two hallmarks, amyloid-beta (Aβ) plaques and neurofibrillary tangles (NFTs). Plaques and tangles are primarily observed in brain regions involved in memory, learning and emotional behaviors, such as the hippocampus, entorhinal cortex, and amygdala[6]. Plaques are generated due to improper cleavage of amyloid precursor protein (*APP*), resulting in the formation of Aβ monomers. These monomeric peptides aggregate to form the amyloid plaques that may damage the synapses and neurites. The *APP* gene is implicated in the causal pathway of at least a subgroup of patients because mutation carriers of the *APP* gene develop early-onset AD. NFTs are formed as a result of hyperphosphorylation of the microtubule-associated protein tau, leading to the disruption of axonal transport and neuronal damage[7,8]. Alongside the two hallmarks of AD pathology, it has been recognized for long that AD pathology is accompanied by neuro-inflammation[9,10]. Recent genetic studies have implicated the microglia and astrocytes as key players in the AD pathophysiology[11]. Despite extensive research about the role of these core pathologies, unraveling the underpinning mechanism of AD pathology which contributes to the onset and progression of AD remains a challenge in achieving a suitable treatment and prevention of AD.

For decades, the diagnosis of AD was based on the criteria of Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV-TR)[12] and the National Institute of Neurological Disorders and Stroke–Alzheimer Disease and Related Disorders (NINCDS-ADRDA)[13] working group, which were and are still commonly used in clinical research. These criteria rely on the initial identification of dementia symptoms followed by the assessment of clinical features of AD phenotype[14]. Clinically, these criteria support the diagnosis of a probable and possible AD, but definitive AD diagnosis relies on the histopathological confirmation of the clinical diagnosis[15]. However, advances in biomarker research have paved the way for a more accurate and reliable diagnosis of AD using structural magnetic resonance imaging (MRI) features, neuroimaging with positron emission topography (PET) and cerebrospinal fluid (CSF) pathology biomarkers and thus have forced a reconsideration of the established clinical diagnostic criteria[14].

Recently, the National Institute on Aging and Alzheimer's Association (NIA-AA) Research Framework proposed A/T/N biomarker criteria to identify high-risk AD subjects. Where "A" indicates the Aβ biomarker (CSF Aβ-42 or cortical amyloid positron emission tomography [PET]); "T" refers to tau biomarker (CSF levels of phosphorylated tau or tau PET); and "N" represents the biomarker values of neuronal injury or neurodegeneration (structural MRI,

[18F]-fluorodeoxyglucose–PET, or CSF total tau)[16]. Although A/T/N biomarkers may assist to identify the high-risk subjects during the prodromal stage or even the preclinical phase of AD, CSF and MRI biomarkers are costly and invasive, and therefore not suitable for screening large populations. Discovery of less invasive blood-based biomarkers is needed which would allow screening of high-risk populations. A/T/N biomarkers are also increasingly been used to evaluate the relevance of new molecular pathways in AD pathophysiology.

Technological advancement has fueled multi-omics research in AD to unravel the underlying etiology of AD to enable treatment and early diagnosis of disease. AD is determined by a complex interplay of genetic and environmental risk factors[17]. Mounting evidence suggests the co-occurrence of cardiovascular risk factors (e.g., diabetes, hypertension and obesity) and classic AD neuropathology[18-20]. Smoking, obesity, low education attainment, physical activity, and diet may contribute to an increased risk of AD[21]. A report from the *Lancet commission* acknowledged the epidemiological findings and suggested that 35% of the dementia risk can be reduced by modifying these cardiovascular and lifestyle-related risk factors[22].

Preclinical and prodromal phase of AD

The AD pathogenesis can be conceptualized as a trajectory including the preclinical stage (presymptomatic), mild cognitive impairment (MCI) (early symptomatic or predementia), and AD dementia [23]. The early preclinical stage of AD starts decades before the symptomatic phase[24] and it can be characterized by altered levels of biomarkers[25,26] (Figure 1). Biomarkers may include both structural brain features at neuroimaging and biochemical changes in blood, and CSF[27,28]. With regard to neuroimaging, MRI is employed to detect volumetric and structural changes to brain morphology and vascular features[29]. MRI measures such as global cortical atrophy, hippocampal atrophy, and white matter hyperintensities are shown to be associated with cognitive measures and risk of dementia[30] and have increasingly been applied in healthy populations as biomarkers of AD diagnosis and its progression[31]. A recent development in PET tracers has allowed *in vivo* detection of Aβ abnormalities (imaging radiotracer: [^{18}F]-florbetaben) and tau accumulations (^{18}F-THK5351 and ^{18}F-AV-1451) with high accuracy[32,33]. Earlier studies have shown the association of tau PET binding with an increased amyloid pathology[34], brain structural atrophy[35], and glucose hypometabolism in preclinical AD individuals. Ongoing progress in tau PET scanning further holds the potential to monitor tau deposition during the early stages of neurodegeneration[36,37]. The preclinical stage is regarded as the best stage for evaluating the genetic and metabolic risk factors involved in disease progression to AD. Advances in CSF and blood-based biomarkers detection made it possible to monitor pathophysiological processes in the preclinical stage of AD (Figure 1) but higher costs of MRI or PET scanning and invasive nature of CSF based methods make them less attractive for high-throughput population screening. There have been recent developments in research of blood-based markers[38]. Future

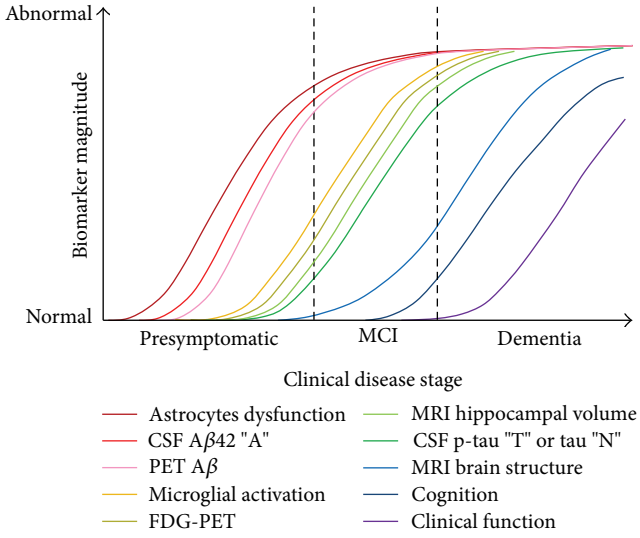


Figure 1: Biomarkers of AD trajectory over time, presenting three stages of AD, presymptomatic stage, MCI and eventually AD dementia. Reprinted from Leclerc *et al.*[28] Abbreviations: MCI, mild cognitive impairment; MRI, magnetic resonance imaging; CSF, cerebrospinal fluid.

research should focus blood-based methods to discover biomarkers of preclinical stage which are relatively less invasive and cost-effective.

The second stage of AD is described by MCI due to AD, which is an early symptomatic predementia phase[23] (Figure 1). Although there is an ongoing discussion, it is of note that 32% of the MCI patients progress into AD within five years; therefore, MCI is often viewed as a prodromal phase of AD dementia[27,39]. Due to the high prevalence of MCI in the general population and the higher risk of MCI patient progression into AD compared with cognitively normal people[40,41], understanding the genetic and biochemical risk factors of MCI may unravel molecular pathways that are relevant for preventive intervention. MCI patients exhibit a 20% to 30% higher prevalence of cerebral amyloid pathology compared with cognitively normal subjects, which is two to three times higher in *APOE* ε4 carriers compared to the noncarriers[42]. Multiple evidence suggests that CSF levels of Aβ-42, p-tau and total tau can be used to identify MCI patients who are at higher risk of developing into AD[43]. Prediction sensitivity of CSF levels of Aβ-42 is 79% and specificity of 65%. P-tau has a sensitivity of 84% and specificity of 47%, while total-tau predicts AD conversion with a sensitivity of 86% and specificity of 56%[43]. A recent study has reiterated the generalizability and robustness of CSF biomarker-based models for prediction of dementia in MCI patients (Aβ-42, p-tau and hippocampal volume)[44]. The study basically constructed and tested various CSF biomarker-based models in multiple cohorts and shown the potential clinical implications of these models in predicting MCI to AD progression[44].

Omics research of Alzheimer's disease

Omics approaches refer to high-throughput technologies aiming to assess the molecular components of biological systems[45]. Several layers of omics including genomics, epigenomics, proteomics, metabolomics, and microbiomics, are used to improve the understanding of etiology and pathophysiology of disease at the molecular level (Figure 2)[46]. The aim of this thesis is to use a multi-omics approach to decipher the molecular pathways of AD.

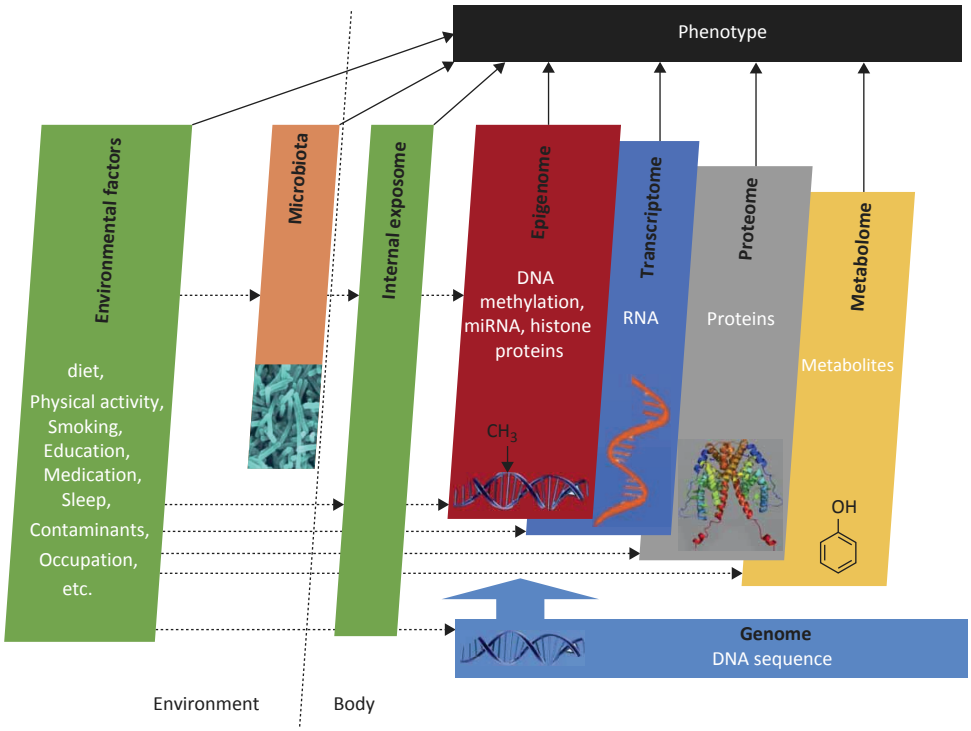


Figure 2: Multi-omics layers and their interaction with environmental factors in biological systems. Source: Adapted from Siroux *et al*[47].

Genetics

In the last decade, substantial advances in the discovery of genetic determinants of AD have been made. The genetic component of AD constitutes a major driving force of AD pathophysiology, and genetic discoveries over the years played a pivotal role in our current understanding of disease molecular mechanism. Based on the age at onset, AD patients who develop dementia before 65 years of age are classified as early-onset AD[48], of which almost half is contributed by mutations in three genes i.e., *PSEN1*[49], *PSEN2*[50] and *APP*[51]. Early-onset AD accounts for only 1-5 % of the total AD cases, of which 10-15 % cases follow autosomal dominant inheritance[52]. Most of the AD cases are sporadic in nature and often develop dementia after 65 years of age[48]. The genetic

etiology of the late-onset AD patients is more complex and heritability is estimated to vary from 60% to 80%[53]. To date, the ε4 allele of the *APOE* gene is the strongest common genetic risk variant for late-onset AD[54] and it confers AD risk in a dose-dependent manner. The *APOE* gene has three allelic variants ε4, ε3, and ε2, of which homozygote carriers of *APOE* ε4 allele have 50% higher lifetime risk of AD compared with 10% for non-carriers by age 85, while *APOE* ε2 allele is considered protective for AD[55,56]. Although *APOE* ε4 allele has a frequency of nearly 25% in the general population[57], it is neither necessary nor sufficient to develop AD therefore not used alone for AD diagnosis[58,59]. It is of note that at least one-third of the AD patients are *APOE* ε4 non-carriers, and nearly 50% homozygotes of *APOE* ε4 do not develop AD by age 80[57]. In the Rotterdam Study, van der Lee *et al.,*[60] have shown that the other common genetic variants may also affect the risk of dementia, particularly in the *APOE* ε4 carriers. Despite the discovery of the *APOE* gene a two decades ago, its role in AD pathology remains unclear as there is a wide range of mechanisms of action of *APOE* in AD, and many questions remain to be answered. A recent study has suggested that *APOE* may contribute to the amyloid disposition in preclinical subjects whereas other common genetic factors may drive the progression of AD in amyloid positive individuals[61]. Disentangling the complex interaction of the *APOE* gene with lifestyle and systemic risk factors would provide new insights into the *APOE* molecular mechanism, and pathophysiology and risk of AD (see Chapter 3.2).

Technical and methodological advances in genotyping and genetic imputations have paved the way for genome-wide association studies (GWAS), which test the association of millions of single nucleotide polymorphism (SNPs) in the whole genome to disease status in individuals. In parallel, several international groups including the Alzheimer's Disease Genetics Consortium (ADGC), Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE), and the Genetic and Environmental Risk in Alzheimer's Disease (GERAD) have joined their efforts under the umbrella of the International Genomics of Alzheimer's Project (IGAP) to push forward the growth in the genetics of AD. IGAP reported 11 novel loci in addition to replicating eight previously known genetic loci from earlier GWAS studies[62-67]. More recently, in January 2019, IGAP published the largest GWAS to date and reported five novel AD risk variants[68] (Figure 3). Moreover, Jansen *et al.*, conducted a GWAS largely based on the UK Biobank data (~635,000) using the family history of AD as 'proxy AD' and identified nine genetic variants, in addition to replicating few already existing risk loci[69]. Together, various meta-analyses of GWAS studies have identified more than 39 risk loci for late-onset AD[70].

One of the major contributions of GWAS is providing insights into the biological pathways involved in AD. Pathway enrichment analysis based on data integration of ribonucleic acid (RNA) expression with AD GWAS meta-analysis results, identified eight gene pathways implicated in AD[71] including *immune response, endocytosis, cholesterol transport, hematopoietic cell lineage, protein ubiquitination, hemostasis, clathrin/AP2 adaptor complex* and *protein folding* [72]. These diverse biological pathways may be involved in the clinically heterogeneous manifestation

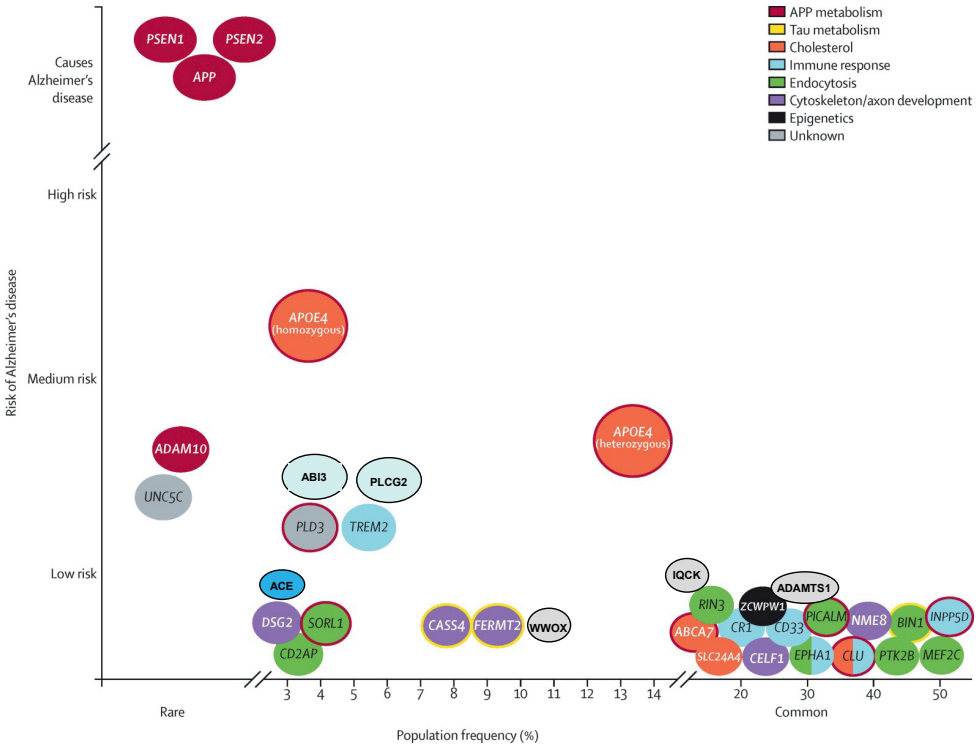


Figure 3: Alzheimer's disease implicated genetic variants with respect to population frequency, effect size and implicated biological pathways. Modified reprint from Scheltens *et al*[3].

of MRI endophenotypes[73-75], and may also modulate the prodromal stages of AD[76-78]. More recently Kunkle *et al.,*[68] is the first one to report the enrichment of the tau binding proteins and *APP* metabolism-related pathways in late-onset AD using GWAS data. Disentangling the role of AD implicated biological pathways in early AD pathology may help to improve our understanding of the pathogenesis of AD during the predementia stage. In chapter 3.1 of this thesis, I have addressed several unanswered questions, including whether AD-related pathways are associated with brain structural and volumetric changes during the preclinical stage of AD, the incidence of MCI in the Rotterdam Study.

All known common SNPs explain only 16% of the variation in the clinical manifestation and 31% of the genetic variance of AD[68,70,79], leaving nearly 60% of the genetic risk uncharacterized[80]. It is likely that rare variants missed out in imputation based GWAS may explain the remaining clinical and genetic variance of AD[81], which reiterates the need to detect these rare variants using advanced genome sequencing techniques. Recently reported rare variant discoveries in *PLD3*, *APP*, *ABI3, PLCG2,* and *TREM2* genes[82-85] strengthen the notion that rare variants may fill the space of missing heritability of late-onset AD[86]. Moreover, Alzheimer's disease sequencing project

(ADSP) has reported the association of three novel rare genetic variants in *IGHG3*, *AC099552.4* and in *ZNF655* genes[87]. Advances in next-generation sequencing techniques made it possible to couple genetic linkage analysis with deep genetic information to identify rare genetic variants using family-based data[81]. Genetic linkage analysis is a classic approach to genetic discovery to identify genomic regions segregating with disease status in families with multiple affected individuals. Genetic linkage is based on the principle that genes that are located physically close to each other on the chromosome, also segregate together during meiosis[88]. The availability of suitable multigenerational families of AD cases is one of the major shortcomings in the discovery of rare variants using linkage analysis[89]. Extended pedigrees with multiple AD cases are expected to harbor highly penetrant variants, therefore they are ideally suited to identify disease loci[90]. In chapter 2 of this thesis, I have studied the genetics of AD in complex multi-generation families of the highly inbred Genetic Research in Isolated Population (GRIP).

Proteomics

Unlike genes, protein levels can be influenced by several factors such as environment, disease stage, medication use, diet patterns. Since disease-related molecular changes are reflected at transcriptome and proteome level, proteomics can be exploited to discover disease biomarkers, targets for treatment, and to understand the disease pathophysiology[91]. One of the major contributions of the genomic studies is to provide insight into the protein functions[92]. Evidence suggests that the majority of AD implicated genes are expressed in blood-derived macrophages[68,93,94] and can modulate the protein expression levels in the blood circulation. Moreover, the interaction of *APOE* with common genetic pathways[60] and its impact on blood-brain barrier integrity[95] can lead to altered blood protein levels which may represent AD brain pathophysiology[96]. Large scale proteomic studies are now emerging in AD research[97-102] and already have identified altered levels of proteins in the circulation, representing signaling, synaptic and oxidative stress-related pathophysiology[103,104]. There is an increasing interest in the relationship between altered levels of proteins and AD in the presymptomatic stage. Prospective studies targeting various protein pathways and their interaction with genetic risk factors such as *APOE* would provide new insight into biomarkers and pathways altered prior to the onset of AD. In this thesis (Chapter 4.1), I studied the association of plasma levels of proteins profiled during the predementia stage of AD and their interaction with *APOE* in the Rotterdam Study.

Metabolomics

Advances in high-throughput metabolomics allowed the detection of hundreds of biochemical compounds (metabolites) in blood, CSF, urine and brain tissues[105]. Metabolic imprints represent complex interactions of genes, proteins, and environmental factors, thus helping to provide deep

insight into disease pathophysiology[105,106]. Metabolomics may provide promising biomarkers for disease progression due to the complex and dynamic nature of the AD continuum, and because blood metabolic repertoire usually reflects CSF and brain biochemical changes[107,108]. Metabolomic studies on tissue samples of MCI and AD patients have suggested altered metabolic function in the preclinical[109-111] and clinical stages of AD[111-113]. Increasing evidence suggests the role of altered levels of metabolites in several key AD related pathways including lipid metabolism, amino acid metabolism, energy metabolism oxidative stress, synaptic function, cell signaling, and inflammation[114,115]. Until now, among all pathways identified by metabolomic studies, lipid-related metabolites provide the strongest and most consistent evidence of association with AD which is also supported by the role of *APOE* in lipid uptake and transport[111]. Several lipid classes have been linked to dementia so far, such as phospholipids, sphingolipids, sterols, sphingomyelins, and phosphatidylcholines[4,108,111,115-117]. Due to the huge risk attributed by *APOE* and its role in lipid transport, the interaction of *APOE* with lipid metabolites need special attention. Moreover, studies are needed that address the role of signaling lipids in AD pathophysiology and in AD progression. In chapter 4.2 of this thesis, I performed an investigation which evaluate the role of signaling lipids (lysophosphatidic acids) in AD. This study was conducted in cohorts participating in the Alzheimer's Disease Apolipoprotein Pathology for Treatment Elucidation and Development (ADAPTED) consortium including the Barcelona-based memory clinic Fundació ACE and the Department of Geriatric Psychiatry at the Medical Faculty Mannheim, University of Heidelberg.

Gut-liver-brain axis

Evidence of metabolic dysfunction in AD trajectory[118] and increased risk of AD contributed by metabolic disorders (diabetics, hypertension, and obesity), strengthen the notion that AD metabolic disorders play a key role in either etiology or progression of disease[119-122]. Disturbed energy metabolism[123,124] and altered gut microbiota in AD animal models[125-127], all point to the complex interplay of liver and gut in AD pathophysiology. Studies addressing the determinants of the liver, gut microbiome, and brain biochemical communication and their relationship to AD pathogenesis could help to understand the unknown role of the Gut-Liver-Brain axis in AD. The liver being a major metabolic organ can directly influence the metabolic milieu of circulation[128,129] or indirectly via the production of bile acids which can modulate the gut microbiota composition[130,131] (Figure 4). Although several studies suggest the association of liver disease with brain structural features, cognition and dementia[132-134], there are many questions to answer on the role of liver function biomarkers in AD pathophysiology. In this thesis (chapter 5.1), I describe the association of liver function biomarkers with biomarkers of AD pathophysiology in the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort.

Another integral component of gut-liver brain axis is the gut microbiota, which makes up to 95% of the total human microbiota[135]. The gut microbiota can regulate the levels of metabolites, cytokines, immunological and neuronal signals[135-138], thus physiologically connecting the gut, the

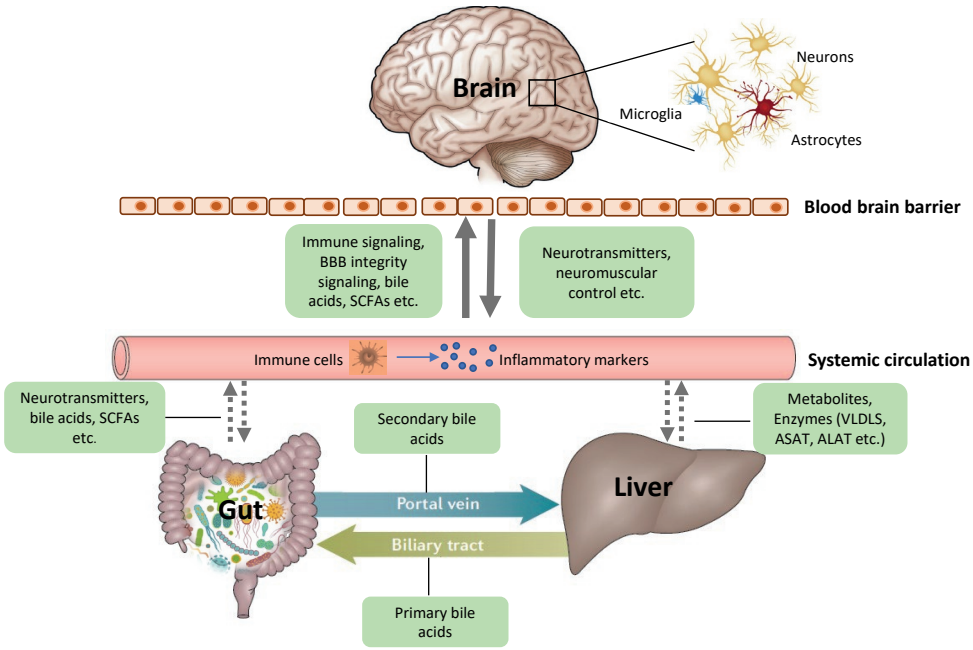


Figure 4: Gut-liver-brain axis. Abbreviations: BBB, blood-brain barrier; SCFA, Short-chain fatty acids; ASAT, Aspartate aminotransferase; ALAT, Alanine aminotransferase. Source: Adapted from Tripathi *et al*[148].

brain, and liver[139] (Figure 4). Secondary bile acids represent an important microbial-derived class of metabolites[140] which can cross blood-brain barrier[141] and act as signaling molecules to regulate the molecular processes in the central nervous system including the energy homeostasis[142]. In Chapter 5.2 of this thesis, I have studied the role of bile acid metabolites in AD in the ADNI and the Rotterdam Study cohorts. Gut microbial diversity is also known to affect levels of short-chain fatty acids such as butyrate, acetate, and propionate in circulation, which may affect inflammatory, *immune response* and the blood-brain barrier function[143,144] (Figure 4). Moreover, increasing evidence from genetic studies implicate *immune response* and brain resident microglia in early phase of AD[145,146]. Function of microglial cells can also be modulated by signaling molecules originating from the gut microbiota, thus, microglia can act at the junction of the gut brain-axis in AD[146,147]. The complex relationship between AD genes, microglia, and microbiota poses another unanswered question i.e., whether, AD risk genes can modify human gut microbial abundance in a non-demented population. Answering this question would help to identify bacterial taxa involved in AD pathophysiology.

Scope of this thesis

The aim of this thesis is to identify the genetic, metabolic and proteomic determinants of AD and to evaluate the interaction of lifestyle factors with the genetic risk of AD.

In **chapter 2**, I present a genetic linkage analysis to identify the genomic regions linked with AD in complex inbred Dutch families and subsequently used whole-genome sequencing to identify the rare genetic variants in the identified linkage region.

Chapter 3 explores the role of AD pathways in the predementia phase of AD and the interaction of lifestyle factors in AD genetic risk. In **chapter 3.1,** I study the association of AD implicated biological pathways with MCI and brain structural features in a healthy population. **Chapter 3.2** provides insight into the role of modifiable lifestyle factors in the genetic risk of dementia in the Rotterdam Study.

Chapter 4 studies the metabolic and proteomic determinants of AD. **Chapter 4.1** explores the altered levels of brain-specific proteins in the blood prior to the onset of AD in a prospective manner and whether *APOE* can influence this association. I studied the role of lysophosphatidic acids in AD pathophysiology in the MCI population and their role in progression to AD in **Chapter 4.2.**

Chapter 5 studies the determinants of the liver-gut-brain axis in AD. **Chapter 5.1** highlights the role of liver function in AD pathology and their relationship with various endophenotypes of AD. **Chapter 5.2** focusses on the relationship of bile acids with AD, cognition and AD genetic variants. In **chapter 5.3**, I study the impact of AD genetic risk factors on the abundance of various gut microbiota in the healthy population.

The findings are integrated and discussed in general discussion (**Chapter 6**)

References

1. Winblad, B. *et al.* Defeating Alzheimer's disease and other dementias: a priority for European science and society. *Lancet Neurol* **15**, 455-532 (2016).
2. Rosenthal, S.L. & Kamboh, M.I. Late-Onset Alzheimer's Disease Genes and the Potentially Implicated Pathways. *Curr Genet Med Rep* **2**, 85-101 (2014).
3. Scheltens, P. *et al.* Alzheimer's disease. *Lancet* (2016).
4. Tochel, C. *et al.* What outcomes are important to patients with mild cognitive impairment or Alzheimer's disease, their caregivers, and health-care professionals? A systematic review. *Alzheimers Dement (Amst)* **11**, 231-247 (2019).
5. Guerreiro, R. & Bras, J. The age factor in Alzheimer's disease. *Genome Med* **7**, 106 (2015).
6. Mattson, M.P. Pathways towards and away from Alzheimer's disease. *Nature* **430**, 631 (2004).
7. Iqbal, K., Liu, F., Gong, C.-X. & Grundke-Iqbal, I. Tau in Alzheimer disease and related tauopathies. *Current Alzheimer Research* **7**, 656-664 (2010).
8. Eid, A., Mhatre, I. & Richardson, J.R. Gene-environment interactions in Alzheimer's disease: A potential path to precision medicine. *Pharmacology & therapeutics* **199**, 173-187 (2019).
9. Heneka, M.T. *et al.* Neuroinflammation in Alzheimer's disease. *The Lancet Neurology* **14**, 388-405 (2015).
10. Bradburn, S., Murgatroyd, C. & Ray, N. Neuroinflammation in mild cognitive impairment and Alzheimer's disease: A meta-analysis. *Ageing research reviews* (2019).
11. Morales, I., Jiménez, J.M., Mancilla, M. & Maccioni, R.B. Tau oligomers and fibrils induce activation of microglial cells. *Journal of Alzheimer's Disease* **37**, 849-856 (2013).
12. Association, A.P. *Diagnostic criteria from dsM-iV-tr*, (American Psychiatric Pub, 2000).
13. McKhann, G.M. *et al.* The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & dementia* **7**, 263-269 (2011).
14. Dubois, B. *et al.* Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS–ADRDA criteria. *The Lancet Neurology* **6**, 734-746 (2007).
15. Morris, E. *et al.* Diagnostic accuracy of 18 F amyloid PET tracers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis. *European journal of nuclear medicine and molecular imaging* **43**, 374-385 (2016).
16. Jack, C.R., Jr. *et al.* NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimers Dement* **14**, 535-562 (2018).
17. Borenstein, A.R., Copenhaver, C.I. & Mortimer, J.A. Early-life risk factors for Alzheimer disease. *Alzheimer Dis Assoc Disord* **20**, 63-72 (2006).
18. Qiu, C., Xu, W. & Fratiglioni, L. Vascular and psychosocial factors in Alzheimer's disease: epidemiological evidence toward intervention. *J Alzheimers Dis* **20**, 689-97 (2010).
19. Barnes, D.E. & Yaffe, K. The projected effect of risk factor reduction on Alzheimer's disease prevalence. *Lancet Neurol* **10**, 819-28 (2011).

20. Attems, J. & Jellinger, K.A. The overlap between vascular disease and Alzheimer's disease-lessons from pathology. *BMC medicine* **12**, 206 (2014).
21. Scheltens, P. *et al.* Alzheimer's disease. *Lancet* **388**, 505-17 (2016).
22. Broce, I.J. *et al.* Dissecting the genetic relationship between cardiovascular risk factors and Alzheimer's disease. *Acta Neuropathologica* **137**, 209-226 (2019).
23. Sperling, R.A. *et al.* Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7**, 280-92 (2011).
24. Morris, J. Early-stage and preclinical Alzheimer disease. *Alzheimer disease and associated disorders* **19**, 163 (2005).
25. Jack Jr, C.R. *et al.* Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology* **9**, 119-128 (2010).
26. Godbolt, A.K. *et al.* A decade of pre-diagnostic assessment in a case of familial Alzheimer's disease: tracking progression from asymptomatic to MCI and dementia. *Neurocase* **11**, 56-64 (2005).
27. Jack, C.R., Jr. *et al.* Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol* **9**, 119-28 (2010).
28. Leclerc, B. & Abulrob, A. Perspectives in molecular imaging using staging biomarkers and immunotherapies in Alzheimer's disease. *ScientificWorldJournal* **2013**, 589308 (2013).
29. Lockhart, S.N. & DeCarli, C. Structural imaging measures of brain aging. *Neuropsychology review* **24**, 271-289 (2014).
30. Debette, S. *et al.* Association of MRI markers of vascular brain injury with incident stroke, mild cognitive impairment, dementia, and mortality: the Framingham Offspring Study. *Stroke* **41**, 600-606 (2010).
31. Frisoni, G.B., Fox, N.C., Jack Jr, C.R., Scheltens, P. & Thompson, P.M. The clinical use of structural MRI in Alzheimer disease. *Nature Reviews Neurology* **6**, 67 (2010).
32. Presotto, L. *et al.* Low-dose CT for the spatial normalization of PET images: A validation procedure for amyloid-PET semi-quantification. *NeuroImage: Clinical* **20**, 153-160 (2018).
33. Jang, Y.K. *et al.* Head to head comparison of [(18)F] AV-1451 and [(18)F] THK5351 for tau imaging in Alzheimer's disease and frontotemporal dementia. *Eur J Nucl Med Mol Imaging* **45**, 432-442 (2018).
34. Mishra, S. *et al.* AV-1451 PET imaging of tau pathology in preclinical Alzheimer disease: Defining a summary measure. *Neuroimage* **161**, 171-178 (2017).
35. LaPoint, M.R. *et al.* The association between tau PET and retrospective cortical thinning in clinically normal elderly. *Neuroimage* **157**, 612-622 (2017).
36. Leuzy, A. *et al.* Tau PET imaging in neurodegenerative tauopathies—still a challenge. *Molecular psychiatry* **24**, 1112-1134 (2019).
37. Saint-Aubert, L. *et al.* Tau PET imaging: present and future directions. *Molecular neurodegeneration* **12**, 19 (2017).
38. Nakamura, A. *et al.* High performance plasma amyloid-β biomarkers for Alzheimer's disease. *Nature* **554**, 249 (2018).

39. Ward, A., Tardiff, S., Dye, C. & Arrighi, H.M. Rate of conversion from prodromal Alzheimer's disease to Alzheimer's dementia: a systematic review of the literature. *Dementia and geriatric cognitive disorders extra* **3**, 320-332 (2013).
40. Taragano, F.E., Allegri, R.F. & Lyketsos, C. Mild behavioral impairment: A prodromal stage of dementia. *Dement Neuropsychol* **2**, 256-260 (2008).
41. Wilson, R.S., Leurgans, S.E., Boyle, P.A. & Bennett, D.A. Cognitive decline in prodromal Alzheimer disease and mild cognitive impairment. *Arch Neurol* **68**, 351-6 (2011).
42. Jansen, W.J. *et al.* Prevalence of cerebral amyloid pathology in persons without dementia: a meta-analysis. *JAMA* **313**, 1924-38 (2015).
43. Mattsson, N. *et al.* CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment. *JAMA* **302**, 385-93 (2009).
44. van Maurik, I.S. *et al.* Biomarker-based prognosis for people with mild cognitive impairment (ABIDE): a modelling study. *The Lancet Neurology* **18**, 1034-1044 (2019).
45. Bersanelli, M. *et al.* Methods for the integration of multi-omics data: mathematical aspects. *BMC bioinformatics* **17**, S15 (2016).
46. Hasin, Y., Seldin, M. & Lusis, A. Multi-omics approaches to disease. *Genome biology* **18**, 83 (2017).
47. Siroux, V., Agier, L. & Slama, R. The exposome concept: a challenge and a potential driver for environmental health research. *Eur Respir Rev* **25**, 124-9 (2016).
48. Rossor, M.N., Fox, N.C., Mummery, C.J., Schott, J.M. & Warren, J.D. The diagnosis of young-onset dementia. *The Lancet Neurology* **9**, 793-806 (2010).
49. Sherrington, R. *et al.* Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature* **375**, 754 (1995).
50. Levy-Lahad, E. *et al.* Candidate gene for the chromosome 1 familial Alzheimer's disease locus. *Science* **269**, 973-977 (1995).
51. Goate, A. *et al.* Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature* **349**, 704 (1991).
52. Jarmolowicz, A.I., Chen, H.Y. & Panegyres, P.K. The patterns of inheritance in early-onset dementia: Alzheimer's disease and frontotemporal dementia. *Am J Alzheimers Dis Other Demen* **30**, 299-306 (2015).
53. Alzheimer's, A. 2015 Alzheimer's disease facts and figures. *Alzheimer's & dementia: the journal of the Alzheimer's Association* **11**, 332 (2015).
54. 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).
55. Corder, E.H. *et al.* Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. *Nat Genet* **7**, 180-4 (1994).
56. Seshadri, S., Drachman, D.A. & Lippa, C.F. Apolipoprotein E epsilon 4 allele and the lifetime risk of Alzheimer's disease. What physicians know, and what they should know. *Arch Neurol* **52**, 1074-9 (1995).
57. Farrer, L.A. *et al.* Effects of Age, Sex, and Ethnicity on the Association Between Apolipoprotein E Genotype and Alzheimer Disease: A Meta-analysis. *Jama* **278**, 1349-1356 (1997).

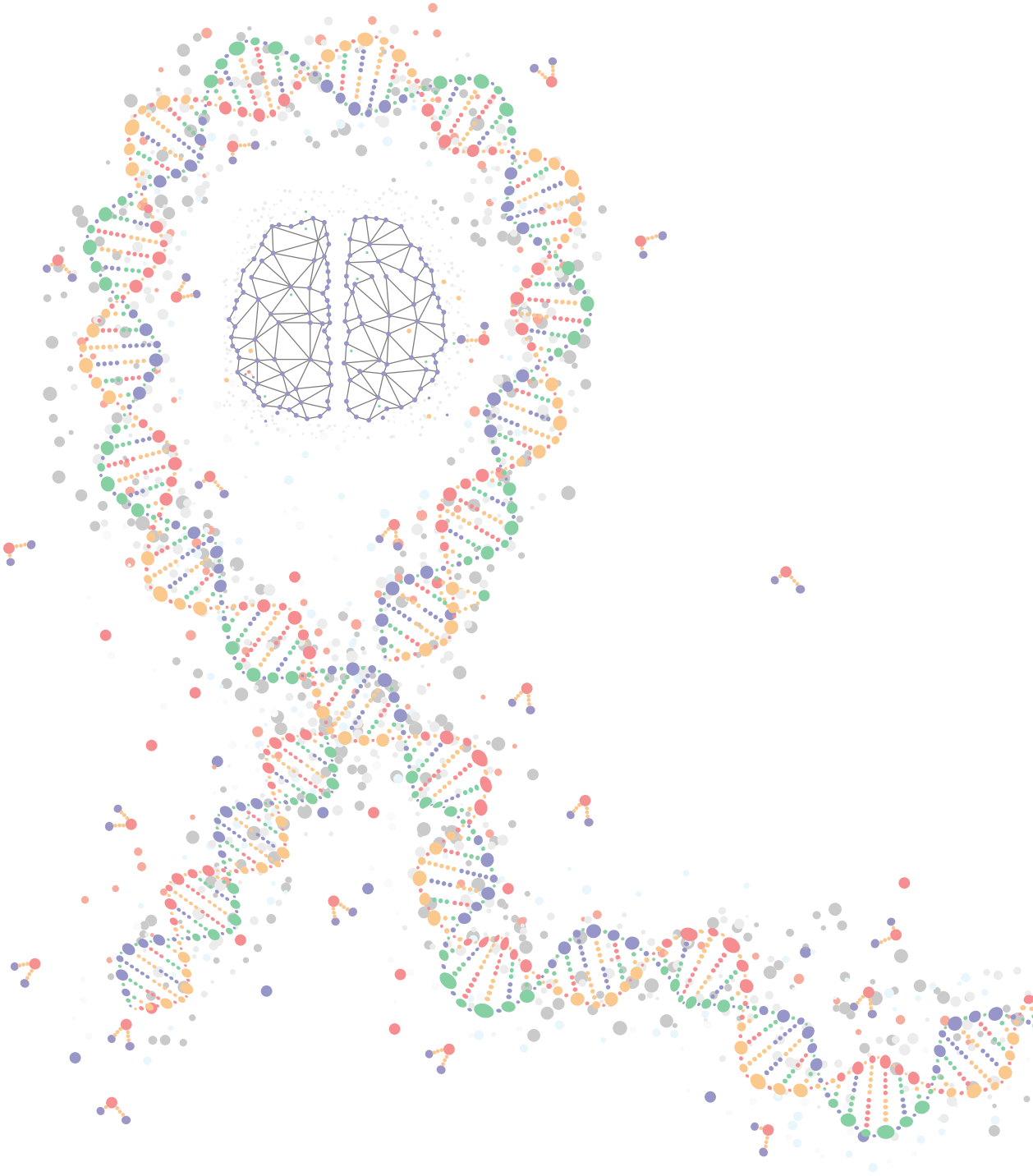
58. Meyer, M.R. *et al.* APOE genotype predicts when—not whether—one is predisposed to develop Alzheimer disease. *Nature genetics* **19**, 321 (1998).
59. Verghese, P.B., Castellano, J.M. & Holtzman, D.M. Apolipoprotein E in Alzheimer's disease and other neurological disorders. *The Lancet Neurology* **10**, 241-252 (2011).
60. van der Lee, S.J. *et al.* The effect of APOE and other common genetic variants on the onset of Alzheimer's disease and dementia: a community-based cohort study. *The Lancet Neurology* **17**, 434-444 (2018).
61. Leonenko, G. *et al.* Genetic risk for Alzheimer's disease is distinct from genetic risk for amyloid deposition. *Annals of neurology* (2019).
62. 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).
63. 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).
64. Genin, E. *et al.* APOE and Alzheimer disease: a major gene with semi-dominant inheritance. *Mol Psychiatry* **16**, 903-7 (2011).
65. Seshadri, S. *et al.* Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA* **303**, 1832-40 (2010).
66. 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).
67. 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).
68. Kunkle, B.W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet* **51**, 414-430 (2019).
69. Jansen, I.E. *et al.* Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. (2019).
70. Kamboh, M.I. A brief synopsis on the genetics of Alzheimer's disease. *Current genetic medicine reports* **6**, 133-135 (2018).
71. Karch, C.M. & Goate, A.M. Alzheimer's disease risk genes and mechanisms of disease pathogenesis. *Biol Psychiatry* **77**, 43-51 (2015).
72. International Genomics of Alzheimer's Disease, C. Convergent genetic and expression data implicate immunity in Alzheimer's disease. *Alzheimers Dement* **11**, 658-71 (2015).
73. Van der Flier, W.M. Clinical heterogeneity in familial Alzheimer's disease. *Lancet Neurol* **15**, 1296-1298 (2016).
74. Becker, J.T., Huff, F.J., Nebes, R.D., Holland, A. & Boller, F. Neuropsychological function in Alzheimer's disease. Pattern of impairment and rates of progression. *Arch Neurol* **45**, 263-8 (1988).
75. Mann, U.M., Mohr, E., Gearing, M. & Chase, T.N. Heterogeneity in Alzheimer's disease: progression rate segregated by distinct neuropsychological and cerebral metabolic profiles. *J Neurol Neurosurg Psychiatry* **55**, 956-9 (1992).
76. Petersen, R.C. *et al.* Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol* **56**, 303-8 (1999).

77. Gauthier, S. *et al.* Mild cognitive impairment. *Lancet* **367**, 1262-70 (2006).
78. Petersen, R.C. Mild cognitive impairment as a diagnostic entity. *Journal of internal medicine* **256**, 183-194 (2004).
79. Ridge, P.G. *et al.* Assessment of the genetic variance of late-onset Alzheimer's disease. *Neurobiology of aging* **41**, 200. e13-200. e20 (2016).
80. Grozeva, D., Saad, S., Menzies, G.E. & Sims, R. Benefits and Challenges of Rare Genetic Variation in Alzheimer's Disease. *Current Genetic Medicine Reports* **7**, 53-62 (2019).
81. Ott, J., Wang, J. & Leal, S.M. Genetic linkage analysis in the age of whole-genome sequencing. *Nat Rev Genet* **16**, 275-84 (2015).
82. Guerreiro, R. *et al.* TREM2 variants in Alzheimer's disease. *N Engl J Med* **368**, 117-27 (2013).
83. Jonsson, T. *et al.* Variant of TREM2 associated with the risk of Alzheimer's disease. *N Engl J Med* **368**, 107-16 (2013).
84. Cruchaga, C. *et al.* Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. *Nature* **505**, 550-4 (2014).
85. Sims, R. *et al.* Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. *Nature genetics* **49**, 1373 (2017).
86. Kunkle, B.W. *et al.* Genome-wide linkage analyses of non-Hispanic white families identify novel loci for familial late-onset Alzheimer's disease. *Alzheimers Dement* (2015).
87. Bis, J.C. *et al.* Whole exome sequencing study identifies novel rare and common Alzheimer's-associated variants involved in immune response and transcriptional regulation. *Molecular psychiatry*, 1 (2018).
88. Pulst, S.M. Genetic linkage analysis. *Archives of neurology* **56**, 667-672 (1999).
89. 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).
90. Jun, G. *et al.* Evaluating the contribution of rare variants to type 2 diabetes and related traits using pedigrees. *Proc Natl Acad Sci U S A* **115**, 379-384 (2018).
91. Lovestone, S. *et al.* Proteomics of Alzheimer's disease: understanding mechanisms and seeking biomarkers. *Expert review of proteomics* **4**, 227-238 (2007).
92. Braun, P. & LaBaer, J. High throughput protein production for functional proteomics. *Trends in biotechnology* **21**, 383-388 (2003).
93. Henstridge, C.M., Hyman, B.T. & Spires-Jones, T.L. Beyond the neuron–cellular interactions early in Alzheimer disease pathogenesis. *Nature Reviews Neuroscience*, 1 (2019).
94. Colonna, M. TREMs in the immune system and beyond. *Nat Rev Immunol* **3**, 445-53 (2003).
95. Bell, R.D. *et al.* Apolipoprotein E controls cerebrovascular integrity via cyclophilin A. *Nature* **485**, 512-6 (2012).
96. Zipser, B.D. *et al.* Microvascular injury and blood–brain barrier leakage in Alzheimer's disease. *Neurobiology of aging* **28**, 977-986 (2007).
97. Ray, S. *et al.* Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins. *Nature medicine* **13**, 1359 (2007).

98. Soares, H.D. *et al.* Plasma biomarkers associated with the apolipoprotein E genotype and Alzheimer disease. *Archives of neurology* **69**, 1310-1317 (2012).
99. Ringman, J.M. *et al.* Plasma signaling proteins in persons at genetic risk for Alzheimer disease: influence of APOE genotype. *Archives of neurology* **69**, 757-764 (2012).
100. Hu, W.T. *et al.* Plasma multianalyte profiling in mild cognitive impairment and Alzheimer disease. *Neurology* **79**, 897-905 (2012).
101. Llano, D.A., Devanarayan, V., Simon, A.J. & Alzheimer's Disease Neuroimaging, I. Evaluation of plasma proteomic data for Alzheimer disease state classification and for the prediction of progression from mild cognitive impairment to Alzheimer disease. *Alzheimer Disease & Associated Disorders* **27**, 233-243 (2013).
102. Hye, A. *et al.* Proteome-based plasma biomarkers for Alzheimer's disease. *Brain* **129**, 3042-50 (2006).
103. Swomley, A.M. *et al.* Abeta, oxidative stress in Alzheimer disease: evidence based on proteomics studies. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* **1842**, 1248-1257 (2014).
104. Schonberger, S.J., Edgar, P.F., Kydd, R., Faull, R.L. & Cooper, G.J. Proteomic analysis of the brain in Alzheimer's disease: molecular phenotype of a complex disease process. *PROTEOMICS: International Edition* **1**, 1519-1528 (2001).
105. Johnson, C.H., Ivanisevic, J. & Siuzdak, G. Metabolomics: beyond biomarkers and towards mechanisms. *Nature reviews Molecular cell biology* **17**, 451-459 (2016).
106. Menni, C., Zierer, J., Valdes, A.M. & Spector, T.D. Mixing omics: combining genetics and metabolomics to study rheumatic diseases. *Nature Reviews Rheumatology* **13**, 174 (2017).
107. Fiandaca, M.S. *et al.* Plasma 24-metabolite panel predicts preclinical transition to clinical stages of Alzheimer's disease. *Frontiers in neurology* **6**, 237 (2015).
108. Varma, V.R. *et al.* Brain and blood metabolite signatures of pathology and progression in Alzheimer disease: A targeted metabolomics study. *PLoS medicine* **15**, e1002482 (2018).
109. Koal, T., Klavins, K., Seppi, D., Kemmler, G. & Humpel, C. Sphingomyelin SM (d18: 1/18: 0) is Significantly Enhanced in Cerebrospinal Fluid Samples Dichotomized by Pathological Amyloid-β 42, Tau, and Phospho-Tau-181 Levels. *Journal of Alzheimer's Disease* **44**, 1193-1201 (2015).
110. Snowden, S.G. *et al.* Association between fatty acid metabolism in the brain and Alzheimer disease neuropathology and cognitive performance: A nontargeted metabolomic study. *PLoS medicine* **14**, e1002266 (2017).
111. Toledo, J.B. *et al.* Metabolic network failures in Alzheimer's disease: A biochemical road map. *Alzheimer's & Dementia* **13**, 965-984 (2017).
112. Paglia, G. *et al.* Unbiased metabolomic investigation of Alzheimer's disease brain points to dysregulation of mitochondrial aspartate metabolism. *Journal of proteome research* **15**, 608-618 (2016).
113. Graham, S.F. *et al.* Untargeted metabolomic analysis of human plasma indicates differentially affected polyamine and L-arginine metabolism in mild cognitive impairment subjects converting to Alzheimer's disease. *PLoS One* **10**, e0119452 (2015).

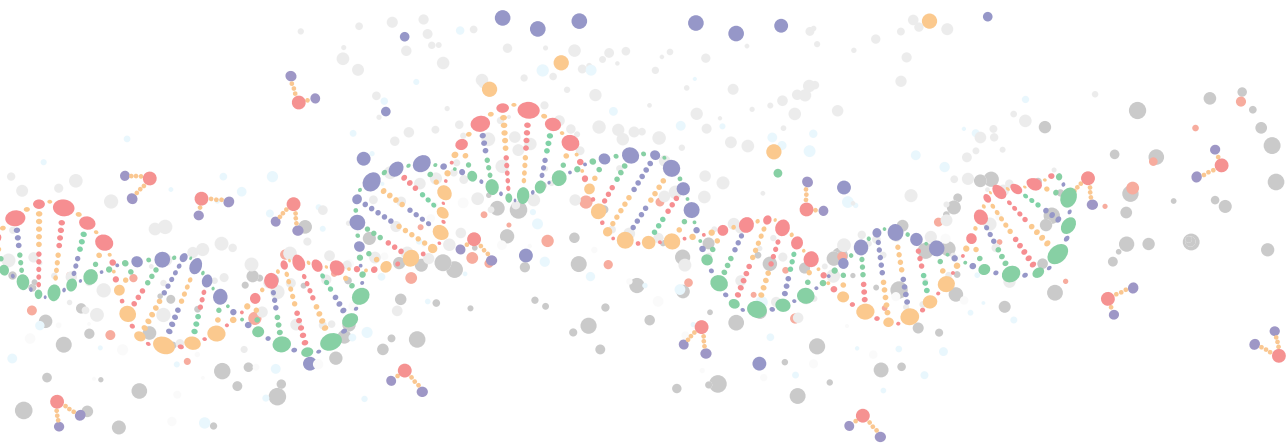
114. Wilkins, J.M. & Trushina, E. Application of metabolomics in Alzheimer's disease. *Frontiers in neurology* **8**, 719 (2018).
115. Jiang, Y. *et al.* Metabolomics in the development and progression of dementia: A systematic review. *Frontiers in Neuroscience* **13**, 343 (2019).
116. van der Lee, S.J. *et al.* Circulating metabolites and general cognitive ability and dementia: Evidence from 11 cohort studies. *Alzheimer's & Dementia* **14**, 707-722 (2018).
117. Tynkkynen, J. *et al.* Association of branched-chain amino acids and other circulating metabolites with risk of incident dementia and Alzheimer's disease: A prospective study in eight cohorts. *Alzheimer's & Dementia* **14**, 723-733 (2018).
118. Willette, A. Does metabolic syndrome impact cognition and emotion in A lzheimer's disease? *European journal of neurology* **23**, 237-238 (2016).
119. Craft, S. The role of metabolic disorders in Alzheimer disease and vascular dementia: two roads converged. *Archives of neurology* **66**, 300-305 (2009).
120. Wang, X.-F. *et al.* Linking Alzheimer's disease and type 2 diabetes: Novel shared susceptibility genes detected by cFDR approach. *Journal of the neurological sciences* **380**, 262-272 (2017).
121. Rios, J.A., Cisternas, P., Arrese, M., Barja, S. & Inestrosa, N.C. Is Alzheimer's disease related to metabolic syndrome? A Wnt signaling conundrum. *Progress in neurobiology* **121**, 125-146 (2014).
122. Demetrius, L.A. & Driver, J. Alzheimer's as a metabolic disease. *Biogerontology* **14**, 641-649 (2013).
123. Kapogiannis, D. & Mattson, M.P. Disrupted energy metabolism and neuronal circuit dysfunction in cognitive impairment and Alzheimer's disease. *The Lancet Neurology* **10**, 187-198 (2011).
124. Clarke, J.R., Ribeiro, F.C., Frozza, R.L., De Felice, F.G. & Lourenco, M.V. Metabolic dysfunction in Alzheimer's disease: from basic neurobiology to clinical approaches. *Journal of Alzheimer's Disease* **64**, S405-S426 (2018).
125. Harach, T. *et al.* Reduction of Abeta amyloid pathology in APPPS1 transgenic mice in the absence of gut microbiota. *Scientific reports* **7**, 41802 (2017).
126. Harach, T. *et al.* Reduction of Alzheimer's disease beta-amyloid pathology in the absence of gut microbiota. *arXiv preprint arXiv:1509.02273* (2015).
127. Bäuerl, C., Collado, M.C., Diaz Cuevas, A., Viña, J. & Pérez Martínez, G. Shifts in gut microbiota composition in an APP/PSS 1 transgenic mouse model of Alzheimer's disease during lifespan. *Letters in applied microbiology* **66**, 464-471 (2018).
128. Jin, R. *et al.* Amino acid metabolism is altered in adolescents with nonalcoholic fatty liver disease—An untargeted, high resolution metabolomics study. *The Journal of pediatrics* **172**, 14-19. e5 (2016).
129. Kalhan, S.C. *et al.* Plasma metabolomic profile in nonalcoholic fatty liver disease. *Metabolism* **60**, 404-413 (2011).
130. Kakiyama, G. *et al.* Modulation of the fecal bile acid profile by gut microbiota in cirrhosis. *Journal of hepatology* **58**, 949-955 (2013).
131. Frimmer, M. & Ziegler, K. The transport of bile acids in liver cells. *Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes* **947**, 75-99 (1988).
132. Bertolotti, M. *et al.* Nonalcoholic fatty liver disease and aging: epidemiology to management. *World journal of gastroenterology: WJG* **20**, 14185 (2014).

133. Yilmaz, Y. & Ozdogan, O. Liver disease as a risk factor for cognitive decline and dementia: An Under-recognized issue. *Hepatology* **49**, 698-698 (2009).
134. Weinstein, G. *et al.* Association of nonalcoholic fatty liver disease with lower brain volume in healthy middle-aged adults in the Framingham study. *JAMA neurology* **75**, 97-104 (2018).
135. Sochocka, M. *et al.* The gut microbiome alterations and inflammation-driven pathogenesis of Alzheimer's disease—a critical review. *Molecular neurobiology* **56**, 1841-1851 (2019).
136. Le Chatelier, E. *et al.* Richness of human gut microbiome correlates with metabolic markers. *Nature* **500**, 541 (2013).
137. Foster, J.A. & McVey Neufeld, K.A. Gut-brain axis: how the microbiome influences anxiety and depression. *Trends Neurosci* **36**, 305-12 (2013).
138. Eid, A., Mhatre, I. & Richardson, J.R. Gene-environment interactions in Alzheimer's disease: A potential path to precision medicine. *Pharmacol Ther* **199**, 173-187 (2019).
139. Nicholson, J.K. *et al.* Host-gut microbiota metabolic interactions. *Science* **336**, 1262-1267 (2012).
140. Shea, H.C., Head, D.D., Setchell, K.D. & Russell, D.W. Analysis of HSD3B7 knockout mice reveals that a 3α-hydroxyl stereochemistry is required for bile acid function. *Proceedings of the National Academy of Sciences* **104**, 11526-11533 (2007).
141. Ackerman, H.D. & Gerhard, G.S. Bile Acids in Neurodegenerative Disorders. *Front Aging Neurosci* **8**, 263 (2016).
142. Chiang, J.Y. Bile acid metabolism and signaling. *Compr Physiol* **3**, 1191-212 (2013).
143. Correa-Oliveira, R., Fachi, J.L., Vieira, A., Sato, F.T. & Vinolo, M.A. Regulation of immune cell function by short-chain fatty acids. *Clin Transl Immunology* **5**, e73 (2016).
144. Stanley, C.A. New genetic defects in mitochondrial fatty acid oxidation and carnitine deficiency. *Adv Pediatr* **34**, 59-88 (1987).
145. McQuade, A. & Blurton-Jones, M. Microglia in Alzheimer's Disease: Exploring How Genetics and Phenotype Influence Risk. *J Mol Biol* **431**, 1805-1817 (2019).
146. Katsumoto, A., Takeuchi, H., Takahashi, K. & Tanaka, F. Microglia in Alzheimer's Disease: Risk Factors and Inflammation. *Front Neurol* **9**, 978 (2018).
147. Abdel-Haq, R., Schlachetzki, J.C.M., Glass, C.K. & Mazmanian, S.K. Microbiome-microglia connections via the gut-brain axis. *J Exp Med* **216**, 41-59 (2019).
148. Tripathi, A. *et al.* The gut-liver axis and the intersection with the microbiome. *Nat Rev Gastroenterol Hepatol* **15**, 397-411 (2018).



Chapter 2

Genetics of Alzheimer's disease



Chapter 2.1

Variation in cadherin genes in the 5p14.3 genomic region is implicated in Alzheimer's disease in an extended ADSP pedigree

Shahzad Ahmad, Rafael A. Nafikov, Elizabeth Blue, Patrick A. Navas, Hata Karamujić-Čomić, Dina Vojinovic, Sven Van der Lee, Ashley van der Spek, Bowen Wang, Andrea R. Horimoto, Alejandro Q. Nato Jr, Harkirat K. Sohi, Mohamad Saad, Michael O. Dorschner, Debby Tsuang, Li-San Wang, CHARGE, ADSP Data Flow Work Group, ADSP Family Analysis Work Group, Lindsay A. Farrer, Jonathan L. Haines, Margaret Pericak-Vance, Gerard Schellenberg, Eric Boerwinkle, Richard P. Mayeux, M. Arfan Ikram, Anita DeStefano, Sudha Seshadri, Najaf Amin, Ellen M. Wijsman*, Cornelia M. Van Duijn* from the Alzheimer's disease sequencing project (ADSP)

**Senior authors contributed equally to the work*

This chapter is submitted for publication

Supplementary information for this chapter is available at:
https://drive.google.com/open?id=12MnrKNOB_1ahWOpQ29usR9XIkUox-nSe

Abstract

Alzheimer's disease (AD) is a complex neurodegenerative disorder characterized by progressive loss of memory, cognition and executive function. As part of the AD Sequencing Project (ADSP), we sequenced whole genomes of 29 AD cases from two large multigenerational families from the Genetic Research in Isolated Populations program (GRIP). We performed a genome-wide multipoint linkage scan to identify genomic regions harboring risk variants for AD. We obtained significant evidence of linkage of AD to chromosome 5p14.3 (logarithm of odds (LOD) score = 3.3) in one of the two extended families that had 189 members and identified a 9.14 cM risk haplotype spanning 24,578 polymorphic variants that drive the linkage signal. The region containing the risk haplotype includes a cluster of cadherin genes (*CDH*) involved in brain function. Filtering of rare genetic variants on the risk haplotype identified only non-coding variants. Further focus on DNase-I hypersensitive (DHS) and transcription factor (TF) binding for these sites together with evidence of differential expression in brain identified four rare potential regulatory variants in *CDH18* and one such variant in *CDH12*. The present study implicates *CDH* genes in AD, a gene family that is highly expressed in the nervous system and is involved in synaptic function and presenilin-mediated signaling.

Recent decades have seen major progress in the discovery of rare and common variants that determine the risk and onset of sporadic Alzheimer's disease (AD)[1-3]. These common and rare genetic variants are implicated in various biological pathways[4]. Moreover, rare mutations in five genes (*PSEN1, PSEN2, APP, SORL1, and TREM2*) involved in amyloid metabolism and *presenilin signaling* are implicated in early-onset AD[5]. As part of the AD Sequencing Project (ADSP) study, we aimed to find new AD genes using extended families derived from the Genetic Research in Isolated Populations program (GRIP) [6]. GRIP is situated in a genetically isolated community in the South West of the Netherlands[6]. The population is descended from fewer than 400 founders who were present in the region during middle of the 18th century[1] with ~20,000 individuals now distributed across eight neighboring communities[6]. These communities are largely endogamous with minimal immigration, where individuals of GRIP are related via multiple generation lines of descent and consanguineous loops. Our AD study here is based primarily on two large pedigrees segregating AD across five generations and extracted from the GRIP population. The two extended pedigrees contributing to the ADSP are the 47-member ERF201 pedigree with five probable AD cases and the 189-member ERF203 pedigree with 30 probable AD cases. We obtained genomic data for 29 cases with probable AD (5 and 24 in pedigrees ERF201 and ERF203, respectively), and for 1 subject in ERF201 with unknown clinical phenotype (Table 1 and Supplementary Figure 1). The AD status of other individuals in the pedigrees was not clinically evaluated and was treated as unknown for analysis purposes.

A genome-wide parametric multipoint linkage scan on the full pedigrees was used to identify the region(s) likely to carry rare risk variants. Compared to genome-wide association studies (GWAS), which require large sample sizes, parametric linkage analysis has the power needed for initial localization of rare risk variants with a small sample size of related individuals[7]. We identified one strong positive linkage signal with a maximum logarithm of odds (LOD) = 3.3 in ERF203 on chromosome (chr) 5p14.3 (Figures 1 and 2). An approximate 95% confidence linkage region of interest (ROI) was defined as[8] LOD ≥ 2.3 that spans the position of the maximum LOD score. There were 49,740 variants in the whole-genome sequence (WGS) data that were polymorphic in the two GRIP pedigrees within our 24.1 cM ROI.

Table 1: Descriptive of Alzheimer's disease cases with genomic data in overall and haplotype carrier versus non-carrier sample

			Age of onset (years)			*APOE* genotype		
AD cases	**N**	**Female (%)**	**Mean**	**SD**	**Range (years)**	***ε2/ε3***	***ε3/ε3***	***ε3/ε4***
Overall sample	29	26(90%)	72.6	7.79	47-85	1	6	22
Haplotype carriers	10	9(90%)	70.3	5.76	61-79	1	4	5
Haplotype non-carriers	19	17(89%)	73.6	8.60	47-85	0	2	17

Abbreviations: AD, Alzheimer's disease; SD, Standard deviation; *APOE*, Apolipoprotein E

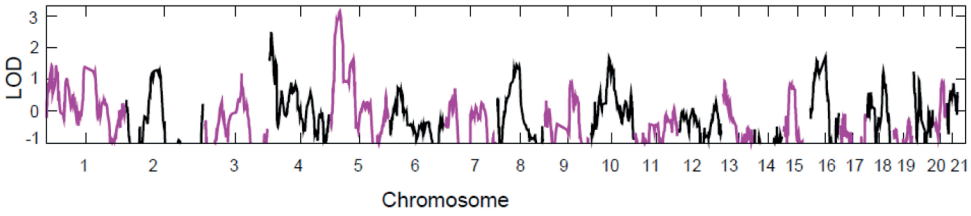


Figure 1: Genome-wide parametric linkage analysis in extended family ERF203, based on the original marker panel (thinned v1 panel).

Following a classical linkage analysis approach[6,9-11], we reduced the ROI by identifying a risk haplotype that was shared among 10 cases in ERF203 out of the 29 AD sequenced cases (details in Materials and Methods). This haplotype spanned 9.14 cM and includes 24,578 sequence variants out of the original 49,740. Interestingly, the 10 AD cases who carry the risk haplotype had somewhat earlier age-at-onset than did the remaining 19 AD cases in both ERF201 and ERF203 (Figure 3), despite a lower frequency of apolipoprotein E (*APOE)* ε*4* allele carriers (50%) in the former than in the latter (89%) group (Table 1). *APOE* ε*4*, being the strongest risk factor for late-onset AD (LOAD), significantly reduces age-at-onset of AD in *APOE* ε*4* carriers compared with none-carriers[12]. This finding suggests that the identified risk haplotype may harbor an age-at-onset modifying locus, as is the case of other well-established LOAD genes such as *APOE* and *TREM2*[13,14]. No additional clinical characteristics of AD showed obvious differences between haplotype carriers and non-carriers (Supplementary Table 1). The risk haplotype region contains 6 coding genes (Supplementary Table 2) including four members of the cadherin II gene family[15]: *CDH18, CDH12, CDH10,* and *CDH9.*

To identify the underlying risk variant(s), we applied three initial filters to restrict attention to 40 genetic variants that were compatible with the linkage signal and have potential functional consequences (Figure 4 and Supplementary Table 3). Filter 1 limited consideration to the 142 variants out of 24,578 that were found only on the risk haplotype, as unique mutations at the base-pair level are both exceedingly rare[16,17] and lead to a unique haplotypic context [18,19]. All of these 142 variants were annotated[20] to non-coding DNA. Filter 2 used a low minor allele frequency (MAF) threshold to restrict attention to 109 rare variants out of these 142 variants. Filter 3 limited consideration to 55 variants out of the 142 variants that were annotated to a gene or other potentially functional element in the region that also showed evidence of preferential expression in brain[21,22]. Joint application of these three filters yielded 40 variants, all in introns of *CDH12* or *CDH18*, and with a MAF < 0.014. See Materials and Methods for further details.

We used additional published information to further prioritize these 40 variants. These bioinformatics filters were predicated on the assumption that both accessible chromatin and transcription factor (TF) binding are involved in transcriptional regulation. This could allow

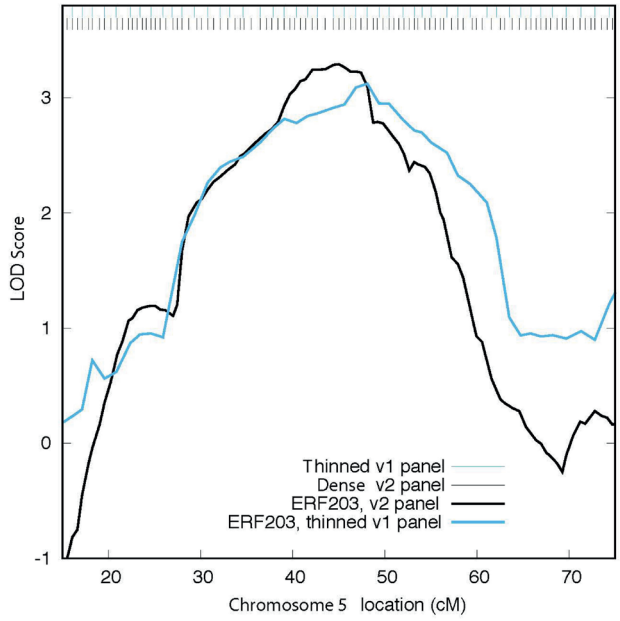


Figure 2: Parametric linkage analysis results on chromosome 5 in region with linkage evidence, with both the original genome scan marker panel (thinned v1 panel), a denser marker panel that included replacement of markers with evidence of genotyping error (dense v2 panel). Positions of the markers are indicated by tick marks on the top horizontal axis, with the color of the tics matching the color of the line(s) showing LOD score curves based on a particular set of markers.

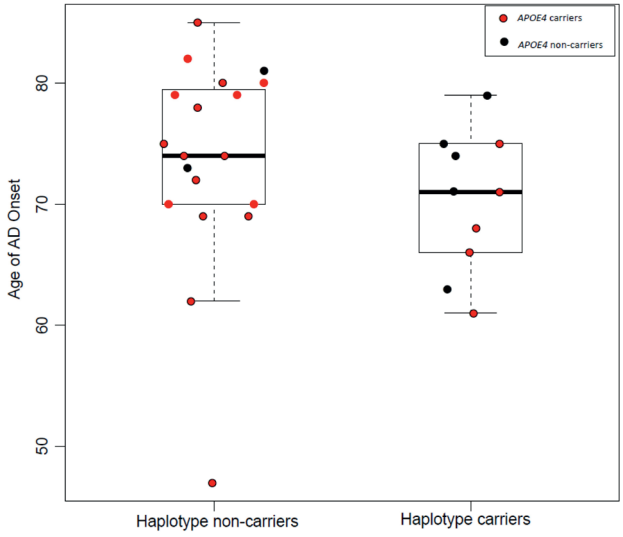


Figure 3: Age at onset of Alzheimer's disease patients in carriers versus non-carriers of identified risk haplotype and their *APOE* ε4 genotype. Red dots indicate *APOE* ε4 carriers and black dots indicate *APOE* ε3ε3/*APOE* ε2ε3 carriers.

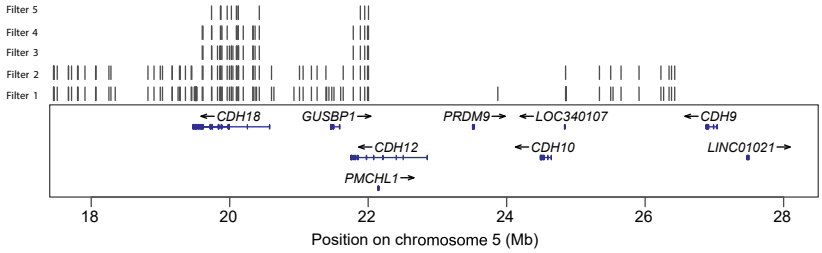


Figure 4: Genomic location of single nucleotide polymorphisms (SNP) in risk haplotype region, filtered based on minor allele frequency, gene expression data, evidence for accessible chromatin, and presence of the variant in a critical position of a predicted transcription-factor binding site. Each line in the filtering step row represents a SNP. First filtering step retained only 142 SNPs unique to risk haplotype from total ~ 24,578 variants. Second filtering step retained only 109 SNPs with minor allele frequency <0.014 in European sample of 1000 genome and filtering step three retained only 40 genetic variants with preferential brain expression[24,25]. Fourth (32) and fifth (16) filter retains only genetic variants that overlap with transcriptional factor binding sites (TFBS) and potentially disrupt the TFBS motif respectively. LocusZoom software[75] was used to create the genomic location map of filtered SNPs.

non-coding variants to influence AD risk or age-at-onset. In Filter 4, we prioritized 32 out of the 40 variants because they overlapped predicted TF binding sites (TFBSs)[23,24] or were located near potentially open chromatin. With Filter 5, we added the constraint that the variants of greatest interest are those for which the minor allele is predicted to disrupt a TFBS motif, yielding 16 variants for follow-up (Figure 4). For these 16 variants, we evaluated predicted TFBSs at the nucleotide level for quality of alignment of the TF to the TFBS, and, together with the site-frequency matrix, the importance of each nucleotide. We also again used gene expression data to eliminate those variants with no predicted TFs that were expressed in brain[22]. See Materials and Methods for more details.

Ten out of the sixteen variants were most highly-prioritized as potential regulatory variants by this nucleotide-level evaluation (Table 2 and Supplemental Table 4). Of these, a particularly compelling candidate variant in *CDH18*, rs199862039, is in a region unique to primates[25], and the position-weight matrices of the two unrelated TFs, *FOXO1* and *ZNF384*, both indicate that the variant occurs at a site that is virtually monomorphic (posterior probabilities for the Reference (Ref) allele of 1 and 0.999, respectively) among all motifs that lead to either TFBS. It is also notable that the Alternative allele for this variant, C, is the only allele at this position that has never been observed in a sample of 30,433 potential binding sites for *ZNF384*, strengthening the support for potential functional effects for this variant site. Three additional variants in introns of *CDH18* are also strong candidates. Variant rs80189338 has probability 1 for the Ref allele for the *CRX* (*OTX3*) TF, and 0.93 for the Ref allele for the *OTX1* TF, with further support from proximity to open chromatin. However, the sample size used to define the motifs is more moderate than for rs199862039. Two variants in introns of *CDH18* implicate a single family of TFs, also with moderate sample size used to generate the motifs: rs569459988 and

rs757428201. One of the predicted TFs in this family of TFs, *NFAT5*, is part of a proposed gene module, Module 3, from a recent transcriptomic analysis[22]. Module 3 includes AD GWAS genes. From the position weight matrix[24], the marginal probabilities of the Ref allele are between 0.92 and 0.97 for the different related TFs at these two variants. Finally, the most compelling site in *CDH12* is rs185858581, with posterior probabilities for the Ref allele of 0.99 and 1 for TFs *SOX15* and *SOX6*, respectively. The remaining five variants in Table 2 are less compelling primarily because the putative TFBS has one or more discrepancies relative to the reference sequence.

The fact that 5p14.3 has not been prioritized by earlier GWAS of AD[2,3,26,27] suggests that the variant driving AD risk or age-at-onset is rare in populations used in GWAS to date (largely European ancestry). However, the genomic region of interest has previously been associated with cognitive function[28,29], and one of the highly-prioritized variants is predicted to affect bindings of a TF (*NFAT5*) that is part of a set of genes also associated with cognition. Additional evidence for this region emerged from the Cohorts for Heart and Aging Research in the Genomic Epidemiology (CHARGE) consortium GWAS of verbal short-term memory that includes 44,874 participants (28 cohorts). Jari *et al.,* 2019 [unpublished), found genome-wide significant evidence that a variant in *CDH18* was associated with *verbal* short-term memory ($P < 5x10^{-8}$) (Supplementary

Table 2: Ten most-highly prioritized genetic variants on chromosome 5 based on allele frequency filter (MAF < 0.014), gene expression data, evidence for accessible chromatin, and presence of the variant in a critical position of a predicted transcription-factor binding site

Position (bp)	Ref allele	Alt allele	rsID	EUR MAF	GMAF	Gene symbol	TFBS*	DHS site	Module†
5:19754993	G	A	rs569459988	NA	NA	*CDH18*	*NFAT5*	N	3
5:20024487	T	A	rs184729824	0.001	0.001	*CDH18*	*SREBF2*	N	1
5:20096655	T	C	rs6881427	0.004	0.108	*CDH18*	*POU2F2*	N	NA
5:20121532	T	C	rs199862039	0.004	0.1	*CDH18*	*FOXO1*	N	1
5:20124597	G	T	rs80189338	0.004	0.1	*CDH18*	*CRX*	Y	NA
5:20254285	G	A	rs538554203	NA	0.004	*CDH18*	*TBX15*	N	NA
5:20305291	C	A	rs757428201	NA	0.0002	*CDH18*	*NFATC3*	N	NA
5:21887060	A	G	rs185858581	0.008	0.002	CDH12	SOX6	N	NA
5:21949961	A	T	rs181135067	0.0139	0.004	CDH12	NR2E1	N	NA
5:22004495	A	G	rs143782484	NA	0.006	CDH12	RXRB	N	NA

Abbreviations: bp, base pair; EUR MAF, European minor allele frequency; GMAF, Global minor allele frequency; TFBS, transcription factor binding site

Note: Position is on build GRCh37/hg19; alleles are called to the forward strand

* Only one Transcription factor binding site for a variant is listed in this table; other predicted TFBSs, when available, are listed in Supplementary Tables 4 and 5

†Genes for which expression is highly correlated with other genes used to define cell type are indicated with the module number from Mathys et al[74]

Figure 7). Loss of short-term memory is also considered one of the early clinical symptoms of AD[30] and verbal short-term memory is also associated with the risk of AD in the Rotterdam study ($\beta = 0.0019$, SE = 0.0002, $P < 2x10^{-16}$ in Rotterdam study) and other studies[31,32].

Discussion

We found significant evidence of linkage of AD to 5p14.3 in an extended GRIP family. Filtering of genetic variants in the shared risk haplotype region based on MAF and overlap with DHS sites and transcription factor binding sites identified 4 potentially regulatory variants affecting brain *CDH18* expression. These are very rare in European populations and would have been missed in prior GWAS of AD. Within the CHARGE consortium, Jari *et al.,* 2019 [unpublished] also found the association of intronic variant in *CDH18* with a measure of verbal short-term memory, strengthening the potential role of regulatory variation on *CDH18* and its role in Alzheimer's disease and memory.

Our current findings suggest that disrupted or altered regulation of *CDH18* is the most likely explanation for the AD phenotype in the carriers of the risk haplotype in ERF203. Expression of *CDH18* is essentially brain specific[20], and key variants in our analyses both potentially alter TF binding and are preferentially expressed in neurons. *CDH18* encodes a type II classical cadherin from the cadherin superfamily, which plays a crucial role in calcium-dependent cell-cell adhesion[33], synaptogenesis and plasticity in central neurons[34,35]. It is of note that *CDH18* is entirely contained within a single Transcriptionally Active Domain in brain tissues compared to other tissues where its expression can be measured[36], leaving open the possibility that implicated variants in *CDH18* are involved in altered gene expression through coordinated transcription. Although evidence from our study supports the linkage of *CDH18* with AD in our sample, there are also three other cadherin's (*CDH9, CDH10, and CDH12*) covered by the shared haplotype, all of which, like *CDH18*, show brain specificity in expression, with little activity in other tissues[20]. *CDH12*, in particular, also contains one rare variant that has characteristics as a potential variant, and cannot be excluded as a candidate. It is also possible that there is coordinated regulation of more than one of these genes, where multiple variants in the region contribute jointly to altered gene regulation and therefore confers increased risk of AD.

All the cadherin genes in our identified linkage region 5p14.3 belong to the cadherin type II of the cadherin superfamily[37]. The type II cadherins in our identified linkage region interact with each other as indicated in the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) network[38] (Supplementary Figure 8). Although the cadherin superfamily is a highly diverse set of transmembrane proteins, each of their sub-families contain highly conserved extracellular cytoplasmic domains that define the nature of their molecular interactions[39]. Studies have shown that their overlapping and redundant nature of expression and temporal distribution

in brain regions may contribute to the role of type II cadherins in the formation of neuronal circuits[40]. The combinatorial nature of regulation and thus expression of the cadherins may define the specificity of neuronal circuits which are responsible for various neuronal functions such as synaptogenesis and synaptic plasticity[41,42]. This may explain why some cadherins have been associated with schizophrenia and autism spectrum disorder[43,44], while others are associated with cognitive function (*CDH9*)[28,29] or verbal short term memory (*CDH18*).

Cognitive reserve is one of the most important protective factors in the epidemiology of AD and dementia[45,46]. Before now, the high genetic correlation (R=-0.27, $P = 2.03x10^{-5}$) between cognitive reserve and AD reported in a recent genetic study of intelligence[28] appeared to be at odds with the fact that there is little to no overlap between the top genes that drive cognitive reserve and AD. For instance, *APOE ε4* (rs429358), the major risk allele for AD in the general population, is only modestly associated with cognitive reserve[47]. Our study highlights the possibility that genes in the 5p14.3 region may explain the part of the genetic, clinical and epidemiologic correlation between AD and cognition.

In summary, in this family-based study, we have found that AD was linked to a region (5p14.3) on chromosome 5 enriched for cadherin genes involved in neuronal development and synaptic function. Moreover, we identified a small number of genetic variants of *CDH18* that overlap with DHS sites in AD-related brain regions. A limitation of our study is that we cannot link the region to AD in the general population, but there is convincing, independent, evidence for role of the region in cognitive reserve and short term memory.

Acknowledgments

The Alzheimer's Disease Sequencing Project (ADSP) is comprised of two Alzheimer's Disease (AD) genetics consortia and three National Human Genome Research Institute (NHGRI) funded Large Scale Sequencing and Analysis Centers (LSAC). The two AD genetics consortia are the Alzheimer's Disease Genetics Consortium (ADGC) funded by NIA (U01 AG032984), and the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) funded by NIA (R01 AG033193), the National Heart, Lung, and Blood Institute (NHLBI), other National Institute of Health (NIH) institutes and other foreign governmental and non-governmental organizations. The Discovery Phase analysis of sequence data is supported through UF1AG047133 (to Drs. Schellenberg, Farrer, Pericak-Vance, Mayeux, and Haines); U01AG049505 to Dr. Seshadri; U01AG049506 to Dr. Boerwinkle; U01AG049507 to Dr. Wijsman; and U01AG049508 to Dr. Goate and the Discovery Extension Phase analysis is supported through U01AG052411 to Dr. Goate, U01AG052410 to Dr. Pericak-Vance and U01 AG052409 to Drs. Seshadri and Fornage. Data generation and harmonization in the Follow-up Phases is supported by U54AG052427 (to Drs. Schellenberg and Wang).

The ADGC cohorts include: Adult Changes in Thought (ACT), the Alzheimer's Disease Centers (ADC), the Chicago Health and Aging Project (CHAP), the Memory and Aging Project (MAP), Mayo Clinic (MAYO), Mayo Parkinson's Disease controls, University of Miami, the Multi-Institutional Research in Alzheimer's Genetic Epidemiology Study (MIRAGE), the National Cell Repository for Alzheimer's Disease (NCRAD), the National Institute on Aging Late Onset Alzheimer's Disease Family Study (NIA-LOAD), the Religious Orders Study (ROS), the Texas Alzheimer's Research and Care Consortium (TARC), Vanderbilt University/Case Western Reserve University (VAN/CWRU), the Washington Heights-Inwood Columbia Aging Project (WHICAP) and the Washington University Sequencing Project (WUSP), the Columbia University Hispanic- Estudio Familiar de Influencia Genetica de Alzheimer (EFIGA), the University of Toronto (UT), and Genetic Differences (GD).

The CHARGE cohorts are supported in part by National Heart, Lung, and Blood Institute (NHLBI) infrastructure grant HL105756 (Psaty), RC2HL102419 (Boerwinkle) and the neurology working group is supported by the National Institute on Aging (NIA) R01 grant AG033193. The CHARGE cohorts participating in the ADSP include the following: Austrian Stroke Prevention Study (ASPS), ASPS-Family study, and the Prospective Dementia Registry-Austria (ASPS/PRODEM-Aus), the Atherosclerosis Risk in Communities (ARIC) Study, the Cardiovascular Health Study (CHS), the Erasmus Rucphen Family Study (ERF), the Framingham Heart Study (FHS), and the Rotterdam Study (RS). ASPS is funded by the Austrian Science Fond (FWF) grant number P20545-P05 and P13180 and the Medical University of Graz. The ASPS-Fam is funded by the Austrian Science Fund (FWF) project I904),the EU Joint Programme–Neurodegenerative Disease Research (JPND) in frame of the BRIDGET project (Austria, Ministry of Science) and the Medical University of Graz and the Steiermärkische Krankenanstalten Gesellschaft. PRODEM-Austria is supported by the Austrian Research Promotion agency (FFG) (Project No. 827462) and by the Austrian National Bank (Anniversary Fund, project 15435. ARIC research is carried out as a collaborative study supported by NHLBI contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C). Neurocognitive data in ARIC is collected by U01 2U01HL096812, 2U01HL096814, 2U01HL096899, 2U01HL096902, 2U01HL096917 from the NIH (NHLBI, NINDS, NIA and NIDCD), and with previous brain MRI examinations funded by R01-HL70825 from the NHLBI. CHS research was supported by contracts HHSN268201200036C, HHSN268200800007C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, and grants U01HL080295 and U01HL130114 from the NHLBI with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided by R01AG023629, R01AG15928, and R01AG20098 from the NIA. FHS research is supported by NHLBI contracts N01-HC-25195 and HHSN268201500001I. This study was also supported by additional grants from the NIA (R01s AG054076, AG049607

and AG033040 and NINDS (R01 NS017950). 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 the Netherlands Organization for Scientific Research and the Russian Foundation for Basic Research (NWO-RFBR 047.017.043). The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, the Netherlands Organization for Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the municipality of Rotterdam. Genetic data sets are also supported by the Netherlands Organization of Scientific Research NWO Investments (175.010.2005.011, 911-03-012), the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), and the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) Netherlands Consortium for Healthy Aging (NCHA), project 050-060-810. All studies are grateful to their participants, faculty and staff. The content of these manuscripts is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the U.S. Department of Health and Human Services.

The four LSACs are: the Human Genome Sequencing Center at the Baylor College of Medicine (U54 HG003273), the Broad Institute Genome Center (U54HG003067), The American Genome Center at the Uniformed Services University of the Health Sciences (U01AG057659), and the Washington University Genome Institute (U54HG003079).

Biological samples and associated phenotypic data used in primary data analyses were stored at Study Investigators institutions, and at the National Cell Repository for Alzheimer's Disease (NCRAD, U24AG021886) at Indiana University funded by NIA. Associated Phenotypic Data used in primary and secondary data analyses were provided by Study Investigators, the NIA funded Alzheimer's Disease Centers (ADCs), and the National Alzheimer's Coordinating Center (NACC, U01AG016976) and the National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site (NIAGADS, U24AG041689) at the University of Pennsylvania, funded by NIA, and at the Database for Genotypes and Phenotypes (dbGaP) funded by NIH. This research was supported in part by the Intramural Research Program of the National Institutes of health, National Library of Medicine. Contributors to the Genetic Analysis Data included Study Investigators on projects that were individually funded by NIA, and other NIH institutes, and by private U.S. organizations, or foreign governmental or nongovernmental organizations.

This work was funded also by the European Union Innovative Medicine Initiative (IMI) programme under grant agreement No. 115975 as part of the Alzheimer's Disease Apolipoprotein Pathology for Treatment Elucidation and Development (ADAPTED, https://www.imi-adapted.eu); and the European Union's Horizon 2020 research and innovation programme as part of the Common mechanisms and pathways in Stroke and Alzheimer's disease CoSTREAM project (www.costream.eu, grant agreement No. 667375).

References

1. 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).
2. Sims, R. *et al.* Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. *Nat Genet* **49**, 1373-1384 (2017).
3. Kunkle, B.W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet* **51**, 414-430 (2019).
4. International Genomics of Alzheimer's Disease, C. Convergent genetic and expression data implicate immunity in Alzheimer's disease. *Alzheimers Dement* **11**, 658-71 (2015).
5. Chouraki, V. & Seshadri, S. Genetics of Alzheimer's disease. *Adv Genet* **87**, 245-94 (2014).
6. 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).
7. Ott, J., Wang, J. & Leal, S.M. Genetic linkage analysis in the age of whole-genome sequencing. *Nat Rev Genet* **16**, 275-84 (2015).
8. Conneally, P.M. *et al.* Report of the Committee on Methods of Linkage Analysis and Reporting. *Cytogenet Cell Genet* **40**, 356-9 (1985).
9. van Duijn, C.M. *et al.* Park7, a novel locus for autosomal recessive early-onset parkinsonism, on chromosome 1p36. *Am J Hum Genet* **69**, 629-34 (2001).
10. Bonifati, V. *et al.* Localization of autosomal recessive early-onset parkinsonism to chromosome 1p36 (PARK7) in an independent dataset. *Ann Neurol* **51**, 253-6 (2002).
11. Schellenberg, G.D. *et al.* Genetic linkage evidence for a familial Alzheimer's disease locus on chromosome 14. *Science* **258**, 668-671 (1992).
12. Poirier, J. *et al.* Apolipoprotein E polymorphism and Alzheimer's disease. *The Lancet* **342**, 697-699 (1993).
13. Slattery, C.F. *et al.* R47H TREM2 variant increases risk of typical early-onset Alzheimer's disease but not of prion or frontotemporal dementia. *Alzheimers Dement* **10**, 602-608 e4 (2014).
14. Wijsman, E.M. *et al.* APOE and other loci affect age-at-onset in Alzheimer's disease families with PS2 mutation. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* **132**, 14-20 (2005).
15. Basu, R., Taylor, M.R. & Williams, M.E. The classic cadherins in synaptic specificity. *Cell Adh Migr* **9**, 193-201 (2015).
16. Keightley, P.D. Rates and fitness consequences of new mutations in humans. *Genetics* **190**, 295-304 (2012).
17. Haldane, J.B.S. The rate of spontaneous mutation of a human gene. *Journal of genetics* **83**, 235-244 (2004).
18. Kan, Y.W. & Dozy, A. Antenatal diagnosis of sickle-cell anaemia by DNA analysis of amniotic-fluid cells. *The Lancet* **312**, 910-912 (1978).

19. Reiss, J. *et al.* Discrimination between recurrent mutation and identity by descent: application to point mutations in exon 11 of the cystic fibrosis (CFTR) gene. *Human genetics* **87**, 457-461 (1991).
20. Yates, A. *et al.* Ensembl 2016. *Nucleic acids research* **44**, D710-D716 (2015).
21. Fagerberg, L. *et al.* Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Molecular & Cellular Proteomics* **13**, 397-406 (2014).
22. Mathys, H. *et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature*, 1 (2019).
23. Hume, M.A., Barrera, L.A., Gisselbrecht, S.S. & Bulyk, M.L. UniPROBE, update 2015: new tools and content for the online database of protein-binding microarray data on protein–DNA interactions. *Nucleic acids research* **43**, D117-D122 (2014).
24. Jolma, A. *et al.* DNA-binding specificities of human transcription factors. *Cell* **152**, 327-339 (2013).
25. Kent, W.J. *et al.* The human genome browser at UCSC. *Genome research* **12**, 996-1006 (2002).
26. 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).
27. Jansen, I.E. *et al.* Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet* (2019).
28. Savage, J.E. *et al.* Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat Genet* **50**, 912-919 (2018).
29. Davies, G. *et al.* Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nat Commun* **9**, 2098 (2018).
30. Peters, F. *et al.* The neural correlates of verbal short-term memory in Alzheimer's disease: an fMRI study. *Brain* **132**, 1833-46 (2009).
31. Sullivan, E.V., Corkin, S. & Growdon, J.H. Verbal and nonverbal short-term memory in patients with Alzheimer's disease and in healthy elderly subjects. *Developmental Neuropsychology* **2**, 387-400 (1986).
32. Peters, F. *et al.* The neural correlates of verbal short-term memory in Alzheimer's disease: an fMRI study. *Brain* **132**, 1833-1846 (2009).
33. Gul, I.S., Hulpiau, P., Saeys, Y. & van Roy, F. Evolution and diversity of cadherins and catenins. *Exp Cell Res* **358**, 3-9 (2017).
34. Seong, E., Yuan, L. & Arikkath, J. Cadherins and catenins in dendrite and synapse morphogenesis. *Cell Adh Migr* **9**, 202-13 (2015).
35. Arikkath, J. & Reichardt, L.F. Cadherins and catenins at synapses: roles in synaptogenesis and synaptic plasticity. *Trends Neurosci* **31**, 487-94 (2008).
36. Wang, Y. *et al.* The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome biology* **19**, 151 (2018).
37. Suzuki, S.C. & Takeichi, M. Cadherins in neuronal morphogenesis and function. *Dev Growth Differ* **50 Suppl 1**, S119-30 (2008).

38. Szklarczyk, D. *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* **43**, D447-52 (2015).
39. Takeichi, M. The cadherin superfamily in neuronal connections and interactions. *Nat Rev Neurosci* **8**, 11-20 (2007).
40. Lefkovics, K., Mayer, M., Bercsenyi, K., Szabo, G. & Lele, Z. Comparative analysis of type II classic cadherin mRNA distribution patterns in the developing and adult mouse somatosensory cortex and hippocampus suggests significant functional redundancy. *J Comp Neurol* **520**, 1387-1405 (2012).
41. Suzuki, S.C., Inoue, T., Kimura, Y., Tanaka, T. & Takeichi, M. Neuronal circuits are subdivided by differential expression of type-II classic cadherins in postnatal mouse brains. *Mol Cell Neurosci* **9**, 433-47 (1997).
42. Matsunaga, E. & Okanoya, K. Cadherins: potential regulators in the faculty of language. *Curr Opin Neurobiol* **28**, 28-33 (2014).
43. Wang, K. *et al.* Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* **459**, 528-33 (2009).
44. Singh, S.M., Castellani, C. & O'Reilly, R. Autism meets schizophrenia via cadherin pathway. *Schizophr Res* **116**, 293-4 (2010).
45. van Loenhoud, A.C. *et al.* A neuroimaging approach to capture cognitive reserve: Application to Alzheimer's disease. *Hum Brain Mapp* **38**, 4703-4715 (2017).
46. Clare, L. *et al.* Potentially modifiable lifestyle factors, cognitive reserve, and cognitive function in later life: A cross-sectional study. *PLoS Med* **14**, e1002259 (2017).
47. Rawle, M.J. *et al.* Apolipoprotein-E (Apoe) epsilon4 and cognitive decline over the adult life course. *Transl Psychiatry* **8**, 18 (2018).
48. Liu, F. *et al.* A study of the SORL1 gene in Alzheimer's disease and cognitive function. *J Alzheimers Dis* **18**, 51-64 (2009).
49. McKhann, G. *et al.* Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* **34**, 939-44 (1984).
50. 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).
51. Thompson, E. The structure of genetic linkage data: from LIPED to 1M SNPs. *Hum Hered* **71**, 86-96 (2011).
52. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
53. Naj, A.C. *et al.* Quality control and integration of genotypes from two calling pipelines for whole genome sequence data in the Alzheimer's disease sequencing project. *Genomics* (2018).
54. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-95 (2010).
55. Challis, D. *et al.* An integrative variant analysis suite for whole exome next-generation sequencing data. *BMC Bioinformatics* **13**, 8 (2012).

56. Evani, U.S. *et al.* Atlas2 Cloud: a framework for personal genome analysis in the cloud. *BMC genomics* **13**, 1 (2012).
57. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122 (2016).
58. McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069-70 (2010).
59. Hubbard, T. *et al.* The Ensembl genome database project. *Nucleic acids research* **30**, 38-41 (2002).
60. Lander, E.S. & Green, P. Construction of multilocus genetic linkage maps in humans. *Proc Natl Acad Sci U S A* **84**, 2363-7 (1987).
61. Lange, K. & Sobel, E. A random walk method for computing genetic location scores. *Am J Hum Genet* **49**, 1320-34 (1991).
62. 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).
63. Koepke, H. & Thompson, E. Efficient identification of equivalences in dynamic graphs and pedigree structures. *J Comput Biol* **20**, 551-70 (2013).
64. Nato, A.Q., Jr. *et al.* PBAP: a pipeline for file processing and quality control of pedigree data with dense genetic markers. *Bioinformatics* **31**, 3790-8 (2015).
65. Haldane, J.B.S. The combination of linkage values and the calculation of distances between the loci of linked factors. *J Genet* **8**, 299-309 (1919).
66. Matise, T.C. *et al.* A second-generation combined linkage physical map of the human genome. *Genome Res* **17**, 1783-6 (2007).
67. Genomes Project, C. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).
68. Cheung, C.Y., Thompson, E.A. & Wijsman, E.M. GIGI: an approach to effective imputation of dense genotypes on large pedigrees. *Am J Hum Genet* **92**, 504-16 (2013).
69. Nachman, M.W. & Crowell, S.L. Estimate of the mutation rate per nucleotide in humans. *Genetics* **156**, 297-304 (2000).
70. Karczewski, K.J. *et al.* Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *BioRxiv*, 531210 (2019).
71. Neph, S. *et al.* An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489**, 83-90 (2012).
72. Fullard, J.F. *et al.* An atlas of chromatin accessibility in the adult human brain. *Genome research* **28**, 1243-1252 (2018).
73. Khan, A. *et al.* JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res* **46**, D260-D266 (2018).
74. Mathys, H. *et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature*, 1 (2019).
75. Pruim, R.J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336-7 (2010).

Methods

Study population

The sample was obtained as a part of the Genetic Research in Isolated Population (GRIP) program. GRIP includes subjects from a genetically isolated community in the southwest of The Netherlands. GRIP maintains genealogy records up to approximately year 1600 A.D for 107,091 study participants, spanning 23 generations[6,48]. Subjects were traced through church records to form pedigrees. The Erasmus Medical Center, Rotterdam, medical ethics committee has approved this study. The study population contains 109 late-onset Alzheimer's disease (LOAD) patients. Alzheimer's disease (AD) cases in the cohort were ascertained through neurologists, general medical practitioners, and nursing homes. Research clinicians collected the relevant data for AD diagnosis among the participants. The final diagnosis of probable AD was confirmed by two independent neurologists based on the criteria of the National Institute of Neurological and communicative diseases and Stroke/Alzheimer's Disease and related disorder association[49]. A family history questionnaire was employed among first to third-degree relatives in order to collect information about the presence of AD, Parkinsonism, and dementia in study subjects and their relatives, but only the clinically-determined diagnosis was used in the current study.

The analysis was based primarily on two large five-generation pedigrees with multiple loops extracted from the GRIP population as provided to the ADSP (Supplementary Figure 1). These two pedigrees consist of 47 (ERF201) and 189 (ERF203) pedigree members, along with five and 24 subjects with a diagnosis of probable AD, respectively.

Genotyping

Array genotyping

For genotyping purposes, DNA of peripheral leukocytes was extracted utilizing a standard protocol. The AD cases were also genotyped for *APOE* ε2/ε3/ε4 using TaqMan allelic discrimination technology on an ABI prism 7900HT sequence detection system with SBS version 2.1 (Applied Biosystems). LOAD cases underwent screening for mutations in *APP, PSEN1* and *PSEN2*[6] and no mutation was identified. Each of the 29 LOAD cases included in this study were genotyped on two 2 Illumina arrays: 24 subjects with the HumanCNV370-Quad-v3c, and 5 subjects on the Human OmniExpress-12-v1.1b. Initial quality control was carried out, with a call rate threshold of >98% used for SNP variant screening. Only SNPs having a minor allele frequency (MAF)> 1% were included in the data set. PedCheck[50] and/or gl_auto from the MORGAN package[51] (https://www.stat.washington.edu/thompson/Genepi/MORGAN/) were used to detect Mendelian errors. SNPs with obligate Mendelian inconsistencies and monomorphic SNPs were excluded from the analysis. For the linkage analyses, the SNP array data were merged with PLINK[52] to create an initial joint SNP array, with alleles called in forward strand orientation, from which smaller panels of markers (linkage panels) were chosen (Supplementary Material).

Whole Genome Sequencing

A detailed description of the sequencing protocols and associated QC analyses are provided elsewhere[53]. In brief, the 29 probable AD cases used in the current analysis had WGS data generated at the Human Genome Sequencing Center at Baylor College of Medicine as part of the ADSP project. Sequencing reads were aligned by Burrows-Wheeler Aligner (BWA) 0.6.2 tool[54], using the GRCh37-lite human reference genome. Variant calling was performed using Atlas V2[55] and GATK[56], with the QC analysis adjudicating discrepancies. The QC protocol applied in the creation of these files was implemented in a series of scripts written in Python, Perl, and R (v2.15 and v3.1.1). The software packages PLINK[52] (v1.07 and v1.9), and PedCheck (v1.2) were also used. Variant call format (VCF) and Sequence Read Archives (SRA) files were generated by the ADSP and were made available for analysis. For the analyses presented here, all base-pair positions are provided on GRCh37.p13/hg19. The final Quality controlled WGS data contains 27,896,774 bi-allelic autosomal single nucleotide variants (SNVs)[53]. VCF files were annotated using the Ensembl Variant Effect Predictor (VEP)[57,58] based on Ensembl version 80[59].

Statistical Analyses

Genome-scan linkage analysis

We carried out an affected-only multipoint genome scan on the two complete multiplex pedigrees as supplied to the ADSP. Analyses were carried out on each pedigree in isolation, and also jointly across the pedigrees, under the assumption that different potentially rare variants could be segregating in the two pedigrees, given the complex genetic basis of LOAD. To achieve computationally feasible results with the complete large pedigrees and the many markers needed here for informative analyses, we used programs from the MORGAN package v3.3. We used gl_auto, which employs Markov chain Monte Carlo (MCMC) on large pedigrees to sample possible inheritance vectors (IVs)[60] at each marker position from the posterior distribution, given the complete pedigree structure and observed marker panel genotype data. We followed this with gl_lods to compute parametric LOD scores at each marker position[51], given the estimated inheritance vectors and observed trait data (Supplementary Figure 2). This approach has been described earlier[61,62] and is computationally efficient[63]. For parametric LOD score computations, we used an affected-only analysis under a dominant mode-of-inheritance with a penetrance of 0.9 in the high-risk and 0.001 in the low-risk genotypes. We used a relatively high disease allele frequency of 0.05 to accommodate the possibility of genetic heterogeneity though allowance for multiple carrier founders.

Conditions of the linkage analysis were as follows. We used the program PBAP, v1.0[64] to carry out relationship checking and construct input files for MORGAN and to select a highly-informative, uniformly-spaced, marker panel with very low (r^2<0.04) linkage disequilibrium (r^2, squared correlation) between markers (Supplementary Methods). Basic relationship checking identified no evidence for sample mix-ups or misspecified relationships. The Haldane map function[65] together with positions from the Rutgers map (v3)[66] defined expected recombination

rates between markers. Marker allele frequencies and inter-marker r^2 were obtained from the 1000 Genomes Project data[67] using the EUR ancestry subjects.

Follow-up analyses of the strongest linkage signal

The evidence for linkage is weaker than expected in the ERF203 pedigree if all affected subjects shared a single haplotype carrying a high-risk allele. This implies genetic and/or etiological heterogeneity. We, therefore, carried out analyses directed at identifying risk variants within the region with strong linkage evidence on chromosome 5. These analyses involved two basic steps. First, we carried out genotype phasing of the WGS onto segregating haplotypes in the pedigree as determined by the multipoint marker computations with marker panel, and identified variants that were specific to the putative risk haplotype(s). Second, we used the information regarding the potential risk variants together with functional annotation to nominate the most plausible variants. These steps are elaborated as Haplotype identification and analysis, and variant filtering, below.

Haplotype identification and analysis

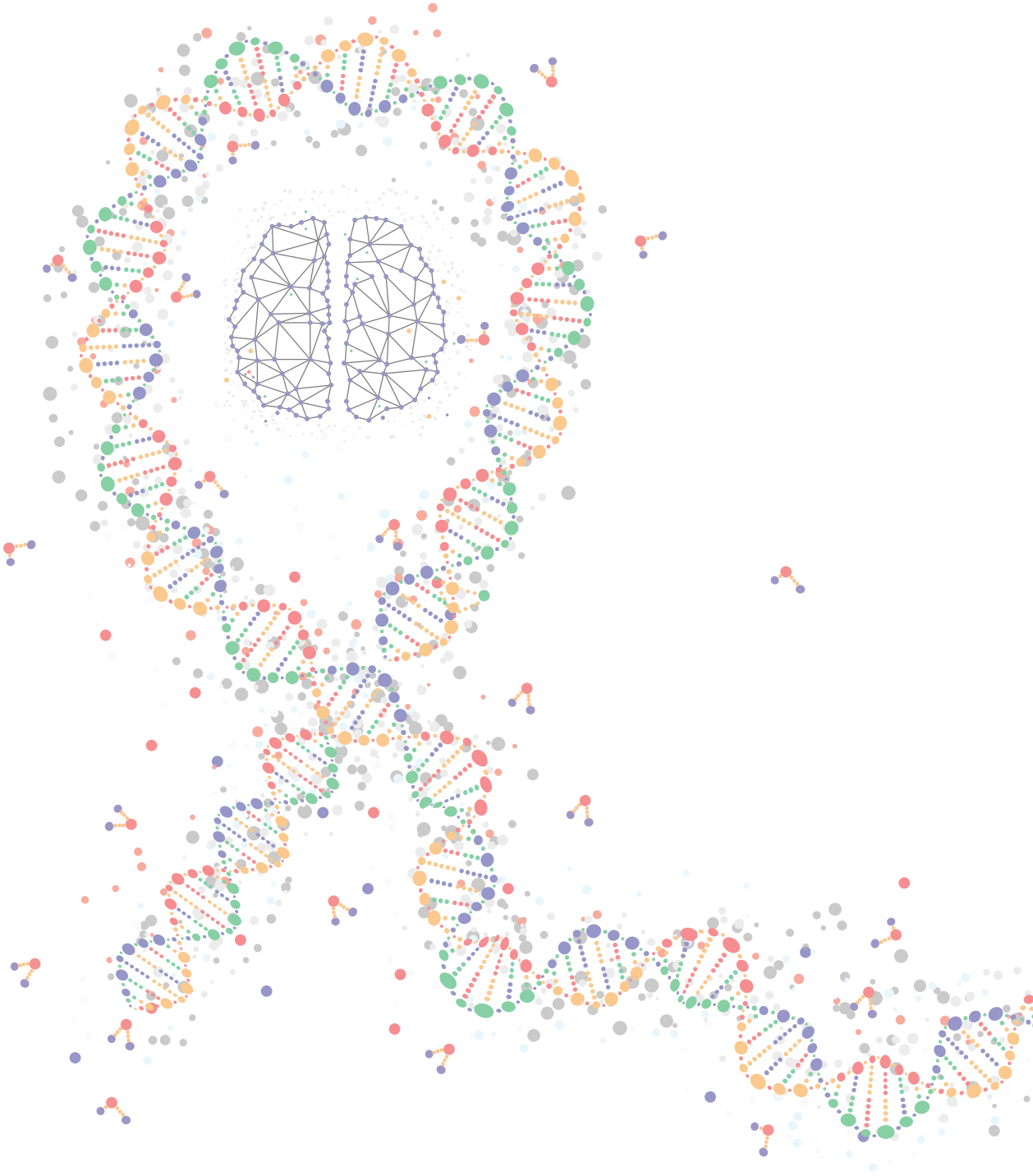
Haplotype analysis was used to follow-up the strongest linkage signal on chromosome 5. A contrast between cases and controls was not possible because of lack of verified unaffected subjects in the family. Therefore, the goal was to use contrasts between the "high risk" vs. "other" haplotypes to reduce the SNVs of interest in this region of interest. Haplotype analysis was carried out with non-overlapping, sequential, 20-variant blocks of SNVs from the variants in the WGS data that spanned the region of interest. For a particular set of SNVs, we used an internally-modified version of the pedigree-based imputation program, GIGI (ver. 1.06.1)[68], to phase the SNVs for each sampled set of inheritance vectors generated as part of the linkage analysis using the final, dense linkage panel for this purpose. This allowed us to obtain posterior distributions of haplotypes. For a particular block of SNVs, we determined the set of most probable haplotypes in those with the high risk and low-risk haplotype. The variants that met this criterion and that were polymorphic in the ERF pedigrees were extracted. To identify variants carried only on the risk haplotype, the allele of each selected variant on the high-risk haplotype was contrasted with the alleles on the 38 haplotypes in the 19 "low risk" subjects in the lowest tertile group. Only variants that were carried on the high-risk haplotype were retained for further consideration, under the assumption that independent mutation at the same base-pair position on two different haplotypes is a vanishingly rare occurrence in a genetically heterogeneous disease.

Variant filtering

We used three initial filters to identify variants for further consideration. Variants not carried forward by these three filters were excluded from all additional analyses. Filter 1 defined variants of interest as those that were only on the risk haplotype in the European population under study. The rationale behind this derives from population genetics data and principles. This includes information that the per-nucleotide mutation rate is exceedingly low on the order of 10^{-8} [17,69]

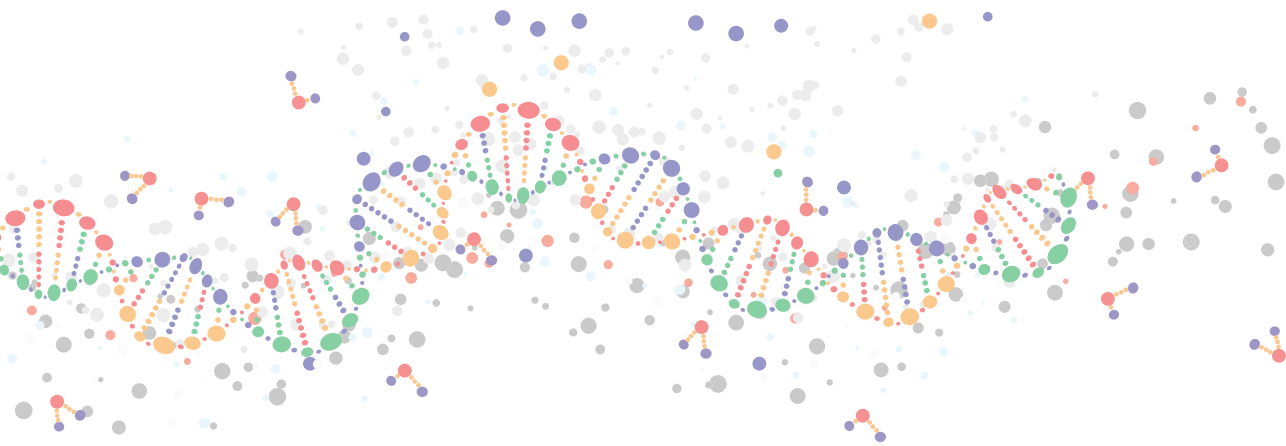
leading to only very rare recurrent mutation at the same site, and a unique haplotypic context for each unique mutation[18,19]. For Filter 2, we retained only variants with MAF ≤0.01 in either the general European (EUR) population or across geographic populations. We used the 1000 Genomes Project sample[67] data to estimate MAFs, except when the variant did not exist in that dataset, in which case we used the Genome Aggregation Database (gnomAD) composite allele frequencies[70]. If there was no data on allele frequency of a particular variant in any of these data sets, we assumed a MAF<0.01. The use of a low MAF threshold is justified by the observation that common variants have already been tested in GWAS of AD in populations of European ancestry, with the 5p14.3 region not emerging as genome-wide significant[2,3,26,27]. Filter 3 restricted attention to variants that had evidence for interpretable function through annotation to a gene or other potentially functional element in the region[20] as well as evidence of expression in brain either in bulk samples[21] or in brain-derived single-cell assays[22].

Bioinformatics filters provided additional non-exclusive prioritization of variants. We prioritized variants located near potentially open chromatin defined by overlap of variants with "peak" signals indicating DNase-I hypersensitive sites (DHS) or hotspots from the ENCODE project[71] or a recent study of DHS in brain-derived single-cell neurons and non-neurons[72]. We also prioritized variants in predicted TFBSs, including making use of the predicted motifs and the position-weight matrix (PWM) data[73] We made use of five features. (1) At each position in the PWM, we considered the total count of bases. (2) For alignment of a TF to a predicted TFBS, we focused on the positions in which the total count was at least 1/3 of the maximum total count across all positions. (3) At the position of a variant, we computed the probability of the Reference (or Alternative) allele by using the counts in the PWM at that site, only. (4) We defined the sample size of experiments used to define the motif as the total counts in the PWM at the variant position. A bigger sample carried more weight than a smaller sample. (5) When judging a particular Alt allele, the relative frequency of potential alleles other than the Reference allele was assumed to be represented by the remaining nucleotides, with their respective probabilities inversely related to the likely effect of the alleles.



Chapter 3

Pathways implicated in Alzheimer's disease and lifestyle factors



Chapter 3.1

Disentangling the biological pathways involved in early features of Alzheimer's disease in the Rotterdam Study

Shahzad Ahmad, Christian Bannister, Sven J. van der Lee, Dina Vojinovic, Hieab H.H. Adams, Alfredo Ramirez, Valentina Escott-Price, Rebecca Sims, Emily Baker, Julie Williams, Peter Holmans, Meike W. Vernooij, M. Arfan Ikram, Najaf Amin, Cornelia M. van Duijn

This chapter is published in the Alzheimer's and dementia: The Journal of the Alzheimer's Association. 2018 July;14(7): 848–857

Supplementary information for this chapter is available at:
https://doi.org/10.1016/j.jalz.2018.01.005

Abstract

Introduction: Exploring the role of Alzheimer's disease (AD) implicated pathways in predementia phase may provide new insight for preventive and clinical trials targeting disease specific pathways.

Methods: We constructed weighted Genetic risk scores, first based on 20 genome-wide significant AD risk variants and second clustering these variants within pathways. Risk scores were investigated for their association with AD, mild cognitive impairment and brain magnetic resonance imaging phenotypes including white matter lesions, hippocampal volume, and brain volume.

Results: The risk score capturing *endocytosis* pathway was significantly associated with mild cognitive impairment ($P = 1.44 \times 10^{-4}$). *Immune response* ($P = 0.016$) and *clathrin/AP2 adaptor complex* pathway ($P = 3.55 \times 10^{-3}$) excluding apolipoprotein E (*APOE*) also showed modest association with white matter lesions but did not sustain Bonferroni correction ($P = 9.09 \times 10^{-4}$).

Discussion: Our study suggests that the clinical spectrum of early AD pathology is explained by different biological pathways, in particular, the *endocytosis, clathrin/AP2 adaptor complex* and *immune response* pathways that are independent of *APOE*.

Introduction

Alzheimer's disease (AD) is a heterogeneous and genetically complex disease with high heritability (56-79 %)[1]. It has been known since the end of the previous century that a polymorphism in the apolipoprotein E (*APOE*) gene is the strongest common genetic risk factor[2-4]. This finding fueled speculations on the role of *lipid metabolism* and *cholesterol transport* pathway in AD in addition to the amyloid cascade and tau phosphorylation mechanism[5,6]. Furthermore, large-scale genome-wide association studies (GWAS) have discovered more than 20 novel common genetic variants that influence the risk of late-onset AD[7-13]. These common genetic variants have been mapped to eight biological pathways including *immune response, endocytosis, cholesterol transport, hematopoietic cell lineage, protein ubiquitination, hemostasis, clathrin/AP2 adaptor complex* and *protein folding*, each having a distinct biological function[14-16]. These eight pathways are not independent in a way that genes may be part of more than one biological pathway. For instance, *APOE* is part of four of the eight pathways namely *cholesterol transport, hematopoietic cell lineage, clathrin/AP2 adaptor complex* and *protein folding* pathways; clusterin (*CLU*) encoding for apolipoprotein J is involved in six pathways; phosphatidylinositol binding clathrin assembly protein (*PICALM*) and complement factor 1 (*CR1*) are involved in two pathways[14-16].

These diverse biological pathways may be responsible for the clinically heterogeneous manifestation of AD[17-19], which include endophenotypes such as changes in structural and functional magnetic resonance imaging (MRI) phenotypes, most notably hippocampal volume, total brain volume, and white matter lesions[20-23]. Furthermore, these biological pathways may also modulate the prodromal stages of AD such as mild cognitive impairment (MCI)[24-26]. Owing to heterogeneity during the predementia phase, one important unanswered question is whether the different biological pathways that are implicated in AD relate to the pleiotropy of clinical endophenotypes. We hypothesized that some biological pathways are involved in distinct clinical endophenotypes whereas others may be involved in multiple or even all. Disentangling the connection of biological pathways to various aspects of AD-related early pathology may be a crucial step towards improving our understanding of the pathogenesis of AD during the predementia stage and the first step toward a more informative and powerful readout for preventive and therapeutic trials targeting specific pathways.

The present study aims to capture the different biological pathways involved in AD using genetic risk scores to evaluate their role in AD and predementia endophenotypes including MCI, white matter lesions, total brain, and hippocampal volume.

Methodology

Study population

This study included samples from the Rotterdam Study (RS). The RS is a prospective population-based study[27] designed to investigate the etiology of age-related disorders. At the baseline examination in 1990-1993, study recruited 7983 subjects ≥ 55 years of age from the Ommoord district of Rotterdam (RS-I). At the baseline entry and after every 3 to 4 years, all the study participants were extensively interviewed and physically examined at the dedicated research center. During 2000 to 2001, the baseline cohort (RS-I) was expanded by adding 3011 subjects ≥55 years of age, who were not yet part of RS-I (RS-II). Second expansion of RS was performed by recruiting 3932 persons having ≥45 years of age during 2006 to 2008 (RS-III). The study has been approved by the Medical Ethical Committee of Erasmus Medical Center and by the Ministry of Health, Welfare and Sport of the Netherlands. Written informed consent was also obtained from each study participant to participate and to collect information from their treating physicians. Details of AD, dementia and MCI diagnosis are provided in the supplementary information. In the present study for AD cross-sectional analysis, we included in total 1270 late-onset AD cases and 7623 controls (age at last follow-up ≥ 65 years and dementia-free) whose follow-up information is complete until 2009-2013 in RS-I, RS-II and RS-III cohorts. This AD sample includes 1057 incident and 213 prevalent AD cases. For prospective AD analysis, 10370 dementia-free (normal) participants were also included in the study from all three RS cohorts at their baseline and were subsequently followed until 2009-2013, to analyze their progression into AD (average 11 years of follow-up). In the MCI data set, we included 360 MCI cases and 3245 cognitively normal controls from the first extensive cognitive assessment conducted between 2002 and2005 in RS-I and RS-II cohorts. MRI was implemented in 2005 in RS cohorts and 5899 persons came for MRI scanning until 2015. After excluding subjects with stroke and/or dementia (n = 251) at time of scanning, poor imaging quality (n = 313) and missing genotyping information (n = 814), we retained 4521 cognitively normal individuals in the MRI sample (Table 1).

Genotyping

Blood was drawn for genotyping from participants of RS cohorts during their first visit and DNA genotyping was performed at the internal genotyping facility of the Erasmus Medical Center, Rotterdam. All samples were genotyped with the 550K, 550K duo, or 610K Illumina arrays. Genotyping quality control criteria include, call rate < 95%, Hardy-Weinberg equilibrium $P < 1.0x10^{-6}$ and minor allele frequency (MAF) < 1%. Moreover, study samples with excess autosomal heterozygosity, call rate < 97.5%, ethnic outliers, and duplicate or family relationships were excluded during quality control analysis. Genetic variants were imputed from the Haplotype Reference Consortium reference panel (version 1.0)[28], using the Michigan imputation server[29]. The server uses SHAPEIT2 (v2.r790)[30] to phase the genotype data and performs imputation

Table 1: Cohort characteristics

Characteristics	RS-I	RS-II	RS-III	Total
AD data set (N)	5854	2062	977	8893
Late-onset AD	1118	134	18	1270
AD free controls	4736	1928	959	7623
Age-at-onset (SD), years	84.58 (6.8)	82.75 (6.7)	78.54 (9.5)	84.30 (6.8)
Age of controls (SD), years	82.87 (6.9)	76.52 (6.4)	69.15 (5.7)	79.53 (8.2)
Female (%)	3526 (60%)	1133 (55%)	569 (58%)	5228 (59%)
MCI data set (N)	2178	1427		3605
MCI cases	235	125	-	360
Controls	1943	1302	-	3245
Age (SD), years	74.79(5.7)	67.53(6.9)	-	71.9 (7.2)
Female (%)	1271(58.4)	786(55.1)	-	2057 (57%)
MRI data set (N)	968	1068	2485	4521
Age (SD), years	78.89 (4.9)	69.34 (5.9)	57.21 (6.4)	64.72 (10.8)
Female (%)	556 (58%)	565 (53%)	1390 (56%)	2511 (56%)

Abbreviation: RS, Rotterdam Study (Cohort I, II, II); AD, Alzheimer's disease;
MCI, Mild cognitive impairment; MRI, Magnetic resonance imaging; SD, Standard deviation.

with Minimac 3 software[31]. For this study, we used genetic variants that had imputation quality (R^2) > 0.5.

MRI scanning

Image acquisition

MRI scanning is assessed on a 1.5-T MRI unit with a dedicated eight-channel head coil (Signa HD platform, GE Healthcare, Milwaukee, USA) since the induction of a dedicated MRI machine in the RS. The MRI protocol was based on several high-resolution axial sequences, including a T1-weighted (slice thickness 0.8 mm), T2-weighted (1.6 mm), and fluid-attenuated inversion recovery sequence (2.5 mm). A detailed description of the MRI protocol is described previously[32].

Image processing

We excluded 251 persons with stroke and/or dementia from the total 5899 subjects, because this may affect image processing. All T1 images were segmented into the supratentorial gray matter, white matter and cerebrospinal fluid using a k-nearest neighbor algorithm[33]. White matter lesions were segmented based on T1 tissue maps and an automatically detected threshold for the intensity of fluid-attenuated inversion recovery scans[34]. The hippocampus was segmented using a fully automated method, as described previously[35]. Semiquantitative MRI postprocessing software was used to measure intracranial volume and brain volume which included Elastix

and custom-built software[36]. To calculate intracranial volume, non-brain tissues (skull, eyes and dura) were removed by nonlinearly registering all brain scans to a manually created template in which non-brain tissues were masked[33,36,37]. In all MRI scans, after visual inspection of all segmentations, additional 313 subjects were excluded because of poor quality.

Statistical analysis

Genetic risk score computation

To construct the risk score, we selected late-onset AD-associated single nucleotide polymorphisms (SNPs) reaching genome-wide significance level ($P < 5.0x10^{-8}$; Supplementary Table 1), including one rare *TREM2* variant[7,38]. In common variants, we considered only variants identified by the International Genomics of Alzheimer's Project (IGAP) meta-analyses. In addition, we considered *APOE* ε4 (rs429358) variant for risk score construction. From a total of 21 SNPs, the *HLA-DRB1-HLA-DRB5* (rs9271192) variant was excluded from risk score calculation because of its low imputation quality ($R^2 = 0.31$) in the RS. This led to a final selection of 20 independent genome-wide significant AD-associated variants. Weighted genetic risk score was constructed using the effect sizes (log of odds ratio) of the genome-wide significant variants from the IGAP meta-analysis [7] as weights and their respective allele dosages from imputed genotype data of our study cohorts. Risk score was constructed as the sum of the products of SNP dosages and their corresponding weights in R software (https://www.R-project.org/). We constructed genetic risk score in two ways: (1) combining all 20 selected variants and (2) clustering the variants into their respective pathways.

Combined genetic risk score

Combined genetic risk score (GRS1) was constructed in two ways, that is, (1) using all the 20 selected SNPs and (2) excluding the *APOE* ε4 variant to identify the joint independent effect of all other genome-wide significant SNPs.

Pathway-specific genetic risk score

For pathway-specific genetic risk score (GRS2), the genome-wide significant AD SNPs were divided into pathways (*immune response, endocytosis, cholesterol transport, hematopoietic cell lineage, protein ubiquitination, hemostasis, clathrin/AP2 adaptor complex and protein folding pathway*) identified by Jones *et al.*[16] (Supplementary Table 2). Classifying genome-wide significant AD SNPs into pathways, we also used information from Guerreiro *et al.*,[14] in which the authors reviewed the possible division of known AD-associated genes into biological pathways[14]. Further, the GeneNetwork database (http://genenetwork.nl/) was used to confirm the allocated pathways. Of the 20 SNPs, 14 could be clustered into seven nonmutually exclusive pathways (Supplementary Table 2). Similar to GRS1, we also constructed GRS2 with and without the *APOE* ε4 variant. *APOE* ε4 variant was grouped under four pathways including *cholesterol transport*[14], *hematopoietic cell lineage*, *clathrin/AP2 adaptor complex* and *protein folding*[16]. GRS2 was constructed for only those pathways, which could be assigned at least two SNPs, therefore *protein ubiquitination*

pathway, which contained only one SNP, was excluded from all analyses, while *hematopoietic cell lineage and protein folding* pathways were also not considered in the analyses excluding the *APOE* ε4 variant.

Association analyses of GRS1, and GRS2

To test the association of AD and MCI with the risk scores, we used logistic regression analysis in R software (www.R-project.org), using disease status as the outcome, risk scores as predictor and age and sex as covariates. To assess the possible inflation of association results between AD and risk scores, we repeated the association analysis excluding 625 AD cases who were part of the IGAP meta-analysis[7] from total 1,270 AD cases of the RS cohort. Furthermore, we performed prospective analysis using the Cox-proportional hazards model (N=1057 incident AD cases) in R software using 'survival' package[39] and reported results as hazard ratio (HR) per 1 standard deviation increase in risk score and 95% confidence interval. The association of single variants with AD and MCI was assessed using a logistic regression model adjusted for age and sex. Results of association analyses were reported as unstandardized regression coefficient and *P* values.

To test the association of MRI phenotypes including total brain volume, white matter lesions and hippocampal volume with the risk scores we used linear regression adjusted for age, sex, and intracranial volume in MRI scans. Single variant association analysis was also performed for MRI phenotypes. Bonferroni correction (0.05/(11 risk scores x 5 phenotypes); $P = 9.09 \times 10^{-4}$ was used to correct for multiple testing.

Results

Association of the GRS1 with AD, MCI, and MRI endophenotypes

The risk score containing all SNPs, this is, GRS1 both including *APOE* ε4 (effect = 0.73, $P = 6.53 \times 10^{-74}$) and excluding *APOE* ε4 (effect = 0.69, $P = 1.12 \times 10^{-11}$) was significantly associated with an increased risk of AD (Table 2). This association remained significant (*APOE* excluding; effect = 0.66, $P = 8.47 \times 10^{-7}$) after removing the patients who were included in the IGAP meta-analysis[7] (Supplementary Table 3). GRS1 was also significantly associated with progression from normal subjects into AD patients both including (HR = 1.69, $P = 6.64 \times 10^{-83}$) and excluding *APOE* ε4 (HR = 1.27, $P = 4.88 \times 10^{-15}$; Supplementary Table 4). GRS1 was associated with MCI when *APOE* ε4 was included (effect = 0.19, $P = 0.012$) but the association was stronger when *APOE* ε4 was excluded from the analysis (effect = 0.59, $P = 9.51 \times 10^{-4}$; Table 3) however, these associations did not pass multiple testing correction. No association of GRS1 was observed with any of the MRI phenotypes: white matter lesions, hippocampal volume, and total brain volume (Table 4).

Table 2: Results of association of Alzheimer's disease with risk scores

	Including *APOE*			Excluding *APOE*		
SNP Cluster*	**β**	**SE**	***P-value***	**β**	**SE**	***P-value***
GRS1 (Combined)	0.73	0.040	$6.53x10^{-74}$	0.69	0.101	$1.12x10^{-11}$
Immune response	-	-	-	0.69	0.166	$3.20x10^{-5}$
Endocytosis	-	-	-	0.75	0.171	$1.28x10^{-5}$
Cholesterol Transport	0.71	0.042	$3.22x10^{-64}$	0.39	0.219	0.077
Hematopoietic cell lineage†	0.73	0.042	$5.16x10^{-66}$	-	-	-
Hemostasis			-	0.50	0.292	0.090
Clathrin/AP2 Adaptor complex	0.72	0.042	$4.68x10^{-65}$	0.50	0.236	0.036
Protein folding†	0.72	0.042	$2.96x10^{-64}$	-	-	-

Abbreviations: GRS1, Combined genetic risk score; SNP, Single nucleotide polymorphism; β, Regression coefficient; SE, Standard error

Note: Multiple testing correction by Bonferroni 0.05/ (5 phenotypes x 11 risk scores); $P < 9.09x10^{-4}$ was considered significant

* Logistic regression model adjusted for age and sex in RS (N=1270 cases)

†Only one SNP available in excluding *APOE* GRS2

Association of the GRS2 with AD

Among GRS2 of which *APOE* ε4 is a part, *cholesterol transport*, *hematopoietic cell lineage*, *clathrin/AP2 adaptor complex,* and *protein folding* were significantly associated with AD (effect ≥ 0.71, $P < 3.22x10^{-64}$) only when *APOE* ε4 was included in the risk scores. Among the non-*APOE* pathways, AD was significantly associated with GRS2 capturing *immune response* (effect = 0.69, $P = 3.20x10^{-5}$) and *endocytosis pathway* (effect = 0.75, $P = 1.28x10^{-5}$; Table 2) and association sustained (*Immune response*; effect = 0.68, $P = 2.22x10^{-3}$ and *endocytosis*; effect = 0.79, $P = 5.37x10^{-4}$) even after removing the patients who were included in the IGAP meta-analysis[7] (Supplementary Table 3). GRS2 capturing *immune response* (HR = 1.14, $P = 1.19x10^{-5}$), *endocytosis* (HR = 1.19, $P = 5.16x10^{-8}$) and *APOE* ε4 excluded *clathrin/AP2 adaptor complex* (HR = 1.09, $P = 5.98x10^{-3}$) pathway showed association with progression from normal into AD. Both *Immune response* and *endocytosis* pathways were significant after correcting for multiple testing. GRS2 including *APOE* ε4 were also significantly associated with normal to AD progression (HR ≥ 1.60, $P \leq 1.44x10^{-69}$; Supplementary Table 4). Comparatively, except for *APOE* ε4, and the variants in *CR1* and *BIN1* genes, no single variant showed significant evidence of association with AD (Supplementary Table 5). The variant rs6733839 in the *BIN1* gene partially explains the association between the *endocytosis* pathway and AD, whereas *APOE* ε4 mainly explains the association of all pathways of which *APOE* ε4 is a part.

Association of the GRS2 with MCI

In GRS2, only the *endocytosis* pathway showed significant evidence for association (effect = 1.16, $P = 1.44x10^{-4}$; Table 3) with MCI. Although the significance of the association is similar to that of

Table 3: Results of association of mild cognitive impairement with risk scores

	Including *APOE*			Excluding *APOE*		
SNP Cluster*	**β**	**SE**	***P-value***	**β**	**SE**	***P-value***
GRS1 (combined)	0.19	0.075	0.012	0.59	0.179	$9.51x10^{-4}$
Immune response	-	-	-	0.46	0.295	0.116
Endocytosis	-	-	-	1.16	0.305	$1.44x10^{-4}$
Cholesterol Transport	0.11	0.082	0.164	0.39	0.392	0.322
Hematopoietic cell lineage†	0.09	0.084	0.269	-	-	-
Hemostasis	-	-	-	-0.08	0.524	0.872
Clathrin/AP2 Adaptor complex	0.12	0.082	0.128	0.72	0.423	0.089
Protein folding†	0.10	0.083	0.218	-	-	-

Abbreviations: GRS1, Combined genetic risk score; SNP, Single nucleotide polymorphism; β, Regression coefficient; SE, Standard error.
Note: Multiple testing correction by Bonferroni 0.05/ (5 phenotypes x 11 risk scores); $P < 9.09x10^{-4}$ was considered significant
* Logistic regression model adjusted for age and sex in RS (N=360 cases)
† Only one SNP available in excluding *APOE* pathway-based GRS2

the overall risk score (GRS1), the effect estimate is considerably higher (1.16 vs. 0.59 overall). In the single variant analysis, the strongest association of MCI was observed with rs6733839 in the *BIN1* gene (effect = 0.262, $P = 1.12x10^{-3}$; Supplementary Table 5). Although this association was not significant after correcting for multiple testing, however, it partially explains the association between MCI and GRS2 capturing *endocytosis*.

Association of the GRS2 with MRI phenotypes

White matter lesions were associated with GRS2 capturing *immune response* (effect = 0.15, $P = 0.016$) and *clathrin/AP2 adaptor complex* excluding *APOE* ε4 (effect = 0.26, $P = 3.55x10^{-3}$). If we consider multiple testing, both these associations lose significance after accounting for all tested phenotypes and risk scores. Of note is that no association of white matter lesions with the GRS2 capturing the *clathrin/AP2 adaptor complex* is observed when *APOE* ε4 is included in the GRS2 (effect = 0.011, $P = 0.507$; Table 4). We did not observe association of GRS2 with hippocampal volume and total brain volume. In the single variant analysis association of white matter lesions are seen with variants in *PICALM, CLU* genes ($P \leq 0.05$). Hippocampal volume shows association with variants in *BIN1* and *CELF1* genes ($P < 0.05$; Supplementary Table 6). None of the single variant association sustained Bonferroni correction for multiple testing.

Table 4: Results for association of risk scores with magnetic resonance imaging phenotypes

	Including *APOE*									Excluding *APOE*								
SNP cluster*	**White matter lesions**			**Hippocampal volume**			**Brain volume**			**White matter lesions**			**Hippocampal volume**			**Brain volume**		
	β	**SE**	***P***	β	**SE**	***P***	β	**SE**	***P***	β	**SE**	***P***	β	**SE**	***P***	β	**SE**	***P***
GRS1 (combined)	0.012	0.016	0.448	-0.001	0.016	0.929	0.002	0.007	0.806	0.059	0.037	0.114	-0.009	0.038	0.810	-0.006	0.016	0.724
Immune response	-	-	-	-	-	-	-	-	-	0.149	0.062	0.016	-0.024	0.062	0.706	-0.010	0.026	0.692
Endocytosis	-	-	-	-	-	-	-	-	-	0.071	0.062	0.254	-0.046	0.063	0.462	0.004	0.026	0.865
Cholesterol Transport	0.005	0.017	0.785	0.001	0.017	0.964	0.004	0.007	0.574	0.063	0.080	0.434	0.013	0.080	0.875	0.023	0.033	0.497
Hematopoietic cell lineage†	0.002	0.017	0.901	0.001	0.017	0.976	0.004	0.007	0.556	-	-	-	-	-	-	-	-	-
Hemostasis	-	-	-	-	-	-	-	-	-	0.228	0.108	0.034	-0.077	0.109	0.479	-0.009	0.045	0.835
Clathrin/AP2 Adaptor complex	0.011	0.017	0.507	-0.002	0.017	0.924	0.003	0.007	0.658	0.258	0.088	3.55×10^{-3}	-0.077	0.109	0.479	-0.009	0.045	0.835
Protein folding†	0.007	0.017	0.700	-0.001	0.017	0.970	0.004	0.007	0.619	-	-	-	-	-	-	-	-	-

Abbreviations: GRS1, Combined genetic risk score; SNP, Single nucleotide polymorphism; β, Regression coefficient; SE, Standard error; *P*, P-value

Note: Multiple testing correction by Bonferroni 0.05/ (5 phenotypes x 11 risk scores); $P < 9.09 \times 10^{-4}$ was considered significant

*Linear regression model with MRI phenotype as outcome and risk score as predictor, adjusted for age at MRI scan, sex in RS (N=4521)

† Only one SNP available in excluding *APOE* pathway-based GRS2

Discussion

Combined risk score is significantly associated with AD and normal to AD progression but not with any of the early features of AD tested in our study including MCI and MRI markers. However, our pathway-based risk score analysis shows that the *endocytosis* pathway significantly associates with MCI in addition to AD and normal to AD progression (Supplementary Figure 1).

The association of GRS1 with AD is consistent with other similar studies on AD[40-42]. However, while others observed significant association of combined risk score with MCI[43,44], in our study the association of GRS1 with MCI was not significant after correcting for multiple testing. We did not find association of GRS1 with any of the studied MRI endophenotypes. These findings are consistent with those of Mormino *et al.*,[45] and Lupton *et al.*,[46]; both studies did not find association of hippocampal volume with combined GRS1 based on genome-wide significant AD variants, but Mormino *et al.*,[45] observed this association only with risk score based on non-genome wide significant AD variants. The largest study so far that included RS, however, reported significant evidence of association of risk score based on all genome-wide significant AD variants with hippocampal volume and total brain volume[47].

This is the first study that addressed the role of specific pathways in AD and its early clinical manifestations, that is., MCI and MRI phenotypes. Our study shows that GRS2 based on the *immune response* pathway was significantly associated with AD and normal to AD progression. Furthermore, we observed association of *immune response* with white matter lesions at MRI but this association did not survive Bonferroni correction. The genes clustered in the *immune response* pathway (*CLU, CRI, INPP5D, MS4A6A, TREM2, MEF2C, EPHA1*) are mainly expressed in microglial cells and play a part in the innate *immune response* in central nervous system[48-52]. Microglial cells are also thought to play a role in amyloid plaque clearance[53,54]. It has been hypothesized that the activation of the immune system and the subsequent inflammatory response are involved in neuronal damage including axonal loss and white matter pathology due to demyelination[55]. White matter lesions are associated with increased risk of cognitive decline, developing dementia[21] and AD[22,56]. White matter lesions are also more frequently observed in AD patients than controls[57,58].

The most interesting finding of the present study is that the genes capturing the *endocytosis* pathway significantly associate with MCI, AD and with progression from normal (dementia-free) to AD. This pathway is independent of *APOE* and includes the *BIN1, PICALM, CD2AP, and SORL1* genes. We show that the association of GRS1 with MCI status is mainly attributed to the genes involved in the *endocytosis* pathway. Omitting the AD genes not related to the *endocytosis* pathway makes the association of the pathway with MCI even stronger. This suggests that the *endocytosis* pathway plays a critical role in an early prodromal phase of AD. Our findings are in line with previous studies suggesting the activation of the *endocytic* pathway is the earliest reported

intracellular manifestation of AD[59-61]. Furthermore, the effect estimate of the *endocytosis* pathway was larger for MCI (1.16) compared with AD (0.75) suggesting a stronger association with MCI; however, this difference in effect estimates was not significant (P = 0.12). The *endocytosis* pathway is involved in neuronal uptake of macromolecules and secretory vesicles during synaptic transmission. As efficient uptake of extracellular cholesterol is critical for neuronal functions such as repair, synapse formation and exon elongation[62], normal neuronal work needs smooth functioning of *endocytosis* pathway[63]. Postmortem studies have also demonstrated reduced brain cholesterol levels in the brain areas responsible for memory and learning, among late-onset AD cases and age-matched controls[64]. These facts suggest that defects in *endocytosis,* which derive the cholesterol uptake, could lead to impaired neurotransmitter release and synaptic function[65]. Dysfunction in *endocytosis* can also contribute to the accumulation of abnormal Aβ peptide[66]. Based on this finding, we can suggest that the *endocytosis* pathway is a common molecular mechanism between MCI and AD that starts manifesting at early stages of disease. Risk contributed by variants clustered in this pathway at various stages of AD progression can possibly provide clue about disease trajectory.

Our study further shows association of the *clathrin/AP2 adaptor complex* pathway with white matter lesions. Although the association failed to pass the multiple testing, it is interesting to note that no association was detected with the combined risk score either in our study or a larger study performed earlier by Chauhan *et al.*,[47] that included up to 11550 individuals. This suggests that pathway-based risk scores may be more sensitive in picking association signals that may be relevant for specific AD pathologies. Two variants tagging *PICALM* and *CLU* genes cluster in the c*lathrin/AP2 adaptor complex* pathway. Each variant independently shows nominal association with white matter lesions in our analyses but combining their effects are additive and improve the strength of association. There is strong evidence that the two proteins encoded by the genes interact at molecular level[67,68]. *PICALM* is involved in VAMP2 trafficking that is a crucial process to maintain functional integrity of synapses which are crucial to cognitive function[69,70]. *PICALM* is also found to be expressed in the white matter and, immunolabeling of human brain tissue shows that *PICALM* is mainly found in blood vessel walls[71]. *CLU* clustered in the *clathrin/AP2 adaptor* is involved in efflux of free insoluble Aβ peptides through blood-brain barrier[72]. Increased plasma levels of *CLU* were associated with increased burden of Aβ peptide in healthy elderly population and brain atrophy in AD[73,74] and decreased integrity of white matter in young adults[75]. Demyelination of white matter is reported to occur even before the accumulation of Aβ plaques and neurofibrillary tangles[76]. The findings of the present study suggest that the increased genetic burden of risk variants in the *clathrin/AP2 adaptor* complex (*clathrin-mediated endocytosis*) and *immune response* pathway may play a role in early pathogenesis of AD through white matter pathology.

Among pathways including *APOE* (*Cholesterol transport, hematopoietic cell lineage, clathrin/AP2 adaptor complex and protein folding*), significant association with AD and normal to AD progression suggest that *APOE* ε4 appears to be the driving genetic factor for these associations.

Our study provides a readout of pathway-based risk score association with AD and its predementia endophenotypes. Our findings are important from a clinical perspective as these will aid in determining whether a certain biological pathway is involved in a patient. This will permit targeted interventions based on predicted pathological pathways. Similar to the case of cardiovascular diseases[77], a heterogeneous disease treatment can be followed based on pathway biomarkers (e.g., glucose level, total cholesterol and high-density lipid levels, and liver enzymes in case of cardiovascular disease)[78] but rather on genetic basis. This requires reference pathways and treatment portfolio. In the meantime, the pathway-based genetic risk score will allow stratification of patients in clinical trials based on causal pathways involved in patients. This may improve both the power and efficiency of future clinical and preventive trials.

Our study is a step forward to use known genetic and pathway information for disentangling the mechanisms of AD but it has one major limitation that pathway information is based on known AD variants identified so far. This will further improve in future with improved genetic risk information that can better capture the underlying pathways. Another possible limitation of our study is that 625 cases of RS-I were a part of the meta-analysis performed by the IGAP [7] which can contribute to possible inflation in our results of association of risk score with AD. However, excluding these patients, the results of this study largely remained unchanged.

To conclude, our study provides strong evidence that the *endocytosis* pathway is relevant in the prodromal phase of AD, that is., in subjects with MCI. Furthermore, the pathways including *immune response* and *clathrin/AP2 adaptor complex* pathways may be relevant for brain-related early endophenotypes of AD, such as white matter lesions; this, however, needs further investigation in larger samples. Interestingly, all the observed associations with early AD pathology are shown by *APOE* excluding pathways. Future findings from genomic research will improve the quality of the pathway-specific genetic scores.

Acknowledgments

This study was funded by the Netherlands Organisation for Health Research and Development (ZonMW) as part of the Joint Programming for Neurological Disease (JPND) as part of the PERADES Program (Defining Genetic, Polygenic, and Environmental Risk for Alzheimer's disease, using multiple powerful cohorts, focused Epigenetics and Stem cell metabolomics), Project number 733051021. This work was funded also by the European Union Innovative Medicine Initiative (IMI) programme under grant agreement No. 115975 as part of the Alzheimer's Disease Apolipoprotein Pathology for Treatment Elucidation and Development (ADAPTED, https://www.imi-adapted.eu); and the European Union's Horizon 2020 research and innovation programme as part of the Common mechanisms and pathways in Stroke and Alzheimer's disease CoSTREAM project (www.costream.eu, grant agreement No. 667375); European Union's Horizon 2020 research and innovation programme Marie Skłodowska-Curie Research and Innovation Staff Exchange (RISE) under the grant agreement No. 645740 as part of the Personalized pREvention of Chronic DIseases (PRECeDI) project. This study was also supported by MRC Dementias Platform UK (DPUK) and the UK Dementia Research Institute (UK DRI). The Rotterdam Study is funded by the 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 authors are grateful to the study participants, the staff from the Rotterdam Study, and the participating general practitioners and pharmacists. The generation and management of GWAS genotype data for the Rotterdam Study (RS-I, RS-II, RS-III) was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. The GWAS data sets are supported by the Netherlands Organization of Scientific Research NWO Investments (Project number 175.010.2005.011, 911-03-012), the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), and the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) Netherlands Consortium for Healthy Aging (NCHA), project number 050-060-810. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera, and Marjolein Peters, MSc, and Carolina Medina-Gomez, MSc, for their help in creating the GWAS database, and Karol Estrada, PhD, Yurii Aulchenko, PhD, and Carolina Medina-Gomez, MSc, for the creation and analysis of imputed data. I would like to thanks Alzheimer Nederland for providing me travel grant to present my research at AAIC2017

References

1. Gatz, M. *et al.* Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry* **63**, 168-74 (2006).
2. 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).
3. Bu, G. Apolipoprotein E and its receptors in Alzheimer's disease: pathways, pathogenesis and therapy. *Nat Rev Neurosci* **10**, 333-44 (2009).
4. Huang, Y. & Mucke, L. Alzheimer mechanisms and therapeutic strategies. *Cell* **148**, 1204-22 (2012).
5. Kopke, E. *et al.* Microtubule-associated protein tau. Abnormal phosphorylation of a non-paired helical filament pool in Alzheimer disease. *J Biol Chem* **268**, 24374-84 (1993).
6. Hardy, J.A. & Higgins, G.A. Alzheimer's disease: the amyloid cascade hypothesis. *Science* **256**, 184-5 (1992).
7. Lambert, J.C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* **45**, 1452-8 (2013).
8. Genin, E. *et al.* APOE and Alzheimer disease: a major gene with semi-dominant inheritance. *Mol Psychiatry* **16**, 903-7 (2011).
9. 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).
10. Seshadri, S. *et al.* Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA* **303**, 1832-40 (2010).
11. 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).
12. 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).
13. Desikan, R.S. *et al.* Polygenic Overlap Between C-Reactive Protein, Plasma Lipids, and Alzheimer Disease. *Circulation* **131**, 2061-9 (2015).
14. Guerreiro, R., Bras, J. & Hardy, J. SnapShot: genetics of Alzheimer's disease. *Cell* **155**, 968-968 e1 (2013).
15. Medway, C. & Morgan, K. Review: The genetics of Alzheimer's disease; putting flesh on the bones. *Neuropathol Appl Neurobiol* **40**, 97-105 (2014).
16. International Genomics of Alzheimer's Disease, C. Convergent genetic and expression data implicate immunity in Alzheimer's disease. *Alzheimers Dement* **11**, 658-71 (2015).
17. Van der Flier, W.M. Clinical heterogeneity in familial Alzheimer's disease. *Lancet Neurol* **15**, 1296-1298 (2016).
18. Becker, J.T., Huff, F.J., Nebes, R.D., Holland, A. & Boller, F. Neuropsychological function in Alzheimer's disease. Pattern of impairment and rates of progression. *Arch Neurol* **45**, 263-8 (1988).

19. Mann, U.M., Mohr, E., Gearing, M. & Chase, T.N. Heterogeneity in Alzheimer's disease: progression rate segregated by distinct neuropsychological and cerebral metabolic profiles. *J Neurol Neurosurg Psychiatry* **55**, 956-9 (1992).
20. Dubois, B. *et al.* Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. *Alzheimers Dement* **12**, 292-323 (2016).
21. Debette, S. *et al.* Association of MRI markers of vascular brain injury with incident stroke, mild cognitive impairment, dementia, and mortality: the Framingham Offspring Study. *Stroke* **41**, 600-6 (2010).
22. Debette, S. & Markus, H.S. The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. *BMJ* **341**, c3666 (2010).
23. Kaye, J.A. *et al.* Volume loss of the hippocampus and temporal lobe in healthy elderly persons destined to develop dementia. *Neurology* **48**, 1297-304 (1997).
24. Petersen, R.C. *et al.* Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol* **56**, 303-8 (1999).
25. Gauthier, S. *et al.* Mild cognitive impairment. *Lancet* **367**, 1262-70 (2006).
26. Petersen, R.C. Mild cognitive impairment as a diagnostic entity. *Journal of internal medicine* **256**, 183-194 (2004).
27. Hofman, A. *et al.* The Rotterdam Study: 2016 objectives and design update. *Eur J Epidemiol* **30**, 661-708 (2015).
28. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-83 (2016).
29. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-7 (2016).
30. Delaneau, O., Marchini, J., Genomes Project, C. & Genomes Project, C. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat Commun* **5**, 3934 (2014).
31. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G.R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* **44**, 955-9 (2012).
32. Ikram, M.A. *et al.* The Rotterdam Scan Study: design update 2016 and main findings. *European journal of epidemiology* **30**, 1299-1315 (2015).
33. Vrooman, H.A. *et al.* Multi-spectral brain tissue segmentation using automatically trained k-Nearest-Neighbor classification. *Neuroimage* **37**, 71-81 (2007).
34. De Boer, R. *et al.* White matter lesion extension to automatic brain tissue segmentation on MRI. *Neuroimage* **45**, 1151-1161 (2009).
35. Fischl, B. FreeSurfer. *Neuroimage* **62**, 774-81 (2012).
36. Klein, S., Staring, M. & Pluim, J.P.W. Comparison of gradient approximation techniques for optimisation of mutual information in nonrigid registration. Vol. 5747 192-203 (2005).
37. Ikram, M.A. *et al.* Common variants at 6q22 and 17q21 are associated with intracranial volume. *Nat Genet* **44**, 539-44 (2012).
38. Guerreiro, R. *et al.* TREM2 variants in Alzheimer's disease. *N Engl J Med* **368**, 117-27 (2013).

39. Therneau, T. A Package for Survival Analysis in S. R package version 2.37-7. 2014. *URL* http://CRAN. *R-project. org/package= survival* (2015).
40. Chouraki, V. *et al.* Evaluation of a Genetic Risk Score to Improve Risk Prediction for Alzheimer's Disease. *J Alzheimers Dis* **53**, 921-32 (2016).
41. Escott-Price, V., Shoai, M., Pither, R., Williams, J. & Hardy, J. Polygenic score prediction captures nearly all common genetic risk for Alzheimer's disease. *Neurobiol Aging* (2016).
42. Sleegers, K. *et al.* A 22-single nucleotide polymorphism Alzheimer's disease risk score correlates with family history, onset age, and cerebrospinal fluid Abeta42. *Alzheimers Dement* **11**, 1452-60 (2015).
43. Adams, H.H. *et al.* Genetic risk of neurodegenerative diseases is associated with mild cognitive impairment and conversion to dementia. *Alzheimers Dement* **11**, 1277-85 (2015).
44. Lacour, A. *et al.* Genome-wide significant risk factors for Alzheimer's disease: role in progression to dementia due to Alzheimer's disease among subjects with mild cognitive impairment. *Mol Psychiatry* **22**, 153-160 (2017).
45. Mormino, E.C. *et al.* Polygenic risk of Alzheimer disease is associated with early- and late-life processes. *Neurology* **87**, 481-8 (2016).
46. Lupton, M.K. *et al.* The effect of increased genetic risk for Alzheimer's disease on hippocampal and amygdala volume. *Neurobiol Aging* **40**, 68-77 (2016).
47. Chauhan, G. *et al.* Association of Alzheimer's disease GWAS loci with MRI markers of brain aging. *Neurobiol Aging* **36**, 1765 e7-16 (2015).
48. Zhang, Y. *et al.* An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J Neurosci* **34**, 11929-47 (2014).
49. Kofler, J., Bissel, S., Wiley, C., Stauffer, M. & Murdoch, G. Differential microglial expression of new Alzheimer's disease associated genes MS4A4A and MS4A6A. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* **8**, P253.
50. Colonna, M. TREMs in the immune system and beyond. *Nat Rev Immunol* **3**, 445-53 (2003).
51. Crehan, H., Hardy, J. & Pocock, J. Blockage of CR1 prevents activation of rodent microglia. *Neurobiol Dis* **54**, 139-49 (2013).
52. Zhang, B. *et al.* Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* **153**, 707-20 (2013).
53. Lee, C.Y. & Landreth, G.E. The role of microglia in amyloid clearance from the AD brain. *J Neural Transm (Vienna)* **117**, 949-60 (2010).
54. Heneka, M.T. *et al.* Neuroinflammation in Alzheimer's disease. *Lancet Neurol* **14**, 388-405 (2015).
55. Heppner, F.L., Ransohoff, R.M. & Becher, B. Immune attack: the role of inflammation in Alzheimer disease. *Nat Rev Neurosci* **16**, 358-72 (2015).
56. Jeong, H. *et al.* Age, hypertension, and genetic polymorphisms and their relative associations with white matter hyperintensities in Korean patients with Alzheimer's disease. *Neurology Asia* **20**, 35-41 (2015).
57. Smith, E.E. *et al.* Magnetic resonance imaging white matter hyperintensities and brain volume in the prediction of mild cognitive impairment and dementia. *Arch Neurol* **65**, 94-100 (2008).

58. Wolf, H., Ecke, G.M., Bettin, S., Dietrich, J. & Gertz, H.J. Do white matter changes contribute to the subsequent development of dementia in patients with mild cognitive impairment? A longitudinal study. *Int J Geriatr Psychiatry* **15**, 803-12 (2000).
59. Funderburk, S.F., Marcellino, B.K. & Yue, Z. Cell "self-eating" (autophagy) mechanism in Alzheimer's disease. *Mt Sinai J Med* **77**, 59-68 (2010).
60. Ginsberg, S.D. *et al.* Upregulation of select rab GTPases in cholinergic basal forebrain neurons in mild cognitive impairment and Alzheimer's disease. *J Chem Neuroanat* **42**, 102-10 (2011).
61. Cataldo, A.M. *et al.* Endocytic pathway abnormalities precede amyloid beta deposition in sporadic Alzheimer's disease and Down syndrome: differential effects of APOE genotype and presenilin mutations. *Am J Pathol* **157**, 277-86 (2000).
62. Vance, J.E., Campenot, R.B. & Vance, D.E. The synthesis and transport of lipids for axonal growth and nerve regeneration. *Biochim Biophys Acta* **1486**, 84-96 (2000).
63. Keating, D.J., Chen, C. & Pritchard, M.A. Alzheimer's disease and endocytic dysfunction: clues from the Down syndrome-related proteins, DSCR1 and ITSN1. *Ageing Res Rev* **5**, 388-401 (2006).
64. Svennerholm, L. & Gottfries, C.G. Membrane lipids, selectively diminished in Alzheimer brains, suggest synapse loss as a primary event in early-onset form (type I) and demyelination in late-onset form (type II). *J Neurochem* **62**, 1039-47 (1994).
65. Woodruff, G. *et al.* Defective Transcytosis of APP and Lipoproteins in Human iPSC-Derived Neurons with Familial Alzheimer's Disease Mutations. *Cell Reports* **17**, 759-773 (2016).
66. Maxfield, F.R. Role of endosomes and lysosomes in human disease. *Cold Spring Harb Perspect Biol* **6**, a016931 (2014).
67. Yang, X., Li, J., Liu, B., Li, Y. & Jiang, T. Impact of PICALM and CLU on hippocampal degeneration. *Hum Brain Mapp* **37**, 2419-30 (2016).
68. Tan, L. *et al.* Effect of CLU genetic variants on cerebrospinal fluid and neuroimaging markers in healthy, mild cognitive impairment and Alzheimer's disease cohorts. *Sci Rep* **6**, 26027 (2016).
69. Harel, A., Wu, F., Mattson, M.P., Morris, C.M. & Yao, P.J. Evidence for CALM in directing VAMP2 trafficking. *Traffic* **9**, 417-29 (2008).
70. Yao, P.J. Synaptic frailty and clathrin-mediated synaptic vesicle trafficking in Alzheimer's disease. *Trends in neurosciences* **27**, 24-29 (2004).
71. Baig, S. *et al.* Distribution and expression of picalm in Alzheimer disease. *J Neuropathol Exp Neurol* **69**, 1071-7 (2010).
72. Zlokovic, B.V. Neurovascular pathways to neurodegeneration in Alzheimer's disease and other disorders. *Nat Rev Neurosci* **12**, 723-38 (2011).
73. Thambisetty, M. *et al.* Association of plasma clusterin concentration with severity, pathology, and progression in Alzheimer disease. *Arch Gen Psychiatry* **67**, 739-48 (2010).
74. Weinstein, G. *et al.* Plasma clusterin levels and risk of dementia, Alzheimer's disease, and stroke. *Alzheimers Dement (Amst)* **3**, 103-9 (2016).
75. Braskie, M.N. *et al.* Common Alzheimer's disease risk variant within the CLU gene affects white matter microstructure in young adults. *J Neurosci* **31**, 6764-70 (2011).

76. Sachdev, P.S., Zhuang, L., Braidy, N. & Wen, W. Is Alzheimer's a disease of the white matter? *Curr Opin Psychiatry* **26**, 244-51 (2013).
77. Sun, X. & Jia, Z. A brief review of biomarkers for preventing and treating cardiovascular diseases. *J Cardiovasc Dis Res* **3**, 251-4 (2012).
78. Upadhyay, R.K. Emerging risk biomarkers in cardiovascular diseases and disorders. *J Lipids* **2015**, 971453 (2015).

Chapter 3.2

Genetic predisposition, modifiable risk factor profile and long-term dementia risk in the general population

Silvan Licher, Shahzad Ahmad, Hata Karamujić-Čomić, Trudy Voortman, Maarten J.G. Leening, M. Arfan Ikram and M. Kamran Ikram

This chapter is published in Nature Medicine. 2019 August; 25:1364–1369

Supplementary information for this chapter is available at:
https://doi.org/10.1038/s41591-019-0547-7

Abstract

The exact etiology of dementia is still unclear, but both genetic and lifestyle factors are thought to be key drivers of this complex disease. The recognition of familial patterns of dementia has led to the discovery of genetic factors that have a role in the pathogenesis of dementia, including the apolipoprotein E (*APOE)* genotype and a large and still growing number of genetic variants[1,2]. Beyond genetic architecture, several modifiable risk factors have been implicated in the development of dementia[3]. Prevention trials of measures to halt or delay cognitive decline are increasingly recruiting older individuals who are genetically predisposed to dementia. However, it remains unclear whether targeted health and lifestyle interventions can attenuate or even offset increased genetic risk. Here, we leverage long-term data on both genetic and modifiable risk factors from 6,352 individuals aged 55 years and older within the population-based Rotterdam Study. In this study, we demonstrate that in individuals at low- and intermediate genetic risk, favorable modifiable risk profiles are related to a lower risk of dementia compared to unfavorable profile. In contrast, these protective associations were not found in those at high genetic risk.

Main

Recent analyses have shown that if currently known modifiable risk factors were to be eliminated at a population level, over one-third of all dementia cases could be prevented[3]. In view of these findings, several dementia prevention trials have been conducted to investigate the efficacy of lifestyle interventions, but have yielded inconsistent results so far[4-6]. Such interventions have been proposed to be more effective when targeted at individuals who have an increased risk of dementia, as identified through their genetic or clinical profile, or a combination thereof[7].

The genetic risk of developing dementia or Alzheimer's disease (AD) might be detrimental for some individuals[8], yet this risk may be mitigated for most when adhering to a healthy lifestyle. A recent subgroup analysis from a 2-year multi-domain intervention trial found that healthy lifestyle changes had beneficial effects on cognitive performance, even in *APOE* ε4 carriers, which is the variant of the gene associated with the highest risk of dementia[9]. Evidence from randomized controlled trials is necessary to determine whether positive effects, including those beyond cognitive performance, also occur in dementia. However, since treatment effects from lifestyle interventions at an individual level are generally small, large trials with a long follow-up are needed. Such trials are expensive and prone to high attrition rates. Instead, data from long-term prospective cohort studies can be leveraged to gain insights into the interplay between genetic and lifestyle factors, with the potential to inform the design of future clinical trials.

Prior studies have mostly focused on the risk of dementia associated with an individual protective factor[10,11], yet the combination of multiple factors may yield more beneficial effects than the individual parts[12], Combining data about a number of factors is also important because it takes into account the multi-factorial nature of late life-dementia[13]. We used data from the Rotterdam Study to determine to what extent a favorable profile based on modifiable risk factors is associated with a lower risk of dementia among individuals at low, intermediate or high genetic risk.

In 6,352 participants, we determined genetic risk using two approaches: (1) *APOE* genotype and (2) a weighted polygenic risk score based on 27 genetic variants (excluding *APOE*) that showed genome-wide significant evidence for associations with AD (Supplementary Table 1). For the first approach, we grouped participants into high *APOE*-risk (ε2ε4, ε3ε4 or ε4ε4 genotypes), intermediate-risk (ε3ε3) or low risk (ε2ε2 or ε2ε3); for the second, the distribution of the polygenic risk scores was divided into tertiles, and individuals were grouped into these tertiles. In these participants, we also measured several health and/or lifestyle factors that have been implicated in lower the risk of dementia[11,14]. These include: (1) abstaining from smoking, (2) absence of depression, (3) absence of diabetes, (4) regular physical activity, (5) avoiding social isolation, and (6) adherence to a healthy diet, which included limited alcohol consumption (see Methods for additional information). Using these six factors, we computed an overall score based on modifiable risk factors, ranging from zero to six. A higher score reflects a more favorable

profile. Subsequently, we categorized participants into three groups: an unfavorable profile (a score of ≤2 protective factors); an intermediate profile (a score of 3 or 4 factors); and a favorable profile (a score of ≥5 factors). Alternatively, we determined modifiable risk on the basis of the Ideal Cardiovascular Health score defined using a seven-item tool from the American Heart Association comprising four lifestyle measures (abstaining from smoking and ideal body weight, regular physical activity and diet at optimal levels) and three health measures (optimal plasma cholesterol and glucose levels, and blood pressure at optimal levels—and the 10-year predicted risk of fatal cardiovascular diseases[15]. We subsequently stratified participants by both their genetic and modifiable risk. We calculated the risk of developing dementia for each stratum separately, on both a relative and absolute scale using Cox and competing risk models, respectively.

This study included more women (56.2%) than men. Baseline characteristics were roughly similar across categories of *APOE*-risk (Table 1). As expected, *APOE* ε4 carriers were generally diagnosed with dementia at a younger age ($P = 1.38\times10^{-12}$), more often had a parental history of dementia ($P = 4.55\times10^{-4}$), and had higher total cholesterol levels ($P = 1.24\times10^{-19}$), compared with non-carriers. The median follow-up was 14.1 years among the 6,352 participants; in this timeframe, 915 individuals were diagnosed with dementia, of whom 739 were diagnosed with AD, and 2,644 participants died free from dementia.

Dementia risk was significantly higher among participants at a high or intermediate *APOE* risk compared with those at low *APOE* risk (Figure 1 and Table 2). The risk of dementia also increased in participants who had fewer protective factors (*P* for trend= 0.0044), such that those with an unfavorable profile (≤2 out of 6 factors) had a 32% higher risk of dementia than participants with a favorable one (≥5 out of 6 factors) (Figure 1 and Table 2). The strength of this association

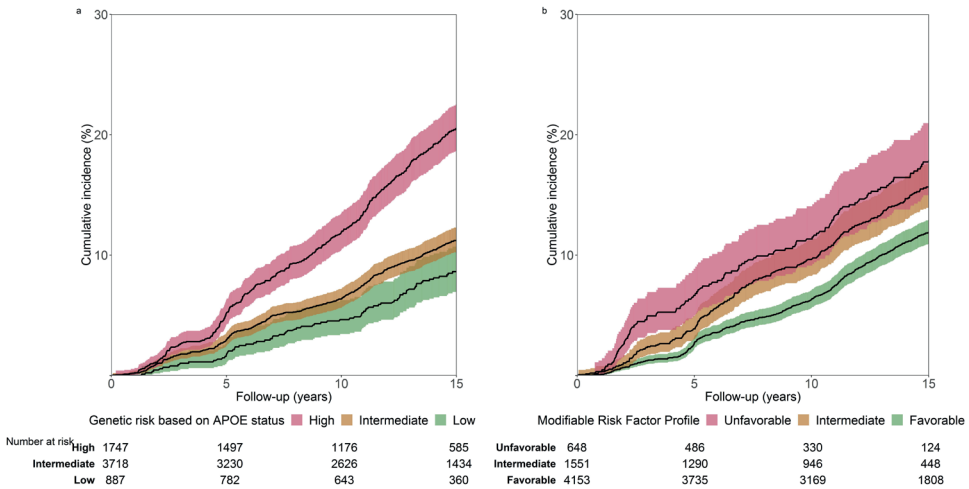


Figure 1: Cumulative incidence of dementia during follow-up. a, according to genetic risk based on *APOE* genotyping, and b, according to modifiable risk factor profiles. For a and b, shaded areas represent 95% confidence intervals.

Table 1: Baseline characteristics per genetic risk category based on *APOE* carrier status

	Low risk (ε2ε2 or ε2ε3) N = 887	Intermediate risk (ε3ε3) N = 3,718	High risk (ε2ε4, ε3ε4 or ε4ε4) N = 1,747	*P* for difference
Age, years	69.4 (8.5)	69.2 (8.3)	68.7 (7.9)	0.042
Women	529 (59.6)	2072 (55.8)	971 (55.5)	0.102
Educational years, median (IQR)	10 (7-13)	10 (7-13)	10 (7-13)	0.325
Parental history of dementia	53 (8.1)	219 (7.8)	155 (11.6)	4.55×10^{-4}
History of stroke	33 (3.7)	138 (3.7)	61 (3.5)	0.929
Body mass index, kg/m^2	27.4 (4.0)	27.0 (4.0)	26.9 (3.9)	0.014
Systolic blood pressure, mmHg	145 (22)	143 (21)	143 (21)	0.045
Diastolic blood pressure, mmHg	77 (11)	77 (12)	77 (11)	0.239
Total cholesterol, mmol/L	5.6 (1.0)	5.8 (1.0)	5.9 (1.0)	1.24×10^{-19}
High-density lipoprotein cholesterol, mmol/L	1.43 (0.4)	1.39 (0.4)	1.35 (0.4)	1.00×10^{-5}
Fasting glucose, mmol/L	5.6 (1.5)	6.0 (1.6)	6.0 (1.6)	0.902
Baseline MMSE score, median (IQR)	28 (27-29)	28 (27-29)	28 (27-29)	0.049
Age of dementia diagnosis	85.5 (5.9)	84.1 (6.3)	81.3 (6.5)	1.38×10^{-12}
Modifiable health and lifestyle factors				
No current smoking	714 (81.0)	2961 (80.1)	1389 (79.9)	0.801
Absence of depression	799 (90.1)	3351 (90.1)	1567 (89.7)	0.881
Absence of diabetes	744 (88.3)	3096 (88.3)	1461 (89.0)	0.750
Regular physical activity	484 (56.0)	2109 (58.5)	1005 (59.4)	0.501
Absence of social isolation	588 (66.7)	2664 (72.4)	1256 (72.1)	0.005
Adherence to a healthy diet	141 (15.9)	559 (15.0)	252 (14.4)	0.649
Modifiable risk profile category				
Favorable: 5-6 health and lifestyle factors	568 (64.0)	2453 (66.0)	1132 (64.8)	0.648
Intermediate: 3-4 health and lifestyle factors	224 (25.2)	884 (23.8)	443 (25.4)	
Unfavorable: 0-2 health and lifestyle factors	95 (10.7)	381 (10.2)	172 (9.8)	

Abbreviations: N, number of individuals at risk; IQR, interquartile range; MMSE, Mini-Mental State Examination.

Data are presented as the frequency in the stratum (percentage) for categorical values and as mean±the standard deviation for continuous variables unless indicated otherwise. We compared baseline characteristics across *APOE* strata using analysis of variance (ANOVA) tests. In the case of frequency distributions or when data were non-normally distributed, we compared variables between groups using non-parametric tests (chi-squared, Mann–Whitney or Kruskal–Wallis). Two-sided *P* values were uncorrected for multiple testing.

Table 2: Risk of incident dementia according to *APOE*-related risk and lifestyle categories

***APOE*-related risk**	**N/n**	**Model 1**	**Model 2**
		HR (95% CI)	**HR (95% CI)**
Low risk (ε2ε2/ ε2ε3)	887/85	Reference	Reference
Intermediate risk (ε3ε3)	3718/456	1.45 (1.15;1.83)	1.45 (1.15;1.83)
High risk (ε2ε4/ε3ε4/ ε4ε4)	1747/374	3.02 (2.38;3.82)	3.02 (2.38;3.83)
P for trend		2.10×10^{-30}	1.87×10^{-30}
Lifestyle risk category			
Favorable	4153/538	Reference	Reference
Intermediate	1551/259	1.15 (0.98;1.34)	1.14 (0.98;1.33)
Unfavorable	648/118	1.32 (1.08;1.63)	1.29 (1.05;1.59)
P for trend		0.0044	0.0087

Abbreviations: N, number of individuals at risk; n, number of dementia cases during follow-up; HR, hazard ratio; CI, confidence interval
Model 1 - adjusted for: age, sex and education
Model 2 - additionally adjusted for: parental history of dementia, history of stroke, systolic blood pressure, total and high-density lipoprotein cholesterol

remained nearly identical after adjusting for parental history of dementia and cardiovascular risk factors (hazard ratio (HR) = 1.29 (95% confidence interval (CI) = 1.05–1.59), *P* for trend = 0.0087).

APOE genotype significantly modified the association between protective factors and dementia (*P* for interaction = 0.023). In those at low or intermediate *APOE* risk, the risk of dementia in participants with an unfavorable profile was higher than that for those with a favorable one (HR = 2.51, 95% CI = 1.40–4.48 and HR = 1.39, CI = 1.04–1.85, respectively; Table 3). In those at high *APOE* risk, we did not find differences in dementia risk between individuals with unfavorable or intermediate profiles compared with those who had a favorable profile (HR = 1.00, 95% CI = 0.79–1.28 and HR = 1.05, 95% CI = 0.73–1.50, respectively).

Among those at low *APOE*-risk, the mean anticipated absolute risk of developing dementia within 15-year ranged from 32.1% (95% CI = 0.0-59.9) for those with an unfavorable profile to 12.6% (95% CI = 4.5-26.8) for those with a favorable one (Supplementary Table 2). Individuals at intermediate *APOE*-risk with an unfavorable profile had a 22.0% (95% CI = 8.3-39.2) anticipated risk, which was 13.5% (95% CI = 8.9-15.6) for those with a favorable profile. Among participants at high *APOE* risk, the anticipated 15-year risk of dementia remained largely unchanged across the different profiles (ranging from 18.2% for a favorable to 19.5% for an unfavorable profile).

Table 3: Risk of incident dementia while stratifying participants on both their *APOE*-related risk and modifiable risk factor profile

APOE-related risk	Risk factor profile	N/n	HR (95% CI)
Low (ε2ε2/ε2ε3)	Favorable	568/44	Reference
	Intermediate	224/23	1.14 (0.66;1.96)
	Unfavorable	95/18	2.51 (1.40;4.48)
	P for trend		0.0059
Intermediate (ε3ε3)	Favorable	2453/253	Reference
	Intermediate	884/139	1.27 (1.02;1.57)
	Unfavorable	381/64	1.39 (1.04;1.85)
	P for trend		0.0087
High (ε2ε4/ε3ε4/ ε4ε4)	Favorable	1132/241	Reference
	Intermediate	443/97	1.00 (0.79;1.28)
	Unfavorable	172/36	1.05 (0.73;1.50)
	P for trend		0.8300

Abbreviations: n, number of cases; N, number of people at risk; HR, hazard ratio; CI, confidence interval
Adjusted for: age, sex and education

Stratified analyses showed that protective associations of favorable risk profiles against dementia tended to be stronger in younger individuals than in older individuals, and were most pronounced for younger individuals at low *APOE*-risk (Supplementary Tables 3 and 4). In all of these analyses, no significant protective associations were found in *APOE* ε4 carriers. There was no effect modification of the association between risk profiles and dementia risk by sex (Supplementary Table 5).

In sensitivity analyses, using a different approach, namely a polygenic risk score for AD (without *APOE*), we also found that associations between protective factors were modified (*P* for interaction = 0.0003); patterns across polygenic risk strata were attenuated yet largely comparable to those of the *APOE* (Supplementary Tables 6 and7).

These patterns also remained unchanged when we varied the composition of modifiable risk factors. For instance, similar results were found when we stratified participants on their Ideal Cardiovascular Health score (*P* for interaction = 0.026, Supplementary Table 8), and when we stratified participants on the basis of their predicted absolute 10-year risk of fatal cardiovascular diseases using the SCORE equation (European Coronary Risk Equation, including age, sex, current smoking, level of total cholesterol and systolic blood pressure) (*P* for interaction = 7.82×10^{-5}, Supplementary Table 9). All of the individual health and lifestyle factor-specific associations with dementia risk that were included in the different profiles are presented in Supplementary Tables 10-12, respectively.

Most evidence on the interaction between specific genetic and modifiable factors with dementia comes from observational studies. These studies primarily examined the associations and interactions of single health or lifestyle factors, such as diabetes, physical activity, alcohol use, smoking or diet, with different *APOE* alleles. Most of these studies reported a lower risk of dementia among individuals[16-19], who had optimal levels of one of these single risk factors[16-19] compared with those who had more adverse ones. These effects were generally more pronounced in *APOE* ε4 carriers during midlife[16-19]. In contrast, studies conducted in older individuals (aged ³60 years), primarily found beneficial effects of these factors on risk of dementia among non-carriers[20-26], or reported no interaction[27,28].

More evidence comes from the Cardiovascular Risk Factors, Aging, and Incidence of Dementia (CAIDE) Study of middle-aged individuals that took multiple protective factors into consideration[29]. *APOE* genotype modified the association between several lifestyle factors and the risk of dementia. The results from CAIDE indicate that *APOE* ε4 carriers are particularly prone to developing dementia if they have hazardous health and lifestyle factors during midlife. The participants in CAIDE were younger (mean age 50.6 years) than the individuals in this study (mean age 69.1 years). This may have led to survival bias in the current study because dementia risk in older *APOE* ε4 carriers is potentially less affected by modifiable risk factors later in life.

To our knowledge, only the Prevention of Dementia by Intensive Vascular Care (preDIVA) trial has assessed the effects of health and lifestyle interventions on dementia[16]. This trial showed that intensive vascular care management in a primary-care setting had no overall benefit on dementia incidence. In a subgroup analysis in this trial, no significant differences were observed between *APOE* ε4 carriers and non-carriers. The Finnish Geriatric Intervention Study to Prevent Cognitive Impairment and Disability (FINGER) trial assessed the effects of multiple lifestyle interventions on cognitive performance in older individuals[4]. A pre-specified subgroup analysis in this trial uncovered similar beneficial effects on cognitive performance in both *APOE* ε4 carriers and non-carriers after two years of follow-up[9].

In the current study, we aimed to complement evidence from these clinical trials with long-term observational data. Such an approach has been previously undertaken to study potential interaction between genetic and modifiable factors and other chronic diseases, such as heart disease and stroke[30,31]. Our results confirm that individuals with a favorable profile have a lower risk of dementia than those with intermediate or unfavorable profiles based on modifiable risk factors. In contrast with the FINGER subgroup analysis, this study found that a favorable profile could not offset high *APOE*-risk. The FINGER trial intervened in multiple lifestyle factors simultaneously, whereas, in our observational study, data on health and lifestyle factors were used to establish risk factor profiles. Non-differential misclassification may, in part, have led to an underestimation of the benefits of a favorable risk factor profile in our study. This may, for instance, apply to physical activity and diet quality because we have chosen cut-off

points to depict 'favorable', 'intermediate' or 'unfavorable' for these variables on the basis of guideline recommendations[32,33]. Moreover, we lacked validated questionnaires to measure social engagement, so we may not have fully captured its beneficial effect on dementia. Misclassification may have occurred in genetic risk stratifications because we did not have enough participants to divide them into strata based on specific genotype. We therefore had to collate some categories of genetic risk to estimate meaningful effect sizes. As an example, we grouped individuals who are either heterozygous or homozygous for *APOE* ε4. This may have led to an underestimation of the benefits of a favorable profile for individuals who are heterozygous for *APOE* ε4, as this group resembles an intermediate risk group between those who are homozygous for *APOE* ε3 and for *APOE* ε4[1].

Furthermore, we studied the interplay between genetic and lifestyle factors in the long-term risk of developing dementia, whereas the FINGER subgroup analysis assessed effects on cognitive performance after a 2-year follow-up. Although cognitive performance was improved in the group that received the multi-domain lifestyle intervention compared with that of the control group, which received standard health advice, in the short-term, it remains questionable whether such effects also occur in the longterm. For instance, participants in the FINGER trial may already have developed essential *APOE*-related brain changes earlier in life[34], making them vulnerable to development of dementia later in life, irrespective of their modifiable risk factor profile.

In this study, a high risk of developing dementia based on *APOE* carrier status was not offset by a favorable profile. These findings contrast with those from other large, observational population-based studies that examined the interaction between genetic and modifiable factors for other chronic diseases, including for instance heart disease and stroke[30,31], These studies found protective associations of favorable modifiable profiles across levels of genetic risk, even for those at the highest genetic risk. Several reasons may underlie this discrepancy.

First, the harmful effects of *APOE* ε4 on cholesterol metabolism are apparent throughout life, and there are cumulative effects on dementia risk as age advances[1]. Second, *APOE* ε4 alters neuronal functioning which may lead to irreversible neuronal cell loss[35]. With advancing age, these effects build up and may, in the absence of disease-modifying drugs and proven preventive strategies, ultimately have a more detrimental effect on the risk of dementia in older individuals. Third, the risk for competing diseases at older ages, such as coronary heart disease and stroke, may be mitigated or even reversed, by having a favorable risk profile through the reduction of atherosclerotic disease[36,37]. Fourth, potential epigenetic changes, such as methylation effects of *APOE* or additional variants, may be age-dependent and exert their effects in midlife or even earlier but this notion deserves further study. Finally, the *APOE* ε4 allele triggers cascades that may be more independent of the profiles that we studied, for example pathways in inflammatory response. Such a response may lead to blood-barrier breakdown, which in turn causes

neurovascular dysfunction[38]. In summary, the interaction between genetic and environmental factors plays an important role in the pathophysiology of dementia.

Our findings provide a less optimistic outlook for individuals at high genetic risk of dementia, yet they may have important implications for the design of future clinical trials. Considering the earlier age at onset of dementia among *APOE* ε4 carriers compared with non-carriers, our results imply that these individuals need to be targeted earlier in the disease process (for example midlife or even earlier) to influence their risk[1]. Additionally, other interventions beyond lifestyle improvements warrant further study. For instance, drugs that lower lipid levels might be considered to lower dementia risk in these individuals, yet evidence for such interventions is still inconclusive[39]. On the positive side, results from this study show that avoiding an unhealthy lifestyle could potentially prevent or postpone the onset of dementia in most individuals in the population (73%), namely those at low and intermediate genetic risk. Among those, the majority were categorized as having a favorable profile (66%), yet room for improvement is still substantial because potential relative risk reductions of up to 30% can be achieved when individuals adhere to the lifestyle factors that confer a favorable risk profile.

Several limitations of this study need to be addressed. First, we lacked data waves on hearing impairment, a potentially important modifiable risk factor for dementia, because assessments to measure hearing were implemented in the study protocol from 2011 onwards. Second, the components used to compute the modifiable risk factor profile were measured at baseline, which does not capture the possibility of shifting from a more adverse risk profile to a more optimal one during follow-up, or vice versa. Third, by stratifying participants using both genetic and environmental information, results are based on small samples in each stratum, resulting in wide confidence intervals around point estimates. Results of this single-cohort study therefore warrant replication in other population-based studies. Nevertheless, we were able to show the robustness of our findings in several sensitivity analyses; thus, it was less likely that our findings were a result of chance. Finally, members of this older population are predominantly of European descent (97%), limiting the generalizability of these findings to younger populations and to other ethnicities. Strengths of this study include the availability of genetic data in combination with the meticulous assessment of several health and lifestyle factors, along with long-term and consistent dementia follow-up.

In conclusion, this large population-based study demonstrates that among those at low- and intermediate genetic risk, a favorable modifiable risk profile is related to a risk of dementia lower than that in individuals with an unfavorable one. In contrast, these protective associations were not found among those at high genetic risk. These results may inform clinical trial design, because dementia prevention trials increasingly recruit individuals genetically predisposed to dementia.

Acknowledgments

We acknowledge the dedication, commitment, and contribution of inhabitants, general practitioners, and pharmacists of the Ommoord district who took part in the Rotterdam Study. We acknowledge Frank van Rooij as data manager, and Brenda C.T. Leening-Kieboom as study coordinator. We thank Jolande Verkroost-van Heemst, for her invaluable contribution to the collection of the 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. Further support was obtained from the Netherlands Consortium for Healthy Ageing and the Dutch Heart Foundation (2012T008) and the Dutch Cancer Society (NKI-20157737). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (project: ORACLE, grant agreement No: 678543).

References

1. van der Lee, S.J. *et al.* The effect of APOE and other common genetic variants on the onset of Alzheimer's disease and dementia: a community-based cohort study. *Lancet Neurol* **17**, 434-444 (2018).
2. Kunkle, B.W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet* **51**, 414-430 (2019).
3. de Bruijn, R.F. *et al.* The potential for prevention of dementia across two decades: the prospective, population-based Rotterdam Study. *BMC Med* **13**, 132 (2015).
4. Ngandu, T. *et al.* A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial. *Lancet* **385**, 2255-63 (2015).
5. Andrieu, S. *et al.* Effect of long-term omega 3 polyunsaturated fatty acid supplementation with or without multidomain intervention on cognitive function in elderly adults with memory complaints (MAPT): a randomised, placebo-controlled trial. *Lancet Neurol* **16**, 377-389 (2017).
6. Moll van Charante, E.P. *et al.* Effectiveness of a 6-year multidomain vascular care intervention to prevent dementia (preDIVA): a cluster-randomised controlled trial. *Lancet* **388**, 797-805 (2016).
7. Licher, S. *et al.* Development and Validation of a Dementia Risk Prediction Model in the General Population: An Analysis of Three Longitudinal Studies. *Am J Psychiatry*, appiajp201818050566 (2018).
8. Guerreiro, R., Bras, J. & Hardy, J. SnapShot: genetics of Alzheimer's disease. *Cell* **155**, 968-968 e1 (2013).
9. Solomon, A. *et al.* Effect of the Apolipoprotein E Genotype on Cognitive Change During a Multidomain Lifestyle Intervention: A Subgroup Analysis of a Randomized Clinical Trial. *JAMA Neurol* (2018).
10. Baumgart, M. *et al.* Summary of the evidence on modifiable risk factors for cognitive decline and dementia: A population-based perspective. *Alzheimers Dement* **11**, 718-26 (2015).
11. Livingston, G. *et al.* Dementia prevention, intervention, and care. *Lancet* **390**, 2673-2734 (2017).
12. Peters, R. *et al.* Combining modifiable risk factors and risk of dementia: a systematic review and meta-analysis. *BMJ Open* **9**, e022846 (2019).
13. Boyle, P.A. *et al.* Attributable risk of Alzheimer's dementia attributed to age-related neuropathologies. *Ann Neurol* **85**, 114-124 (2019).
14. Organization, W.H. Risk reduction of cognitive decline and dementia: WHO guidelines. Vol. CC BY-NC-SA 3.0 IGO (WHO, Geneva, 2019).
15. Conroy, R.M. *et al.* Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J* **24**, 987-1003 (2003).
16. Rovio, S. *et al.* Leisure-time physical activity at midlife and the risk of dementia and Alzheimer's disease. *Lancet Neurol* **4**, 705-11 (2005).
17. Anttila, T. *et al.* Alcohol drinking in middle age and subsequent risk of mild cognitive impairment and dementia in old age: a prospective population based study. *BMJ* **329**, 539 (2004).

18. Laitinen, M.H. *et al.* Fat intake at midlife and risk of dementia and Alzheimer's disease: a population-based study. *Dement Geriatr Cogn Disord* **22**, 99-107 (2006).
19. Karlsson, I.K. *et al.* Apolipoprotein E epsilon4 genotype and the temporal relationship between depression and dementia. *Neurobiol Aging* **36**, 1751-1756 (2015).
20. Podewils, L.J. *et al.* Physical activity, APOE genotype, and dementia risk: findings from the Cardiovascular Health Cognition Study. *Am J Epidemiol* **161**, 639-51 (2005).
21. Huang, T.L. *et al.* Benefits of fatty fish on dementia risk are stronger for those without APOE epsilon4. *Neurology* **65**, 1409-14 (2005).
22. Barberger-Gateau, P. *et al.* Dietary patterns and risk of dementia: the Three-City cohort study. *Neurology* **69**, 1921-30 (2007).
23. Luchsinger, J.A., Tang, M.X., Siddiqui, M., Shea, S. & Mayeux, R. Alcohol intake and risk of dementia. *J Am Geriatr Soc* **52**, 540-6 (2004).
24. Merchant, C. *et al.* The influence of smoking on the risk of Alzheimer's disease. *Neurology* **52**, 1408-12 (1999).
25. Ott, A. *et al.* Smoking and risk of dementia and Alzheimer's disease in a population-based cohort study: the Rotterdam Study. *Lancet* **351**, 1840-3 (1998).
26. Lindsay, J. *et al.* Risk factors for Alzheimer's disease: a prospective analysis from the Canadian Study of Health and Aging. *Am J Epidemiol* **156**, 445-53 (2002).
27. Ritchie, K. *et al.* Designing prevention programmes to reduce incidence of dementia: prospective cohort study of modifiable risk factors. *BMJ* **341**, c3885 (2010).
28. Samieri, C. *et al.* Association of Cardiovascular Health Level in Older Age With Cognitive Decline and Incident Dementia. *JAMA* **320**, 657-664 (2018).
29. Kivipelto, M. *et al.* Apolipoprotein E epsilon4 magnifies lifestyle risks for dementia: a population-based study. *J Cell Mol Med* **12**, 2762-71 (2008).
30. Khera, A.V. *et al.* Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. *N Engl J Med* **375**, 2349-2358 (2016).
31. Rutten-Jacobs, L.C. *et al.* Genetic risk, incident stroke, and the benefits of adhering to a healthy lifestyle: cohort study of 306 473 UK Biobank participants. *BMJ* **363**, k4168 (2018).
32. Voortman, T. *et al.* Adherence to the 2015 Dutch dietary guidelines and risk of non-communicable diseases and mortality in the Rotterdam Study. *Eur J Epidemiol* **32**, 993-1005 (2017).
33. Craig, C.L. *et al.* International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc* **35**, 1381-95 (2003).
34. Chang, L. *et al.* Gray matter maturation and cognition in children with different APOE epsilon genotypes. *Neurology* **87**, 585-94 (2016).
35. Kim, J., Basak, J.M. & Holtzman, D.M. The role of apolipoprotein E in Alzheimer's disease. *Neuron* **63**, 287-303 (2009).
36. Ornish, D. *et al.* Can lifestyle changes reverse coronary heart disease? The Lifestyle Heart Trial. *Lancet* **336**, 129-33 (1990).

37. Licher, S. *et al.* Lifetime risk and multimorbidity of non-communicable diseases and disease-free life expectancy in the general population: A population-based cohort study. *PLoS Med* **16**, e1002741 (2019).
38. Bell, R.D. *et al.* Apolipoprotein E controls cerebrovascular integrity via cyclophilin A. *Nature* **485**, 512-6 (2012).
39. McGuinness, B., Craig, D., Bullock, R. & Passmore, P. Statins for the prevention of dementia. *Cochrane Database Syst Rev*, CD003160 (2016).
40. Ikram, M.A. *et al.* The Rotterdam Study: 2018 update on objectives, design and main results. *Eur J Epidemiol* **32**, 807-850 (2017).

Methods

Study population

We used data from participants of the Rotterdam Study, a prospective population-based cohort study. In 1990, all residents aged 55 and older living in Ommoord, a district of Rotterdam, the Netherlands, were invited. Of 10215 invited inhabitants, 7983 (78%) agreed to participate in the baseline examination. In 2000, the cohort was extended: all residents who turned 55 or moved into the research area. Of the 4472 invitees, 3011 (67%) agreed to participate. Follow-up examinations take place every 3 to 4 years[40].

Analyses of this study are based on data obtained from the third examination of the original cohort in 1997-1999 (N=4797) and the first examination of the extended cohort in 2000-2001 (N=3011). These two cohorts were similar in design and examinations were identical. After excluding participants who did not complete the interview and research center visit in these rounds (N=873), had dementia or insufficient screening for dementia at baseline (N=170), did not undergo genotyping (N=365), did not provide informed consent to access medical records and hospital discharge letters (N=33), or if there was no follow-up due to logistic reasons (N=15), 6352 participants were included for analysis in this study (study flowchart displayed in Supplementary Figure 1. A comparison of baseline characteristics for in- and excluded participants is presented in Supplementary Table 13).

Ethics statement

The Rotterdam Study has medical ethics committee approval per the Population Study Act: Rotterdam Study, executed by the Ministry of Health, Welfare and Sport of the Netherlands. Written informed consent was obtained from all participants.

APOE genotyping

DNA was extracted from blood samples drawn at baseline. *APOE* genotype was determined using a polymerase chain reaction in the original cohort and was determined with a bi-allelic TaqMan assay (rs7412 and rs429358) in the extended cohort on coded DNA samples. *APOE* ε4 carrier status was defined as carrier of one or two ε4 alleles.

Calculation of a polygenic risk score

The majority of samples (81.1%) were further genotyped with the Illumina 610K and 660K chips and imputed to the Haplotype Reference Consortium reference panel (version 1.0) with Minimac 3. We included 27 genetic variants that showed genome-wide significant evidence for associations with AD to calculate a weighted polygenic risk score. Supplementary Table 1 provides an overview of the included variants. The polygenic risk score was calculated as the sum of the products of single nucleotide polymorphism dosages of the 27 genetic variants (excluding

APOE) and their respective reported effect estimates. All 27 variants selected for the calculation of the polygenic risk score were well imputed (imputation score R2 > 0.6, median 0.99).

Modifiable risk factor profile

We adapted six health and/or lifestyle factors shown to be important during later life to lower dementia risk, as set out by a recent meta-analysis and endorsed by the World Health Organization (WHO)[11,14]. Among these are: (1) abstaining from smoking, (2) absence of depression, (3) absence of diabetes, (4) regular physical activity, (5) avoiding social isolation, and (6) adherence to a healthy dietary pattern, which included limited alcohol consumption. Based on these six factors, we computed an overall profile of modifiable risk factors ranging from zero to six, and we subsequently grouped participants into three categories of modifiable risk (unfavorable-: ≤2, intermediate-: 3-4, and favorable profiles: ≥5 protective factors)

Assessment of individual health and lifestyle factors

During a structured home interview, participants were enquired about their smoking habits. Participants were classified as never, former or, current smokers. In addition, the participants were screened with the Center for Epidemiologic Studies Depression Scale during the interview. Presence of depressive symptoms was defined as a score of >16 points on scale of 0 to 60. Diabetes was defined as fasting serum glucose levels ≥7.0 mmol/L, non-fasting serum glucose levels ≥11.0 mmol/L (if fasting samples were unavailable), and/or the use of blood glucose-lowering medication. Physical activity levels were assessed using a validated adapted version of Zutphen Physical Activity Questionnaire and expressed in Metabolic Equivalent of Task hours (METhours) per week. The METhours/week are the product of MET-values of specific activities (walking, cycling, domestic work, sports, and gardening) with time in hours per week spent in that activity. Categories were calculated based on International Physical Activity Questionnaire and also expressed in METhours/week. We defined being physically active based on a minimum of ≥40 minutes of exercise per week with a MET intensity of ≥4[33]. Social engagement was constructed using three domains based on various questionnaires. We included marital status, living arrangements (living alone, with spouse or with others) and asked if the participant felt lonely during the past week. If a participant lived alone and felt lonely during the past week – we considered them as being less socially engaged. A validated food frequency questionnaire was used to measure the dietary pattern of participants[32]. A healthy dietary pattern was ascertained on the basis of adherence to at least half of the following dietary guidelines: consumption of an sufficient amount of fruits, nuts, vegetables, whole grains, fish, and dairy products and a limited amount of refined grains, processed meats, unprocessed red meats, sugar-sweetened beverages, trans fats, sodium and alcohol, for which further details and cut-off values are described elsewhere[32].

Other covariates

Participants were questioned about parental history of dementia during the interview. During the center visit, blood pressure was assessed at the right upper arm with the participant in sitting

position. The mean of two measurements was used in the analyses. Serum total cholesterol, and high-density lipoprotein cholesterol were acquired by an automated enzymatic procedure (Boehringer Mannheim System). Glucose levels were measured after overnight fasting (8–14 h). The history of clinical stroke was assessed by both self-report and continuous monitoring of medical records through digitized linkage of files from general practitioners with the study database. All strokes were adjudicated by a panel of study physicians.

Ascertainment of dementia

Participants were screened for dementia at baseline and subsequent center visits with the Mini-Mental State Examination and the Geriatric Mental Schedule organic level[3]. Those with a Mini-Mental State Examination score <26 or Geriatric Mental Schedule score >0 underwent further investigation and informant interview, including the Cambridge Examination for Mental Disorders of the Elderly. All participants also underwent routine cognitive assessment. In addition, the entire cohort was continuously under surveillance for dementia through electronic linkage of the study database with medical records from general practitioners and the regional institute for outpatient mental health care. Available information on cognitive testing and clinical neuroimaging was used when required for diagnosis of dementia subtype. An event adjudication panel led by a consultant neurologist established the final diagnosis according to standard criteria for dementia (DSM-III-R) and AD (NINCDS–ADRDA). Follow-up until 1st of January 2016 was virtually complete (95.5% of potential person-years). Within this period, participants were followed until the date of dementia diagnosis, death, loss to follow-up, or 1st of January 2016, whichever came first.

Statistical analysis

We used Cox proportional hazard models to assess the association of *APOE*-risk and the health and lifestyle factors with incident dementia. We verified the proportionality assumption with use of Schoenfeld residuals. We tested for interaction between *APOE* carrier status and the level of the modifiable risk factor profiles on a multiplicative scale. We subsequently evaluated the hazard ratios (HRs) for participants with a high *APOE* risk status (ε2ε4, ε3ε4 or ε4ε4 genotypes), and compared those with hazard ratios from those with an intermediate risk (ε3ε3 genotype) or low risk status (ε2ε2 or ε2ε3 genotypes). Similarly, we calculated hazard ratios for participants with a favorable profile (which was defined as the presence of at least five of the six health and/or lifestyle factors) with an intermediate profile (three or four factors) or an unfavorable profile (two or less factors). Models were adjusted for age, sex and level of attained education. In extended models, we additionally adjusted for parental history of dementia and cardiovascular risk factors. Finally, we used a competing risk framework based on the Fine & Gray model to calculate the 15-year absolute dementia risk for participants within each category of genetic risk and modifiable risk profile separately. Confidence intervals were computed based on 1000 bootstrap samples.

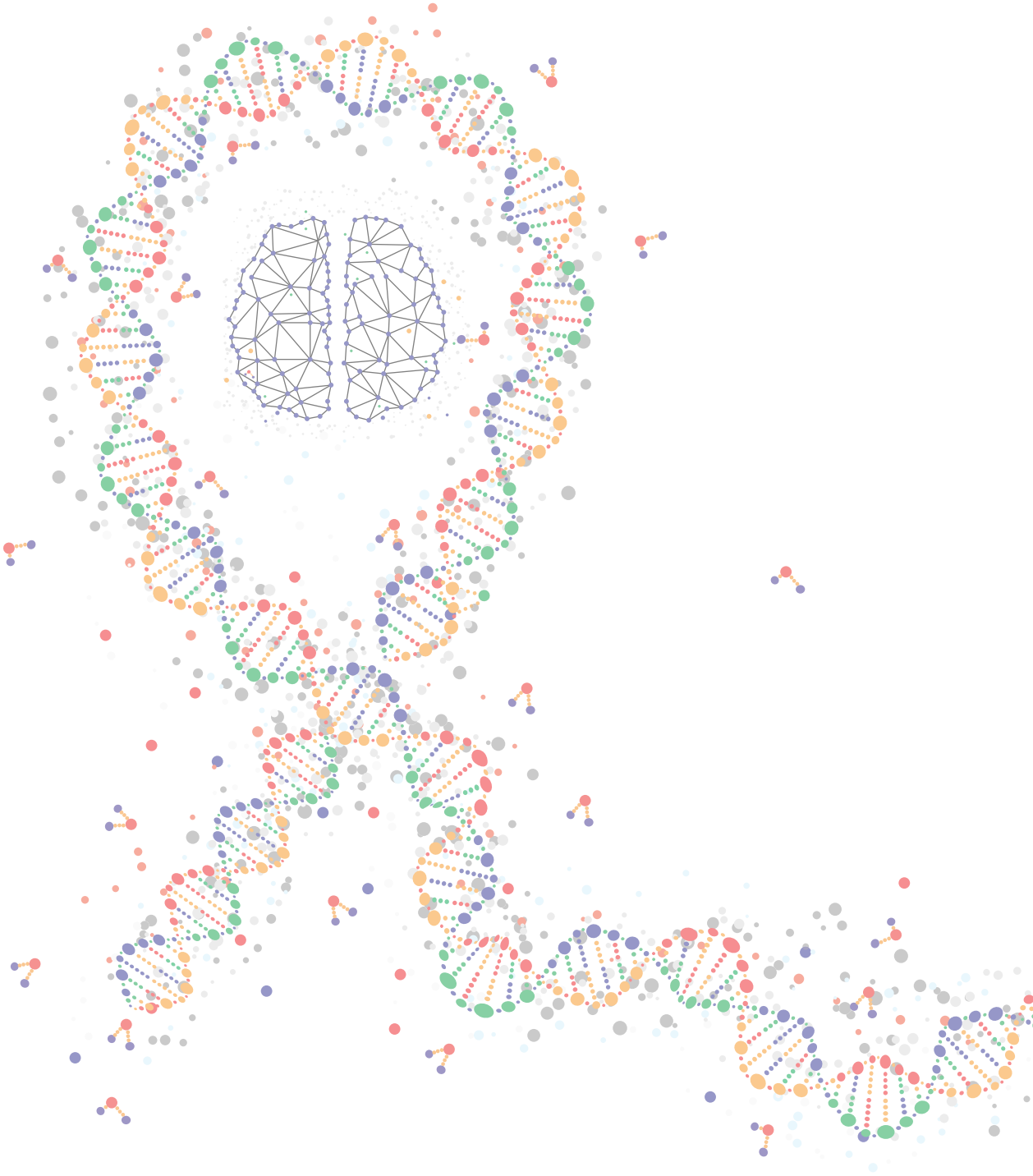
In stratified analyses, we explored whether associations differed among younger and older participants by stratifying on the median age of this study (68.2 years) and additionally on the age of 70 years as this age range is often used as an eligibility criterion to recruit individuals for preventative trials that assess multi-domain lifestyle interventions[4-6]. Finally, we stratified on sex.

In sensitivity analyses, we studied the robustness of our findings by varying the definitions and compositions of both modifiable risk factors and genetic risk factors that were used in the main analyses. This is in particular important since some of the included modifiable risk factors, such as depression and social isolation, may have been altered by pre-clinical dementia. Similar to the main analyses, we first explored statistical interaction on a multiplicative scale between the studied genetic component (*APOE* or the polygenic score), and the risk factor profile under study. We subsequently repeated the main analyses while categorizing participants on the Ideal Cardiovascular Health metric (favorable, intermediate and unfavorable), instead of the currently employed modifiable risk factors. Similarly, we replaced the current modifiable risk factors by the 10-year predicted absolute risk of fatal cardiovascular disease and subsequently categorized participants into low-to-moderate <5%, high risk 5-10%, and very high risk > 10%, based on the European Coronary Risk Equation (SCORE), that includes age, sex, and several adverse risk factors namely current smoking, level of cholesterol and systolic blood pressure[15].

Regarding genetic risk factors, we repeated the main analyses stratified for an AD polygenic risk score that included 27 genome-wide significant variants (excluding *APOE*), comparing participants at high genetic risk (i.e. highest tertile of the polygenic score) with those at intermediate risk (middle tertile), or low risk (lowest tertile).

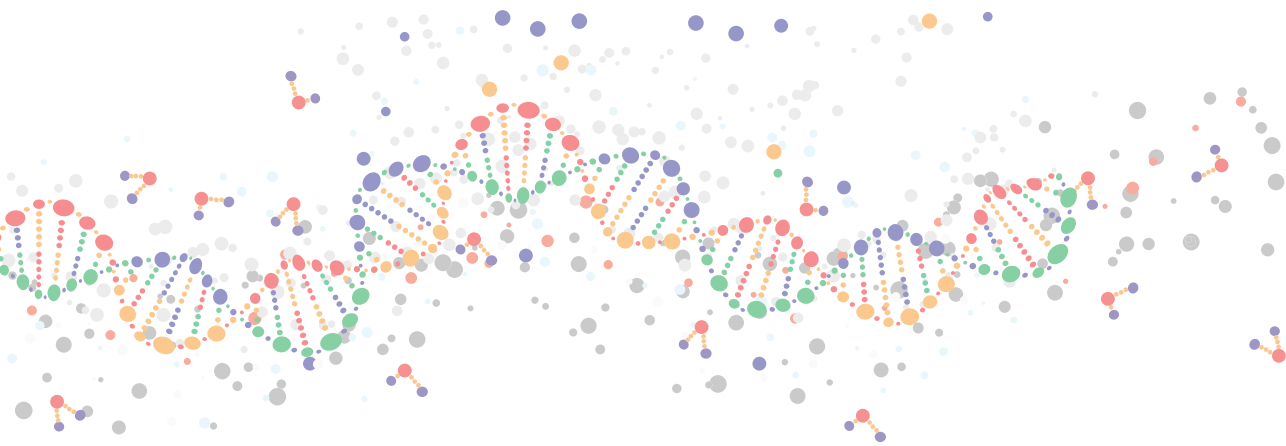
We compared baseline characteristics across *APOE* strata using analysis of variance (anova) tests. In the case of frequency distributions or when data were non-normally distributed, we compared variables between groups using non-parametric tests (chi-square, Mann-Whitney or Kruskal Wallis).

Data were handled and analyzed with SPSS Statistics version 24.0.0.1 (IBM Corp., Armonk, NY) and R, CRAN version 3.5.1, with packages survival, rms and cmprsk. All analyses were performed at the significance level of 0.05 (2-tailed). P values were uncorrected for multiple testing.



Chapter 4

Proteomics and metabolomics of Alzheimer's disease



Chapter 4.1

CDH6 and HAGH protein levels in plasma associate with Alzheimer's disease in *APOE* ε4 carriers

Shahzad Ahmad, Marta del Campo Milan, Oskar Hansson, Ayse Demirkan, Ruiz Agustin, Maria E Sáez, Nikolaos Giagtzoglou, Alfredo Cabrera-Socorro, Margot H.M. Bakker, Alfredo Ramirez, Thomas Hankemeier, Erik Stomrud, Niklas Mattsson-Carlgren, Philip Scheltens, Wiesje M. van der Flier, M. Arfan Ikram, Anders Malarstig, Charlotte E Teunissen, Najaf Amin, Cornelia M. van Duijn

This chapter is published in Scientific Reports. 2020 May; 10:8233

The Supplementary information for this chapter is available at: https://doi.org/10.1038/s41598-020-65038-5

Abstract

Many Alzheimer's disease (AD) genes including Apolipoprotein E (*APOE*) are found to be expressed in blood-derived macrophages and thus may alter blood protein levels. We measured 91 neuro-proteins in plasma from 316 participants of the Rotterdam Study (incident AD = 161) using Proximity Extension Ligation assay. We studied the association of plasma proteins with AD in the overall sample and stratified by *APOE*. Findings from the Rotterdam study were replicated in 186 AD patients of the BioFINDER study. We further evaluated the correlation of these protein biomarkers with total tau (t-tau), phosphorylated tau (p-tau) and amyloid-beta (Aβ) 42 levels in cerebrospinal fluid (CSF) in the Amsterdam Dementia Cohort (N = 441). Finally, we conducted a genome-wide association study (GWAS) to identify the genetic variants determining the blood levels of AD-associated proteins. Plasma levels of the proteins, CDH6 ($\beta = 0.638$, $P = 3.33 \times 10^{-4}$) and HAGH ($\beta = 0.481$, $P = 7.20 \times 10^{-4}$), were significantly elevated in *APOE* ε4 carrier AD patients. The findings in the Rotterdam Study were replicated in the BioFINDER study for both CDH6 ($\beta = 1.365$, $P = 3.97 \times 10^{-3}$) and HAGH proteins ($\beta = 0.506$, $P = 9.31 \times 10^{-7}$) when comparing cases and controls in *APOE* ε4 carriers. In the CSF, CDH6 levels were positively correlated with t-tau and p-tau in the total sample as well as in *APOE ε4* stratum ($P < 1 \times 10^{-3}$). The HAGH protein was not detected in CSF. GWAS of plasma CDH6 protein levels showed significant association with a cis-regulatory locus (rs111283466, $P = 1.92 \times 10^{-9}$). CDH6 protein is implicated in cell adhesion and synaptogenesis while HAGH protein is related to the oxidative stress pathway. Our findings suggest that these pathways may be altered during presymptomatic AD and that CDH6 and HAGH may new blood-based biomarkers.

Introduction

Apolipoprotein E (*APOE*) is the most common genetic risk factor for Alzheimer's disease (AD)[1,2] and an important driver of the lifetime risk for AD[3,4]. *APOE* interacts with other common genetic determinants of AD[2,5], suggesting an interaction with specific protein pathways. Despite two decades of research, the role of *APOE* in determining the risk of AD is far from being understood[6]. The IMI ADAPTED (The Alzheimer's Disease Apolipoprotein Pathology for Treatment Elucidation and Development) is an Innovative Medicine Initiative (IMI) that aims to improve the understanding about the role of *APOE* gene in AD.

AD pathology is characterized by the extracellular deposition of amyloid-beta (Aβ)-42 and intracellular accumulation of phosphorylated tau in the brain. Cerebrospinal fluid (CSF) levels of Aβ-42, phosphorylated (p-tau) and total tau (t-tau) are well-established biomarkers of the central nervous system and brain AD pathology[7]. However, there is a growing evidence for a relation between other pathologies and AD, such as vascular pathology[8]. For example, studies integrating epidemiological and vascular research showed that vascular pathology may affect brain function and increase the risk of AD[9]. *APOE* and many of the novel genes implicated in AD are expressed in monocytes/macrophages[10-12] in the blood, and thus these genes may alter the protein signatures in blood. There is also a growing body of evidence indicating that Aβ may disrupt the cerebral microcirculation regulation[13-15], endothelial function[16,17], and brain perivascular macrophages function[18]. Thus, protein and metabolite homeostasis in blood may also be altered as a consequence of (early) amyloid pathology. Indeed, there is an increasing interest in the relation between protein levels in plasma and AD during presymptomatic stages of AD[19]. Multiple studies have investigated the association of a range of proteins with AD in plasma, but few have addressed the effect of *APOE*[19-25]. Furthermore, there is lack of investigations connecting molecular signatures of AD in blood to neuropathological AD markers in CSF.

Advances in high-throughput omics technologies have allowed the detection and quantification of several classes of plasma-based biomolecular compounds including circulating metabolites and proteins[26]. In the present study, we aimed to identify altered levels of proteins in the circulation of presymptomatic AD patients in the overall population and among various genetic risk groups based on the *APOE* gene, with a view to obtaining insights into molecular signatures in the circulation. To this end, we have examined the association of neurology relevant proteins in a prospective population-based, the Rotterdam Study. Proteins associated with AD were further tested for replication in the BioFINDER study. Next, we conducted a genome-wide association study to find the genetic variants determining the blood levels of AD-associated proteins. Finally, we studied the association of the protein consistently associated with AD to amyloid and tau levels in CSF in the Amsterdam Dementia Cohort (ADC).

Results

Association of plasma proteins with AD

Detailed results of overall and *APOE* stratified association analysis of proteins with AD are provided in Table 1 and Figure 1. No significantly associated protein to AD was identified in the overall analyses at an *FDR* < 0.05. Overall, there is a tendency that protein levels are more likely increased (positive effect size, β) than decreased (negative effect size, β) in AD patients that carry the *APOE* ε4 allele (Figure 1a) and those homozygous for *APOE* ε3 allele (Figure 1c) but not for *APOE* ε2 patients (Figure 1d). In *APOE* stratified discovery analysis, we observed that levels of CDH6 (β = 0.638, $P = 3.33 \times 10^{-4}$, *FDR*=0.030) and HAGH (β = 0.481, $P = 7.20 \times 10^{-4}$, *FDR* = 0.033) were significantly increased in AD patients who carry the *APOE* ε4 allele (see Table 1).

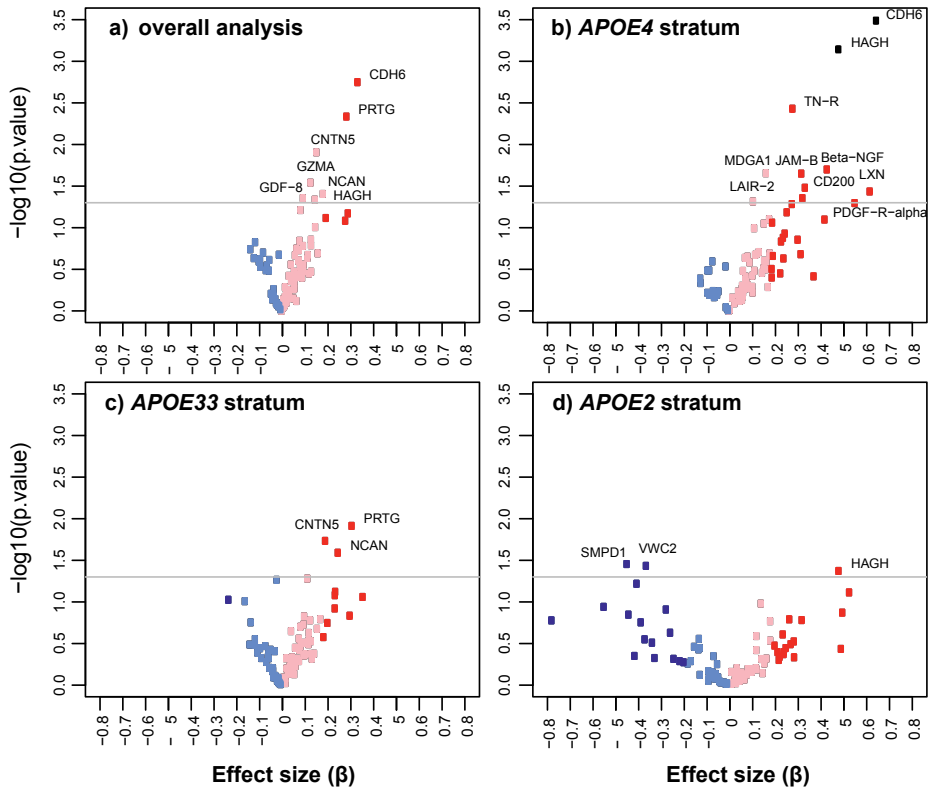


Figure 1: Volcano plots representing the association of plasma protein levels with Alzheimer's disease (AD) in (a) overall analysis; (b) *APOE4* stratum; (c) *APOE33* stratum and (d) *APOE2* stratum. Each dot represents a protein with regression coefficient (β) of association plotted on x-axis and -log10 of *P-values* on y-axis. Proteins showing nominal association (*P-value* < 0.05) are annotated in overall and stratified analysis. Light blue color of dot indicates decreased protein levels with β -0.0 to -0.184 and dark blue to indicate with β < -0.184 while pink color indicates increased protein levels with β ranging from 0.0 to 0.184 and red color shows β > 0.184. Black dots are used for proteins which pass the multiple testing (false discovery rate <0.05).

Table 1: Results of plasma-based proteome association with Alzheimer's disease

Uniprot id	Annotation	Effect size (β)	OR	SE	*P-value*	*FDR-value*
Overall population						
P55285	CDH6	0.334	1.397	0.106	$1.78x10^{-3}$	0.162
Q2VWP7	PRTG	0.286	1.331	0.100	$4.62x10^{-3}$	0.210
O94779	CNTN5	0.155	1.168	0.061	$1.24x10^{-2}$	0.377
P12544	GZMA	0.129	1.138	0.059	$2.88x10^{-2}$	0.598
O14594	NCAN	0.183	1.201	0.088	$3.93x10^{-2}$	0.598
O14793	GDF-8	0.095	1.100	0.047	$4.42x10^{-2}$	0.598
Q16775	HAGH	0.147	1.158	0.074	$4.60x10^{-2}$	0.598
***APOE4* stratum**						
P55285	CDH6	0.638	1.893	0.171	$3.33x10^{-4}$	0.030
Q16775	HAGH	0.481	1.618	0.138	$7.20x10^{-4}$	0.033
Q92752	TN-R	0.280	1.323	0.094	$3.72x10^{-3}$	0.113
P01138	Beta-NGF	0.431	1.539	0.182	$2.00x10^{-2}$	0.340
Q8NFP4	MDGA1	0.164	1.178	0.070	$2.22x10^{-2}$	0.340
P57087	JAM-B	0.318	1.374	0.137	$2.24x10^{-2}$	0.340
P41217	CD200	0.335	1.398	0.155	$3.31x10^{-2}$	0.394
Q9BS40	LXN	0.619	1.857	0.292	$3.67x10^{-2}$	0.394
P16234	PDGF-R-alpha	0.324	1.383	0.159	$4.42x10^{-2}$	0.394
Q6ISS4	LAIR-2	0.107	1.113	0.053	$4.84x10^{-2}$	0.394
***APOE33* stratum**						
Q2VWP7	PRTG	0.309	1.362	0.122	$1.22x10^{-2}$	0.778
O94779	CNTN5	0.193	1.213	0.081	$1.85x10^{-2}$	0.778
O14594	NCAN	0.248	1.281	0.110	$2.57x10^{-2}$	0.778
P55285	CDH6	0.202	1.224	0.150	$1.79x10^{-1}$	0.842
Q16775	HAGH	-0.018	0.982	0.092	$8.42x10^{-1}$	0.957
***APOE2* stratum**						
P17405	SMPD1	-0.447	0.640	0.203	$3.51x10^{-2}$	0.966
Q2TAL6	VWC2	-0.362	0.696	0.166	$3.69x10^{-2}$	0.966
Q16775	HAGH	0.482	1.619	0.227	$4.25x10^{-2}$	0.966
P55285	CDH6	0.498	1.645	0.324	$1.35x10^{-1}$	0.966

Abbreviations: β, regression coefficient; OR, odds ratio; SE, standard error; *APOE*, apolipoprotein E; FDR, False discovery rate

Note: Multiple testing correction by false discovery rate (FDR) < 0.05 was considered significant

Both CDH6 (β = 0.624, $P = 5.52x10^{-4}$, *FDR* = 0.030) and HAGH (β = 0.491, $P = 6.62x10^{-4}$, *FDR* = 0.030) proteins remained significantly associated with AD even after adjusting for other covariates in model 2 (Supplementary Table 1).

In the replication analysis in the BioFINDER study (Table 2), plasma levels of CDH6 and HAGH were significantly associated with AD in the overall sample (CDH6: β = 1.212, $P = 5.18x10^{-4}$; HAGH: β = 0.631, $P = 7.56x10^{-15}$) as well as in *APOE* ε4 carriers (CDH6: β = 1.365, $P = 3.97x10^{-3}$; HAGH: β = 0.506, $P = 9.31x10^{-7}$) but not in *APOE* ε2 carriers. Plasma levels of HAGH protein were also associated with AD in *APOE* ε33 carriers (β = 0.739, $P = 3.76x10^{-7}$) but in this subgroup no association with CDH6 was seen.

Figure 2 shows that the *APOE* genotype modifies the association between proteins and AD based on nominal statistical significance. In discovery analysis, eight additional proteins (TN-R, Beta-NGF, MDGA1, JAM-B, CD200, LXN, PDGF-R-alpha, and LAIR-2) were also positively associated with AD *APOE* ε4 carriers (β > 0.107, $P < 0.05$), but they did not survive multiple testing. In the *APOE2* stratum, the levels of two proteins including SMPD1 and VWC2 were reduced in AD cases compared to the *APOE* genotype matched controls (Supplementary Figure 1). In *APOE33* stratum, PRTG, CNTN5 and NCAN proteins, that do not emerge in the *APOE4* or *APOE2* stratum (see Figure 2), showed suggestive associations but did not survive multiple testing (β > 0.193, $P < 2.57x10^{-2}$). Both CDH6 (β = 0.202, $P = 1.79x10^{-1}$) and HAGH (β = -0.018, $P = 8.42x10^{-1}$) did not show association with AD in *APOE33* carriers while HAGH showed nominal association in *APOE* ε2 carriers (β = 0.482, $P = 4.25x10^{-2}$) (See Table 1).

Sensitivity analyses

Sensitivity analyses were performed in the Rotterdam Study to test the robustness of our findings. In the first sensitivity analysis (Supplementary Table 2), we adjusted for follow-up time, taking into account that some cases or controls may die of other diseases. This analysis showed that levels of both HAGH (β = 0.477, $P = 6.47x10^{-4}$) and CDH6 (β = 0.661, $P = 1.48x10^{-4}$) proteins were significantly increased in AD patients compared to controls in *APOE* ε4 carriers. In the second sensitivity analysis, we only analyzed protein that were assessed directly (non-imputed data). Similarly, the association of HAGH and CDH6 proteins with AD remained significant in *APOE* ε4 carriers when analyzing non-imputed proteomics data (Supplementary Table 3). Last but not least, we performed a formal interaction test to evaluate the interaction of *APOE* with each of the 91 proteins (Supplementary Table 4). Only three of 91 proteins showed interaction with *APOE* (*P-value* < 0.05) including HAGH ($\beta_{interaction} = 0.414$, $P_{interaction} = 1.70x10^{-2}$), G-CSF ($\beta_{interaction} = 0.276$, $P_{interaction} = 2.78x10^{-2}$) and CRTAM ($\beta_{interaction} = -0.221$, $P_{interaction} = 3.77x10^{-2}$). Except, HAGH other two proteins (G-CSF, CRTAM) did not show association with AD in any of the analyzed *APOE* stratum (*P-value* > 0.05). For CDH6, the test for was not significant ($\beta_{interaction} = 0.078$, $P_{interaction} = 7.23x10^{-1}$).

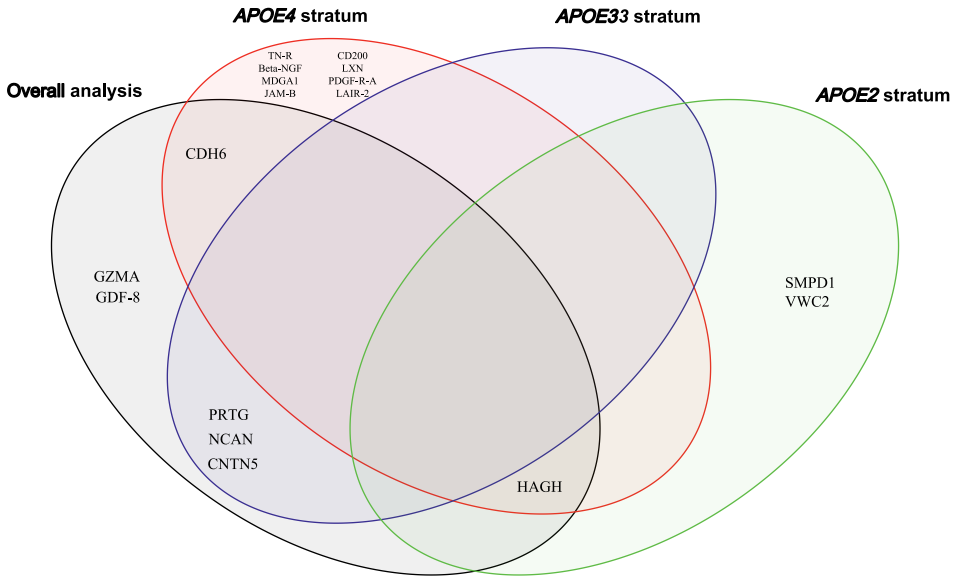


Figure 2: Venn diagram showing the overlap of proteins identified in the association analysis results of overall and *APOE* stratified analysis.

Association of CDH6 and HAGH protein levels with Aβ-42, p-tau, and t-tau in CSF

Among the two proteins that were associated to the future risk of AD, CDH6 and HAGH, the latter was not detected in the CSF in > 90% of the subjects in ADC cohort. CSF CDH6 protein levels were not associated with AD (β = 0.329, SE = 0.220, P = 0.136) in the overall as well as in *APOE* stratified analysis (P > 0.114; see Supplementary Table 5). However, multiple regression analysis adjusted for age and sex revealed a significant association of CDH6 CSF levels with both p-tau (β = 23.2, SE = 3.4, $P = 3.48 \times 10^{-11}$) and t-tau (β = 207.4, SE = 36.4, $P = 2.40 \times 10^{-8}$) when pooling AD patients and controls (Table 3 and Figure 3). In the *APOE* stratified analysis, levels of CDH6 were significantly associated with p-tau and t-tau levels but not with Aβ-42 levels in CSF in three *APOE* strata (see Table 3). When stratifying by case-control status (Supplementary Table 6), CDH6 levels were significantly associated with p-tau and t-tau levels in both cases and controls. In controls, also Aβ-42 was positively associated with CDH6 ($P < 1x10^{-3}$; see Supplementary Table 6).

Association of proteins with *APOE*

Next, we associated the *APOE* genotype to the protein levels significantly associated with AD (CDH6 and HAGH). Results of the association of protein levels with *APOE* genotypes are provided in Supplementary Table 7 and Figure 2. In the overall sample, CDH6 protein levels were increased in the *APOE* ε4 carriers compared to *APOE* ε2 carriers (β = 0.163, $P = 3.79 \times 10^{-3}$). In controls, levels of CDH6 protein were decreased (β= -0.131, P = 0.026) in *APOE* ε2 carriers

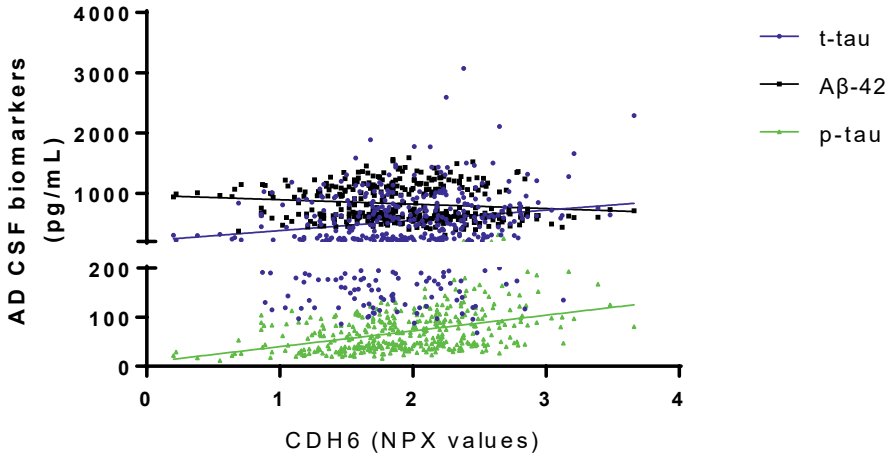


Figure 3: Correlation plot between cerebrospinal fluid (CSF) levels of CDH6 with Aβ-42, p-tau and t-tau.

compared to *APOE 33* carriers. In the controls, levels of HAGH were decreased in the *APOE* ε4 (β = -0.192, *P* = 0.028) and *APOE* ε2 carriers (β = -0.214, *P* = 0.042) compared to *APOE* 33 carriers.

Genome-Wide Association Study (GWAS)

The GWAS was conducted to determine the genetic drivers of the CDH6 and HAGH protein levels (Supplementary Figure 3:A and 4 respectively). We identified 13 genome-wide significant cis protein quantitative trait loci (pQTLs) located at 5p13.3 locus of chromosome 5 for CDH6 protein levels. All genome-wide significant pQTLs are located in the intergenic region at 5' UTR region of the *CDH6* gene. Among the 13 identified pQTLs, rs111283466 was the lead pQTL with the effect estimate (β) of 1.068 and *P*-value 1.92×10^{-9} (Supplementary Table 8). Q-Q plot (Supplementary Figure 5:A) indicates that the results are well adjusted for population stratification (λ = 1.0056). Further lookups in the GTEx database showed that the lead pQTL (rs111283466) also affects the expression of *CDH6* gene in various body tissues. GWAS analysis of HAGH protein levels did not identify any genome-wide significant pQTLs. Manhattan plot and Q-Q plot for GWAS results of HAGH protein levels are provided in the Supplementary materials (Supplementary Figure 4 and 5:B).

Discussion

In our study, plasma levels of CDH6 and HAGH proteins are significantly increased in presymptomatic AD patients compared to controls in the *APOE4* stratum. In the replication analysis, both CDH6 and HAGH proteins showed significant association with AD in the BioFINDER study in *APOE ε4* carriers. CDH6 protein levels were significantly correlated with

Table 2: Association of plasma levels of CDH6 and HAGH proteins with Alzheimer's disease in the BioFINDER Study

	Overall			*APOE4* **stratum**			*APOE33* **stratum**			*APOE2* **stratum**		
Biomarkers	β	SE	*P-value*	β	SE	*P-value*	β	SE	*P-value*	β	SE	*P-value*
CDH6	1.212	0.349	$5.18x10^{-4}$	1.365	0.474	$3.97x10^{-3}$	0.961	0.585	$1.01x10^{-1}$	2.612	3.210	$4.16x10^{-1}$
HAGH	0.631	0.081	$7.56x10^{-15}$	0.506	0.103	$9.31x10^{-7}$	0.739	0.145	$3.76x10^{-7}$	5.565	3.979	$1.62x10^{-1}$

Abbreviations: AD, Alzheimer's disease; β, regression coefficient; SE, standard error
*Logistic regression analysis adjusting for age and sex

Table 3: Association of CSF based CDH6 protein levels with AD biomarkers in Amsterdam Dementia Cohort

	Overall			*APOE4* **stratum**			*APOE33* **stratum**			*APOE2* **stratum**		
Biomarkers	β	SE	*P-value*	β	SE	*P-value*	β	SE	*P-value*	β	SE	*P-value*
Aβ-42	16.574	26.827	$5.37x10^{-1}$	33.701	35.001	$3.37x10^{-1}$	14.499	44.966	$7.48x10^{-1}$	-103.767	69.354	$1.45x10^{-1}$
p-tau	23.189	3.404	$3.48x10^{-11}$	26.674	5.781	$8.36x10^{-6}$	21.981	5.564	$1.17x10^{-4}$	25.663	7.697	$2.29x10^{-3}$
T-tau	207.396	36.437	$2.40x10^{-8}$	235.009	59.158	$1.10x10^{-4}$	193.859	64.068	$2.90x10^{-3}$	298.437	86.637	$1.66x10^{-3}$

Abbreviations: AD, Alzheimer's disease; β, regression coefficient; SE, standard error
*Linear regression analysis adjusting for age and sex

p-tau and t-tau measurements in CSF of the ADC. In GWAS analysis, we have also identified a genome-wide significant pQTL for CDH6 protein levels in the blood (rs111283466), which also affects the expression levels of *CDH6* transcripts in several tissues.

We observed a significant increase of CDH6 protein levels in the plasma of presymptomatic AD cases carrying the *APOE* ε4 allele which was also replicated in the BioFINDER study. When comparing our findings to the other studies[19-22], we do not have an overlap in understudy proteins. However, like previous studies we do find an effect of the *APOE* gene on plasma level of prorteins[21,22]. In the *APOE4* stratum, we see that the volcano plot (Figure 1) is clearly asymmetric suggesting increased levels of most neuronal proteins in AD patients carrying this allele before the clinical onset of disease. This might be explained by an increase in the blood-brain barrier permeability in *APOE* ε4 carriers[27], which may lead to increased levels of CDH6 in the blood as a result of higher levels of CDH6 in the brain. We found that CDH6 levels in the blood are driven by a genetic variant (rs111283466) in the cis-regulatory region. This may determine the CDH6 levels in both brain and blood cells, leaving the possibility open that elevated CDH6 has a blood-derived origin. Yet, such a mechanism does not explain why elevated levels in the blood are only seen in patients carrying the *APOE* ε4 allele. It is of note that the CDH6 coding gene is part of a larger cluster of cadherin (CDH) genes including *CDH9, CDH10, CDH12 and CDH18*. As all of the CDH genes are paralogues and share homology, it is crucial to exclude cross-reactions of the antibodies[28] used by Olink across CDH proteins. Our GWAS benchmarks that the protein assessed in our plasma is indeed CDH6, as we found that the most important driver of the protein is in the promoter region of *CDH6*. None of the recently published GWAS of proteins reported significant pQTLs for CDH6 protein based on SomaLogic[29,30]. The aptomere based measurement of the SomaLogic yields a different protein spectrum than that of the antibody-based method of Olink[31]. Our identified pQTLs did not show any association with AD (*P-value* < 0.05) in the largest AD GWAS[11]. However, we find in our study that levels of CDH6 were increased in *APOE* ε4 carriers compared to *APOE* ε2 carriers. We do find that the region is associated to postcentral gyri in GWAS[32] and a study has reported reduced volume of postcentral gyri in dementia patients[33].

Interestingly, we found significant positive associations of CSF levels of CDH6 protein with p-tau and t-tau levels in overall as well as in AD and controls only analyses, which are considered as biomarkers of neuronal injury and tau pathology[34,35]. The upregulation of CDH6 protein in cerebral cortices of AD mice models (*APP/PS1*) compared to wild type has been reported by Lu *et al.,*[36] which is in line with our observation of positive correlation between CDH6 levels and AD pathological markers in CSF. Yet, we did not observe an association of the CDH6 protein with AD in CSF. We also observed positive association between CDH6 levels with amyloid-beta 42 in CSF of the controls, which might indicate disturbance in the amyloid-beta 42 metabolism which precedes decades before the buildup of Aβ in the brain[37]. Alternatively, it may point towards a similar mechanism of production of the Aβ-42 and CDH6 proteins in a

healthy state[38]. Increased levels of phosphorylated CDH6 protein levels were reported upon the addition of amyloid-beta in cortical neuronal cells[39], which adds evidence to the role of CDH6 in AD pathology. Taken together, these findings suggest that CSF levels of CDH6 protein may be associated with neuronal and axonal cell injury and neurofibrillary tangles in AD.

CDH6 is a cell surface glycoprotein that belongs to type II cadherin's[40]. Cadherins are highly expressed in the brain and other tissues. They strongly interact with other molecules to perform molecular processes including synaptic functions[41-43], synaptogenesis[44], TGF-B signaling[45], neural crest differentiation[46], presenilin-mediated signaling and integrity of blood-brain barrier[47]. Although it is not possible to infer whether the correlation of AD pathology with CDH6 in plasma and CSF, are cause or consequence of the disease, several pieces of evidence favor the role of CDH6 in the pathogenesis of AD[39,48]. A recent study showed that the ADAM10 enzyme, whose coding gene is associated with AD[11], is involved in proteolytic cleavage of the *CDH6* protein, resulting in the formation of C-terminal fragment[49], in a similar manner as it cleaves the amyloid precursor protein (*APP*)[50,51]. The transmembrane N-cadherin (*CDH2*), a paralogue of CDH6 and functionally related to CDH6[43], is also known to be cleaved by ADAM10 into N-cadherin C-terminal fragment 1 (NcadCTF1). Andreyeva *et al.,*[52] have demonstrated that NcadCTF1 leads to accelerated amyloid-β-induced synaptic impairment, a process that characterizes an early stage event in AD[53,54]. Increased levels of NcadCTF1 were also found in postmortem AD brain tissues compared with controls, suggesting that cadherins might induce synaptic dysfunction in a synergistic manner[52].

In addition to CDH6, increased plasma levels of HAGH (Hydroxyacylglutathione hydrolase, mitochondrial) protein also showed significant association with AD in those who carry the *APOE* ε4 variant and suggestive association in overall and in *APOE* ε2 carriers. This finding is in line with the recently published findings of the BioFINDER study[25] and further supported by the *APOE* stratified analysis in the BioFINDER study that was conducted for the present study. In the replication analysis, plasma levels of HAGH showed significant association in both *APOE* ε4 and *APOE* ε33 carriers while in the discovery analysis in Rotterdam Study HAGH only showed significant association in *APOE* ε4 carriers which may be due to the lack of power. The HAGH protein is also known as glyoxalase-2, an enzyme, which is involved in the glyoxalase system along with glyoxalase-1 and its cofactor glutathione, a key player is oxidative stress control[55,56]. Overall, the glyoxalase system is involved in the detoxification of glycolysis by-products particularly cytotoxic metabolite methylglyoxal[57]. Levels of methylglyoxal in plasma are elevated during various disease conditions including hyperglycemia, which leads to the formation of reactive oxygen species (ROS) and causes oxidative stress. Moreover, methylglyoxal is also the precursor of glycation end products (AGEs) which are implicated in neurodegeneration and AD[58,59]. The most compelling evidence for the role of the glyoxalase-2 protein in AD is that the AGEs and glyoxalase system is implicated in the regulation of amyloid precursor protein (*APP*) expression[60,61]. Although glyoxalase system attributes protection against methylglyoxal mediated

oxidative stress, earlier studies have also observed increased levels of glyoxalase-1 enzyme (involved in the first step of methylglyoxal detoxification) in early AD stages[62,63]. Increased levels of glyoxalase-2 (involved in the second step of methylglyoxal detoxification) in plasma might be a compensatory mechanism to increased levels of methylglyoxal during the early phase of disease or a general stress response[55]. The growing number of studies have suggested the involvement of oxidative stress during the prodromal stage of AD[64-66], which is in line with our finding of increased levels of glyoxalase-2 observed before the onset of AD.

The strength of the current study includes that it is conducted in the prospective population-based RS cohort, where samples were selected with mean 6.9 years of follow-up preceding the diagnosis of AD. It allowed us to study the plasma proteomics changes prior to the development of AD clinical symptoms. As AD is a disorder of the brain, we have validated that CSF levels of CDH6 are also associated with biomarkers of AD in CSF in an independent cohort. Further, we used the Olink neurology proteomic panel of 91 proteins for the quantification of proteins in the plasma, which estimates targeted proteins expressed in the brain from different pathways. One of the major limitations of our study is the limited sample size, including a small number of *APOE ε4* carrier controls in the stratified analysis.

In conclusion, we observed elevated protein levels of CDH6 in plasma of AD patients carrying *APOE ε4* allele in the discovery and replication analysis, a protein that plays a role in synaptogenesis. Positive correlation of CSF CDH6 levels with p-tau and t-tau may also indicate the association of CDH6 with neurodegeneration. We further found the association of the plasma levels of HAGH protein to AD in those carrying the *APOE* ε4 allele. Association of HAGH with AD further suggest the involvement of the glyoxalase and oxidative stress pathways in the pathogenesis of AD.

Methods

Study populations

Rotterdam Study

The Rotterdam Study (RS) is a prospective population-based study comprising of 14,926 middle and older aged (≥ 45 years) individuals from the Ommoord district of Rotterdam. The RS consists of three cohorts including RS-I (started in 1990, N=7983 participants), RS-II (started in 2000, N=3011) and RS-III (started in 2006, N=3932)[67]. Study participants were extensively interviewed and physically examined at baseline and after every 3 to 4 years. For each participant fasting blood was collected at a dedicated center, centrifuged (Speed=3500g for 20 min at 4^{0}C) within 4 hours of venipuncture to collect plasma and stored at -80^{0}C. The study has been approved by the Medical Ethical Committee of Erasmus Medical Center and by the Ministry of Health, Welfare and Sport of the Netherlands. Written informed consent was obtained from

each study participant to participate and to collect information from their treating physicians. All methods were performed in accordance with the relevant guidelines and regulations. In current nested case-control proteomics analysis, we chose 161 incident AD cases and 155 controls match with respect to their age and sex, from the fifth visit of RS-I (RS-I-5) cohort. Table 4 shows the baseline characteristics of the selected sample. There were no significant differences in age, sex and body mass index (BMI). AD patients were more often carriers of the *APOE* ε4 variant and less often of the *APOE* ε2 variant. Blood for the proteome profiling was collected on average 6.9 years (standard deviation [SD] = 1.7) before the onset of clinical dementia in patients and mean 8.7 years (SD = 3.2) before the latest follow-up in controls.

Table 4: Population descriptive

	Total participants	Incident AD cases	Controls
N	316	161	155
Age (SD) blood collection, years	77.16(5.39)	77.43(5.21)	76.89(5.59)
Age at onset/last follow up	84.99(5.33)	84.37(5.01)	85.63(5.56)
Female (%)	201(63%)	104 (65%)	97 (63%)
Body Mass index	27.32(4.10)	27.29(3.75)	27.37(4.46)
Follow-up (SD) years	7.82(2.71)	6.94(1.71)	8.74(3.22)
APOE genotype			
APOE 44/34/24	98	68	30
APOE 33	171	76	95
APOE 22/23	34	13	21

Abbreviations: AD, Alzheimer's disease, SD, Standard deviation, *APOE*, apolipoprotein E gene

Dementia diagnosis

Over time, all participants were screened for dementia using the Mini-Mental State Examination (MMSE)[68] and Geriatric Mental Schedule (GMS)[69] organic level for all participants. Screen-positive subjects (MMSE<26 or GMS organic level > 0) underwent the Cambridge examination for mental disorders of the elderly (CAMDEX)[70] and participants suspected of having dementia were extensively examined with neuropsychological testing and neuroimaging biomarkers when available. Patients were further ascertained by linking them with their medical records from general practitioners, the regional institute for outpatient mental health care and municipality. Dementia of all patients was diagnosed based on the internationally accepted Diagnostic and Statistical Manual of Mental Disorders (DSM-III-R) criteria and AD using the National Institute of Neurological Disorders and Stroke–Association Internationale pour la Recherche et l'Enseignement en Neurosciences (NINCDS-ADRDA)[71] criteria for possible, probable and definite AD. NINCDS-ADRDA criteria were also used to diagnose vascular dementia. The final diagnosis was confirmed by a panel of neurologists, neurophysiologists, and research physicians[72]. AD diagnosis in RS is also provided in more detail in earlier publications[72].

Proteome profiling

Proteomics profiling of the 316 plasma samples was performed using neurology panel of OLINK's Proximity Extension Assay (ProSeek, OLINK AB, Uppsala, Sweden), which includes 91 proteins involved in various pathways including axon development, axon guidance, cell adhesion, cell death, cell differentiation, cell growth, cellular metabolic process, immune response, MAPK cascade, neurogenesis, proteolysis, signal transduction and synapse assembly (https://www.olink.com/products/neurology/). This method uses affinity-based assay, in which a pair of oligonucleotide-labeled antibody probes bind to a target protein. Proximity-dependent DNA polymerization event forms a polymerase chain reaction (PCR) target sequence between two probes bound in close proximity. The generated PCR target sequence is detected and quantified using real-time PCR method. The resultant protein abundance is provided as NPX (Normalized Protein Expression), which is an arbitrary unit on log2 scale. Lower limit of detection is estimated based on negative controls inserted in each run and measurements below this limited were treated as missing. None of the detected markers in our dataset reach missingness more than 10 percent. Protein markers with missing values less than 10% were imputed with the lowest detected limit for further analysis. More detailed information about detection limits, assay performance and validation methods are available from the service provider (www.olink.com)[73].

APOE genotyping

In the RS *APOE* genotyping was performed using Polymerase chain reaction (PCR) and amplified PCR product was digested with *HhaI* enzyme. Restriction fragments of enzyme products were visualized by silver staining after getting them separated with precast ExcelGel gels (Pharmacia Biotech, Uppsala, Sweden). Genotype results were examined by three independent persons. In the case of non-agreement *APOE* genotype was repeated[74,75].

Genotyping and imputations

In the RS participant's blood was collected during baseline and follow-up visit. DNA genotyping was performed for all the participants with proper DNA quality with the 550K, 550K duo, or 610K Illumina arrays. In genotyping quality control, genetic variants exclusion criteria include, call rate < 95%, Hardy-Weinberg equilibrium $P < 1.0x10^{-6}$ and Minor Allele Frequency (MAF) < 1%. Sample exclusion criteria include excess autosomal heterozygosity (0.336), call rate < 97.5%, duplicate or family relationships and ethnic outliers identified by the identity-by-state clustering analysis (having identity-by-state probability <97% or >3 standard deviation from population mean)[76]. Further, genetic variants were imputed with the Haplotype Reference Consortium (HRC) reference panel (version 1.0)[77], using the Michigan imputation server [78]. The server uses SHAPEIT2 (v2.r790)[79] to phase the genotype data and performs imputation with Minimac 3 software[80]. Genotyping information was available for 281 among 316 participants included in the current study.

BioFINDER Study

In the current study, replication analysis was performed in 671 participants (AD patients = 186, Controls = 485) of the BioFINDER (Biomarkers For Identifying Neurodegenerative Disorders Early and Reliably) study. Characteristics of the BioFINDER study participants included in the replication analysis are provided in Supplementary Table 9. The BioFINDER study includes participants from southern Sweden recruited between 2009 and 2014 (www.biofinder.se). The study participants were assessed by experienced physicians including the neurological, psychiatric and cognitive assessments[81]. The NINCDS-ADRDA criteria were used to classify Alzheimer's disease dementia patients for probable Alzheimer's disease patients. All dementia due to Alzheimer's disease patients had pathological CSF Aβ42/Aβ40 ratio of < 0.1. The inclusion criteria for the cognitively normal elderly participants included i) aged 60–80 years, ii) MMSE scores ranging between 28–30 at their baseline screening visit, iii) no cognitive impairment symptoms assessed by a physician, and iv) not fulfilling the criteria for mild cognitive impairment or dementia. Exclusion criteria included i) refused lumbar puncture, ii) significant neurological or psychiatric disease, iii) current alcohol or substance misuse, or iv) systematic illness preventing them from participating in the study[25,81]. Written Informed consents were collected from each study participant and the study has been approved by the Regional Ethics Committee in Lund, Sweden.

Protein profiling

During the baseline visit of the BioFINDER study, plasma and lumbar CSF samples were collected from non-fasting participants. Standardized protocol was followed to analyze the plasma and CSF samples. All samples were centrifuged at 2000 g (+4 °C for 10 min), and aliquoted into 1 ml polypropylene tubes (Sarstedt AG & Co., Nümbrecht, Germany), and stored at -80°C. Before the proteomics profiling, plasma and CSF samples underwent one cycle of freeze-thaw, and further aliquoted into 200L Lobind tubes (Eppendorf Nordic A/S, Denmark). Protein concentrations were quantified using the ProSeek multiplex immunoassay, developed by Olink Proteomics (Uppsala, Sweden)[25].

Amsterdam Dementia Cohort (ADC)

In the validation analysis of most interesting proteins, we used 441 participants from the ADC cohort whose CSF samples were already profiled for neurology related proteins using the OLINK's Proximity Extension Assay (ProSeek, OLINK AB, Uppsala, Sweden). Information about characteristics of patients included in current analysis as a part of the validation dataset is listed in Supplementary Table 9.

The ADC is a prospective memory-clinic cohort that was established in September 2000 at the Alzheimer Center Amsterdam of Amsterdam UMC. The cohort has included 6000 individuals until September 2017[82] [83,84]. All participants underwent standardized cognitive screening including neurological and cognitive examination, blood sampling, a lumbar puncture to collect

CSF and brain magnetic resonance imaging. All CSF samples were stored in agreement with the JPND-BIOMARKAPD guidelines[85]. All subjects provided written informed consent for use of biomaterial and clinical data for research and the study was approved by the local medical ethical review board. All methods were performed in accordance with the relevant guidelines and regulations. A sample of 441 participants selected for our validation analysis consists of 242 AD and 199 cognitively normal controls who were presented at the memory clinic with subjective cognitive decline (i.e., Criteria for mild cognitive impairment and dementia not fulfilled)). As additional inclusion criteria, controls were required to have normal AD CSF biomarkers profile: low CSF β-amyloid 1–42 (Aβ42) and high p- or t-tau level (applying local laboratory cut-offs) and to remain cognitively stable for 2 years. All participants underwent standard neurological and cognitive assessments and the diagnosis was assigned according to consensus AD criteria[86]. Global Mini-Mental State Examination (MMSE) was used to examine global cognition. The levels of CSF AD-related biomarkers (Aβ42, total and phosphorylated tau [t-Tau and p-Tau_{181}]) were analyzed at Amsterdam UMC as part of the routine diagnostic procedure using commercially available kits (Innotest Aβ(1-42), total Tau, phospho-Tau(181P); Fujirebio, Ghent, Belgium)[10,24].

Statistical analysis

Plasma protein association with AD

To identify AD-associated proteins, plasma levels of 91 proteins were compared between incident AD cases and non-demented controls using logistic regression, adjusted for age and sex in the first model. In the second model, we additionally adjusted for body mass index (BMI), smoking, educational status and medication use (lipid-lowering medications, antihypertensive and anti-inflammatory medication). To identify *APOE* specific associations of proteins with AD, we performed stratified association analysis based on *APOE* genotype carrier status. All participants were divided into *APOE4* stratum (*APOE* 44/34/24), *APOE3* stratum (*APOE* 33) and *APOE2* stratum (*APOE* 22/23). Participants with *APOE 24* genotypes were pooled within the *APOE4* stratum because an earlier study has demonstrated that the risk profiles of *APOE 24* genotype to AD and dementia is similar to those with *APOE 34* genotype carriers[5]. The association results were corrected for multiple testing using false discovery rate (*FDR*) by Benjamini and Hochberg method separately for the overall analysis, and in each *APOE* stratum [87] and association tests with *FDR* < 0.05 were considered significant. All analyses were performed using R software (https://www.r-project.org).

Sensitivity analyses

Moreover, we performed sensitivity analyses. In the first sensitivity analysis, we repeated the overall and *APOE* stratified regression analysis (Model 1: age and sex) additionally adjusting for the follow-up time (the time between blood collection and onset of AD or last follow-up for controls). In the second sensitivity analysis, to assess the differential bias due to missingness, we performed the overall and *APOE* stratified association analysis in the non-imputed proteomics

data adjusting for age and sex. We also tested the interaction of *APOE* genotype (ε4 carriers and non-carriers) and proteins levels using logistic regression model adjusting for age and sex.

Additional analysis of proteins showing association with AD

A detailed flowchart of the analysis is provided in Figure 4 about the discovery, replication and validation analysis. Proteins that appeared significantly altered in overall or *APOE* stratified analysis were further tested for association with *APOE* genotypes; second GWAS was performed to identify pQTLs, regulating the levels of protein in blood.

Replication analysis

Replication analysis of two proteins was performed in an independent BioFINDER study. We performed association of plasma levels of proteins with AD versus controls (AD cases = 186, controls = 485) in the overall sample and stratified by *APOE* genotype: *APOE4* stratum (*APOE* 44/34), *APOE3* stratum (*APOE* 33) and *APOE2* stratum (*APOE* 22/23). We used logistic regression analysis adjusted for age, sex and date of sample collection.

Validation analysis: Association of CSF protein levels with Aβ-42, p-tau, and t-tau

In the validation analysis of specific proteins in an independent ADC cohort (N = 441), we performed association of CSF protein levels with AD versus control group and with Aβ-42, p-tau and t-tau levels in CSF. All the validation analyses were performed in the overall sample and stratified by *APOE* genotype: *APOE4* stratum (*APOE* 44/34), *APOE3* stratum (*APOE* 33) and *APOE2* stratum (*APOE* 22/23). We used linear regression analysis adjusted for age and sex to evaluate the association of proteins measured in CSF with AD brain pathology biomarkers in the overall sample and stratified by clinical diagnosis (AD and controls).

Association of plasma protein levels with APOE genotype

To further evaluate the association of proteins with *APOE* genotypes, we compared protein levels, among *APOE* genotype groups (*APOE* 44/34/24=*1* versus *APOE* 33=*0*, *APOE* 44/34=1 versus *APOE* 22/23=0 and *APOE* 22/23=1 versus *APOE* 33=0) in the overall study sample, in AD patients, and in control groups separately. Linear regression analysis was performed using protein levels as outcome and *APOE* status as predictor, adjusted for age and sex.

Genome-wide association study

Further, we performed the genome-wide association study (GWAS) to identify protein quantitative trait loci (pQTLs) for candidate proteins. We regressed out protein levels against age, sex and principal components to calculate residuals. To normalize the calculated residuals we applied Rank-inverse transformation on residuals. Principal components derived from genotypes were used in the association analysis to adjust for population stratification. GWAS of rank-inverse normalized residuals was performed using score test option in RVTEST software [88]. Variants with

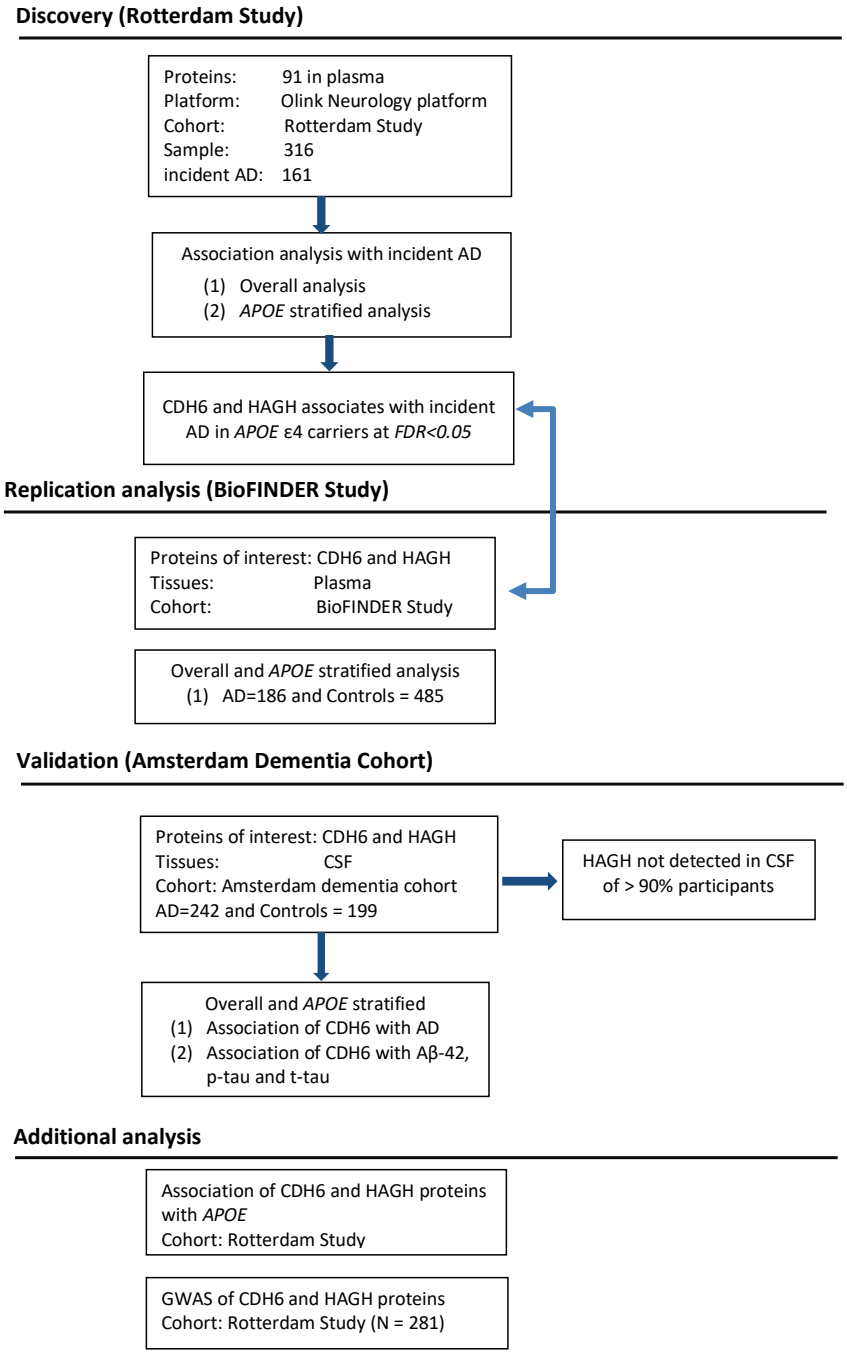


Figure 4: Flowchart of the analyses. Rotterdam Study was used as discovery cohort in plasma-based proteomics analysis. Altered proteins in plasma analysis were replicated in the BioFINDER study and further validated in Amsterdam Dementia Cohort participants. Abbreviations; AD: Alzheimer's disease; *APOE*: apolipoprotein E; GWAS: genome-wide association study; CSF: cerebrospinal fluid.

low imputation quality R-squared < 0.3 and minor allele count less than five were excluded from the results. Manhattan and quantile-quantile (Q-Q) plots for GWAS results were generated with web-based utility Functional mapping and annotation of genetic associations (FUMA)[89] and regional association plots using LocusZoom (http://locuszoom.org). pQTLs with a *P*-value < 2.5 $x10^{-8}$ ($5x10^{-8}$/2 tested proteins) were considered genome-wide significant. To check the overlap of identified pQTL with expression quantitative loci (eQTLs) we used GTEx data base[90].

Data access and availability

Current study used data from RS and ADC, where sharing of participants data is not allowed publicly due to legal and ethical permissions. Informed consents collected for both studies do not allow to share individual participants data in public repository. Data access can be made available for interested researchers upon request to corresponding author Cornelia M. van Duijn (Cornelia.vanDuijn@ndph.ox.ac.uk).

Acknowledgement

This study was funded by the ADAPTED: Alzheimer's Disease Apolipoprotein Pathology for Treatment Elucidation and Development consortium which has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under Grant Agreement No 115975. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and the European Federation of Pharmaceutical Industries and Associations. This study was also funded by the PERADES Program (Defining Genetic, Polygenic and Environmental Risk for Alzheimer's disease, using multiple powerful cohorts, focused Epigenetics and Stem cell metabolomics), Project number 733051021 and Memorabel supported by ZonMW (Project number733050814). 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 authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists. The generation and management of GWAS genotype data for the Rotterdam Study (RS-I, RS-II, RS-III) was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. The GWAS datasets are supported by the Netherlands Organization of Scientific Research NWO Investments (Project number 175.010.2005.011, 911-03-012), the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/ Netherlands Organization for Scientific Research (NWO) Netherlands Consortium for Healthy Aging (NCHA), project number 050-060-810. We thank Pascal Arp, Mila Jhamai, Marijn

Verkerk, Lizbeth Herrera and Marjolein Peters, MSc, and Carolina Medina-Gomez, MSc, for their help in creating the GWAS database, and Karol Estrada, PhD, Yurii Aulchenko, PhD, and Carolina Medina-Gomez, MSc, for the creation and analysis of imputed data.

Amsterdam Dementia Cohort
The Alzheimer Center Amsterdam is supported by Stichting Alzheimer Nederland and Stichting VUmc fonds. The clinical database structure was developed with funding from Stichting Dioraphte. CSF analysis were supported by the Weston Brain Institute and Stichting Alzheimer Nederland.

Competing interests
S.A., M.C.M., A.D., R.A., M.E.S., A.R., T.H., E.S., N.M., P.S., M.A.I., N.A., C.M.D. declare no competing interests. O.H. has acquired research support (for the institution) from Roche, Pfizer, GE Healthcare, Biogen, AVID Radiopharmaceuticals and Euroimmun. In the past 2 years, O.H. has received consultancy/speaker fees (paid to the institution) from Biogen and Roche. Dr. Nikolaos Giagtzoglou was an employee at Biogen Idec, USA. Dr. Alfredo Cabrera-Socorro is an employee of Janssen Pharmaceutical NV, Turnhoutseweg 30, 2340 Beerse, Belgium. Dr. Margot H.M. Bakker is a full time employee of AbbVie GmbH & Co KG and owns AbbVie stock. AbbVie contributed to the study design, research, and interpretation of data, reviewing, and approving the publication. Prof. dr. Teunissen has functioned in advisory boards of Roche, received non-financial support in the form of research consumables from ADxNeurosciences and Euroimmun, performed contract research or received grants from Probiodrug, Biogen, Esai, Toyama, Janssen prevention center, Boehringer, AxonNeurosciences, EIP farma, PeopleBio, Roche.Research programs of Prof. dr. Wiesje van der Flier have been funded by ZonMW, NWO, EU-FP7, Alzheimer Nederland, CardioVascular Onderzoek Nederland, stichting Dioraphte, Gieskes-Strijbis fonds, Pasman stichting, Biogen MA Inc, Boehringer Ingelheim, Piramal Neuroimaging, Roche BV, Janssen Stellar, Combinostics. WF holds the Pasman chair. WF has performed contract research for Biogen MA Inc and Boehringer Ingelheim. WF has been an invited speaker at Boehringer Ingelheim and Biogen MA Inc. All funding is paid to her institution.

References

1. Au, R., Piers, R.J. & Lancashire, L. Back to the future: Alzheimer's disease heterogeneity revisited. *Alzheimers Dement (Amst)* **1**, 368-370 (2015).
2. 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).
3. Corder, E.H. *et al.* Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. *Nat Genet* **7**, 180-4 (1994).
4. Seshadri, S., Drachman, D.A. & Lippa, C.F. Apolipoprotein E epsilon 4 allele and the lifetime risk of Alzheimer's disease. What physicians know, and what they should know. *Arch Neurol* **52**, 1074-9 (1995).
5. van der Lee, S.J. *et al.* The effect of APOE and other common genetic variants on the onset of Alzheimer's disease and dementia: a community-based cohort study. *Lancet Neurol* **17**, 434-444 (2018).
6. Fernandez, C.G., Hamby, M.E., McReynolds, M.L. & Ray, W.J. The Role of APOE4 in Disrupting the Homeostatic Functions of Astrocytes and Microglia in Aging and Alzheimer's Disease. *Frontiers in aging neuroscience* **11**, 14 (2019).
7. Sunderland, T. *et al.* Cerebrospinal fluid beta-amyloid1-42 and tau in control subjects at risk for Alzheimer's disease: the effect of APOE epsilon4 allele. *Biol Psychiatry* **56**, 670-6 (2004).
8. de la Torre, J.C. Is Alzheimer's disease a neurodegenerative or a vascular disorder? Data, dogma, and dialectics. *Lancet Neurol* **3**, 184-90 (2004).
9. Larsson, S.C. *et al.* Modifiable pathways in Alzheimer's disease: Mendelian randomisation analysis. *BMJ* **359**, j5375 (2017).
10. Henstridge, C.M., Hyman, B.T. & Spires-Jones, T.L. Beyond the neuron–cellular interactions early in Alzheimer disease pathogenesis. *Nature Reviews Neuroscience*, 1 (2019).
11. Kunkle, B.W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet* **51**, 414-430 (2019).
12. Colonna, M. TREMs in the immune system and beyond. *Nat Rev Immunol* **3**, 445-53 (2003).
13. Zlokovic, B.V. Neurovascular pathways to neurodegeneration in Alzheimer's disease and other disorders. *Nature Reviews Neuroscience* **12**, 723 (2011).
14. Katusic, Z.S. & Austin, S.A. Endothelial nitric oxide: protector of a healthy mind. *European heart journal* **35**, 888-894 (2013).
15. Koizumi, K., Wang, G. & Park, L. Endothelial dysfunction and amyloid-β-induced neurovascular alterations. *Cellular and molecular neurobiology* **36**, 155-165 (2016).
16. Han, B.H. *et al.* Cerebrovascular dysfunction in amyloid precursor protein transgenic mice: contribution of soluble and insoluble amyloid-β peptide, partial restoration via γ-secretase inhibition. *Journal of Neuroscience* **28**, 13542-13550 (2008).
17. Tong, X.-K., Lecrux, C. & Hamel, E. Age-dependent rescue by simvastatin of Alzheimer's disease cerebrovascular and memory deficits. *Journal of neuroscience* **32**, 4705-4715 (2012).

18. Park, L. *et al.* Brain Perivascular Macrophages Initiate the Neurovascular Dysfunction of Alzheimer Abeta Peptides. *Circ Res* **121**, 258-269 (2017).
19. Hye, A. *et al.* Proteome-based plasma biomarkers for Alzheimer's disease. *Brain* **129**, 3042-50 (2006).
20. Ray, S. *et al.* Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins. *Nature medicine* **13**, 1359 (2007).
21. Soares, H.D. *et al.* Plasma biomarkers associated with the apolipoprotein E genotype and Alzheimer disease. *Archives of neurology* **69**, 1310-1317 (2012).
22. Ringman, J.M. *et al.* Plasma signaling proteins in persons at genetic risk for Alzheimer disease: influence of APOE genotype. *Archives of neurology* **69**, 757-764 (2012).
23. Hu, W.T. *et al.* Plasma multianalyte profiling in mild cognitive impairment and Alzheimer disease. *Neurology* **79**, 897-905 (2012).
24. Llano, D.A., Devanarayan, V., Simon, A.J. & Alzheimer's Disease Neuroimaging, I. Evaluation of plasma proteomic data for Alzheimer disease state classification and for the prediction of progression from mild cognitive impairment to Alzheimer disease. *Alzheimer Disease & Associated Disorders* **27**, 233-243 (2013).
25. Whelan, C.D. *et al.* Multiplex proteomics identifies novel CSF and plasma biomarkers of early Alzheimer's disease. *Acta neuropathologica communications* **7**, 1-14 (2019).
26. Blankenburg, M., Haberland, L., Elvers, H.D., Tannert, C. & Jandrig, B. High-Throughput Omics Technologies: Potential Tools for the Investigation of Influences of EMF on Biological Systems. *Curr Genomics* **10**, 86-92 (2009).
27. Bell, R.D. *et al.* Apolipoprotein E controls cerebrovascular integrity via cyclophilin A. *Nature* **485**, 512-6 (2012).
28. Phizicky, E., Bastiaens, P.I., Zhu, H., Snyder, M. & Fields, S. Protein analysis on a proteomic scale. *Nature* **422**, 208-15 (2003).
29. Emilsson, V. *et al.* Co-regulatory networks of human serum proteins link genetics to disease. *Science* **361**, 769-773 (2018).
30. Sun, B.B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73-79 (2018).
31. Gold, L. *et al.* Aptamer-based multiplexed proteomic technology for biomarker discovery. *Nature Precedings*, 1-1 (2010).
32. Zhao, B. *et al.* Genome-wide association analysis of 19,629 individuals identifies variants influencing regional brain volumes and refines their genetic co-architecture with cognitive and mental health traits. *Nature genetics* **51**, 1637-1644 (2019).
33. Harper, L. *et al.* Patterns of atrophy in pathologically confirmed dementias: a voxelwise analysis. *Journal of Neurology, Neurosurgery & Psychiatry* **88**, 908-916 (2017).
34. Sutphen, C.L. *et al.* Longitudinal decreases in multiple cerebrospinal fluid biomarkers of neuronal injury in symptomatic late onset Alzheimer's disease. *Alzheimers Dement* **14**, 869-879 (2018).
35. Hesse, C. *et al.* Transient increase in total tau but not phospho-tau in human cerebrospinal fluid after acute stroke. *Neuroscience letters* **297**, 187-190 (2001).

36. Lv, J. *et al.* Quantitative proteomics reveals that PEA15 regulates astroglial Abeta phagocytosis in an Alzheimer's disease mouse model. *J Proteomics* **110**, 45-58 (2014).
37. Ye, L. *et al.* Abeta seeding potency peaks in the early stages of cerebral beta-amyloidosis. *EMBO Rep* **18**, 1536-1544 (2017).
38. Asada-Utsugi, M. *et al.* N-cadherin enhances APP dimerization at the extracellular domain and modulates Aβ production. *Journal of neurochemistry* **119**, 354-363 (2011).
39. Henriques, A.G. *et al.* Altered protein phosphorylation as a resource for potential AD biomarkers. *Sci Rep* **6**, 30319 (2016).
40. Gul, I.S., Hulpiau, P., Saeys, Y. & van Roy, F. Evolution and diversity of cadherins and catenins. *Exp Cell Res* **358**, 3-9 (2017).
41. Fannon, A.M. & Colman, D.R. A model for central synaptic junctional complex formation based on the differential adhesive specificities of the cadherins. *Neuron* **17**, 423-34 (1996).
42. Obst-Pernberg, K. & Redies, C. Cadherins and synaptic specificity. *J Neurosci Res* **58**, 130-8 (1999).
43. Yamagata, M., Duan, X. & Sanes, J.R. Cadherins Interact With Synaptic Organizers to Promote Synaptic Differentiation. *Front Mol Neurosci* **11**, 142 (2018).
44. Arikkath, J. & Reichardt, L.F. Cadherins and catenins at synapses: roles in synaptogenesis and synaptic plasticity. *Trends Neurosci* **31**, 487-94 (2008).
45. Sancisi, V. *et al.* Cadherin 6 is a new RUNX2 target in TGF-beta signalling pathway. *PLoS One* **8**, e75489 (2013).
46. Clay, M.R. & Halloran, M.C. Cadherin 6 promotes neural crest cell detachment via F-actin regulation and influences active Rho distribution during epithelial-to-mesenchymal transition. *Development* **141**, 2506-15 (2014).
47. Montagne, A. *et al.* Blood-brain barrier breakdown in the aging human hippocampus. *Neuron* **85**, 296-302 (2015).
48. Inoue, T., Tanaka, T., Suzuki, S.C. & Takeichi, M. Cadherin-6 in the developing mouse brain: expression along restricted connection systems and synaptic localization suggest a potential role in neuronal circuitry. *Dev Dyn* **211**, 338-51 (1998).
49. Kuhn, P.H. *et al.* Systematic substrate identification indicates a central role for the metalloprotease ADAM10 in axon targeting and synapse function. *Elife* **5**(2016).
50. Malinverno, M. *et al.* Synaptic localization and activity of ADAM10 regulate excitatory synapses through N-cadherin cleavage. *Journal of Neuroscience* **30**, 16343-16355 (2010).
51. Marambaud, P. *et al.* A CBP binding transcriptional repressor produced by the PS1/ϵ-cleavage of N-cadherin is inhibited by PS1 FAD mutations. *Cell* **114**, 635-645 (2003).
52. Andreyeva, A. *et al.* C-terminal fragment of N-cadherin accelerates synapse destabilization by amyloid-beta. *Brain* **135**, 2140-54 (2012).
53. Hu, N.W. *et al.* Extracellular Forms of Abeta and Tau from iPSC Models of Alzheimer's Disease Disrupt Synaptic Plasticity. *Cell Rep* **23**, 1932-1938 (2018).
54. Togashi, H. *et al.* Cadherin regulates dendritic spine morphogenesis. *Neuron* **35**, 77-89 (2002).
55. Chen, F. *et al.* Role for glyoxalase I in Alzheimer's disease. *Proc Natl Acad Sci U S A* **101**, 7687-92 (2004).

56. Thornalley, P.J. The glyoxalase system: new developments towards functional characterization of a metabolic pathway fundamental to biological life. *Biochem J* **269**, 1-11 (1990).
57. Xu, Y. & Chen, X. Glyoxalase II, a detoxifying enzyme of glycolysis byproduct methylglyoxal and a target of p63 and p73, is a pro-survival factor of the p53 family. *J Biol Chem* **281**, 26702-13 (2006).
58. Desai, K.M. *et al.* Oxidative stress and aging: is methylglyoxal the hidden enemy? *Can J Physiol Pharmacol* **88**, 273-84 (2010).
59. Takeuchi, M. *et al.* Involvement of advanced glycation end-products (AGEs) in Alzheimer's disease. *Curr Alzheimer Res* **1**, 39-46 (2004).
60. Ko, S.Y., Lin, Y.P., Lin, Y.S. & Chang, S.S. Advanced glycation end products enhance amyloid precursor protein expression by inducing reactive oxygen species. *Free Radic Biol Med* **49**, 474-80 (2010).
61. Wang, J.Z., Wang, Z.H. & Tian, Q. Tau hyperphosphorylation induces apoptotic escape and triggers neurodegeneration in Alzheimer's disease. *Neurosci Bull* **30**, 359-66 (2014).
62. More, S.S., Vartak, A.P. & Vince, R. Restoration of glyoxalase enzyme activity precludes cognitive dysfunction in a mouse model of Alzheimer's disease. *ACS Chem Neurosci* **4**, 330-8 (2013).
63. Kuhla, B. *et al.* Age- and stage-dependent glyoxalase I expression and its activity in normal and Alzheimer's disease brains. *Neurobiol Aging* **28**, 29-41 (2007).
64. Shea, T.B., Rogers, E., Ashline, D., Ortiz, D. & Sheu, M.S. Apolipoprotein E deficiency promotes increased oxidative stress and compensatory increases in antioxidants in brain tissue. *Free Radic Biol Med* **33**, 1115-20 (2002).
65. Baldeiras, I. *et al.* Peripheral oxidative damage in mild cognitive impairment and mild Alzheimer's disease. *J Alzheimers Dis* **15**, 117-28 (2008).
66. Guidi, I. *et al.* Oxidative imbalance in patients with mild cognitive impairment and Alzheimer's disease. *Neurobiol Aging* **27**, 262-9 (2006).
67. Hofman, A. *et al.* The Rotterdam Study: 2016 objectives and design update. *Eur J Epidemiol* **30**, 661-708 (2015).
68. Folstein, M.F., Folstein, S.E. & McHugh, P.R. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12**, 189-98 (1975).
69. Copeland, J.R. *et al.* A semi-structured clinical interview for the assessment of diagnosis and mental state in the elderly: the Geriatric Mental State Schedule. I. Development and reliability. *Psychol Med* **6**, 439-49 (1976).
70. Roth, M. *et al.* CAMDEX. A standardised instrument for the diagnosis of mental disorder in the elderly with special reference to the early detection of dementia. *Br J Psychiatry* **149**, 698-709 (1986).
71. McKhann, G. *et al.* Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* **34**, 939-44 (1984).
72. 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).

73. Brann, E. *et al.* Inflammatory markers in late pregnancy in association with postpartum depression-A nested case-control study. *Psychoneuroendocrinology* **79**, 146-159 (2017).
74. Slooter, A.J. *et al.* Apolipoprotein E genotype and progression of Alzheimer's disease: the Rotterdam Study. *J Neurol* **246**, 304-8 (1999).
75. Slooter, A.J. *et al.* Risk estimates of dementia by apolipoprotein E genotypes from a population-based incidence study: the Rotterdam Study. *Arch Neurol* **55**, 964-8 (1998).
76. Roshchupkin, G.V. *et al.* Fine-mapping the effects of Alzheimer's disease risk loci on brain morphology. *Neurobiol Aging* **48**, 204-211 (2016).
77. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-83 (2016).
78. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-7 (2016).
79. Delaneau, O., Marchini, J., Genomes Project, C. & Genomes Project, C. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat Commun* **5**, 3934 (2014).
80. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G.R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* **44**, 955-9 (2012).
81. Palmqvist, S. *et al.* Accuracy of brain amyloid detection in clinical practice using cerebrospinal fluid beta-amyloid 42: a cross-validation study against amyloid positron emission tomography. *JAMA Neurol* **71**, 1282-9 (2014).
82. van der Flier, W.M. & Scheltens, P. Amsterdam Dementia Cohort: Performing Research to Optimize Care. *J Alzheimers Dis* **62**, 1091-1111 (2018).
83. van der Flier, W.M. *et al.* Optimizing patient care and research: the Amsterdam Dementia Cohort. *Journal of Alzheimer's disease* **41**, 313-327 (2014).
84. Van Der Flier, W.M. & Scheltens, P. Amsterdam dementia cohort: performing research to optimize care. *Journal of Alzheimer's Disease* **62**, 1091-1111 (2018).
85. Reijs, B.L.R. *et al.* The Central Biobank and Virtual Biobank of BIOMARKAPD: a resource for studies on neurodegenerative diseases. *Frontiers in neurology* **6**, 216 (2015).
86. van der Flier, W.M. *et al.* Optimizing patient care and research: the Amsterdam Dementia Cohort. *J Alzheimers Dis* **41**, 313-27 (2014).
87. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300 (1995).
88. Zhan, X., Hu, Y., Li, B., Abecasis, G.R. & Liu, D.J. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics* **32**, 1423-6 (2016).
89. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826 (2017).
90. Consortium, G.T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-5 (2013).

Chapter 4.2

Association of lysophosphatidic acids with cerebrospinal fluid biomarkers and progression to Alzheimer's disease

Shahzad Ahmad, Adelina Orellana, Isabelle Kohler, Lutz Frölich, Itziar de Rojas, Silvia Gil, Mercè Boada, Isabel Hernández, Lucrezia Hausner, Margot H.M. Bakker, Alfredo Cabrera-Socorro, Najaf Amin, Alfredo Ramírez, Agustín Ruiz, Thomas Hankemeier*, Cornelia M. Van Duijn*

This Chapter is under revised submission for publication in Alzheimer's Research and Therapy

**Senior authors contributed equally to the work*

The supplementary information for this chapter is available at:
https://drive.google.com/open?id=1YjXy5cjuJ6X6tzuC3UpuUOu6vg-7aJmF

Abstract

Lysophosphatidic acids (LPAs) are bioactive signaling phospholipids that have been implicated in Alzheimer's disease (AD). It is largely unknown whether LPAs are associated with AD pathology and progression from mild cognitive impairment (MCI) to AD. The current study was performed on cerebrospinal fluid (CSF) and plasma samples of 182 MCI patients from two independent cohorts. We profiled LPA-derived metabolites using liquid chromatography-mass spectrometry. We evaluated the association of LPAs with CSF biomarkers of AD, Aβ-42, p-tau, and total tau levels overall and stratified by *APOE* genotype, and with MCI to AD progression. Five LPAs (C16:0, C16:1, C22:4, C22:6, and isomer-LPA C22:5) showed significant positive association with CSF biomarkers of AD, Aβ-42, p-tau, and total tau, while LPA C14:0 and C20:1 associated only with Aβ-42 and alkyl-LPA C18:1, LPA C20:1 associated with tau pathology biomarkers. Association of cyclic-LPA C16:0 and two LPAs (C20:4, C22:4) with Aβ-42 levels was found only *APOE* ε4 carriers. Furthermore, LPA C16:0 and C16:1 also showed association with MCI to AD dementia progression but results did not replicate in an independent cohort. Our findings provide evidence that LPAs may contribute to early AD pathogenesis. Future studies are needed to determine whether LPAs play a role in upstream of AD pathology or are downstream markers of neurodegeneration.

Background

Lipids play a key role in Alzheimer's disease (AD)[1-3]. Lysophosphatidic acids (LPAs) are bioactive phospholipids representing a significant class of signaling molecules[4]. LPAs regulate a plethora of downstream processes including brain immune response[5], myelination[6], synaptic transmission[7,8], synaptic plasticity[9], as well as in endothelial cells and neurovascular function[10]. A recent study has reported altered levels of LPA C18:2 in AD patients compared to controls in plasma[11] and LPAs have been implicated in amyloid-beta (Aβ) formation[12] and phosphorylation of tau[13] as well, the neuropathological hallmarks of AD. The LPAs may contribute to amyloid pathology, which is supported by their role in enhancing Aβ production through upregulation β-secretase expression[12]. Moreover, as a bioactive component of oxidized LDL (OxLDL), LPAs affect the integrity of blood-brain barrier[14] and are also involved in neuronal cell death[15,16]. The mounting evidence for the role of LPA metabolites as a mediator in AD related molecular process underline their importance in AD pathophysiology. Nevertheless, studies are lacking investigating the relationship between LPA metabolites with AD biomarkers of pathophysiology.

Encompassing a large group of related metabolites, LPA (1 or 2-acyl-sn-glycero-3-phosphate) metabolites comprises of an sn-glycerol-3 phosphate connected to a fatty acid[17-19]. Molecular species of LPAs differ based on their acyl chain length (C14 to C22) and degrees of saturation (C16:1, C18:1, C18:2, C20:3, C20:5, C20:4, C22:4, C22:6)[20]. For example, C18:1 denoting an acyl chain of 18 carbons with a double bond, while C18:2 denotes an acyl chain of 18 carbons with two double bonds[21,22]. This structural diversity of LPAs also imparts them with differential biological activity[23-25]. Differential biological activity of LPAs can also be attributed to their G-coupled-protein-receptors (GCPR) ranging from LPA_{1-5} which differ in their affinity and response to diverse LPA species[20,25]. This structure-activity relationship of LPAs may be relevant to their role in AD pathophysiology[26]. A comprehensive study of LPA metabolites in AD-related pathology is lacking[14], and the interaction of apolipoprotein E (*APOE)* has not been studied. It is also not known whether LPAs play a role in the progression from mild cognitive impairment (MCI) to AD dementia.

Our study aims to delineate the role of various LPA species in AD during the prodromal phase of AD, i.e., MCI. We hypothesize that cerebrospinal fluid (CSF) and plasma levels of LPAs may be associated to markers of AD pathology, including to Aβ-42, phosphorylated tau (p-tau) and total tau (t-tau) in MCI patients, and this association may be modified by *APOE gene*. We further hypothesized that LPAs might contribute to MCI to AD dementia progression. As prior information on which LPA species may be relevant for the hypothesis, we assessed a series of structurally different LPA metabolites in CSF and plasma.

Methods

Study populations

The current study was performed in cohorts participating in the Alzheimer's Disease Apolipoprotein Pathology for Treatment Elucidation and Development (ADAPTED) consortium including the Barcelona-based memory clinic Fundació ACE (142 CSF-plasma paired samples) and the Department of Geriatric Psychiatry at the Medical Faculty Mannheim, University of Heidelberg (40 CSF samples). Both participating studies are approved by the medical ethical committee of their respective institutes and informed consents were collected from all participants, which allow the use of phenotype and biomarker information for research purpose. From both participating cohorts, we selected MCI patients for which complete information was available on age at blood collection, sex, body mass index (BMI), lipid-lowering medication use, as well as AD biomarkers in CSF (i.e., Aβ-42, p-tau, and total tau).

Fundació ACE cohort

All the MCI patients from the Fundació ACE (ACE) cohort were recruited and assessed between 2016 to 2017 at the Memory Disorders Unit from Fundació ACE, Institut Català de Neurociènces Aplicades, Barcelona, Spain[27]. Each patient was assigned a diagnosis after consensus at a case conference attended by neurologists, neuropsychologists, and social workers. MCI patients fulfilled MCI Petersen's diagnostic criteria[28,29] including subjective memory complaints, decline from normal general cognition, preserved performance in activities of daily living, absence of dementia, and a measurable impairment in one or more cognitive functions, with or without deficit in other cognitive domains (amnestic MCI: single domain or amnestic MCI: multiple domain). At follow-up, dementia was defined according to the DSM-V criteria[30].The underlying etiologies of the cognitive deficits within the dementia group were classified according to the following criteria: the 2011 National Institute of Aging- Alzheimer´s Association (NIA-AA)[31] for Alzheimer´s disease; the National Institute of Neurological Disorder and Stroke and Association Internationale pour la Recherche et l'Enseignement in Neurosciences criteria(NINDS-AIREN)[32] for vascular dementia; Frontotemporal Dementia[33], and the for Lewy body dementia[34].

Paired samples of CSF and plasma were collected from patients under fasted conditions. CSF was obtained by lumbar puncture following the established consensus recommendations[35]. Briefly, the lumbar puncture (LP) was performed by an experienced neurologist with the patients in a sitting position. After local anesthesia (1% mepivacaine) was injected subcutaneously, CSF was obtained by LP in the intervertebral space of L3-L4. The fluid was collected passively in two 10-ml polypropylene tubes (Sarstedt ref 62610018). The first tube of CSF was analyzed for basic biochemistry (glucose, total proteins, proteinogram, and cell type and cell number). The second tube was centrifuged (2000xg 10 min at 4^0C), and the fluid was aliquoted into polypropylene tubes (Sarstedt ref 72694007) and stored at -80^0C until analysis. The time delay between CSF collection and storage was less than 2 hours. On the same day as the AD biomarker

analysis (Aβ-42, p-tau, and total tau), an aliquot was thawed at room temperature and vortexed for 5-10 seconds. CSF Aβ1-42, total tau, and p-tau levels were measured using commercially available enzyme-linked immunosorbent assays, namely Innotest Aβ1-42, Innotest hTAU Ag and Innotest PHOSPHO-TAU (181P) (Innotest, Fujirebio Europe)[35,36].

For *APOE* genotyping in the ACE cohort, genomic DNA was obtained from whole blood collected in BD Vacutainer tubes (K2-EDTA). DNA extraction was performed using DNA Chemagen technology (Perkin Elmer). Afterward, the *APOE* genotype was determined by TaqMan probes analysis in a system of Real-Time PCR QuantStudio3 (Thermofisher).

Heidelberg/Mannheim memory clinic sample

Forty MCI patients were recruited and assessed between 2012 to 2016 at the Memory Clinic of the Central Institute of Mental Health (Mannheim, Germany). Neuropsychiatric or general medical causes of impaired cognition were excluded by detailed medical history, physical and neuropsychiatric examination, and standard serum laboratory assessment. Thus, all MCI patients met the MCI Petersen's diagnostic criteria[28,29] including subjective memory complaints, normal general cognition, only minimally impaired performance in instrumental activities of daily living, absence of dementia, and a measurable impairment in one or more cognitive domains. Cognitive impairment was defined as performance below 1.2 standard deviation in one or more cognitive domains in standard neuropsychological test battery[37] (test battery of the Consortium to Establish a Registry for Alzheimer Disease (CERAD)[38] plus the Wechsler memory scale – logical memory (WMS) immediate and delayed recall[39], and the trail making test A (TMT-A) and B (TMT-B)[40]. For biomarker assessments, lumbar puncture was performed to determine amyloid pathology in CSF following the NIA/AA criteria for the diagnosis of MCI due to AD[41]. The results of the clinical assessment for each patient were discussed at a case conference attended by geriatric psychiatrists and neuropsychologists the diagnosis of MCI due to AD or prodromal AD[42] was assigned by consensus using all clinical and biomarker data (CSF Aβ-42, t-tau, and p-tau). Paired samples of CSF and plasma were collected from patients according to the established consensus recommendations[35]. Aliquots were stored in polypropylene tubes at -80°C. Aβ1–42, p-tau, and t-tau were performed in the Neurochemistry Laboratory at the Department of Neurology, University Medical School, Göttingen using established protocols. P-tau levels in CSF were quantified with a commercially available ELISA kit [INNOTEST® PHOSPHO-TAU(181P), Innogenetics]. Aβ1–42 was detected with a commercially available ELISA kit [INNOTEST®β- AMYLOID (1–42) Innogenetics] for quantitative analysis.

APOE genotyping in Heidelberg/Mannheim memory clinic sample was performed on an Illumina GSA1.0 SharedCustom Content bead array according to the manufacturer's instructions. GenomeStudio 2.0 software was used to determine *APOE* genotypes and results were exported in PLINK format.

Metabolomics profiling

All CSF and plasma samples of both cohorts were profiled for the same set of metabolites using a UHPLC-MS/MS approach targeting signaling lipid mediators including LPAs, alkyl-lysophosphatidic acid (aLPAs), and cyclic-lysophosphatidic acids (cLPAs) ranging from C14 to C22 acyl chain length[43].

Samples were stored at -80°C, thawed at room temperature, and randomized prior to analysis. Quality control (QC) samples, consisting of a pool of all samples, and blanks were also analyzed to ensure the quality of the obtained data. For CSF samples, 350 µL of samples were evaporated to dryness, spiked with isotopically-labeled internal standards and antioxidant (BHT:EDTA 1:1, 0.2 mg/mL), and reconstituted in two aliquots using a mixture of methanol:water (70:30, v/v). Plasma samples were first acidified through the addition of 0.2M citric acid and 0.1M disodium hydrogen phosphate buffer at pH 4.5. Metabolites were extracted using liquid-liquid extraction with a mixture of 1-butanol:ethyl acetate (1:1, v/v) prior to mixing, centrifugation, collection of the supernatant, evaporation and reconstitution into two aliquots with a mixture of ice-cold methanol:water (70:30, v/v).

Samples were measured using a Shimadzu LC-30AD system coupled to a LCMS-8050 Triple Quadrupole system (Shimadzu, Japan).

For both plasma and CSF samples, the first aliquot (high pH injection) was analyzed using a Kromasil EternityXT-1.8 C18 column, 2.1 × 50 mm, 1.8 µm (Akzo Nobel, Netherlands) with a mobile phase composed of (A) water with 5 mM ammonium acetate and 0.0625 % ammonium hydroxide and (B) 80 % acetonitrile with 20% isopropanol and 0.1 % ammonium hydroxide. For both matrices, the second aliquot (low pH injection) was analyzed using an Acquity BEH C18 column, 2.1 × 50 mm, 1.7 µm (Waters) with a mobile phase composed of (A) water with 0.1 % acetic acid, (B) 75 % acetonitrile with 25 % methanol and 0.1 % acetic acid, and (C) 100% isopropanol. For both pH injections, polarity switching and dynamic multiple reaction monitoring (dMRM) mode were used for MS acquisition.

To perform the quality control (QC), metabolites showing a relative standard deviation (RSD) higher than 30% on corrected peak areas in QC samples were excluded. After QC correction, 19 and 17 LPAs in CSF and plasma, respectively, were used for further data analysis (Supplementary Table 1). Common metabolites detected in both CSF and plasma included LPAs (C14:0, C14:1, C14:2, C16:0, C18:0, C18:1, C18:2, C20:1, C20:3, C20:5, C22:4, C22:5) and three cyclic-LPAs (C16:0, C18:0, C18:1). Metabolites only detected in CSF samples included some LPAs (C20:4, C22:6, C22:5) and an alkyl-LPA C16:1. LPA C18:3 and two cyclic-LPAs (C18:2, C20:4) were detected only in plasma samples. The inverse rank transformation was used to normalize the distribution of metabolites in both cohorts.

Statistical analysis

Association of LPAs with Aβ-42, p-tau, and t-tau

We performed linear regression analysis to test the association of Aβ-42, p-tau, and t-tau with the profiled metabolites in paired CSF and plasma samples from the ACE cohort and CSF samples from Heidelberg-Manheim memory clinic. Levels of Aβ-42, p-tau, and t-tau in CSF were used as the outcome variable in the regression model. Analyses were adjusted for age, sex, body mass index (BMI) and lipid-lowering medications. The inverse rank transformation was applied to normalize the distribution of both CSF AD biomarkers (Aβ-42, p-tau and t-tau) and LPA metabolite levels in CSF and plasma. A meta-analysis of the regression analysis results of the two cohorts was performed using METAL software[44] using the inverse-variance fixed-effect model. Meta-analysis results of associations were also corrected for multiple testing separately for each AD biomarker using false discovery rate (*FDR*) by the Benjamini and Hochberg method[45] and findings with $FDR < 0.05$ were considered significant in the overall analysis. All analyses were performed in R (https://www.r-project.org/). To test whether conversion from LPA to another was relevant, we have tested all ratios between LPAs.

APOE stratified regression analysis

To identify *APOE* specific associations of metabolites with AD biomarkers, *APOE* stratified analyses were performed in both participating cohorts based on three *APOE* strata including *APOE 44/34/24*, *APOE 33* and *APOE 22/23*. In the stratified analyses, subjects with *APOE 24* genotype were pooled with patients having *APOE 44/34* genotypes based on their similar risk profiles, as reported in an earlier study[46]. *APOE* stratified analyses results were reported as a combined meta-analysis of both datasets included in the current study (ACE cohort and Heidelberg/Manheim cohort). Due to the smaller number of *APOE 22/23* carriers in the two datasets, a combined regression analysis was performed, aggregating all *APOE 22/23* carriers from two cohorts while adjusting for cohort effects. Multiple testing correction was performed using the false-discovery rate *(FDR<0.05)* based on Benjamin and Hochberg[45].

To assess the association of the *APOE* genotype with LPAs, we compared levels of LPAs in CSF of *APOE* ε4 (*APOE 44/34/24*) and *APOE* ε2 (22/23) carriers versus *APOE* ε33 carriers using regression analysis adjusting for the age, sex, BMI and lipid-lowering medications. This regression analysis was conducted for each cohort and their combined meta-analysis.

MCI to AD dementia progression analysis

In the ACE cohort, follow-up information was available for 138 out of 142 MCI patients including 17 non-amnestic and 121 amnestic MCIs. A total of 43 MCI patients progressed into AD dementia (31%) during follow-up, while 95 MCI patients did not convert to AD dementia. The mean follow-up time in converters was 1.42 years (SD = 0.53) and 1.44 years (SD = 0.70) in non-converters. The rate of MCI to AD dementia progression in our sample is similar to other clinic-based studies[47]. We analyzed the association of LPAs with MCI to AD dementia

progression using the cox proportional hazard model adjusting for age at blood collection, sex, BMI and lipid-lowering medication use. In the ACE cohort, 11 MCI patients also progressed to other types of dementia including vascular dementia (n = 6), semantic dementia (n = 1), Parkinson dementia (n = 1), Lewy Body dementia (n = 2) and frontal temporal dementia (n = 1). We repeated the conversion analysis in the Heidelberg/Mannheim samples. Among the 40 MCIs, 23 converted to AD dementia. Mean follow-up time in the Heidelberg/Mannheim samples were 1.80 years (SD = 1.06). Three MCI patients also progressed to frontal temporal dementia in this sample.

Association of cognitive measures with LPA levels

We also assessed the association of cognitive measures, MMSE and CDR with LPAs levels in CSF of both ACE cohort and Heidelberg/Mannheim samples. We used linear regression analysis adjusted for age, sex, BMI and lipid lowering medication. Results were meta-analyzed using METAL software[44] using the inverse-variance fixed-effect model and multiple testing was performed using false discovery rate (*FDR*) by the Benjamini and Hochberg method[45].

Results

General characteristics

The general characteristic of the ACE and Heidelberg/Mannheim cohorts are provided in Table 1. The patients of the ACE cohort of Barcelona are on average three years older (P = 0.042) than the Heidelberg/Mannheim samples. The proportion of women is similar between the two cohorts (P = 0.747). The proportion of patients treated with lipid-lowering medication in the ACE cohort (44%) is 1.6 times (P = 0.055) higher compared to that in the Heidelberg/Mannheim series of patients. The levels of Aβ-42, p-tau, and t-tau in CSF between the two cohorts were not significantly different. In terms of basic cognitive measures, the Mini-Mental State Examination (MMSE) score ($P = 2.67 \times 10^{-3}$) and clinical dementia rating (CDR) score (P = 0.047) was significantly higher in Heidelberg/Mannheim samples compared to ACE cohort.

Association of LPAs with CSF Aβ-42, p-tau, and total tau

Findings of the association of the metabolites with Aβ-42 levels in CSF are provided in Table 2. In a meta-analysis, eight LPAs including C18:1 (β = 0.281, $P = 1.65 \times 10^{-3}$), C16:1 (β = 0.242, $P = 1.88 \times 10^{-3}$), C16:0 (β = 0.234, $P = 2 \times 10^{-3}$), C22:6 (β = 0.231, $P = 2.91 \times 10^{-3}$), C14:0 (β = 0.223, $P = 6.04 \times 10^{-3}$), C22:4 (β = 0.191, $P = 1.20 \times 10^{-2}$), C20:4 (β = 0.203, $P = 1.44 \times 10^{-2}$) and isomer-LPA C22:5 (β = 0.183, $P = 1.78 \times 10^{-2}$) showed positive association with Aβ-42 levels in CSF. The effect estimates (β) of all associated LPAs were in the same direction in both cohorts.

Six LPAs, C20:1 (β = 0.347, $P = 7.11 \times 10^{-6}$), isomer-LPA C22:5 (β = 0.328, $P = 8.68 \times 10^{-6}$), C22:6 (β = 0.270, $P = 4.03 \times 10^{-4}$), C16:0 (β = 0.230, $P = 2.26 \times 10^{-3}$), C16:1 (β = 0.206, P =

Table 1: Population descriptive

	ACE cohort	Heidelberg/Mannheim cohort	*P-value* of difference
MCI patients (N)	142	40	
Metabolomics profiling tissue	CSF and Plasma	CSF	
Age (SD) blood collection, years	71.94(7.74)	68.85(8.50)	0.042
Female (%)	74(52%)	22(55%)	0.747
Body Mass index (SD)	26.46(3.74)	25.85(3.61)	0.353
Lipid lowering medication user (%)	63(44%)	11(27%)	0.055
Amyloid-beta 42 in pg/mL (SD)	791.59 (337.36)	690.84 (397.13)	0.151
P-Tau in pg/mL (SD)	71.37 (37.30)	63.17 (29.96)	0.153
Total tau in pg/mL (SD)	478.82 (253.45)	380.95 (326.97)	0.124
MMSE	24.93(4.07)	26.55(2.51)	$2.67x10^{-3}$
CDR	0.50 (0.06)	0.55 (0.15)	0.047
APOE genotype N (%)			
APOE 44/34/24	50 (35%)	18(45%)	
APOE 33	81(57%)	18(45%)	
APOE 22/23	11(8%)	4(10%)	

Abbreviations: MCI, mild cognitive impairment; SD, Standard deviation; CSF, Cerebrospinal fluid; MMSE, the Mini-Mental State Examination; CDR, clinical dementia rating; *APOE*, apolipoprotein E gene

$8.23x10^{-3}$) and C22:4 (β = 0.186, $P = 1.39x10^{-2}$) showed significant association (*FDR* < 0.05) with p-tau levels in CSF. In terms of the direction of effects, the regression coefficients (β) were very similar across two cohorts (Table 3).

Findings for t-tau levels in CSF were very similar as those for p-tau levels (Table 4), with five LPAs including C20:1, isomer-LPA C22:5, C22:6, C16:0, C16:1, C22:4, C18:1 and alkyl-LPA C18:1 showing significant positive association (*FDR* < 0.05) with total tau levels in CSF. Among the identified metabolites, LPA C18:1 and alkyl-LPA C18:1 showed association with only total tau but not with p-tau levels.

Figure 1 A shows a heatmap comparing Aβ-42, p-tau, and t-tau in CSF. The pattern of association is very similar for amyloid and tau biomarkers except for two LPAs (C14:0, C20:4) showing association to only Aβ-42, and an alkyl-LPA C18:1 showing a unique relation to total tau levels in CSF. Correlation plots are provided in Figure 1 for five LPAs (C16:0. C16:1, C22:4, C22:6, isomer-LPA C22:5), which showed significant association with all three AD biomarkers (Aβ-42, p-tau, and t-tau). The analysis of the ratios did not yield any significant findings.

Table 2: Association of cerebrospinal fluid (CSF) level of metabolites with Aβ-42 levels in CSF

	ACE cohort			Heidelberg/Mannheim samples			Meta-analysis				
	β	**SE**	***P-value***	**β**	**SE**	***P-value***	**β**	**SE**	**Direction**	***P-value***	***FDR***
LPA C18:1	0.286	0.100	$4.85x10^{-3}$	0.260	0.199	$2.00x10^{-1}$	0.281	0.089	++	$1.65x10^{-3}$	$1.26x10^{-2}$
LPA C16:1	0.264	0.087	$2.81x10^{-3}$	0.150	0.175	$3.97x10^{-1}$	0.242	0.078	++	$1.88x10^{-3}$	$1.26x10^{-2}$
LPA C16:0	0.249	0.081	$2.61x10^{-3}$	0.134	0.209	$5.27x10^{-1}$	0.234	0.076	++	$2.00x10^{-3}$	$1.26x10^{-2}$
LPA C22:6	0.199	0.086	$2.19x10^{-2}$	0.372	0.181	$4.76x10^{-2}$	0.231	0.078	++	$2.91x10^{-3}$	$1.38x10^{-2}$
LPA C14:0	0.220	0.092	$1.76x10^{-2}$	0.232	0.175	$1.92x10^{-1}$	0.223	0.081	++	$6.04x10^{-3}$	$2.29x10^{-2}$
LPA C22:4	0.150	0.084	$7.59x10^{-2}$	0.393	0.184	$4.02x10^{-2}$	0.191	0.076	++	$1.20x10^{-2}$	$3.79x10^{-2}$
LPA C20:4	0.211	0.096	$2.91x10^{-2}$	0.178	0.165	$2.89x10^{-1}$	0.203	0.083	++	$1.44x10^{-2}$	$3.90x10^{-2}$
Isomer-LPA C22:5	0.208	0.086	$1.61x10^{-2}$	0.071	0.177	$6.90x10^{-1}$	0.183	0.077	++	$1.78x10^{-2}$	$4.23x10^{-2}$
LPA C18:0	0.172	0.090	$5.65x10^{-2}$	0.211	0.176	$2.39x10^{-1}$	0.180	0.080	++	$2.39x10^{-2}$	$5.05x10^{-2}$
LPA C18:2	0.189	0.083	$2.51x10^{-2}$	0.028	0.170	$8.69x10^{-1}$	0.158	0.075	++	$3.50x10^{-2}$	$6.65x10^{-2}$
cLPA C18:1	0.170	0.093	$7.03x10^{-2}$	0.055	0.186	$7.68x10^{-1}$	0.147	0.083	++	$7.77x10^{-2}$	$1.34x10^{-1}$
LPA C22:5	0.112	0.088	$2.04x10^{-1}$	0.230	0.179	$2.09x10^{-1}$	0.135	0.079	++	$8.70x10^{-2}$	$1.35x10^{-1}$
LPA C20:1	0.164	0.092	$7.59x10^{-2}$	0.038	0.175	$8.29x10^{-1}$	0.137	0.081	++	$9.21x10^{-2}$	$1.35x10^{-1}$
cLPA C16:0	0.199	0.105	$5.91x10^{-2}$	-0.046	0.187	$8.06x10^{-1}$	0.141	0.091	+-	$1.23x10^{-1}$	$1.67x10^{-1}$
aLPA C18:1	0.335	0.127	$9.31x10^{-3}$	-0.225	0.174	$2.04x10^{-1}$	0.140	0.103	+-	$1.72x10^{-1}$	$2.17x10^{-1}$
cLPA C18:0	0.085	0.091	$3.54x10^{-1}$	0.036	0.193	$8.55x10^{-1}$	0.076	0.082	++	$3.58x10^{-1}$	$4.25x10^{-1}$
aLPA C16:1	0.199	0.132	$1.35x10^{-1}$	-0.106	0.166	$5.30x10^{-1}$	0.081	0.103	+-	$4.33x10^{-1}$	$4.84x10^{-1}$
LPA C20:3	0.124	0.103	$2.29x10^{-1}$	-0.104	0.169	$5.41x10^{-1}$	0.062	0.088	+-	$4.77x10^{-1}$	$5.04x10^{-1}$
LPA C20:5	-0.114	0.090	$2.08x10^{-1}$	0.253	0.185	$1.80x10^{-1}$	-0.044	0.081	-+	$5.88x10^{-1}$	$5.88x10^{-1}$

Abbreviations: LPA, lysophosphatidic acid; cLPA, cyclic lysophosphatidic acid; aLPA, alkyl-Lysophosphatidic acid; SE, standard error; *FDR*, false discovery rate
Note: Direction column indicates the direction of regression co-efficient of association in the ACE and Heidelberg/Mannheim cohort respectively

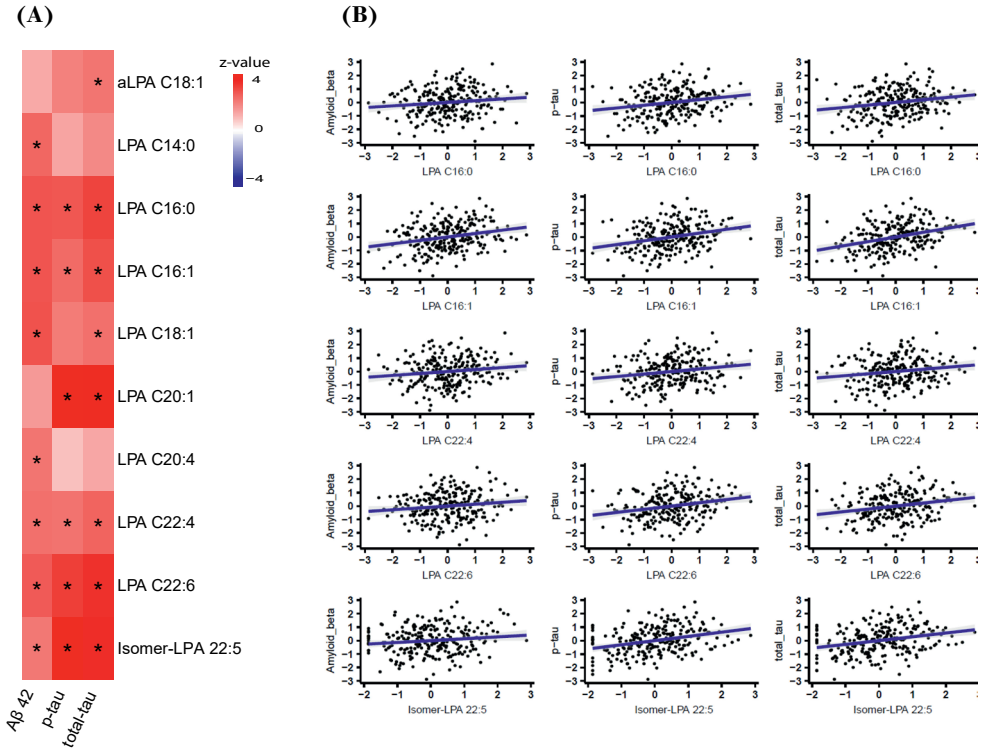


Figure 1: (A) Heatmap of the overall meta-analysis of regression analysis results of metabolites with Aβ-42, p-tau and t-tau levels in cerebrospinal fluid (CSF). Note: Star indicates significant association with false discovery rate <0.05. (B) Scatter plots of correlation between LPAs (C16:0. C16:1, C22:4, C22:6, isomer-LPA C22) with Aβ-42, p-tau and t-tau levels in CSF.

In the regression analysis of LPAs in plasma (Supplementary Table 2), no significant association was found with Aβ-42, p-tau or with t-tau in ACE cohort. Correlation analysis between CSF and plasma levels of LPAs for which paired samples were available (Supplementary Figure 1) shows that levels of 8 LPAs were significantly correlated between CSF and plasma: cyclic-LPA C16:0 (R = -0.28, P = 9.4x10^{-4}), cyclic-LPA C18:1 (R = -0.18, P = 0.029), LPA C18:1 (R = 0.44, P = 4.5x10^{-8}), LPA C20:1 (R = -0.23, P = 6.2x10^{-3}) and LPA C20:5 (R = 0.4, P = 1.2x10^{-6}).

The role of *APOE* in the association between LPA levels and CSF AD biomarkers

Results of meta-analysis of association results of *APOE* stratified analyses are provided in Figure 2. *APOE* stratified analysis of Aβ-42 (Figure 2: A) showed that LPAs species which showed significant association in the overall meta-analysis (C14:0, C16:0, C16:1, C18:1, C20:4, C22:4, C22:6, isomer-LPA C22:5) were restricted to *APOE* ε33 (LPA C16:0, C18:0, C22:6, isomer-LPA C22:5) and *APOE* ε4 (LPA C20:4, C22:4) carriers while in *APOE* 22/23 carriers the association was not significant and in the opposite direction. Based on statistical significance, we observed a few unique associations in the *APOE4* stratum involving a cyclic-LPA C16:0 and in the *APOE33* stratum LPAs C18:0, which did not show significant association in overall regression analysis.

Table 3: Association of cerebrospinal fluid (CSF) level of metabolites with p-tau levels in CSF

	ACE cohort			Heidelberg/Mannheim samples			Meta-analysis				
	β	SE	*P-value*	β	SE	*P-value*	β	SE	Direction	*P-value*	*FDR*
LPA C20:1	0.350	0.086	$8.16x10^{-5}$	0.333	0.174	$6.42x10^{-2}$	0.347	0.077	++	$7.11x10^{-6}$	$8.25x10^{-5}$
Isomer-LPA C22:5	0.302	0.082	$3.15x10^{-4}$	0.438	0.170	$1.46x10^{-2}$	0.328	0.074	++	$8.68x10^{-6}$	$8.25x10^{-5}$
LPA C22:6	0.279	0.083	$9.39x10^{-4}$	0.214	0.198	$2.88x10^{-1}$	0.270	0.076	++	$4.03x10^{-4}$	$2.55x10^{-3}$
LPA C16:0	0.218	0.080	$7.67x10^{-3}$	0.316	0.214	$1.48x10^{-1}$	0.230	0.075	++	$2.26x10^{-3}$	$1.07x10^{-2}$
LPA C16:1	0.226	0.086	$9.49x10^{-3}$	0.112	0.182	$5.43x10^{-1}$	0.206	0.078	++	$8.23x10^{-3}$	$3.13x10^{-2}$
LPA C22:4	0.168	0.082	$4.17x10^{-2}$	0.287	0.196	$1.53x10^{-1}$	0.186	0.076	++	$1.39x10^{-2}$	$4.40x10^{-2}$
LPA C18:1	0.217	0.099	$3.08x10^{-2}$	0.171	0.212	$4.25x10^{-1}$	0.209	0.090	++	$2.04x10^{-2}$	$5.54x10^{-2}$
aLPA C18:1	0.110	0.128	$3.91x10^{-1}$	0.425	0.169	$1.69x10^{-2}$	0.224	0.102	++	$2.79x10^{-2}$	$6.62x10^{-2}$
cLPA C18:0	0.159	0.089	$7.53x10^{-2}$	0.243	0.197	$2.26x10^{-1}$	0.173	0.081	++	$3.23x10^{-2}$	$6.82x10^{-2}$
LPA C20:3	0.072	0.101	$4.79x10^{-1}$	0.380	0.163	$2.62x10^{-2}$	0.157	0.086	++	$6.77x10^{-2}$	$1.29x10^{-1}$
cLPA C18:1	0.115	0.092	$2.14x10^{-1}$	0.265	0.189	$1.69x10^{-1}$	0.144	0.083	++	$8.25x10^{-2}$	$1.42x10^{-1}$
LPA C14:0	0.129	0.091	$1.60x10^{-1}$	0.088	0.187	$6.42x10^{-1}$	0.121	0.082	++	$1.40x10\text{-}^{1}$	$2.12x10^{-1}$
LPA C22:5	0.065	0.087	$4.56x10^{-1}$	0.326	0.184	$8.51x10^{-2}$	0.112	0.078	++	$1.52x10^{-1}$	$2.12x10^{-1}$
aLPA C16:1	0.061	0.131	$6.42x10^{-1}$	0.287	0.168	$9.65x10^{-2}$	0.146	0.103	++	$1.56x10^{-1}$	$2.12x10^{-1}$
LPA C18:0	0.096	0.089	$2.82x10^{-1}$	0.077	0.188	$6.86x10^{-1}$	0.092	0.080	++	$2.50x10^{-1}$	$3.16x10^{-1}$
LPA C18:2	-0.094	0.083	$2.60x10^{-1}$	-0.024	0.175	$8.90x10^{-1}$	-0.081	0.075	--	$2.79x10^{-1}$	$3.30x10^{-1}$
cLPA C16:0	0.102	0.104	$3.28x10^{-1}$	0.075	0.196	$7.04x10^{-1}$	0.096	0.092	++	$2.95x10^{-1}$	$3.30x10^{-1}$
LPA C20:4	0.077	0.095	$4.20x10^{-1}$	0.081	0.176	$6.49x10^{-1}$	0.078	0.084	++	$3.53x10^{-1}$	$3.72x10^{-1}$
LPA C20:5	-0.028	0.088	$7.54x10^{-1}$	0.114	0.193	$5.59x10^{-1}$	-0.003	0.081	-+	$9.68x10^{-1}$	$9.68x10^{-1}$

Abbreviations: LPA, lysophosphatidic acid; cLPA, cyclic lysophosphatidic acid; aLPA, alkyl-Lysophosphatidic acid; SE, standard error; *FDR*, false discovery rate
Note: Direction column indicates the direction of regression co-efficient of association in the ACE and Heidelberg/Mannheim cohort respectively

Table 4: Association of cerebrospinal fluid (CSF) level of metabolites with t-tau levels in CSF

	ACE cohort			Heidelberg/Mannheim samples			Meta-analysis				
	β	SE	*P-value*	β	SE	*P-value*	β	SE	Direction	*P-value*	*FDR*
LPA C20:1	0.318	0.085	$2.89x10^{-4}$	0.391	0.181	$3.76x10^{-2}$	0.331	0.077	++	$1.79x10^{-5}$	$3.40x10^{-4}$
Isomer-LPA C22:5	0.271	0.081	$1.05x10^{-3}$	0.390	0.184	$4.14x10^{-2}$	0.290	0.074	++	$8.88x10^{-5}$	$8.07x10^{-4}$
LPA C22:6	0.268	0.081	$1.22x10^{-3}$	0.409	0.200	$4.84x10^{-2}$	0.288	0.075	++	$1.27x10^{-4}$	$8.07x10^{-4}$
LPA C16:0	0.227	0.079	$4.42x10^{-3}$	0.461	0.218	$4.18x10^{-2}$	0.254	0.074	++	$5.80x10^{-4}$	$2.75x10^{-3}$
LPA C16:1	0.236	0.084	$5.67x10^{-3}$	0.280	0.187	$1.44x10^{-1}$	0.244	0.077	++	$1.48x10^{-3}$	$5.63x10^{-3}$
LPA C22:4	0.185	0.080	$2.21x10^{-2}$	0.361	0.203	$8.51x10^{-2}$	0.209	0.074	++	$5.05x10^{-3}$	$1.60x10^{-2}$
LPA C18:1	0.230	0.097	$1.93x10^{-2}$	0.190	0.223	$3.99x10^{-1}$	0.224	0.089	++	$1.20x10^{-2}$	$3.26x10^{-2}$
aLPA C18:1	0.125	0.125	$3.18x10^{-1}$	0.488	0.175	$8.52x10^{-3}$	0.248	0.102	++	$1.46x10^{-2}$	$3.48x10^{-2}$
LPA C20:3	0.070	0.099	$4.83x10^{-1}$	0.504	0.163	$4.06x10^{-3}$	0.187	0.085	++	$2.78x10^{-2}$	$5.86x10^{-2}$
LPA C14:0	0.156	0.089	$8.13x10^{-2}$	0.196	0.194	$3.19x10^{-1}$	0.163	0.081	++	$4.36x10^{-2}$	$7.58x10^{-2}$
cLPA C18:0	0.166	0.087	$5.88x10^{-2}$	0.140	0.210	$5.11x10^{-1}$	0.162	0.080	++	$4.39x10^{-2}$	$7.58x10^{-2}$
aLPA C16:1	0.093	0.128	$4.70x10^{-1}$	0.377	0.172	$3.54x10^{-2}$	0.194	0.103	++	$5.89x10^{-2}$	$9.32x10^{-2}$
cLPA C18:1	0.110	0.090	$2.26x10^{-1}$	0.202	0.201	$3.21x10^{-1}$	0.126	0.082	++	$1.28x10^{-1}$	$1.87x10^{-1}$
LPA C20:4	0.084	0.094	$3.68x10^{-1}$	0.245	0.180	$1.83x10^{-1}$	0.119	0.083	++	$1.54x10^{-1}$	$2.08x10^{-1}$
LPA C18:0	0.112	0.087	$2.01x10^{-1}$	0.050	0.198	$8.03x10^{-1}$	0.102	0.080	++	$2.01x10^{-1}$	$2.55x10^{-1}$
LPA C22:5	0.026	0.085	$7.59x10^{-1}$	0.407	0.189	$3.88x10^{-2}$	0.090	0.078	++	$2.46x10^{-1}$	$2.92x10^{-1}$
cLPA C16:0	0.110	0.102	$2.83x10^{-1}$	0.021	0.207	$9.18x10^{-1}$	0.092	0.091	++	$3.11x10^{-1}$	$3.48x10^{-1}$
LPA C18:2	-0.041	0.082	$6.19x10^{-1}$	0.106	0.183	$5.66x10^{-1}$	-0.016	0.075	-+	$8.27x10^{-1}$	$8.60x10^{-1}$
LPA C20:5	-0.007	0.087	$9.39x10^{-1}$	0.128	0.203	$5.34x10^{-1}$	0.014	0.080	-+	$8.60x10^{-1}$	$8.60x10^{-1}$

Abbreviations: LPA, lysophosphatidic acid; cLPA, cyclic lysophosphatidic acid; aLPA, alkyl-Lysophosphatidic acid; SE, standard error; *FDR*, false discovery rate
Note: Direction column indicates the direction of regression co-efficient of association in the ACE and Heidelberg/Mannheim cohort respectively

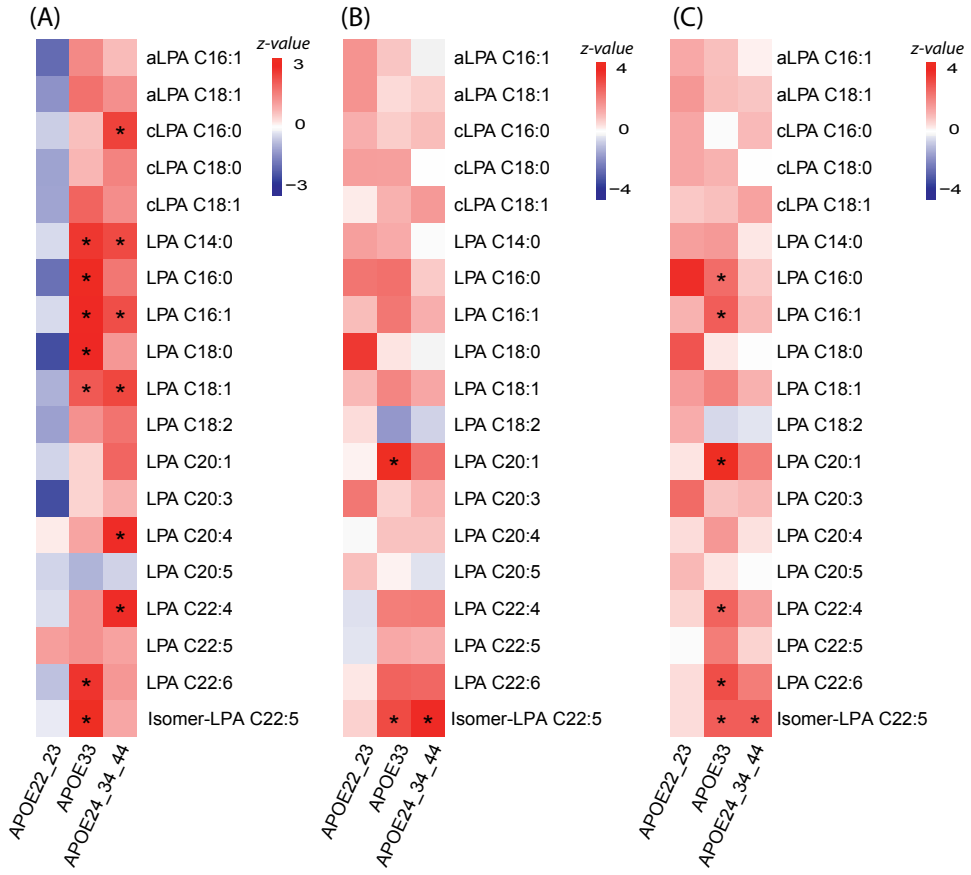


Figure 2: Heatmap of *APOE* stratified meta-analysis of regression analysis results of metabolites with Aβ-42 (A), p-tau (B) and t-tau (C) levels in cerebrospinal fluid. Note: Star indicate significant association with false discovery rate < 0.05 in each stratum.

Although the level of statistical significance differed between *APOE* ε33 and *APOE* ε4 carriers, the direction of association was always similar.

For p-tau and t-tau levels, significant association was observed with isomer-LPA C22:5 in *APOE* 33 and *APOE* ε4 carriers (Figure 2: B, C). Further significant association was observed between LPA C20:1 and p-tau, and three LPAs (C16:0, C16:1, C22:4) and t-tau only in *APOE* 33 carriers.

As *APOE* appeared to modify the association between various metabolites and CSF biomarkers, we tested whether the *APOE* genotype is associated with levels of the metabolites associated with biomarkers in the overall and *APOE* stratified analyses (see Supplementary Table 3 and 4). We did not observe significant differences between *APOE* ε4 versus *APOE* ε33 carriers and *APOE* ε2 versus *APOE 33* carriers in the combined meta-analysis of the two datasets. Although cLPA

C18:1 showed suggestive association with *APOE* ε2 in the combined meta-analysis (β = -0.636, $P = 4.66x10^{-3}$, $FDR = 8.85x10^{-2}$) but did not pass multiple testing.

MCI to AD dementia progression

LPA C16:1 (β = -0.472, $P = 3.25x10^{-3}$, $FDR = 4.41x10^{-2}$) and LPA C16:0 (β = -0.412, $P = 4.64x10^{-3}$, $FDR = 4.41x10^{-2}$) were significantly associated with progression from MCI to AD dementia (Supplementary Table 5). LPA C16:1 and C16:0 levels in CSF are correlated with each other (Supplementary Figure 2), which is also evident from their similar regression coefficient in the progression analysis. As a sensitivity analysis, we also performed cox proportional hazard analysis, additionally correcting for *APOE* (Supplementary Table 6) and Aβ-42 levels in CSF (Supplementary Table 7) to assess the role of LPAs in MCI to AD progression. *APOE* did not affect the association of LPA C16:0 and C16:1 to MCI to AD dementia progression while the association was lost upon adjusting for Aβ-42 levels. In plasma, LPA C20:1 (β = 0.599, $P = 1.84x10^{-2}$) showed evidence of association to MCI to AD dementia progression but significance was lost upon adjusting for multiple testing (Supplementary Table 8). In the Heidelberg/Mannheim sample (Supplementary Table 9), we did not replicate our findings from ACE cohort i.e., LPA C16:1 (β = 0.457, $P = 8.91x10^{-2}$) and LPA C16:0 (β = 0.431, $P = 1.96x10^{-1}$).

Association of cognitive measures with LPA levels

We did not find significant association of LPA levels in CSF with MMSE and CDR score in the combined meta-analysis (Supplementary Tables 10 and 11). In individual cohorts, only alkyl-LPA C16:1 showed evidence of association with MMSE (β = 0.253, $P = 4x10^{-2}$) in ACE cohort and with CDR score (β = 0.254, $P = 6.01x10^{-3}$) in Heidelberg/Mannheim samples.

Discussion

Meta-analysis of the data of two independent cohorts showed a significant association of eight LPAs to Aβ-42, six LPAs to p-tau, and eight LPAs to t-tau levels in CSF. In the *APOE* stratified meta-analysis, one cyclic LPA C16:0 and two LPAs (C20:4, C22:4) associate significantly with Aβ-42 levels in *APOE* ε4 carriers only. The association of LPAs with p-tau and total tau were confined to *APOE* ε33 carriers except for isomer-LPA C22:5, which showed association in both *APOE33* and *APOE4* strata. LPA C16:0 and C16:1 were associated with the progression of MCI to AD in the ACE cohort but the association was no longer significant after adjusting for Aβ-42 in the model.

The positive association between various LPAs and Aβ-42 is in line with the findings of an earlier study, suggesting that LPAs play a role in Aβ production by upregulation of β-secretase (BACE1)[12], a key enzyme involved in the cleavage of amyloid precursor protein (*APP*). Our study pinpoints a key role of LPA C18:1, C16:1, C16:0, C22:6, C14:0, C22:4, C20:4 and isomer-

LPA C22:5 in CSF amyloid levels, detailing which specific LPAs are relevant. The association of LPA to CSF amyloid sheds new light on the role of (signaling) lipids in AD pathogenies. LPAs are a bioactive component of oxLDL, which show positive correlation with CSF levels Aβ[48]. Moreover, Traumatic brain injury (TBI) patients also exhibit increased CSF levels of LPAs[49]. Because amyloid pathology is observed in nearly 30 percent of TBI patients with unknown mechanisms[50], our observed positive association of amyloid pathology and LPA may suggest the role of LPAs as a mediator in the aggregation of amyloid pathology, which needs further investigation. On the other hand, Aβ may increase oxidative stress and inflammation, which results in the production of LPAs. There is a need for functional studies to ascertain whether the positive correlation between amyloid and LPA is a cause or consequence of pathological process.

In our study, CSF levels of LPAs (C16:0 and C16:1) were significantly associated with MCI to AD dementia progression in the ACE cohort. Similar to the inverse relation of CSF Aβ-42 levels in MCI to AD dementia progression, decreased levels of the LPAs are associated with MCI to AD dementia progression[51]. Loss of association of LPAs with MCI to AD dementia progression when accounting for Aβ-42 levels suggests that Aβ-42 mediates the observed association of LPA C16:0 and C16:1 in conversion. This is of note that LPA 16:0 and 16:1 did not show significant association with MCI to AD progression in the smaller Heidelberg/Mannheim samples. However, LPA 16:0 and 16:1 also did not show significant association with Aβ-42 levels in this small cohort, making findings difficult to interpret. The fact that the association LPA C16:0 and C16:1 to conversion loses its significant when adjusting for Aβ-42 levels, suggests the LPA C16:0 and C16:1 are likely preceding the changes in Aβ-42 that predict conversion to AD.

In the *APOE* stratified analysis, LPA C16:1 showed a positive association to Aβ-42 levels in both *APOE* ε4 and *APOE* ε33 carriers, whereas LPA C16:0 showed significant positive association in only *APOE* ε3 stratum. No effect of the ratio of these two correlated LPAs was seen. The findings imply that the association of these LPAs to Aβ-42 levels may be modified by *APOE* genotype of the person. To our knowledge, this is the first study that shows the *APOE* interacts with LPA in humans. Interestingly, all LPA showed a negative association with Aβ-42 in *APOE* ε2 carriers, i.e., in the opposite direction compared to *APOE* ε33 and *APOE* ε4 carrier. Since *APOE* ε2 carriers are protected from AD and have delayed onset of AD [52], LPA modification may be relevant in *APOE* ε33 and *APOE* ε4 carriers. We did not observe association of the interacting LPAs with *APOE* genotypes which may be due to limited sample size in these analyses.

Except LPA C16:0, we found that the majority of unsaturated LPAs (C16:1, C20:1, isomer-LPA C22:5, C22:6, and C22:4) showed significant association to both p-tau and total tau levels in CSF. Due to differential activation of LPA receptors by diverse LPA metabolite species[23], association of unsaturated LPAs with AD biomarkers of pathophysiology can also be explained by their affinity for LPA_3 receptors[53] which are also expressed in hippocampus, frontal cortex and amygdala[54]. Moreover, both saturated and unsaturated LPAs are reported to influence Ca^{2+}

signaling through LPA2 receptors[53] which may also suggest their involvement in the dysregulation of Ca^{2+} signaling in AD. Earlier studies have shown that LPAs acts as mediators to maintain the intracellular Ca^{2+} levels in both astrocytes[55] and microglial cells[5,56]. One of these LPAs (LPA C20:1) only showed significant association with p-tau and total tau levels but not with Aβ-42. These tau specific associations may be explained by the fact that LPAs are involved in the upregulation of glycogen synthase kinase-3 (GSK-3), an enzyme involved in phosphorylation of tau and thus may influence levels of p-tau in CSF[13,57]. The association of LPA C20:1 to only tau pathology may also indicate the specificity of association of LPA species of different acyl chains to different AD pathophysiological mechanisms. In the *APOE* stratified meta-analysis, all the observed associations were largely confined to *APOE* ε33 carriers except for isomer-LPA C22:5. This observation is in line with the studies which demonstrated that *APOE* ε4 may influence amyloid pathology in the brain rather than tau aggregation[58-60].

We did not observe association of LPA levels in plasma with Aβ-42, p-tau, and total-tau in CSF nor did we find association with MCI to AD dementia progression. It is interesting that we observed a significant correlation between CSF and plasma measurements of various LPA molecular species. A negative correlation was observed for cyclic-LPA C16:0, cyclic-LPA C18:1, LPA C20:1. Of these, the association was strongest and most convincing for LPA C20:1 in terms of R (-0.23) and p-value $6.2x10^{-3}$. The positive correlations are more convincing in particular for LPA C18:1 (R = 0.44 and $P = 4.5x10^{-8}$) and C20:5(R = 0.40 and $P = 1.2x10^{-6}$) and LPA 22:4 (R = 0.27 and $P = 1.2x10^{-3}$), (Supplementary Figure 1). LPAs found associated with Aβ, p-tau and t-tau in CSF were not correlated with their counterparts in plasma, which indicates that the LPA role in relation to AD pathology is primarily cerebral and not in the circulation.

Our study provides a comprehensive overview of association of various LPA species including alkyl-LPAs and cyclic-LPAs to biomarkers of AD during the prodromal phase. The inclusion of two independent cohorts is a major strength of our investigation, allowing us to check consistency of effect across cohorts. Moreover, we have also assessed the role of LPAs longitudinally for MCI to AD dementia progression in the ACE cohort.

Limitations

The short follow-up time for MCI patients in the progression study is a limitation of our study and asks for replication in a study with longer follow-up. The small sample size in MCI to AD progression analysis is another major limitation of our study. In the absence of any data on the association between LPA and AD biomarkers in CSF, we did not perform a power calculation a priori, which limits the clinical and predictive implications of the discovery analysis. Future large sample sizes in follow-up studies will also provide more power to perform *APOE* stratified analysis.

Conclusions

Overall findings from our study suggest that various LPAs based on acyl chain length and saturation level are associated with Aβ-42, p-tau and total tau levels. Our study suggests the role of LPAs in the pathophysiology of AD. Future studies are needed to determine whether LPA metabolites triggers various biological pathways leading to increase in biomarkers of AD pathophysiology or are produced as a downstream effect of AD pathology. We further find that *APOE* may influence the association between LPAs and Aβ-42.

Acknowledgments

This study was funded by the ADAPTED: Alzheimer's Disease Apolipoprotein Pathology for Treatment Elucidation and Development consortium which has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under Grant Agreement No 115975. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and the European Federation of Pharmaceutical Industries and Associations. Part of this work was supported by the JPND EADB grant (German Federal Ministry of Education and Research, BMBF: 01ED1619A).

We acknowledge all patients participating in this study from Fundació ACE and Heidelberg-Mannheim memory clinic. We also want to thank the investigators from the Fundació ACE Barcelona Alzheimer Treatment and Research Center, for their close collaboration and continuous intellectual input. We are grateful to Laura Montrreal for her exceptional technical support. A.R. and M.B. receive support from the European Union/EFPIA Innovative Medicines Initiative Joint Undertaking ADAPTED and MOPEAD projects (Grants No. 115975 and 115985, respectively). M.B. S.G. and A.R. are also supported by national grants PI13/02434, PI16/01861, and PI17/01474. Acción Estratégica en Salud is integrated into the Spanish National R + D + I Plan and funded by ISCIII (Instituto de Salud Carlos III)-Subdirección General de Evaluación and the Fondo Europeo de Desarrollo Regional (FEDER- "Una manera de Hacer Europa").

Competing interests:

S.A., A.O., I.K., L.F., I.R., S.G., M.B., I.H., L.H., N.A., A.R., A.R., T.H., C.M.D. declared no competing interests. Margot H.M. Bakker is a full-time employee of AbbVie Deutschland GmbH & Co KG and owns AbbVie stock. Alfredo Cabrera-Socorro is full-time employee of Janssen Pharmaceutical NV, Turnhoutseweg 30, 2340 Beerse, Belgium.

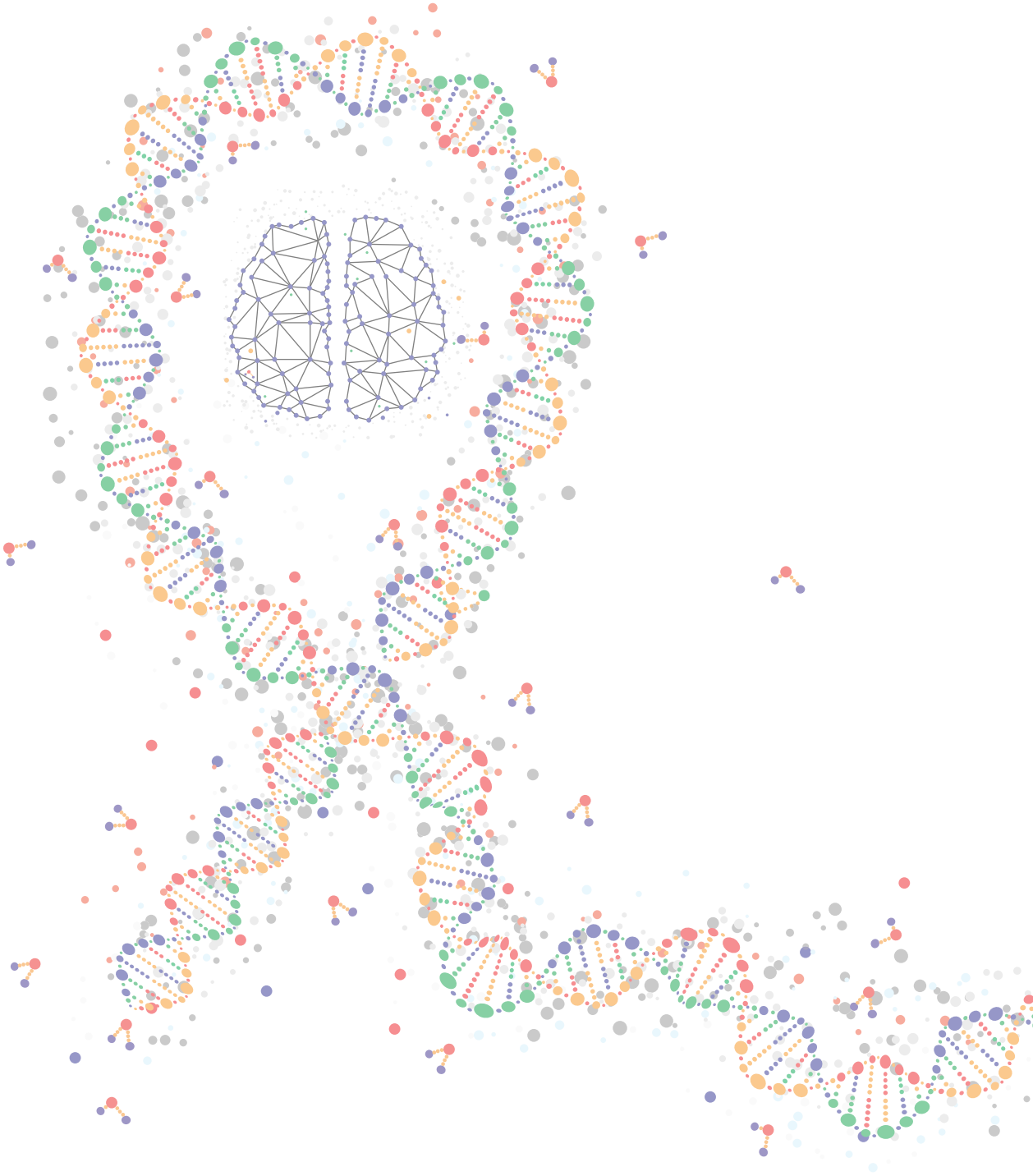
References

1. Varma, V.R. *et al.* Brain and blood metabolite signatures of pathology and progression in Alzheimer disease: A targeted metabolomics study. *PLoS medicine* **15**, e1002482 (2018).
2. van der Lee, S.J. *et al.* Circulating metabolites and general cognitive ability and dementia: Evidence from 11 cohort studies. *Alzheimer's & Dementia* **14**, 707-722 (2018).
3. Liu, Q. & Zhang, J. Lipid metabolism in Alzheimer's disease. *Neurosci Bull* **30**, 331-45 (2014).
4. Yung, Y.C., Stoddard, N.C., Mirendil, H. & Chun, J. Lysophosphatidic Acid signaling in the nervous system. *Neuron* **85**, 669-82 (2015).
5. Schilling, T., Stock, C., Schwab, A. & Eder, C. Functional importance of Ca2+-activated K+ channels for lysophosphatidic acid-induced microglial migration. *European Journal of Neuroscience* **19**, 1469-1474 (2004).
6. Anliker, B. *et al.* Lysophosphatidic acid (LPA) and its receptor, LPA1, influence embryonic schwann cell migration, myelination, and cell-to-axon segregation. *Glia* **61**, 2009-2022 (2013).
7. Armati, P.J. & Mathey, E.K. An update on Schwann cell biology—immunomodulation, neural regulation and other surprises. *Journal of the neurological sciences* **333**, 68-72 (2013).
8. Garcia-Morales, V. *et al.* Membrane-derived phospholipids control synaptic neurotransmission and plasticity. *PLoS biology* **13**(2015).
9. Robinson, R. One lipid, two synaptic plasticity pathways. *PLoS biology* **13**, e1002154-e1002154 (2015).
10. Yukiura, H. *et al.* Autotaxin regulates vascular development via multiple lysophosphatidic acid (LPA) receptors in zebrafish. *Journal of Biological Chemistry* **286**, 43972-43983 (2011).
11. de Leeuw, F.A. *et al.* Blood-based metabolic signatures in Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* **8**, 196-207 (2017).
12. Shi, J., Dong, Y., Cui, M.-Z. & Xu, X. Lysophosphatidic acid induces increased BACE1 expression and Aβ formation. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* **1832**, 29-38 (2013).
13. Sayas, C.L., Moreno-Flores, M.T., Avila, J. & Wandosell, F. The neurite retraction induced by lysophosphatidic acid increases Alzheimer's disease-like Tau phosphorylation. *J Biol Chem* **274**, 37046-52 (1999).
14. Frisardi, V., Panza, F., Seripa, D., Farooqui, T. & Farooqui, A.A. Glycerophospholipids and glycerophospholipid-derived lipid mediators: a complex meshwork in Alzheimer's disease pathology. *Progress in lipid research* **50**, 313-330 (2011).
15. Draczynska-Lusiak, B., Doung, A. & Sun, A.Y. Oxidized lipoproteins may play a role in neuronal cell death in Alzheimer disease. *Mol Chem Neuropathol* **33**, 139-48 (1998).
16. Holtsberg, F.W. *et al.* Lysophosphatidic acid and apoptosis of nerve growth factor-differentiated PC12 cells. *Journal of neuroscience research* **53**, 685-696 (1998).
17. Uranbileg, B. *et al.* Alteration of the lysophosphatidic acid and its precursor lysophosphatidylcholine levels in spinal cord stenosis: A study using a rat cauda equina compression model. *Scientific reports* **9**, 1-13 (2019).

18. Ramesh, S., Govindarajulu, M., Suppiramaniam, V., Moore, T. & Dhanasekaran, M. Autotaxin(-) Lysophosphatidic Acid Signaling in Alzheimer's Disease. *Int J Mol Sci* **19**(2018).
19. Hayakawa, K. *et al.* Lysophosphatidic acids and their substrate lysophospholipids in cerebrospinal fluid as objective biomarkers for evaluating the severity of lumbar spinal stenosis. *Scientific reports* **9**, 1-9 (2019).
20. Aoki, J., Inoue, A. & Okudaira, S. Two pathways for lysophosphatidic acid production. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids* **1781**, 513-518 (2008).
21. Cho, H.-J. *et al.* Bioactive lipids in gintonin-enriched fraction from ginseng. *Journal of ginseng research* **43**, 209-217 (2019).
22. Takeda, A., Umemoto, E. & Miyasaka, M. Lysophosphatidic acid as a regulator of lymphocyte trafficking in the lymph nodes. *Translational Cancer Research* **4**, 537-543 (2015).
23. Bandoh, K. *et al.* Lysophosphatidic acid (LPA) receptors of the EDG family are differentially activated by LPA species: structure–activity relationship of cloned LPA receptors. *FEBS letters* **478**, 159-165 (2000).
24. Yoshida, K. *et al.* Vascular remodeling induced by naturally occurring unsaturated lysophosphatidic acid in vivo. *Circulation* **108**, 1746-1752 (2003).
25. Aikawa, S., Hashimoto, T., Kano, K. & Aoki, J. Lysophosphatidic acid as a lipid mediator with multiple biological actions. *J Biochem* **157**, 81-9 (2015).
26. Moolenaar, W.H. Lysophosphatidic acid, a multifunctional phospholipid messenger. *Journal of Biological Chemistry* **270**, 12949-12952 (1995).
27. Boada, M. *et al.* Design of a comprehensive Alzheimer's disease clinic and research center in Spain to meet critical patient and family needs. *Alzheimers Dement* **10**, 409-15 (2014).
28. Petersen, R.C. Mild cognitive impairment as a diagnostic entity. *J Intern Med* **256**, 183-94 (2004).
29. Petersen, R.C. *et al.* Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol* **56**, 303-8 (1999).
30. Edition, F. Diagnostic and statistical manual of mental disorders. *Am Psychiatric Assoc* (2013).
31. McKhann, G.M. *et al.* The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & dementia* **7**, 263-269 (2011).
32. Román, G.C. *et al.* Vascular dementia: diagnostic criteria for research studies: report of the NINDS-AIREN International Workshop. *Neurology* **43**, 250-250 (1993).
33. Mesulam, M.M., Grossman, M., Hillis, A., Kertesz, A. & Weintraub, S. The core and halo of primary progressive aphasia and semantic dementia. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* **54**, S11-S14 (2003).
34. Halliday, G. *et al.* Diagnosis and management of dementia with Lewy bodies. (2017).
35. Vanderstichele, H. *et al.* Standardization of preanalytical aspects of cerebrospinal fluid biomarker testing for Alzheimer's disease diagnosis: a consensus paper from the Alzheimer's Biomarkers Standardization Initiative. *Alzheimers Dement* **8**, 65-73 (2012).
36. Vanderstichele, H. *et al.* Biomarkers for early diagnosis of Alzheimer's disease. *Hauppauge, NY: Nova Science Publishers* (2008).

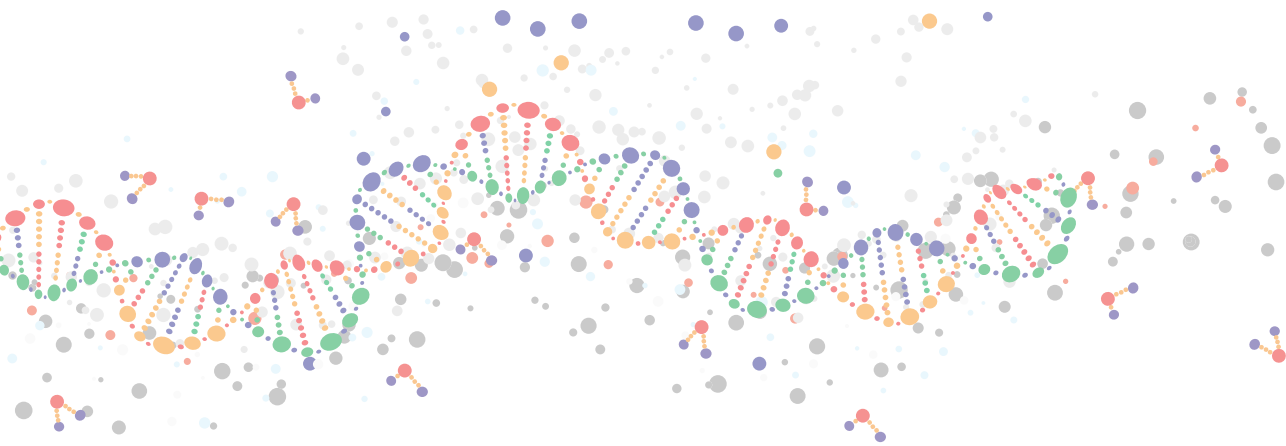
37. Winblad, B. *et al.* Mild cognitive impairment--beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *J Intern Med* **256**, 240-6 (2004).
38. Morris, J.C. *et al.* The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology* **39**, 1159-65 (1989).
39. Walton, D. The diagnostic and predictive accuracy of the Wechsler Memory Scale in psychiatric patients over 65. *Journal of Mental Science* **104**, 1111-1118 (1958).
40. Reitan, R.M. & Wolfson, D. The Trail Making Test as an initial screening procedure for neuropsychological impairment in older children. *Archives of Clinical Neuropsychology* **19**, 281-288 (2004).
41. Albert, M.S. *et al.* The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7**, 270-9 (2011).
42. Kaerst, L. *et al.* Cerebrospinal fluid biomarkers in Alzheimer's disease, vascular dementia and ischemic stroke patients: a critical analysis. *J Neurol* **260**, 2722-7 (2013).
43. Schoeman, J.C. *et al.* Development and application of a UHPLC-MS/MS metabolomics based comprehensive systemic and tissue-specific screening method for inflammatory, oxidative and nitrosative stress. *Anal Bioanal Chem* **410**, 2551-2568 (2018).
44. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
45. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300 (1995).
46. van der Lee, S.J. *et al.* The effect of APOE and other common genetic variants on the onset of Alzheimer's disease and dementia: a community-based cohort study. *Lancet Neurol* **17**, 434-444 (2018).
47. Mitchell, A.J. & Shiri-Feshki, M. Rate of progression of mild cognitive impairment to dementia--meta-analysis of 41 robust inception cohort studies. *Acta Psychiatr Scand* **119**, 252-65 (2009).
48. Sun, Y.-X. *et al.* Inflammatory markers in matched plasma and cerebrospinal fluid from patients with Alzheimer's disease. *Dementia and geriatric cognitive disorders* **16**, 136-144 (2003).
49. McDonald, W.S. *et al.* Matrix-Assisted Laser Desorption Ionization Mapping of Lysophosphatidic Acid Changes after Traumatic Brain Injury and the Relationship to Cellular Pathology. *The American journal of pathology* **188**, 1779-1793 (2018).
50. Sivanandam, T.M. & Thakur, M.K. Traumatic brain injury: a risk factor for Alzheimer's disease. *Neuroscience & Biobehavioral Reviews* **36**, 1376-1381 (2012).
51. Hampel, H. *et al.* Value of CSF β-amyloid 1–42 and tau as predictors of Alzheimer's disease in patients with mild cognitive impairment. *Molecular psychiatry* **9**, 705 (2004).
52. van der Lee, S.J. *et al.* The effect of APOE and other common genetic variants on the onset of Alzheimer's disease and dementia: a community-based cohort study. *The Lancet Neurology* **17**, 434-444 (2018).

53. Bandoh, K. *et al.* Molecular cloning and characterization of a novel human G-protein-coupled receptor, EDG7, for lysophosphatidic acid. *J Biol Chem* **274**, 27776-85 (1999).
54. Im, D.S. *et al.* Molecular cloning and characterization of a lysophosphatidic acid receptor, Edg-7, expressed in prostate. *Mol Pharmacol* **57**, 753-9 (2000).
55. Keller, J.N., Steiner, M.R., Holtsberg, F.W., Mattson, M.P. & Steiner, S.M. Lysophosphatidic acid-induced proliferation-related signals in astrocytes. *Journal of neurochemistry* **69**, 1073-1084 (1997).
56. Schilling, T. *et al.* Lysophospholipids induce membrane hyperpolarization in microglia by activation of IKCa1 Ca2+-dependent K+ channels. *Neuroscience* **109**, 827-835 (2002).
57. Sayas, C., Avila, J. & Wandosell, F. Regulation of neuronal cytoskeleton by lysophosphatidic acid: role of GSK-3. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids* **1582**, 144-153 (2002).
58. Morris, J.C. *et al.* APOE predicts amyloid-beta but not tau Alzheimer pathology in cognitively normal aging. *Annals of neurology* **67**, 122-131 (2010).
59. Sunderland, T. *et al.* Cerebrospinal fluid β-amyloid1–42 and tau in control subjects at risk for Alzheimer's disease: The effect of APOE ε4 allele. *Biological psychiatry* **56**, 670-676 (2004).
60. Huang, Y.A., Zhou, B., Wernig, M. & Sudhof, T.C. ApoE2, ApoE3, and ApoE4 Differentially Stimulate APP Transcription and Abeta Secretion. *Cell* **168**, 427-441 e21 (2017).



Chapter 5

Gut-Liver-Brain axis



Chapter 5.1

Association of altered liver enzymes with Alzheimer's disease diagnosis, cognition, neuroimaging measures, and cerebrospinal fluid biomarkers

Kwangsik Nho*, Alexandra Kueider-Paisley*, Shahzad Ahmad*, Siamak MahmoudianDehkordi, Matthias Arnold, Shannon L. Risacher, Gregory Louie, Colette Blach, Rebecca Baillie, Xianlin Han, Gabi Kastenmüeller, John Q. Trojanowski, Leslie M. Shaw, Michael W. Weiner, P. Murali Doraiswamy, Cornelia van Duijn, Andrew J. Saykin#, Rima Kaddurah-Daouk#, for the Alzheimer's Disease Neuroimaging Initiative and the Alzheimer Disease Metabolomics Consortium

This chapter is published in JAMA Network Open. 2019 Jul 3;2(7): e197978

*First authors contributed equally to the work
#Senior authors contributed equally to the work

Supplementary information for this chapter is available at:
10.1001/jamanetworkopen.2019.7978

Abstract

Importance: Increasing evidence suggests an important role of liver function in the pathophysiology of Alzheimer's disease (AD). The liver is a major metabolic hub; therefore, investigating the association of liver function with AD, cognition, neuroimaging, and CSF biomarkers would improve the understanding of the role of metabolic dysfunction in AD.

Objective: To examine whether liver function markers are associated with cognitive dysfunction and the "A/T/N" (amyloid, tau, and neurodegeneration) biomarkers for AD.

Design, Setting, and Participants: In this cohort study, serum-based liver function markers were measured from September 1, 2005, to August 31, 2013, in 1581 AD Neuroimaging Initiative participants along with cognitive measures, cerebrospinal fluid (CSF) biomarkers, brain atrophy, brain glucose metabolism, and amyloid-β accumulation. Associations of liver function markers with AD-associated clinical and A/T/N biomarkers were assessed using generalized linear models adjusted for confounding variables and multiple comparisons. Statistical analysis was performed from November 1, 2017, to February 28, 2019.

Exposure: Five serum-based liver function markers (total bilirubin, albumin, alkaline phosphatase, alanine aminotransferase, and aspartate aminotransferase) from AD Neuroimaging Initiative participants were used as exposure variables.

Main Outcomes and Measures: Primary outcomes included diagnosis of AD, composite scores for executive functioning and memory, CSF biomarkers, atrophy measured by magnetic resonance imaging, brain glucose metabolism measured by fludeoxyglucose F 18 (18F) positron emission tomography, and amyloid-β accumulation measured by [18F] florbetapir positron emission tomography.

Results: Participants in the AD Neuroimaging Initiative (n = 1581; 697 women and 884 men; mean [SD] age, 73.4 [7.2] years) included 407 cognitively normal older adults, 20 with significant memory concern, 298 with early mild cognitive impairment, 544 with late mild cognitive impairment, and 312 with AD. An elevated aspartate aminotransferase (AST) to alanine aminotransferase (ALT) ratio and lower levels of ALT were associated with AD diagnosis (AST to ALT ratio: odds ratio, 7.932 [95% CI, 1.673-37.617]; $P = .03$; ALT: odds ratio, 0.133 [95% CI, 0.042-0.422]; $P = .004$) and poor cognitive performance (AST to ALT ratio: β [SE], −0.465 [0.180]; $P = .02$ for memory composite score; β [SE], −0.679 [0.215]; $P = .006$ for executive function composite score; ALT: β [SE], 0.397 [0.128]; $P = .006$ for memory composite score; β [SE], 0.637 [0.152]; $P < .001$ for executive function composite score). Increased AST to ALT ratio values were associated with lower CSF amyloid-β 1-42 levels (β [SE], −0.170 [0.061]; $P = .04$) and increased amyloid-β deposition (amyloid biomarkers), higher CSF phosphorylated

tau181 (β [SE], 0.175 [0.055]; *P* = .02) (tau biomarkers) and higher CSF total tau levels (β [SE], 0.160 [0.049]; *P* = .02) and reduced brain glucose metabolism (β [SE], –0.123 [0.042]; *P* = .03) (neurodegeneration biomarkers). Lower levels of ALT were associated with increased amyloid-β deposition (amyloid biomarkers), and reduced brain glucose metabolism (β [SE], 0.096 [0.030]; *P* = .02) and greater atrophy (neurodegeneration biomarkers).

Conclusion and Relevance: Consistent associations of serum-based liver function markers with cognitive performance and A/T/N biomarkers for AD highlight the involvement of metabolic disturbances in the pathophysiology of AD. Further studies are needed to determine if these associations represent a causative or secondary role. Liver enzyme involvement in AD opens avenues for novel diagnostics and therapeutics.

Introduction

Metabolic activities in the liver determine the state of the metabolic readout of peripheral circulation. Mounting evidence suggests that patients with Alzheimer's disease (AD) display metabolic dysfunction[1]. Clinical studies suggest that impaired signaling, energy metabolism, inflammation, and insulin resistance play a role in AD[2,3]. This observation is in line with the observation that many metabolic disorders (e.g., diabetes, hypertension, obesity, dyslipidemia) are risk factors for AD[4]. This evidence highlights the importance of the liver in the pathophysiological characteristics of AD. Focused investigation to assess the role of liver function in AD and its endophenotypes is required to bridge the gap between these observations.

Peripheral blood levels of biochemical markers including albumin, alkaline phosphatase, alanine aminotransaminase (ALT), aspartate aminotransferase (AST), and total bilirubin are used to assess liver function. ALT and AST are used in general clinical practice to measure liver injury[5,6] and are factors associated with cardiovascular and metabolic diseases[7,8], known risk factors of AD and cognitive decline[9,10]. Given this fact, it is conceivable that aminotransferases are surrogate biomarkers of liver metabolic functioning. A systematic search yielded few reports related to research in humans linking peripheral biomarkers of liver functioning to central biomarkers related to AD including amyloid-β and tau accumulation, brain glucose metabolism, and structural atrophy.

We investigated the association of peripheral liver function markers with AD diagnosis, cognition, and biomarkers of AD pathophysiological characteristics including neuroimaging (magnetic resonance imaging [MRI] and positron emission topography [PET]) and cerebrospinal fluid (CSF) from older adults in the AD Neuroimaging Initiative (ADNI) cohort. The AD biomarkers were selected and defined consistent with the National Institute on Aging-Alzheimer Association Research Framework (amyloid, tau and neurodegeneration[A/T/N] for AD biomarkers that defines three general groups of biomarkers based on the nature of pathologic process that each measures[11].

Methods

Study population

Individuals in this study were participants of ADNI. The initial phase (ADNI-1) was launched in 2003 to test whether serial MRI markers, PET markers, other biological markers, and clinical and neuropsychological assessment could be combined to measure the progression of mild cognitive impairment (MCI) and early AD. The initial phase was extended to subsequent phases (ADNI-GO, ADNI-2, and ADNI-3) for follow-up of existing participants and additional new enrollments. Inclusion and exclusion criteria, clinical and neuroimaging protocols, and other

information are reported elsewhere[12,13] (http://adni.loni.usc.edu/). Demographic and clinical information, raw data from neuroimaging scans, CSF biomarkers, information on *APOE* status, and cognitive scores were downloaded from the ADNI data repository. Baseline data were collected from September 1, 2005, to August 31, 2013. Written informed consent was obtained at enrollment which included permission for analysis and data sharing. This study was approved by each participating site's Institutional Review Board. This report followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guidelines for cohort studies.

Liver function markers

Five laboratory tests were downloaded from the ADNI data repository and used in the study: total bilirubin, albumin, alkaline phosphatase, ALT, and AST. The liver function markers followed a normal distribution after log transformation. For each marker, participants with values greater or smaller than 4 SDs from its mean value were considered outliers and were removed. To determine if outliers had a significant effect on our findings, we performed a sensitivity analysis and observed few differences (or slightly more significant), if any, in results when including outliers (Supplementary Table 1).

Dementia diagnosis

Participants in ADNI were classified as cognitively normal controls (CN) or having significant memory concerns (SMC), MCI, or mild clinical AD. Criteria for classification were as follows: Mini-Mental State Examination score range (range, 0 [worst] to 30 [best]) for CN and MCI was 24 to 30, and for AD was 20 to 26; and overall Clinical Dementia Rating score (range for each, 0 [best] to 3 [worst]) for CN was 0, for MCI was 0.5 with a mandatory requirement of memory box score of 0.5 or greater, and for AD 0.5 or 1[14]. Cognitively normal controls did not have any significant impairment in cognition or activities of daily living. Participants with SMC had normal cognition and no significant impairment in activities of daily living, but had a score of 16 or more on the first 12-items of the self-report version of the Cognitive Change Index (range, 12 [no change] to 60 [severe change])[15]. Participants with MCI had cognitive impairments in memory and/or other domains but were able to perform activities of daily living and did not qualify for a diagnosis of dementia[14]. Participants with AD had to meet the National Institute of Neurological and Communicative Disorders and Stroke–AD and Related Disorders Association criteria for probable AD[16]. Participants from the ADNI-1 cohort with MCI were all classified as late MCI, with a memory impairment approximately 1.5 SD below education-adjusted norms. In the ADNI GO and ADNI-2 cohort, participants with MCI were classified as either early MCI, with a memory impairment approximately 1 SD below education-adjusted norms or late MCI (same criteria as in ADNI-1). All ADNI-1 and ADNI-GO and ADNI-2 met the criteria for amnestic MCI, but many in the ADNI GO and ADNI-2 cohort included the earlier stage MCI designation (i.e., early MCI)[17].

Cognition

Composite scores were used to measure memory and executive functioning. A memory composite score was created from the following: memory tasks from the Alzheimer Disease Assessment Scale-cognitive subscale (ADAS-Cog), the Rey Auditory Verbal Learning Test, memory components of the Mini-Mental State Examination, and the Logical Memory task[18]. An executive function composite score included the following: Wechsler Adult Intelligence Scale-Revised Digit Symbol Substitution task and Digit Span backward task, Trail Making Test Parts A and B, category fluency (animals and vegetables), and 5 clock drawing items. Composite scores have a mean of 0 and an SD of 1[19].

Neuroimaging processing

MRI scans

Baseline T_1-weighted brain MRI scans were acquired using a sagittal 3-dimensional magnetization prepared rapid gradient echo scans following the ADNI MRI protocol [20,21]. As previously detailed, FreeSurfer, version 5.1, a widely used automated MRI analysis approach, was used to process MRI scans and extract whole-brain and region-of-interest (ROI) based neuroimaging endophenotypes including volumes and cortical thickness determined by automated segmentation and parcellation[22-24]. The cortical surface was reconstructed to measure thickness at each vertex. The cortical thickness was calculated by taking the Euclidean distance between gray and white boundary and the gray and CSF boundary at each vertex on the surface[25-27].

PET scans

Preprocessed fludeoxyglucose (FDG) F 18 (^{18}F) and [^{18}F] Florbetapir PET scans (coregistered, averaged, standardized image and voxel size, and uniform resolution) were downloaded from the ADNI Laboratory of Neuro Imaging (LONI) site (http://adni.loni.usc.edu/) as described in previously reported methods for acquisition and processing of PET scans[22,28]. For [^{18}F] FDG-PET, scans were intensity-normalized using a pons ROI to create [^{18}F] FDG standardized uptake value ratio (SUVR) images. For [^{18}F] Florbetapir PET, scans were intensity-normalized using a whole cerebellum reference region to create SUVR images.

CSF biomarkers

The ADNI generated CSF biomarkers (amyloid-β 1-42, total tau [t-tau], and phosphorylated tau_{181} [p-tau181]) in pristine aliquots of 2401 ADNI CSF samples using the validated and highly automated Roche Elecsys electrochemiluminescence immunoassays[29,30] and the same reagent lot for each of these 3 biomarkers. Cerebrospinal fluid biomarker data were downloaded from the ADNI LONI site (http://adni.loni.usc.edu).

Statistical analysis

Statistical analysis was conducted from November 1, 2017, to February 28, 2019. Logistic regression analysis was performed to explore the diagnostic group differences between AD

diagnosis and each liver function marker separately. Age, sex, body mass index (BMI), and *APOE ε4* status were used as covariates. We performed a linear regression analysis to assess the association of liver function markers with composite scores for memory and executive functioning using age, sex, years of education, BMI, and *APOE ε4* status as covariates. We also performed a linear regression analysis using age, sex, BMI, and *APOE ε4* status as covariates.

ROI-based analysis of structural MRI and PET scans

Mean hippocampal volume was used as an MRI-related phenotype. For FDG-PET, a mean SUVR value was extracted from a global cortical ROI representing regions where patients with AD show decreased glucose metabolism relative to CN participants from the full ADNI-1 cohort, normalized to pons[28]. For [^{18}F] Florbetapir PET, a mean SUVR value was extracted using MarsBaR from a global cortical region generated from an independent comparison of ADNI-1 [11C] Pittsburgh Compound B SUVR scans (regions where AD > CN). We performed a linear regression analysis using age, sex, BMI, and *APOE* ε4 status as covariates to evaluate the association of liver function markers with AD-related endophenotypes from MRI and PET scans. For hippocampal volume, years of education, intracranial volume (ICV), and magnetic field strength were added as additional covariates[31].

Whole-brain imaging analysis

The SurfStat software package (www.math.mcgill.ca/keith/surfstat/) was used to perform a multivariable analysis of cortical thickness to examine the association of liver function markers with brain structural changes on a vertex-by-vertex basis using a general linear model approach[27]. General linear models were developed using age, sex, years of education, intracranial volume, BMI, *APOE ε4* status, and magnetic field strength as covariates. The processed FDG-PET and [^{18}F] Florbetapir PET images were used to perform a voxelwise statistical analysis of the association of liver function markers with brain glucose metabolism and amyloid-β accumulation across the whole brain using SPM8 (www.fil.ion.ucl.ac.uk/spm/). We performed a multivariable regression analysis using age, sex, BMI, and *APOE ε4* status as covariates. In the whole-brain surface-based analysis, the adjustment for multiple comparisons was performed using the random field theory correction method with $P < 0.05$ adjusted as the level for significance[32-34]. In the voxelwise whole-brain analysis, the significant statistical parameters were selected to correspond to a threshold of $P < 0.05$ (false discovery rate [FDR]-corrected)[35].

Multiple testing correction

Results of the analysis of liver function markers with AD diagnosis groups, cognitive composite measures, and A/T/N biomarkers for AD separately were corrected for multiple testing using the FDR with the Benjamini-Hochberg procedure (p.adjust command in R [R Project for Statistical Computing]).

Results

Study sample

Our analyses included 1581 ADNI participants (407 CN, 20 with SMC, 298 with early MCI, 544 with late MCI, and 312 with AD). Demographic information as well as mean and SD of liver function markers stratified by clinical diagnosis are presented in Supplementary Table 2 in the Supplement.

Diagnostic group difference of liver function markers with AD diagnosis

Levels of ALT were significantly decreased in AD compared with CN (odds ratio, 0.133; 95% CI,0.042-0.422; *P* = .004) (Table 1), while AST to ALT ratio values were significantly increased in AD (odds ratio, 7.932; 95% CI, 1.673-37.617; *P* = .03). There was a trend to suggest that ALT levels were increased and AST to ALT ratio values were decreased in MCI compared with CN, but these became nonsignificant after adjustment for multiple comparisons (Supplementary Table 3).

Table 1: Results of association of liver function biomarkers with Alzheimer's disease diagnosis*

Liver function marker	Odds Ratio (95% CI)	Corrected *P-value*
Albumin (g/dL)	5.789 (0.040-843.993)	4.90×10^{-1}
Alkaline Phosphatase (U/L)	3.620 (0.844-15.529)	1.25×10^{-1}
ALT (U/L)	0.133 (0.042-0.422)	3.63×10^{-3}
AST (U/L)	0.229 (0.045-1.175)	1.25×10^{-1}
AST to ALT ratio	7.932 (1.673-37.617)	2.73×10^{-2}
Total Bilirubin (mg/dL)	1.405 (0.585-3.377)	4.90×10^{-1}

Abbreviations: AST, Aspartate aminotransferase; ALT, alanine aminotransferase; CI, confidence interval.
*Cognitively normal vs Alzheimer's disease. Analyses were adjusted for age, sex, body mass index, and *APOE* ε4 status.

Cognition

After adjusting for multiple comparison correction using FDR, we identified significant associations of liver function markers with cognition (Table 2). Higher levels of alkaline phosphatase and AST to ALT ratio were associated with lower memory scores (alkaline phosphatase: β [SE], –0.416 [0.162]; *P* = .02; AST to ALT ratio: β [SE], – 0.465 [0.180]; *P* = .02) and executive functioning scores (alkaline phosphatase: β [SE], –0.595 [0.193]; *P* = .006; AST to ALT ratio: β [SE], –0.679 [0.215]; *P* = .006). Higher ALT levels were associated with higher memory scores (β [SE], 0.397 [0.128]; *P* = .006) and executive functioning scores (β [SE], 0.637 [0.152]; *P* < .001), whereas higher AST levels were associated with higher executive functioning scores (β [SE], 0.607 [0.215]; *P* = .01).

Table 2: Results of association of liver function biomarkers with composite cognitive performance measures*

Liver function marker	Memory composite score		Executive function composite score	
	β (SE)	Corrected *P-value*	β (SE)	Corrected *P-value*
Albumin (g/dL)	-0.872 (0.576)	1.73×10^{-1}	-0.203 (0.689)	7.68×10^{-1}
Alkaline Phosphatase (U/L)	-0.416 (0.162)	1.74×10^{-2}	-0.595 (0.193)	6.33×10^{-3}
ALT (U/L)	0.397 (0.128)	6.33×10^{-3}	0.637 (0.152)	3.71×10^{-4}
AST (U/L)	0.339 (0.180)	9.00×10^{-2}	0.607 (0.215)	1.16×10^{-2}
AST to ALT ratio	-0.465 (0.180)	1.74×10^{-2}	-0.679 (0.215)	6.33×10^{-3}
Total Bilirubin (mg/dL)	-0.068 (0.103)	6.07×10^{-1}	-0.066 (0.123)	6.5×10^{-1}

Abbreviations: AST, Aspartate aminotransferase; ALT, alanine aminotransferase; SE, standard error
*Analyses were adjusted for age, sex, education, body mass index, and *APOE* ε4 status.

Biomarkers of amyloid-β

We used CSF amyloid-β 1-42 levels and a global cortical amyloid deposition measured from amyloid PET scans as biomarkers of amyloid-β. The regression coefficient of the AST to ALT ratio showed a negative association with CSF amyloid-β 1-42 levels (β [SE], –0.170 [0.061]; $P = .04$), indicating that higher AST to ALT ratio values were associated with CSF amyloid-β 1-42 positivity (Figure 1). However, there was no significant correlation between liver function markers and global cortical amyloid deposition.

In the whole-brain analysis using multivariable regression models to determine the association of liver function markers with amyloid-β load measured from amyloid PET scans on a voxelwise level. We identified significant associations for two liver function markers. Higher ALT levels were significantly associated with reduced amyloid-β deposition in the bilateral parietal lobes (Figure 2A). Increased AST to ALT ratio values were significantly associated with increased amyloid-β deposition in the bilateral parietal lobes and right temporal lobe (Figure 2C).

Biomarkers of fibrillary tau

We used CSF p-tau levels as a biomarker of fibrillary tau. We investigated the association of liver function markers with CSF p-tau, adjusting for *APOE ε4* status as a covariate. Higher AST to ALT ratio values were associated with higher CSF p-tau values (β [SE], 0.175 [0.055]; $P = .02$) (Figure 1).

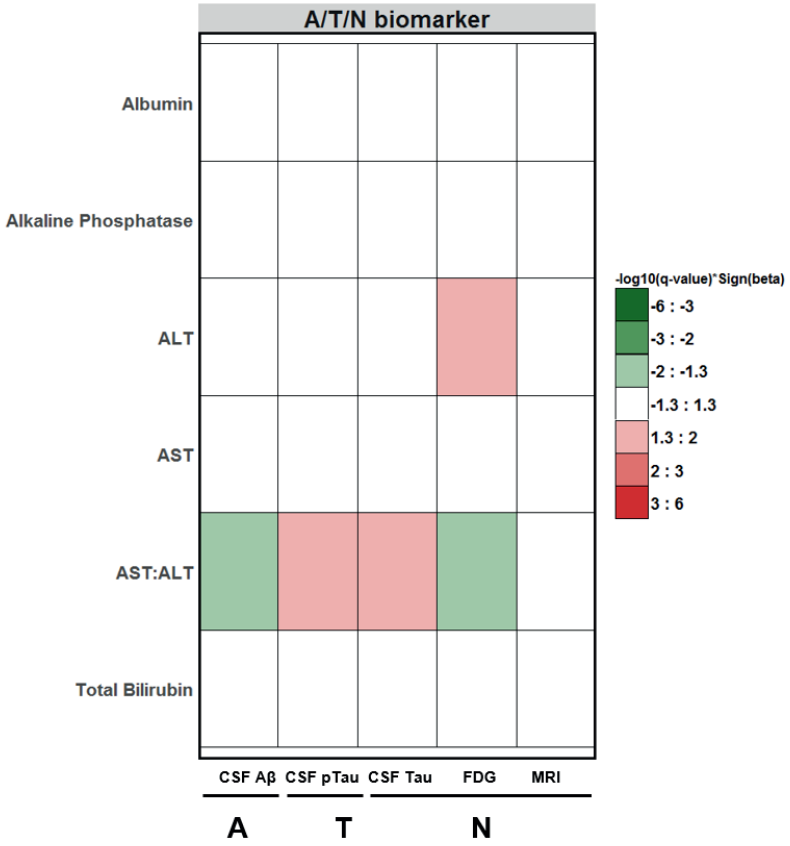


Figure 1: Results of association of liver function biomarkers with A/T/N biomarkers for Alzheimer's disease. Heat map of q-values of association between liver function markers and the A/T/N biomarkers for AD. P-values estimated from linear regression analyses were corrected for multiple testing using FDR (q-value). Color code: white indicates q-value>0.05, red indicates significant positive association, and green indicates significant negative association. Aβ indicates amyloid-β; ALT, alanine aminotransferase; AST, aspartate aminotransferase; CSF, cerebrospinal fluid; FDG, fludeoxyglucose positron emission tomography; MRI, magnetic resonance imaging; and p-tau, phosphorylated tau.

Biomarkers of neurodegeneration or neuronal injury

We used structural atrophy measured from MRI scans, brain glucose metabolism from FDG-PET scans, and CSF total tau (t-tau) levels as biomarkers of neurodegeneration or neuronal injury.

Brain glucose metabolism

We performed an ROI-based association analysis of liver function markers with a global cortical glucose metabolism value measured from FDG-PET scans across 1167 ADNI participants with both FDG-PET scans and measurement of liver function markers. The association analysis

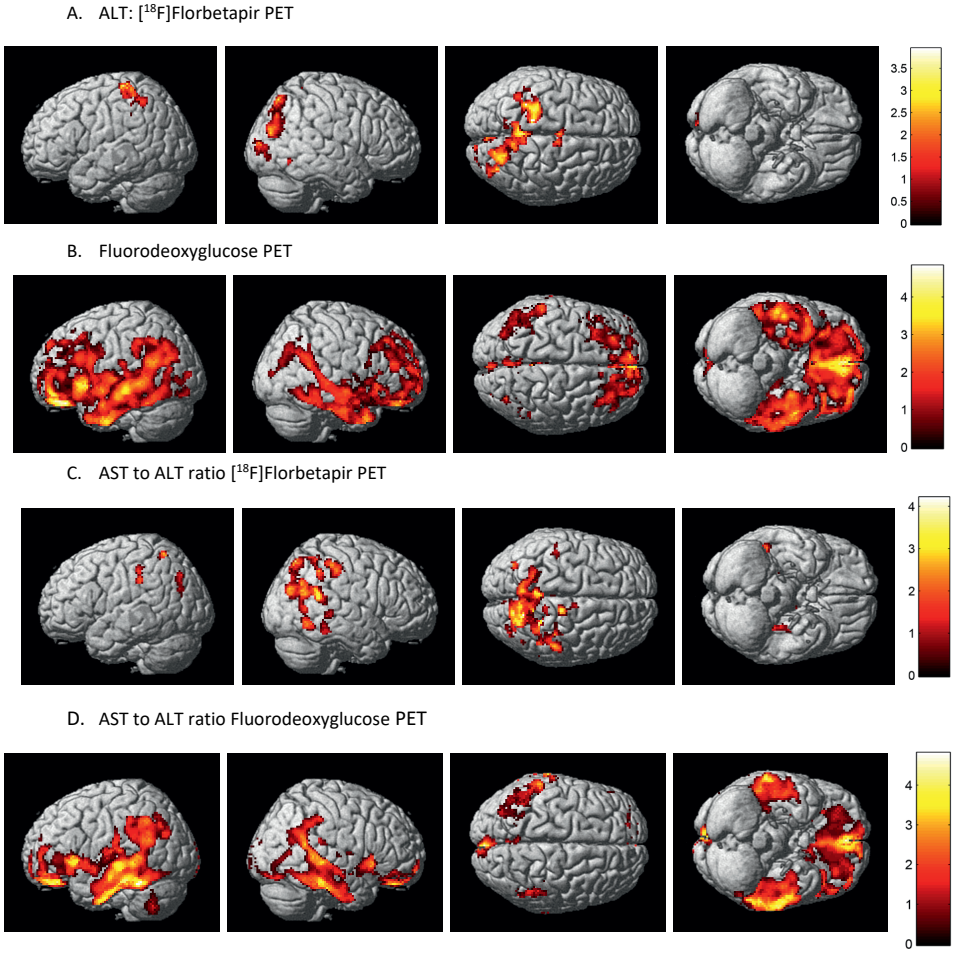


Figure 2: Detailed whole-brain voxel-based imaging analysis for alanine aminotransferase (ALT) and aspartate aminotransferase (AST) to ALT ratio levels using positron emission topography (PET) scans. Whole-brain multivariable analysis was performed to visualize the topography of the association of ALT levels and AST to ALT ratio values with amyloid-β load and glucose metabolism on a voxelwise level (false discovery rate–corrected $P < 0.05$). A: Higher ALT levels were significantly associated with reduced amyloid-β deposition in the bilateral parietal lobes. B: Increased ALT levels were significantly associated with increased glucose metabolism in a widespread manner, especially in the bilateral frontal, parietal, and temporal lobes. C: Increased AST to ALT ratio values were significantly associated with increased amyloid-β deposition in the bilateral parietal lobes and the right temporal lobe. D: Increased AST to ALT ratio values were significantly associated with reduced brain glucose metabolism in the bilateral frontal, parietal, and temporal lobes.

including *APOE ε4* status as a covariate identified 2 markers as significantly associated with brain glucose metabolism after controlling for multiple testing using FDR (Figure 1). For ALT, higher levels were associated with increased glucose metabolism (β [SE], 0.096 [0.030]; *P* = .02), while for the AST to ALT ratio, higher ratio values were associated with reduced glucose metabolism (β [SE], –0.123 [0.042]; *P* = .03).

In the detailed whole-brain analysis to determine the association of liver function markers with brain glucose metabolism on a voxelwise level. Increased ALT levels were associated with increased glucose metabolism in a widespread pattern, especially in the bilateral frontal, parietal, and temporal lobes (Figure 2B). However, higher AST to ALT ratio values were significantly associated with reduced glucose metabolism in the bilateral frontal, parietal, and temporal lobes (Figure 2D).

Structural MRI (atrophy)

In the investigation of the association of liver function markers with mean hippocampal volume with *APOE ε4* status as a covariate. We did not identify any significant association with hippocampal volume after controlling for multiple testing using FDR (Figure 1). Following the detailed whole-brain surface-based analysis of liver function markers using multivariable regression models to assess associations with cortical thickness. Higher ALT levels were significantly associated with larger cortical thickness in the bilateral temporal lobes (Figure 3), which showed consistent patterns in the associations of brain glucose metabolism.

CSF t-tau

Higher AST to ALT ratio values were associated with higher CSF t-tau levels (β [SE], 0.160 [0.049]; *P* = .02) (Figure 1), which showed consistent patterns in the associations of CSF amyloid-β1-42 or p-tau levels and brain glucose metabolism.

Discussion

We investigated the association between serum-based liver function markers and AD diagnosis, cognition, and AD pathophysiological characteristics based on the A/T/N framework for AD biomarkers in the ADNI cohort[36]. Our findings suggest that the decreased levels of ALT and elevated AST to ALT ratio that were observed in patients with AD were associated with poor cognition and reduced brain glucose metabolism. We also found that an increased AST to ALT ratio was associated with lower CSF amyloid-β1-42 levels, greater amyloid-β deposition, and higher CSF p-tau and t-tau levels. Furthermore, we observed that decreased levels of ALT were associated with greater amyloid-β deposition and structural atrophy.

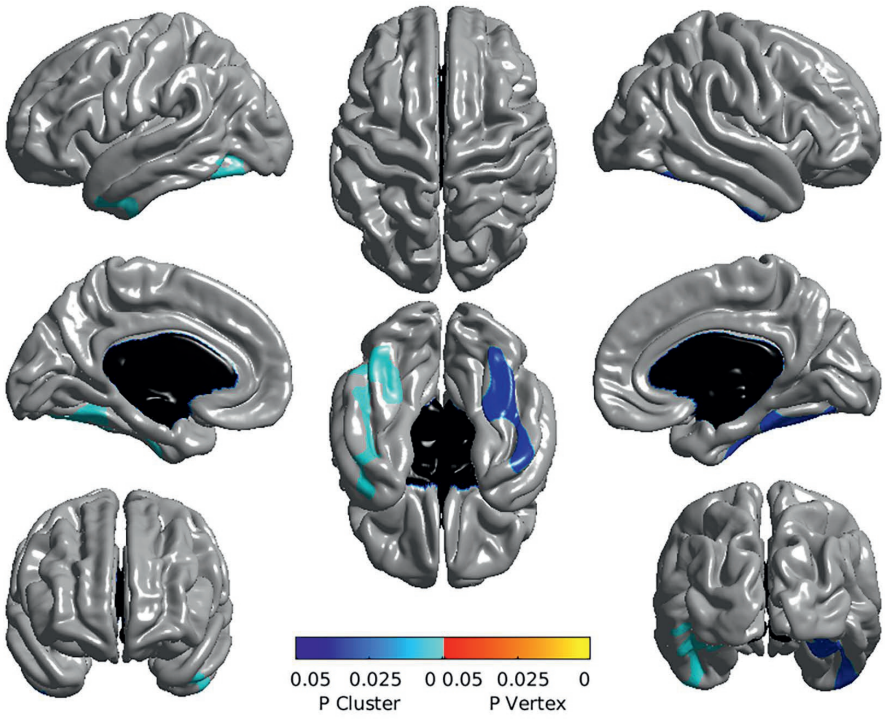


Figure 3: Detailed whole-brain surface-based imaging analysis for alanine aminotransferase (ALT) levels using magnetic resonance imaging (MRI) scans. A whole-brain multivariable analysis of cortical thickness across the brain surface was performed to visualize the topography of the association of ALT levels with brain structure. Statistical maps were threshold using a random field theory for a multiple testing adjustment to a corrected significance level of 0.05. The *P*-value for clusters indicates significant corrected *P* values with the lightest blue color. Higher ALT levels were significantly associated with greater cortical thickness, especially in bilateral temporal lobes.

Decreased levels of ALT and increased AST to ALT ratio values were observed in patients with AD and were associated with lower scores on measures of memory and executive function. Our findings are comparable with those of an earlier study that reported increased AST to ALT ratio values and lower levels of ALT in patients with AD compared with controls, although in that study, the association between AD and ALT levels did not reach statistical significance[37]. Altered liver enzymes lead to disturbances in liver associated metabolites including branched-chain amino acids, ether phosphatidylcholines, and lipids[38], which we and others show are altered in AD[1,39-41] and may play a role in disease pathophysiological characteristics[42]. Disturbed energy metabolism is one of the processes that may explain the observed lower levels of ALT and increased enzyme ratio in individuals with AD and impaired cognition[3,5]. This finding is concordant with our observation that increased AST to ALT ratio values and lower levels of ALT showed a consistent significant association with reduced brain glucose metabolism particularly in the orbitofrontal cortex and temporal lobes, areas of the brain implicated in memory and executive function.

Brain glucose hypometabolism is an early feature of AD and cognitive impairment during the prodromal stage[43,44]. Moreover, ALT and AST are key enzymes in gluconeogenesis in the liver and production of neurotransmitters required in maintaining synapses[45]. Alanine aminotransferase catalyzes a reversible transamination reaction between alanine and α-ketoglutarate to form pyruvate and glutamate, while AST catalyzes a reversible reaction between aspartate and α-ketoglutarate to form oxaloacetate and glutamate[46]. Although exact mechanisms remain unclear, 2 possible mechanisms may explain altered levels of enzymes in AD. First, reduced ALT levels lead to reduced pyruvate, which is required for glucose production via gluconeogenesis in the liver and glucose is distributed in various body tissues as an energy source[47], thus disturbing energy homeostasis. Second, altered levels of ALT and AST may affect levels of glutamate, an excitatory neurotransmitter of the central nervous system involved in synaptic transmission, which also plays an important role in memory[48].

In the case of low glucose metabolism in the brain, as observed in our current study, less α-ketoglutarate is available via the tricarboxylic acid cycle that favors glutamate catabolism vs glutamate synthesis in reversible reaction (catalyzed by AST and ALT) [49]. Glutamate acts as a neurotransmitter in approximately two-thirds of the synapses in neocortical and hippocampal pyramidal neurons and thus is involved in memory and cognition via long-term potentiation [50]. In a sample of healthy adults, plasma ALT and AST levels were significantly positively correlated with plasma glutamate levels[5,51], which indicates that lower levels of ALT will decrease glutamate levels in plasma. Based on evidence from earlier studies that peripheral blood levels of glutamate are positively correlated with levels of glutamate in the CSF[52] and studies that reported lower levels of glutamate in patients with AD compared with controls in both blood[53] and brain tissues[33,54-56], we can infer that lower levels of ALT or AST may affect glutamate levels in AD. In older adults, lower serum ALT levels are associated with mortality[57,58] and are thought to be a biomarker for increased frailty, sarcopenia and/or reduced levels of pyridoxine levels (Vitamin B6)[59]. Pyridoxine phosphate is a coenzyme for the synthesis of amino acids, neurotransmitters (e.g., serotonin and norepinephrine) and sphingolipids. Alanine aminotransferase decreases with age[60] and may be a sign of hepatic aging. Glutamate levels also decrease with increasing age[61]. Together with the fact that age is the strongest risk factor for AD[62], decreasing levels of ALT with age may also indicate a possible biological link between aging and AD. Nevertheless, further research is needed to determine the exact cause of reducing ALT levels with age and the pathway through which it can influence neurologic disorders including AD.

Increased AST to ALT ratio values are observed in individuals with nonalcoholic fatty liver disease which is the hepatic manifestation of metabolic syndrome[63]. In the Framingham Heart Study, nonalcoholic fatty liver disease was associated with smaller total cerebral brain volume even after adjustment for multiple cardiovascular risk factors[64]. Liver dysfunction is also associated with the development of disease including cardiovascular disease and insulin-resistance through disruptions in glucose and lipid metabolism, key physiological functions of

the liver[65,66]. Thus, using the AST to ALT ratio as a marker for overall metabolic disturbance[5], our study provides evidence of an association between altered metabolic status and AD, cognition, and AD endophenotypes.

In addition to ALT levels and the AST to ALT ratio, elevated levels of alkaline phosphatase were significantly associated with poor cognition. This is in line with results from the Oxford Project to Investigate Memory and Aging, which reported increased alkaline phosphatase levels in individuals with AD and an inverse association with cognition[67]. Alkaline phosphatase is an enzyme primarily expressed in the liver and kidneys as well as in endothelial cells in the brain[68,69]. The neuronal form of the alkaline phosphatase plays a role in developmental plasticity and activity-dependent cortical functions via contributing in γ-aminobutyric acid metabolism[70-73]. Changes in plasma levels of alkaline phosphatase may occur as a result of central nervous system injury[74].

Limitations

This study has several limitations. The observational design of this ADNI cohort study limits our ability to make assumptions about causality. There is need to evaluate the association of liver enzymes with AD in prospective manner. Another limitation of our study is that we did not adjust for alcohol consumption which was not available in ADNI. Alcohol consumption is associated with altered liver enzymes. Instead, we used a well-established surrogate marker of alcohol consumption, γ-glutamyltransferase. Elevations in γ-glutamyltransferase generally indicate long-term heavy drinking rather than episodic heavy drinking[75]. Our key findings remained significant after adjustment for γ-glutamyltransferase and statin use (Supplementary Table 4, Table 5, and Figure in the Supplement). However, given the strong associations with liver function measures and A/T/N biomarkers for AD, it appears that liver function may play a role in the pathogenesis of AD, but limitations should be taken into account before further extrapolating our findings.

Conclusions

This study's results suggest that altered liver function markers are associated with AD diagnosis and impaired memory and executive function as well as amyloid-β, tau and neurodegenerative biomarkers of AD pathophysiological characteristics. These results are, to our knowledge, the first to show an association of peripheral markers of liver functioning with central biomarkers associated with AD. Although our results suggest an important role of liver functioning in AD pathophysiological characteristics, the causal pathways remain unknown. The liver-brain biochemical axis of communication should be further evaluated in model systems and longitudinal studies to gain deeper knowledge of causal pathways.

Acknowledgements

The Alzheimer's Disease Neuroimaging Initiative (ADNI): Data used in the preparation of this article were obtained from the ADNI database (http://adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc .edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Funding/Support: Funding for ADMC (Alzheimer's Disease Metabolomics Consortium, led by R.K.-D. at Duke University) was provided by the National Institute on Aging grant R01AG046171, a component of the Accelerated Medicines Partnership for AD (AMP-AD) Target Discovery and Preclinical Validation Project (https://www.nia.nih.gov/research/dn/amp-ad-target-discovery-and-preclinical-validation-project) and the National Institute on Aging grant RF1 AG0151550, a component of the M^{2}OVE-AD Consortium (Molecular Mechanisms of the Vascular Etiology of AD – Consortium https://www.nia.nih.gov/news/decoding-molecular-ties-between-vascular-disease-and-alzheimers).

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

The work of various Consortium Investigators are also supported by various NIA grants [U01AG024904-09S4, P50NS053488, R01AG19771, P30AG10133, P30AG10124,

K01AG049050, R03 AG054936], the National Library of Medicine [R01LM011360, R01LM012535], and the National Institute of Biomedical Imaging and Bioengineering [R01EB022574]. Additional support came from Helmholtz Zentrum, the Alzheimer's Association, the Indiana Clinical and Translational Science Institute, and the Indiana University-IU Health Strategic Neuroscience Research Initiative.

Role of the Funder/Sponsor: [Funders listed above] had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Contributions: The authors are grateful to Lisa Howerton for administrative support and the numerous ADNI study volunteers and their families.

Disclosures

J.Q.T. may accrue revenue in the future on patents submitted by the University of Pennsylvania wherein he is a co-inventor and he received revenue from the sale of Avid to Eli Lily as a co-inventor on imaging-related patents submitted by the University of Pennsylvania. L.M.S. receives research funding from MJFox Foundation for PD Research and is a consultant for Eli Lilly, Novartis, and Roche; he provides QC over-sight for the Roche Elecsys immunoassay as part of responsibilities for the ADNI study. A.J.S. reports investigator-initiated research support from Eli Lilly unrelated to the work reported here. He has received consulting fees and travel expenses from Eli Lilly and Siemens Healthcare and is a consultant to Arkley BioTek. He also receives support from Springer-Nature publishing as Editor-In-Chief of *Brain Imaging and Behavior*. M.W.W. reports stock/stock options from Elan, Synarc, travel expenses from Novartis, Tohoku University, Fundacio Ace, Travel eDreams, MCI Group, NSAS, Danone Trading, ANT Congress, NeuroVigil, CHRU-Hopital Roger Salengro, Siemens, AstraZeneca, Geneva University Hospitals, Lilly, University of California, San Diego–ADNI, Paris University, Institut Catala de Neurociencies Aplicades, University of New Mexico School of Medicine, Ipsen, Clinical Trials on Alzheimer's Disease, Pfizer, AD PD meeting. P.M.D. has received research grants and advisory/speaking fees from several health and technology companies. He serves as a director and/or owns stock in companies/organizations whose products are not discussed here. P.M.D is also a co-inventor, through Duke, on patent applications on metabolomics for Alzheimer disease which may accrue revenue in the future. R.K.D. is inventor on key patents in the field of metabolomics including applications for Alzheimer disease. All other authors report no disclosures.

References

1. Toledo, J.B. *et al.* Metabolic network failures in Alzheimer's disease: A biochemical road map. *Alzheimers Dement* **13**, 965-984 (2017).
2. Clarke, J.R., Ribeiro, F.C., Frozza, R.L., De Felice, F.G. & Lourenco, M.V. Metabolic Dysfunction in Alzheimer's Disease: From Basic Neurobiology to Clinical Approaches. *J Alzheimers Dis* **64**, S405-S426 (2018).
3. Kapogiannis, D. & Mattson, M.P. Disrupted energy metabolism and neuronal circuit dysfunction in cognitive impairment and Alzheimer's disease. *Lancet Neurol* **10**, 187-98 (2011).
4. Craft, S. The role of metabolic disorders in Alzheimer disease and vascular dementia: two roads converged. *Arch Neurol* **66**, 300-5 (2009).
5. Sookoian, S. *et al.* Serum aminotransferases in nonalcoholic fatty liver disease are a signature of liver metabolic perturbations at the amino acid and Krebs cycle level, 2. *The American journal of clinical nutrition* **103**, 422-434 (2016).
6. Sookoian, S. & Pirola, C.J. Alanine and aspartate aminotransferase and glutamine-cycling pathway: Their roles in pathogenesis of metabolic syndrome. *World J Gastroenterol* **18**, 3775-81 (2012).
7. Goessling, W. *et al.* Aminotransferase levels and 20-year risk of metabolic syndrome, diabetes, and cardiovascular disease. *Gastroenterology* **135**, 1935-44, 1944.e1 (2008).
8. Sattar, N. *et al.* Elevated alanine aminotransferase predicts new-onset type 2 diabetes independently of classical risk factors, metabolic syndrome, and C-reactive protein in the west of Scotland coronary prevention study. *Diabetes* **53**, 2855-60 (2004).
9. Santos, C.Y. *et al.* Pathophysiologic relationship between Alzheimer's disease, cerebrovascular disease, and cardiovascular risk: A review and synthesis. *Alzheimer's & Dementia : Diagnosis, Assessment & Disease Monitoring* **7**, 69-87 (2017).
10. Fillit, H., Nash, D.T., Rundek, T. & Zuckerman, A. Cardiovascular risk factors and dementia. *Am J Geriatr Pharmacother* **6**, 100-18 (2008).
11. Jack, C.R., Jr. *et al.* NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimers Dement* **14**, 535-562 (2018).
12. Saykin, A.J. *et al.* Genetic studies of quantitative MCI and AD phenotypes in ADNI: Progress, opportunities, and plans. *Alzheimers Dement* **11**, 792-814 (2015).
13. Weiner, M.W. *et al.* Recent publications from the Alzheimer's Disease Neuroimaging Initiative: Reviewing progress toward improved AD clinical trials. *Alzheimers Dement* **13**, e1-e85 (2017).
14. Petersen, R.C. *et al.* Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology* **74**, 201-9 (2010).
15. Rattanabannakit, C. *et al.* The Cognitive Change Index as a Measure of Self and Informant Perception of Cognitive Decline: Relation to Neuropsychological Tests. *J Alzheimers Dis* **51**, 1145-55 (2016).
16. McKhann, G. *et al.* Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* **34**, 939-44 (1984).

17. Aisen, P.S. *et al.* Clinical Core of the Alzheimer's Disease Neuroimaging Initiative: progress and plans. *Alzheimers Dement* **6**, 239-46 (2010).
18. Crane, P.K. *et al.* Development and assessment of a composite score for memory in the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Brain imaging and behavior* **6**, 502-516 (2012).
19. Gibbons, L.E. *et al.* A composite score for executive functioning, validated in Alzheimer's Disease Neuroimaging Initiative (ADNI) participants with baseline mild cognitive impairment. *Brain imaging and behavior* **6**, 517-527 (2012).
20. Jack, C.R., Jr. *et al.* Update on the magnetic resonance imaging core of the Alzheimer's disease neuroimaging initiative. *Alzheimers Dement* **6**, 212-20 (2010).
21. Jack, C.R., Jr. *et al.* The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *J Magn Reson Imaging* **27**, 685-91 (2008).
22. Kim, S. *et al.* Influence of genetic variation on plasma protein levels in older adults using a multi-analyte panel. *PLoS One* **8**, e70269 (2013).
23. Nho, K. *et al.* Whole-exome sequencing and imaging genetics identify functional variants for rate of change in hippocampal volume in mild cognitive impairment. *Mol Psychiatry* **18**, 781-7 (2013).
24. Nho, K. *et al.* Protective variant for hippocampal atrophy identified by whole exome sequencing. *Ann Neurol* **77**, 547-52 (2015).
25. Fischl, B., Sereno, M.I. & Dale, A.M. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage* **9**, 195-207 (1999).
26. Dale, A.M., Fischl, B. & Sereno, M.I. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* **9**, 179-94 (1999).
27. Chung, M.K., Worsley, K.J., Nacewicz, B.M., Dalton, K.M. & Davidson, R.J. General multivariate linear modeling of surface shapes using SurfStat. *Neuroimage* **53**, 491-505 (2010).
28. Risacher, S.L. *et al.* APOE effect on Alzheimer's disease biomarkers in older adults with significant memory concern. *Alzheimers Dement* **11**, 1417-29 (2015).
29. Bittner, T. *et al.* Technical performance of a novel, fully automated electrochemiluminescence immunoassay for the quantitation of beta-amyloid (1-42) in human cerebrospinal fluid. *Alzheimers Dement* **12**, 517-26 (2016).
30. Hansson, O. *et al.* CSF biomarkers of Alzheimer's disease concord with amyloid-beta PET and predict clinical progression: A study of fully automated immunoassays in BioFINDER and ADNI cohorts. *Alzheimers Dement* (2018).
31. Noble, K.G. *et al.* Hippocampal volume varies with educational attainment across the life-span. *Frontiers in Human Neuroscience* **6**, 307 (2012).
32. Hagler, D.J., Jr., Saygin, A.P. & Sereno, M.I. Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data. *Neuroimage* **33**, 1093-103 (2006).
33. Hayasaka, S., Phan, K.L., Liberzon, I., Worsley, K.J. & Nichols, T.E. Nonstationary cluster-size inference with random field and permutation methods. *Neuroimage* **22**, 676-87 (2004).
34. Worsley, K.J., Taylor, J.E., Tomaiuolo, F. & Lerch, J. Unified univariate and multivariate random field theory. *Neuroimage* **23 Suppl 1**, S189-95 (2004).

35. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological* **57**, 289-300 (1995).
36. Jack, C.R., Jr. *et al.* NIA-AA research framework: Towards a biological definition of Alzheimer's disease. in *Alzheimer's Association International Conference* 1-57 (Alzheimer's Association, London, England, 2017).
37. Giambattistelli, F. *et al.* Effects of hemochromatosis and transferrin gene mutations on iron dyshomeostasis, liver dysfunction and on the risk of Alzheimer's disease. *Neurobiol Aging* **33**, 1633-41 (2012).
38. Pietzner, M. *et al.* Hepatic steatosis associates with adverse molecular signatures in subjects without diabetes. *J Clin Endocrinol Metab* (2018).
39. Varma, V.R. *et al.* Brain and blood metabolite signatures of pathology and progression in Alzheimer disease: A targeted metabolomics study. *PLoS Med* **15**, e1002482 (2018).
40. Kaddurah-Daouk, R. *et al.* Alterations in metabolic pathways and networks in mild cognitive impairment and early Alzheimer's disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* **9**, P571 (2013).
41. Kaddurah-Daouk, R. *et al.* Metabolomic changes in autopsy-confirmed Alzheimer's disease. *Alzheimers Dement* **7**, 309-17 (2011).
42. Botros, M. & Sikaris, K.A. The de ritis ratio: the test of time. *Clin Biochem Rev* **34**, 117-30 (2013).
43. Mosconi, L. *et al.* Hypometabolism exceeds atrophy in presymptomatic early-onset familial Alzheimer's disease. *J Nucl Med* **47**, 1778-86 (2006).
44. Drzezga, A. *et al.* Cerebral metabolic changes accompanying conversion of mild cognitive impairment into Alzheimer's disease: a PET follow-up study. *Eur J Nucl Med Mol Imaging* **30**, 1104-13 (2003).
45. Rui, L. Energy metabolism in the liver. *Compr Physiol* **4**, 177-97 (2014).
46. Ellinger, J.J., Lewis, I.A. & Markley, J.L. Role of aminotransferases in glutamate metabolism of human erythrocytes. *J Biomol NMR* **49**, 221-9 (2011).
47. Qian, K. *et al.* Hepatic ALT isoenzymes are elevated in gluconeogenic conditions including diabetes and suppressed by insulin at the protein level. *Diabetes/metabolism research and reviews* **31**, 562-571 (2015).
48. Reis, H.J. *et al.* Neuro-transmitters in the central nervous system & their implication in learning and memory processes. *Curr Med Chem* **16**, 796-840 (2009).
49. Fleck, M.W., Henze, D.A., Barrionuevo, G. & Palmer, A.M. Aspartate and glutamate mediate excitatory synaptic transmission in area CA1 of the hippocampus. *J Neurosci* **13**, 3944-55 (1993).
50. Francis, P.T. Glutamatergic systems in Alzheimer's disease. *Int J Geriatr Psychiatry* **18**, S15-21 (2003).
51. Kamada, Y. *et al.* Impact of plasma transaminase levels on the peripheral blood glutamate levels and memory functions in healthy subjects. *BBA Clin* **5**, 101-7 (2016).

52. Alfredsson, G., Wiesel, F.A. & Tylec, A. Relationships between glutamate and monoamine metabolites in cerebrospinal fluid and serum in healthy volunteers. *Biol Psychiatry* **23**, 689-97 (1988).
53. Wang, G. *et al.* Plasma metabolite profiles of Alzheimer's disease and mild cognitive impairment. *J Proteome Res* **13**, 2649-58 (2014).
54. Lowe, S.L., Bowen, D.M., Francis, P.T. & Neary, D. Ante mortem cerebral amino acid concentrations indicate selective degeneration of glutamate-enriched neurons in Alzheimer's disease. *Neuroscience* **38**, 571-7 (1990).
55. Fayed, N., Modrego, P.J., Rojas-Salinas, G. & Aguilar, K. Brain glutamate levels are decreased in Alzheimer's disease: a magnetic resonance spectroscopy study. *Am J Alzheimers Dis Other Demen* **26**, 450-6 (2011).
56. Procter, A.W. *et al.* Evidence of glutamatergic denervation and possible abnormal metabolism in Alzheimer's disease. *J Neurochem* **50**, 790-802 (1988).
57. Liu, Z. *et al.* Complex Association between Alanine Aminotransferase Activity and Mortality in General Population: A Systematic Review and Meta-Analysis of Prospective Studies. *PLOS ONE* **9**, e91410 (2014).
58. Peltz-Sinvani, N. *et al.* Low ALT Levels Independently Associated with 22-Year All-Cause Mortality Among Coronary Heart Disease Patients. *Journal of General Internal Medicine* **31**, 209-214 (2016).
59. Vespasiani-Gentilucci, U. *et al.* Low Alanine Aminotransferase Levels in the Elderly Population: Frailty, Disability, Sarcopenia, and Reduced Survival. *The Journals of Gerontology: Series A* **73**, 925-930 (2018).
60. Elinav, E. *et al.* Correlation between serum alanine aminotransferase activity and age: an inverted U curve pattern. *Am J Gastroenterol* **100**, 2201-4 (2005).
61. Kaiser, L.G., Schuff, N., Cashdollar, N. & Weiner, M.W. Age-related glutamate and glutamine concentration changes in normal human brain: 1H MR spectroscopy study at 4 T. *Neurobiol Aging* **26**, 665-72 (2005).
62. Guerreiro, R. & Bras, J. The age factor in Alzheimer's disease. *Genome Med* **7**, 106 (2015).
63. Katsiki, N., Perez-Martinez, P., Anagnostis, P., Mikhailidis, D.P. & Karagiannis, A. Is Nonalcoholic Fatty Liver Disease Indeed the Hepatic Manifestation of Metabolic Syndrome? *Curr Vasc Pharmacol* **16**, 219-227 (2018).
64. Weinstein, G. *et al.* Association of Nonalcoholic Fatty Liver Disease With Lower Brain Volume in Healthy Middle-aged Adults in the Framingham Study. *JAMA Neurol* **75**, 97-104 (2018).
65. Bedogni, G. *et al.* Relationship between glucose metabolism and non-alcoholic fatty liver disease severity in morbidly obese women. *J Endocrinol Invest* **37**, 739-744 (2014).
66. Perla, F.M., Prelati, M., Lavorato, M., Visicchio, D. & Anania, C. The Role of Lipid and Lipoprotein Metabolism in Non-Alcoholic Fatty Liver Disease. *Children (Basel)* **4**(2017).
67. Kellett, K.A.B., Williams, J., Vardy, E., Smith, A.D. & Hooper, N.M. Plasma alkaline phosphatase is elevated in Alzheimer's disease and inversely correlates with cognitive function. *Int J Mol Epidemiol Genet* **2**, 114-21 (2011).

68. Moss, D.W. Physicochemical and pathophysiological factors in the release of membrane-bound alkaline phosphatase from cells. *Clin Chim Acta* **257**, 133-40 (1997).
69. Goldstein, D.J., Rogers, C.E. & Harris, H. Expression of alkaline phosphatase loci in mammalian tissues. *Proc Natl Acad Sci U S A* **77**, 2857-60 (1980).
70. Fonta, C., Negyessy, L., Renaud, L. & Barone, P. Areal and subcellular localization of the ubiquitous alkaline phosphatase in the primate cerebral cortex: evidence for a role in neurotransmission. *Cereb Cortex* **14**, 595-609 (2004).
71. Waymire, K.G. *et al.* Mice lacking tissue non-specific alkaline phosphatase die from seizures due to defective metabolism of vitamin B-6. *Nat Genet* **11**, 45-51 (1995).
72. Narisawa, S., Wennberg, C. & Millan, J.L. Abnormal vitamin B6 metabolism in alkaline phosphatase knock-out mice causes multiple abnormalities, but not the impaired bone mineralization. *J Pathol* **193**, 125-33 (2001).
73. Langer, D., Ikehara, Y., Takebayashi, H., Hawkes, R. & Zimmermann, H. The ectonucleotidases alkaline phosphatase and nucleoside triphosphate diphosphohydrolase 2 are associated with subsets of progenitor cell populations in the mouse embryonic, postnatal and adult neurogenic zones. *Neuroscience* **150**, 863-79 (2007).
74. Yamashita, M. *et al.* Measurement of serum alkaline phosphatase isozyme I in brain-damaged patients. *Neurol Med Chir (Tokyo)* **29**, 995-8 (1989).
75. Gjerde, H., Amundsen, A., Skog, O.-J., Mørland, J. & Aasland, O.G. Serum gamma-glutamyltransferase: An epidemiological indicator of alcohol consumption? *British Journal of Addiction* **82**, 1027-1031 (1987).

Chapter 5.2

Altered bile acid profile associates with cognitive impairment in Alzheimer's disease – An emerging role for gut microbiome

Siamak MahmoudianDehkordi*, Matthias Arnold*, Kwangsik Nho*, Shahzad Ahmad, Wei Jia, Guoxiang Xie, Gregory Louie, Alexandra Kueider-Paisley, M. Arthur Moseley, J. Will Thompson, Lisa St John Williams, Jessica D. Tenenbaum, Colette Blach, Rebecca Baillie, Xianlin Han, Sudeepa Bhattacharyya, Jon B. Toledo, Simon Schafferer, Sebastian Klein, Therese Koal, Shannon L. Risacher, Mitchel Allan Kling, Alison Motsinger-Reif, Daniel M. Rotroff, John Jack, Thomas Hankemeier, David A. Bennett, Philip L. De Jager, John Q. Trojanowski, Leslie M. Shaw, Michael W. Weiner, P. Murali Doraiswamy, Cornelia M. van Duijn, Andrew J. Saykin[#], Gabi Kastenmüller[#], Rima Kaddurah-Daouk[#], for the Alzheimer's Disease Neuroimaging Initiative and the Alzheimer Disease Metabolomics Consortium

**First authors contributed equally to the work*
#Senior authors contributed equally to the work

This chapter is published in the Alzheimer's and dementia: The Journal of the Alzheimer's Association. 2019 January;15(1): 76–92

Supplementary information for this chapter is available at: https://doi.org/10.1016/j.jalz.2018.07.217

Abstract

Introduction: Increasing evidence suggests a role for the gut microbiome in central nervous system disorders and a specific role for the gut-brain axis in neurodegeneration. Bile acids (BAs), products of cholesterol metabolism and clearance, are produced in the liver and are further metabolized by gut bacteria. They have major regulatory and signaling functions and seem dysregulated in Alzheimer's disease (AD).

Methods: Serum levels of 15 primary and secondary BAs and their conjugated forms were measured in 1464 subjects including 370 cognitively normal older adults (CN), 284 with early mild cognitive impairment (MCI), 505 with late MCI, and 305 AD cases enrolled in the AD Neuroimaging Initiative. We assessed associations of BA profiles including selected ratios with diagnosis, cognition, and AD-related genetic variants, adjusting for confounders and multiple testing.

Results: In AD compared to CN, we observed significantly lower serum concentrations of a primary BA (cholic acid [CA]) and increased levels of the bacterially produced, secondary BA, deoxycholic acid (DCA), and its glycine and taurine conjugated forms. An increased ratio of DCA:CA, which reflects 7 α-dehydroxylation of CA by gut bacteria, strongly associated with cognitive decline, a finding replicated in serum and brain samples in the Rush Religious Orders and Memory and Aging Project. Several genetic variants in immune response-related genes implicated in AD showed associations with BA profiles.

Discussion: We report for the first time an association between altered BA profile, genetic variants implicated in AD and cognitive changes in disease using a large multicenter study. These findings warrant further investigation of gut dysbiosis and possible role of gut-liver-brain axis in the pathogenesis of AD.

Introduction

Alzheimer's disease (AD), a progressive neurodegenerative disorder, is the leading cause of dementia in old age affecting over 40 million people worldwide[1]. There are currently no therapies to prevent or slow down AD progression, highlighting our incomplete knowledge of disease mechanisms and the need for new drug targets. A large number of biochemical processes are affected in AD and genes implicated in AD highlight the possible roles for lipid processing, immune function, phagocytosis, (innate) immunity and neurotransmitter function, and biological pathways that may affect metabolism[2,3]. Recent AD hypotheses implicate viral and bacterial contributions to disease pathogenesis[4-6].

Bidirectional biochemical communication between the brain and the gut contribute to a variety of neurodegenerative and psychiatric diseases[7-10]. The gut microbiome and the host collaboratively produce a large array of small molecules that impacts human health[11,12]. Recently, a role for the gut microbiome in motor dysfunction in Parkinson's disease has been highlighted[13] and several animal models of AD showed a possible role of gut bacteria in amyloid-b pathology[14,15]. The *APP* transgenic mouse model of AD had a drastically altered gut microbiome composition compared to wild-type mice[15]. Other studies linked proinflammatory bacteria, such as gram-negative producers of neurotoxic lipopolysaccharides, to brain amyloidosis and systemic inflammation, a central feature of AD[16,17]. These studies suggest microbial dysbiosis or imbalance could potentially contribute to AD pathogenesis.

Cholesterol metabolism in the liver is thought to play a key role in AD[18]. In fact, many cholesterol metabolism-related genes (e.g., *BIN1*, *CLU*, *PICALM*, *ABCA7*, *ABCG1*, and *SORL1)* are among the top AD susceptibility loci identified by genome-wide association studies[2,19]. Cholesterol is cleared through production of bile acids (BAs). Primary BAs, chenodeoxycholic acid (CDCA) and cholic acid (CA), are synthesized from cholesterol in the liver, conjugated with glycine or taurine, secreted into the gallbladder via the bile salt export pump, and transported to the intestine to be metabolized by gut bacteria (Figure 1). Intestinal anaerobic bacteria deconjugate the liver-derived BAs through the action of bile salt hydrolases to their respective free BAs. Subsequently, anaerobe bacteria convert primary BAs to the secondary BAs. That is, CA is converted to deoxycholic acid (DCA). CDCA is converted to lithocholic acid (LCA) and ursodeoxycholic acid through 7α or 7β –dehydroxylation, respectively[20,21]. In the terminal ileum and colon, BAs are reabsorbed by the enterocytes and released into the portal vein for return to the liver where they are conjugated to produce their glycine and taurine forms.

Beyond BA's role in cholesterol clearance, BAs are major regulators for maintaining energy homeostasis through binding to nuclear receptors, including FXR and LXR, among others and seem to be indicators of gut dysbiosis. BAs also modulate the gut microbiome[22,23]. Both primary and secondary BAs are present in the brains of mice and possibly humans with evidence that they

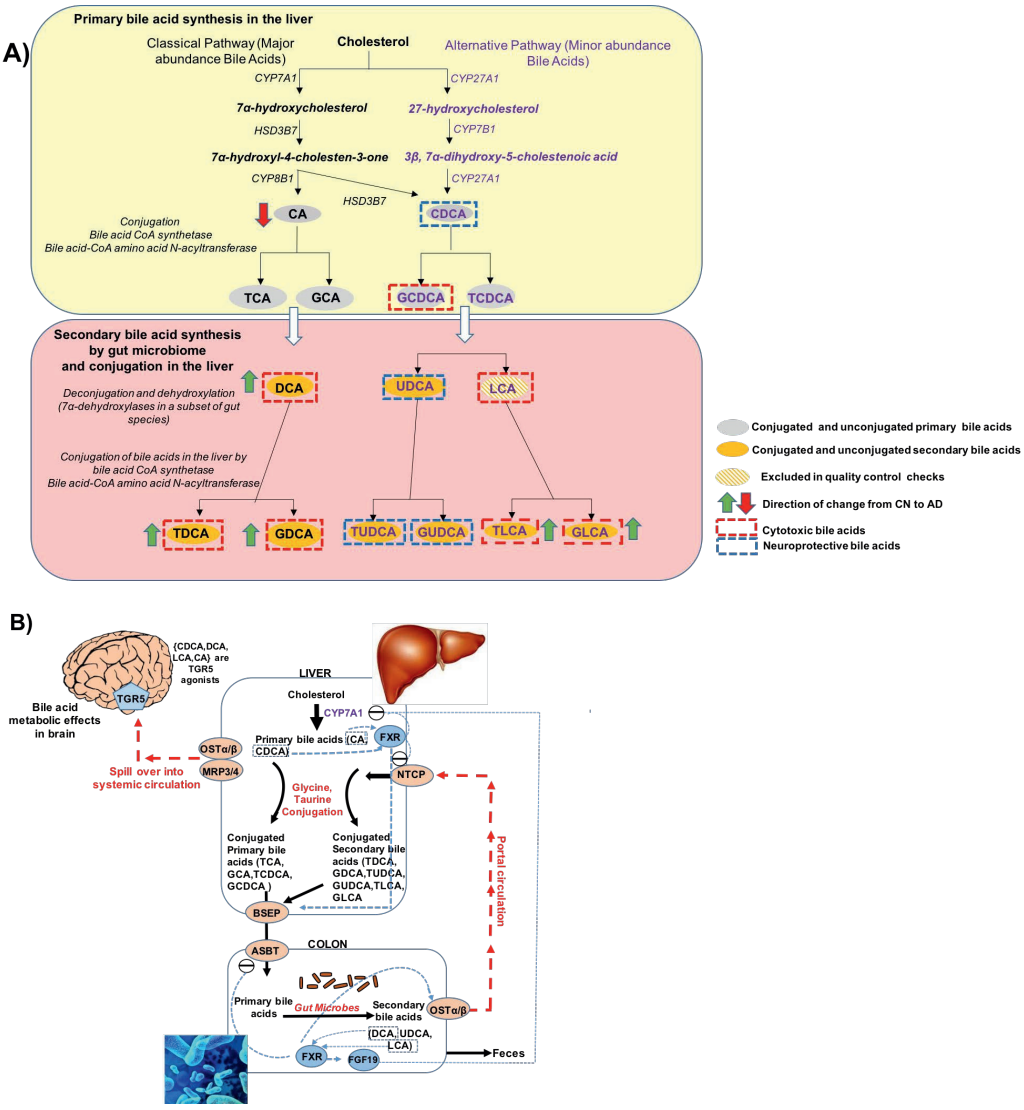


Figure 1: Bile acid synthesis and cholesterol clearance pathway. Regulation of bile acid synthesis by feedback mechanism and bile acid transport through enterohepatic circulation. In the liver, the bile acids (CDCA, DCA, LCA, CA) activate FXR that inhibits (via a repressor SHP, not shown here) the rate-limiting enzyme CYP7A1. The bile acids via FXR/SHP also inhibit the influx transporter NTCP; induce BSEP and canalicular bile acid secretion. In the intestine, bile acids, via FXR, inhibit the uptake transporter ASBT, decreasing absorption and increasing basolateral secretion into portal circulation by inducing OSTα & β. Bile acid activated FXR in the intestine also exerts inhibition on CYP7A1 in the liver via FGF19 pathway. At the basolateral membrane of hepatocytes, transporters OSTα & β, and also MRP3 and MRP4, secrete bile acids into the systemic circulation.
Abbreviations: ASBT: Apical sodium-dependent Bile acid Transporters; BSEP: Bile Salt Export Pump; FXR: Farnesoid X Receptor; NTCP: Sodium/Taurocholate Co-transporting Polypeptide; SHP: Small heterodimer partner.

cross the blood-brain barrier[24-29]. Some BAs such ursodeoxycholic acid exert beneficial effects while others are known to be cytotoxic[30-34]. In particular, DCA's toxicity has been associated with modulating apoptosis involving mitochondrial pathways in a variety of tissues and cell types[35-38].

In recent pilot human studies, BA profiles were shown to be affected in AD[26,38-41]. Here, we used a targeted metabolomics approach to evaluate BA profiles in a large cohort of 1464 individuals enrolled in the AD Neuroimaging Initiative (ADNI) where rich clinical, imaging, and genetic data exist. A schematic representation of study design is shown in Figure 2. We used these data to address the following:

1. Investigate if BA profiles are altered in mild cognitive impairment (MCI) and AD patients and if these differences are related to cognitive decline.
2. Use ratios of BAs to pinpoint possible enzymatic alterations in the liver and in the gut microbiome that directly contribute to altered BA profile.
3. Investigate whether immune-related AD genome-wide significant genes affect levels of BAs in circulation as markers for altered gut microbiome function.

In a subsequent study, we evaluated correlations between BAs and ATN (amyloid, tau and neurodegenerative) biomarkers of AD including cerebrospinal (CSF) biomarkers, brain atrophy, and brain glucose metabolism.

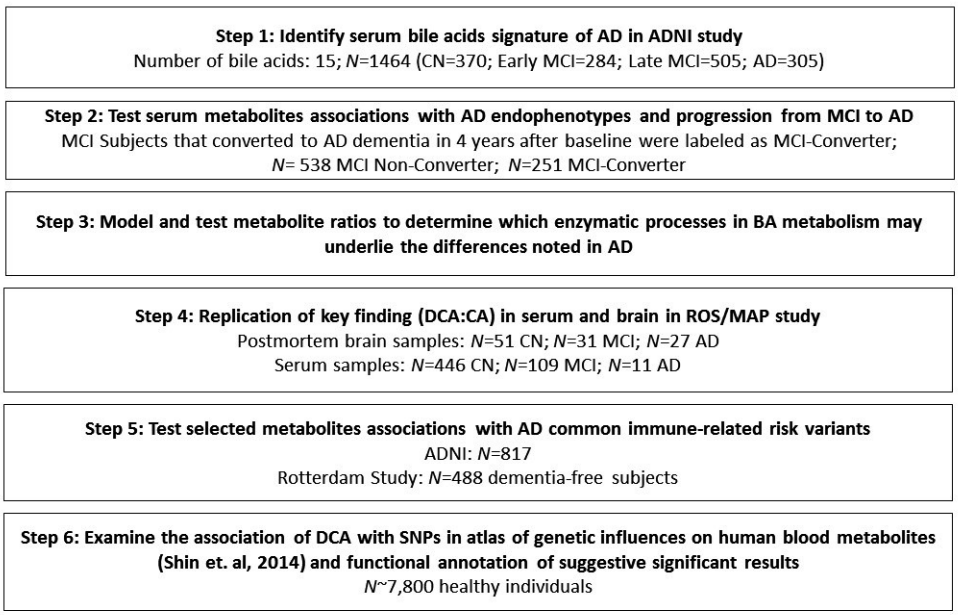


Figure 2: Schematic representation of study design.

Methods

Study cohorts and samples

ADNI baseline samples

Data used in the preparation of this article were downloaded from the ADNI database (http://adni.loni.usc.edu/). The ADNI studies have recruited over 1500 adults, ages 55 to 90, consisting of cognitively normal older individuals (CN), individuals with subjective memory concerns (SMC), subjects with early (EMCI) or late mild cognitive impairment (LMCI), and patients with early probable AD dementia. Subjects categorized as SMC were excluded in this study. For key clinical and demographic variables of ADNI participants included in this study, see Tables 1 and 2.

The Religious Orders Study and the Rush Memory and Aging Project (ROS/MAP) for replication of key finding

The ROS/MAP studies are both longitudinal cohort studies of aging and AD at Rush University and are designed to be used in joint analyses to maximize sample size. ROS enrolled individuals from religious orders (nuns, priests, brothers) across the United States[42]. MAP was designed to complement the ROS study by using a similar structure and design as ROS, but enrolling participants with a wider range of life experiences and socioeconomic status from the Chicago, IL metropolitan area[43]. The entire ROS/MAP cohort consists of approximately 3300 participants, more than 1500 of whom have come to autopsy (www.radc.rush.edu). We measured a subset of serum BAs in 566 subjects (446 CN, 109 MCI, and 11 AD), as well as a subset of BAs in postmortem brain samples from the dorsolateral prefrontal cortex of 111 subjects with brain pathology measured (51 CN, 31 MCI, and 27 AD at time of death), of whom 93 had serum measurements. Key demographic characteristics of the ROS/MAP cohort are in Supplementary Table 1.

Rotterdam Study (RS)

RS was used to examine the association of BAs with AD genetic variants. RS is a prospective population-based study[44]. At the baseline examination in 1990-1993, 7983 subjects≥ 55 years of age were recruited from the Ommoord district of Rotterdam (RS-I). All the study participants were extensively interviewed and physically examined at baseline and after every 3 to 4 years. During 2000 to 2001, the baseline cohort (RS-I) was expanded with 3011 subjects ≥55 years of age, who were not yet part of RS-I (RS-II). In this analysis, fasting serum BAs were measured for 488 dementia-free subjects with mean(SD) age of 73.1(6.3) from RS-I using Metabolon platform (Durham, North Carolina, USA) as described previously[45] (see Supplementary Table 2 for demographics).

Table 1: Demographics of ADNI participants stratified by baseline diagnosis*

Variable	N	CN (N=370)	EMCI (N=284)	LMCI (N=505)	AD (N=305)	*P-value*†
Age	1464	74.58(5.71)	71.12(7.51)	73.95(7.59)	74.70(7.79)	0.001
Sex: Female, No. (%)	1464	190(51%)	130(46%)	197(39%)	139(46%)	0.004
Education, years	1464	16.28(2.92)	15.95(2.66)	15.87(2.90)	15.16(3.00)	0.001
BMI (Kg/M²)	1461	27.05(4.46)	28.06(5.41)	26.54(4.25)	25.83(4.69)	0.001
≥1 *APOE* ε4 allele, No. (%)	1464	104(28%)	121(43%)	273(54%)	202(66%)	0.001
ADAS-Cog13‡	1455	9.19(4.17)	12.64(5.40)	18.67(6.62)	29.67(8.20)	0.001

Abbreviations: AD: Alzheimer's disease; BMI: Body mass index; CN: Cognitively normal; EMCI: Early mild cognitive impairment; LMCI: Late mild cognitive impairment; ADAS-Cog13, Alzheimer Disease Assessment Scale 13-item cognitive subscale.
*Data are reported as mean (SD) unless otherwise indicated. Bolded values indicate statistical significance. SD: Standard deviation.
†Based on 2-sample *t* tests, or Pearson χ^2 tests.
‡Score explanations: ADAS-Cog13 range, 0 (best) to 85 (worst).

Table 2: Demographics of ADNI participants stratified by MCI progression to AD*

Variable	N	MCI-nonconverter (N=538)	MCI-converter (N=251)	*P-value†*
Age	789	72.47(7.90)	73.91(7.08)	0.01
Sex: Female, No. (%)	789	41% (223)	41% (104)	1
Education, years	789	15.95 (2.85)	15.79 (2.76)	0.43
BMI (Kg/M²)	788	27.37 (4.80)	26.47(4.61)	0.005
≥1 *APOE* ε4 allele, No. (%)	789	41% (223)	68% (171)	0.001
ADAS-Cog13	786	14.26 (6.04)	21.31 (5.94)	0.29

Abbreviations: AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; ADAS-Cog13, Alzheimer Disease Assessment Scale 13-item cognitive subscale; MCI, mild cognitive impairment.
*MCI subjects that converted to AD dementia in 4 years after baseline were labeled as MCI converter.
†Based on 2-sample *t* tests, or Pearson χ^2 tests.

Sample collection and quantification of BAs

Targeted metabolomics profiling was performed to measure concentrations of 20 BA metabolites in serum samples of the ADNI cohorts. Morning fasting serum samples from the baseline visit were collected and aliquoted as described in the ADNI standard operating procedures. BA quantification was performed by liquid chromatography tandem mass spectrometry using the Biocrates® Life Sciences Bile Acids Kit (BIOCRATES Life Science AG, Innsbruck, Austria) according to manufacturer's instructions (see Table 3 for list of BAs, abbreviations and their levels across diagnosis groups).

Table 3: Levels of primary and secondary bile acids measured in the ADNI cohort stratified by clinical diagnosis*

Bile Acid	Category	N†	CN (N=370)	EMCI (N=284)	LMCI (N=505)	AD (N=305)
CA	Primary	1446	0.221(0.024)	0.155(0.021)	0.192(0.021)	0.135(0.025)
CDCA	Primary	1357	0.285(0.042)	0.241(0.034)	0.288(0.033)	0.216(0.033)
GCA	Primary Conjugated	1463	0.236(0.019)	0.234(0.021)	0.239(0.014)	0.297(0.037)
GCDCA	Primary Conjugated	1464	0.658(0.035)	0.724(0.059)	0.710(0.037)	0.806(0.049)
TCA	Primary Conjugated	1020	0.068(0.008)	0.057(0.006)	0.068(0.006)	0.066(0.009)
TCDCA	Primary Conjugated	1426	0.090(0.006)	0.088(0.007)	0.091(0.006)	0.097(0.008)
TMCA	Primary Conjugated	1146	0.012(0.001)	0.011(0.001)	0.014(0.002)	0.014(0.002)
DCA	Secondary	1445	0.526(0.041)	0.574(0.043)	0.529(0.026)	0.627(0.045)
UDCA	Secondary	1111	0.065(0.007)	0.072(0.011)	0.091(0.010)	0.087(0.012)
GDCA	Secondary Conjugated	1439	0.440(0.034)	0.488(0.038)	0.502(0.031)	0.672(0.054)
TDCA	Secondary Conjugated	1430	0.058(0.006)	0.059(0.005)	0.065(0.005)	0.077(0.006)
GLCA	Secondary Conjugated	1037	0.027(0.002)	0.034(0.003)	0.030(0.002)	0.039(0.003)
TLCA	Secondary Conjugated	1008	0.005(0.0002)	0.005(0.0003)	0.005(0.0003)	0.006(0.0005)
GUDCA	Secondary Conjugated	1401	0.115(0.010)	0.114(0.012)	0.129(0.012)	0.136(0.015)
TUDCA	Secondary Conjugated	1369	0.008(0.001)	0.008(0.001)	0.008(0.001)	0.008(0.001)

Abbreviations: AD,Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; CN, cognitively normal; EMCI, early mild cognitive impairment;LMCI, late mild cognitive impairment; CA, cholic acid; CDCA, chenodeoxycholic acid; GCA, glycocholic acid; GCDCA, glycochenodeoxycholic acid; TCA, taurocholic acid; TCDCA, taurochenodeoxycholic acid; TMCA, tauromuricholic acid; DCA, deoxycholic acid; UDCA, ursodeoxycholic acid; GDCA, glycodeoxycholic acid; TDCA, taurodeoxycholic acid; GLCA, glycolithocholic acid; TLCA, taurolithocholic acid; GUDCA, glycoursodeoxycholic acid; TUDCA, tauroursodeoxycholic acid.

*Values represent mM in mean (standard error of the mean).

†Number of nonmissing measurements.

In the ROS/MAP, quantification of BA concentrations in 566 serum samples and 111 postmortem brain samples was performed at the University of Hawaii cancer center using ultra-performance liquid chromatography coupled to a tandem mass spectrometry (UPLC-MS/MS) system (ACQUITY UPLC-Xevo TQ-S, Waters Corp., Milford, MA)[46].

In the RS study, serum BAs were measured in 488 serum samples using the non-targeted Metabolon platform (Durham, North Carolina, USA).

Quality control of BA profiles

Metabolomics laboratory staff were blinded to diagnosis and pathological data in all the studies. In ADNI, after unblinding and data release, metabolite profiles went through quality-control (QC) checks and data preprocessing including batch-effect adjustment, missing value imputation, and log-transformation (Supplementary Methods and Supplementary Table 3). After QC correction, the data set included 15 BAs (Five BAs did not pass QC criteria) for a total of 1,464 subjects (after excluding 99 SMC). The preprocessed BA values after QC were used for subsequent association analyses directly or were adjusted to take into account the effect of medications on BA levels[47]. The list of medications selected for adjustment for each BA is shown in Supplementary Table 7. We performed all analyses using both medication adjusted and unadjusted BA levels, results derived from medication-adjusted data and the adjustment process are described in Supplementary Methods and its accompanying tables.

In both RS and ROS/MAP, missing metabolite levels were imputed using half of the limit of detection. Log-transformed values were used in subsequent analyses.

Clinical Outcomes

For ADNI data, continuous response variables included the modified Alzheimer Disease Assessment Scale 13-item cognitive subscale (ADAS-Cog13; range, 0 [best] to 85 [worst] points), an index of general cognitive functioning. Categorical response variables included clinical diagnosis at baseline and MCI conversion (MCI-nonconverter, MCI-converter). For the ROS/MAP cohort, cognition was measured using a battery of tests (details are published[48-51]). A composite measure of global cognition was created by averaging the z-scores of all tests as previously described[51]. Mean and standard deviation at baseline were used to compute z-scores. A negative z-score means that an individual has an overall score that is lower than the average of the entire sample at baseline. Cognitive tests were used from the same cycle as serum, and proximate to death for brain.

Genotype and whole-genome sequencing data

Whole-genome sequencing: For 817 ADNI participants, whole-genome sequencing (WGS) was performed on blood-derived genomic DNA. Samples were sequenced on the Illumina HiSeq2000 using paired-end read chemistry and read-length of 100 bp at 30–40X coverage. For data processing and QC, an established analysis pipeline based on GATK was used. The QC steps included participant sex check, participant identity check, and variant quality check of the Illumina-generated VCF files (see Saykin *et al.,* for details[52]).

DNA genotyping in the participants of the RS cohort was performed using 550K, 550K duo, or 610K Illumina arrays at the internal genotyping facility of the Erasmus Medical Center, Rotterdam. Study samples with excess autosomal heterozygosity, call rate < 97.5%, ethnic outliers, and duplicate or family relationships were excluded during quality control analysis. Genotype exclusion criteria further included call rate < 95%, Hardy-Weinberg equilibrium $p < 1.0x10^{-6}$ and Minor Allele Frequency (MAF) < 1%. Genetic variants were imputed to the Haplotype Reference Consortium (HRC) reference panel (version 1.0)[53] using the Michigan imputation server[54].

Reference genetic associations with BA profiles in healthy individuals were obtained from supplementary data of the atlas of genetic influences on blood metabolites[45]. To obtain genome-wide genetic associations with DCA, we considered all suggestive significant results with $P < 1.0 \times 10^{-5}$. Gene and complex trait annotations of the 13 resulting genetic loci were performed using the SNiPA tool v3.2[55] and the NHGRI-EBI Catalog of published genome-wide association studies (www.ebi.ac.uk/gwas; accessed 02/01/2018, version 1.0)[56]. Lookup of AD genetic associations for DCA candidate variants was performed using the IGAP repository[2].

Statistical analysis

Differences of demographic, clinical, and cognitive measurements among the clinical diagnostic groups were evaluated using two-sample *t*-test (for continuous variables) and Pearson Chi-squared test (for categorical variables). All analyses were performed in a metabolite-wise manner and Bonferroni-adjusted critical values were used to assess statistical significance. All models included age at baseline, sex, *APOE* ε4, and $\log_{10}$-transformed body mass index (BMI). For cognition, number of years of education was added as an additional covariate.

Separate binary logistic regression models were conducted to examine cross-sectional association of each metabolite with baseline diagnosis (six models per metabolite). We performed logistic regression models to compare BA levels between the MCI-nonconverter and MCI-converter groups. Cox proportional hazard models were used to evaluate the association of metabolite levels with progression from MCI (combined EMCI and LMCI subjects) to AD. The cross-sectional association of ADAS-Cog13 with BAs was assessed using linear regression models with square root of ADAS-Cog13 as the dependent variable.

In ROS/MAP, one sample per individual was used. Linear regression models with global cognition score as dependent variable and metabolites as independent variables were used to assess the association of serum BAs with cognition, while adjusting for sex, age, *APOE* ε4, and years of education. Similar analyses were conducted for brain BAs separately.

We restricted our genetic variant analysis to single nucleotide polymorphisms in genes involved in the immune response pathway that were significantly associated with AD genome-wide[2,57-59].

Selected genetic variant included rs616338-T(*ABI3*), rs143332484-T(*TREM2*), rs72824905-C(*PLCG2*), rs9331896-T(*CLU*), rs6656401-A(*CR1*), rs35349669-T(*INPP5D*), rs11771145-G(*EPHA1*), rs983392-A(*MS4A6A*), and rs190982 -A(*MEF2C*). Associations of AD risk variants in immune-related genes with selected metabolic traits in ADNI and RS were computed using sex, age, and BMI as covariates.

Results

Characteristics of ADNI participants are depicted in Tables 1 and 2. Baseline cognitive measurements were significantly different among diagnostic groups, as expected. AD patients were more often carriers of at least one *APOE* ε4 allele. In addition, ADAS-Cog13 scores were not significantly different between the MCI-converter and nonconverter groups. However, the proportion of *APOE* ε4 carriers was higher in the MCI-converter group.

Serum BA profiles are significantly altered in AD

The Bonferroni-corrected threshold for statistical significance was determined as $P < 4.76 \times 10^{-4}$ (0.05 divided by 15 metabolites times seven phenotypes including cognition). When we compared BA profile in AD to CN, we detected a significant decrease in levels of the primary BA, CA ($P = 1.56 \times 10^{-4}$). In contrast, a significant increase of bacterially produced secondary BA, DCA was noted ($P = 1.61 \times 10^{-4}$) along with several secondary conjugated BAs, GDCA, TDCA, and GLCA (Table 4). GDCA and GLCA were significantly associated with ADAS-Cog13 where higher levels indicated worse cognition. Comparing BA levels between AD and both MCI groups yielded similar results, while the comparison of BA levels between the CN and MCI groups did not reach statistical significance (Supplementary Table 4).

Ratios reflective of conversion of BAs by gut microbiome are significantly associated with AD and cognitive performance

To determine which enzymatic processes in BA metabolism may underlie the differences noted in AD, we investigated eight selected ratios reflective of enzymatic activities in the liver and the gut microbiome. These ratios included the following:

1. The CA:CDCA ratio was selected to test if a possible shift in BA synthesis from the primary to the alternative BA pathway occurs in the liver.
2. Ratios of secondary to primary BAs (DCA:CA, GLCA:CDCA, and TLCA:CDCA) to investigate differences in gut microbiome enzymatic activity leading to altered production of secondary BAs. Because LCA was excluded in QC steps, the GLCA:CDCA and TLCA:CDCA ratios were used as proxies for the LCA:CDCA ratio.

Table 4: Cross-sectional association of bile acids with clinical diagnosis and cognition in the ADNI study*

Bile Acid	CN vs. AD (*n*=673) OR (95% CI); *P*-value†	ADAS-Cog13 (*n*=1453) β (95% CI); *P*-value‡
CA	**0.85(0.78,0.92);1.56E-04**	-0.04(-0.07,-0.01);2.81E-03
CDCA	0.94(0.87,1.01);7.19E-02	-0.02(-0.04,0.00);1.07E-01
GCA	1.07(0.96,1.18);2.03E-01	0.01(-0.02,0.05);4.36E-01
GCDCA	1.15(1.02,1.29);2.07E-02	0.06(0.02,0.09);4.60E-03
TCA	1.03(0.94,1.12);5.32E-01	-0.01(-0.03,0.03);7.92E-01
TCDCA	1.04(0.94,1.15);4.29E-01	0.02(-0.02,0.05);3.39E-01
TMCA	1.09(1.00,1.18);4.46E-02	0.029(0.00,0.06);4.21E-02
DCA	**1.24(1.11,1.39);1.61E-04**	0.05(0.01,0.08);9.26E-03
UDCA	0.96(0.90,1.03);2.41E-01	-0.01(-0.03,0.01);2.44E-01
GDCA	**1.30(1.17,1.43);4.20E-07**	**0.07(0.04,0.10);1.05E-05**
TDCA	**1.19(1.08,1.30);3.26E-04**	0.05(0.02,0.08);2.39E-03
GLCA	**1.33(1.20,1.48);9.21E-08**	**0.07(0.04,0.11);1.97E-05**
TLCA	**1.19(1.07,1.31);9.53E-04**	**0.06(0.03,0.1);3.18E-04**
GUDCA	1.09(1.00,1.19);5.39E-02	0.03(-0.00,0.06);6.04E-02
TUDCA	1.08(0.96,1.20);1.86E-01	0.01(-0.02,0.05);4.85E-01

Abbreviations: AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; CN, cognitively normal; EMCI, early mild cognitive impairment; LMCI, late mild cognitive impairment; CA, cholic acid; CDCA, chenodeoxycholic acid; GCA, glycocholic acid; GCDCA, glycochenodeoxycholic acid; TCA, taurocholic acid; TCDCA, taurochenodeoxycholic acid; TMCA, tauromuricholic acid; DCA, deoxycholic acid; UDCA, ursodeoxycholic acid; GDCA, glycodeoxycholic acid; TDCA, taurodeoxycholic acid; GLCA, glycolithocholic acid; TLCA, taurolithocholic acid; GUDCA, glycoursodeoxycholic acid; TUDCA, tauroursodeoxycholic acid.

*Statistically significant associations that passed Bonferroni correction are bolded.

†Odds ratios and *p*-values were obtained from logistic regressions. Models were corrected for age, sex, body mass index, and *APOE* ε4 status; Bonferroni-adjusted critical value was set to 5.76×10^{-4} (0.05 divided by 15 metabolites times 7 phenotypes including cognition)

‡Outcome: Square root of ADASCog-13 (0 [best] to 85 [worst]); Models were corrected for age, sex, years of education, body mass index and *APOE* ε4 status; Bonferroni-adjusted critical value was set to 2.17E-03.

3. GDCA:DCA and TDCA:DCA ratios were used to test if the observed secondary BA dysregulation is related to enzymatic differences related to their taurine and glycine conjugation.

Here, we considered associations as significant at a Bonferroni-corrected $P < 3.11 \times 10^{-4}$ (0.05 divided by all 23 metabolic traits times seven phenotypes, which include cognition). The ratio of the primary BAs (CA:CDCA) showed no significant association with AD. Yet, for the ratio of DCA:CA (i.e., the conversion of unconjugated primary to unconjugated secondary BA), we

observed a highly significant association with AD diagnosis (P=1.53 x 10^{-8}). Ratios between primary and secondary conjugated BAs showed the same effect and direction, including GDCA:CA (P=8.53 x 10^{-10}), TDCA:CA (P=9.83 x 10^{-7}), and GLCA:CDCA (P=3.61 x 10^{-6}).

Ratios modeling the glycine and taurine conjugation step of DCA, i.e. GDCA:DCA, TDCA:DCA, were not significantly associated with diagnosis (Figure 3 and Table 5).

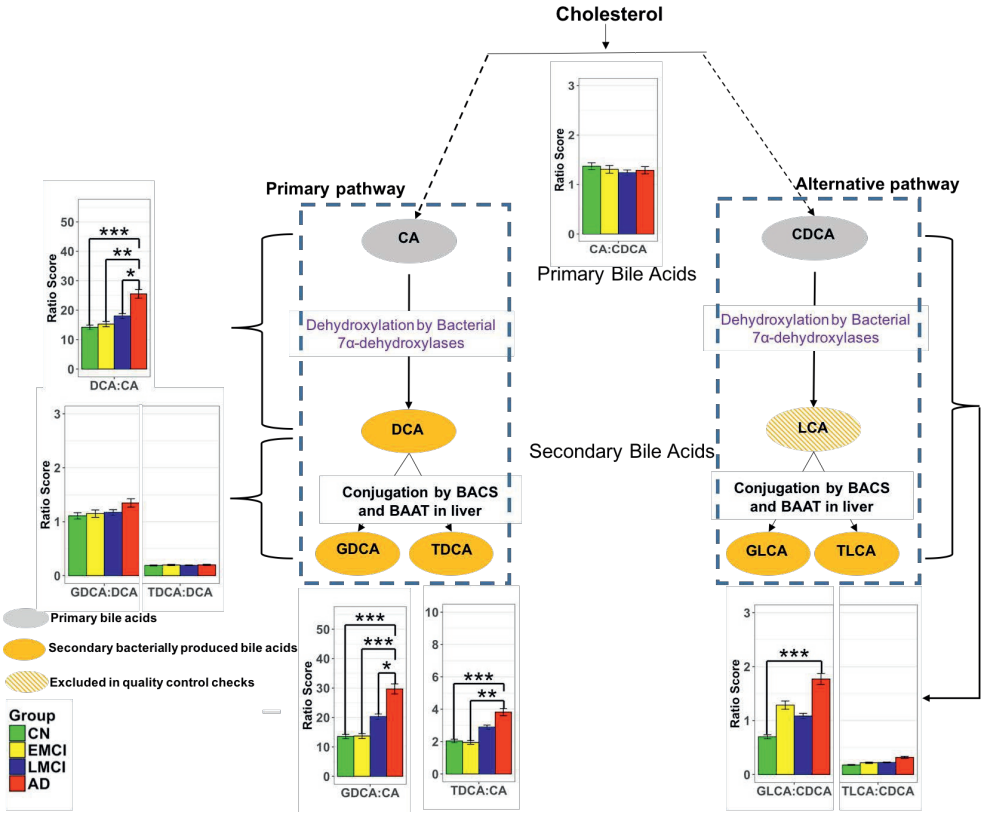


Figure 3: Ratios of bile acids reflective of liver and gut microbiome enzymatic activities in CN, Early MCI, Late MCI, and AD patients. Three types of ratios were calculated to inform about possible enzymatic activity changes in Alzheimer's patients. These ratios reflect one of the following: (1) Shift in bile acid metabolism from primary to alternative pathway. (2) Changes in gut microbiome correlated with production of secondary bile acids. (3) Changes in glycine and taurine conjugation of secondary bile acids. Color code: Green: cognitively normal; Yellow: EMCI; Blue: LMCI; Red: AD. Composition of selected ratios stratified by clinical diagnosis. Error bars indicate standard error of the means; Asterisks indicate statistical significance (*P<10^{-03}, ** P< 10^{-04}, and ***P< 10^{-05}). P-values were estimated from logistic regression models and adjusted for age, sex, body mass index, and *APOE* ε4 status. The significance level was adjusted for multiple testing according to the Bonferroni method to 0.05/138 = 3.62x10^{-4}; LCA was excluded in the quality control steps.

Table 5: Ratios of bile acids reflective of gut microbiome and liver enzymatic activities and their correlation with disease status and cognitive function*

Ratios informative about metabolic processes	Ratios calculated	CN vs. AD (*n*=673) OR (95% CI); *P*-value†	ADAS-Cog13 (*n*=1453) β (95% CI); *P*-value‡
Bile acid synthesis: primary vs. alternative pathway	CA:CDCA	0.87(0.77,0.97);1.67E-02	-0.03(-0.07,0.01);1.27E-01
Conversion from primary to secondary BA by the gut microbiome	DCA:CA	**1.25(1.16,1.35);1.53E-08**	**0.05(0.03,0.08);1.05E-05**
	GDCA:CA	**1.24(1.16,1.33);8.53E-10**	**0.06(0.04,0.08);1.20E-07**
	TDCA:CA	**1.16(1.10,1.24);9.83E-07**	**0.04(0.02,0.06);5.40E-05**
	GLCA:CDCA	**1.16(1.09,1.23);3.61E-06**	**0.04(0.02,0.06);9.15E-05**
	TLCA:CDCA	1.09(1.03,1.16);1.60E-03	0.03(0.01,0.05);1.50E-03
Glycine or Taurine conjugation of secondary bile acids by liver enzymes	GDCA:DCA	1.16(1.02,1.31);2.41E-02	0.05(0.02,0.10);5.49E-03
	TDCA:DCA	1.02(0.93,1.11);7.40E-01	0.01(-0.02,0.04);4.15E-01

Abbreviations: AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; CN, cognitively normal; EMCI, early mild cognitive impairment; LMCI, late mild cognitive impairment; CA, cholic acid; CDCA, chenodeoxycholic acid; DCA, deoxycholic acid; GDCA, glycodeoxycholic GCDCA; TDCA, taurodeoxycholic acid; GLCA, glycolithocholic acid; TLCA, taurolithocholic acid.

*Several ratios were calculated to inform about possible enzymatic activity changes in Alzheimer's patients. These ratios reflect: (1) Shift in bile acid metabolism from primary to alternative pathway. (2) Changes in gut microbiome correlated with production of secondary bile acids. (3) Changes in glycine and taurine conjugation of secondary bile acids.

†Outcome: Baseline diagnosis; Odds ratios and p-values were obtained from logistics regressions. Models were corrected for age, sex, body mass index, and APOE ε4 status; Bonferroni-adjusted critical value was set to 1.04E-03 based on 6 possible pairwise comparison of diagnosis groups (CN, EMCI, LMCI, and AD) for 8 ratios.

‡Outcome: Square root of ADASCog-13 (0 [best] to 85 [worst]); Models were corrected for age, sex, years of education, body mass index, and APOE ε4 status; Bonferroni-adjusted critical value was set to .11 x 10^{-4} (0.05 divided by all 23 metabolic traits times 7 phenotypes, which include cognitive function).

Four ratios (including DCA:CA and GLCA:CDCA) were significantly associated with ADAS-Cog13. For the ratios we observed the same pattern as AD diagnosis, with higher ratios of secondary to primary BAs being highly significantly associated with worse cognitive performance, while neither conjugation, nor a shift between primary and alternative BA pathways in the liver was significantly linked to cognition (Table 5).

Serum BA levels were associated with progression from MCI to AD in ADNI

The nine metabolites and ratios associated with diagnosis were further investigated to assess their relationship with progression from MCI to AD. Out of 789 MCI (EMCI and LMCI) patients with mean (SD) follow-up 3.94 (2.35), 32.2% progressed to AD dementia in 4 years (labeled as MCI-converter (*n*=251) vs. those who did not progress MCI-nonconverter [*n*=528]). BA profiles were compared between the two groups using logistic regression models with conversion status as the dependent variable and metabolite as an independent variable. Models were adjusted for age,

sex, BMI, baseline ADAS-Cog13 score, and *APOE* ε4. The Bonferroni-corrected threshold for statistical significance was determined as $P < 5.56 \times 10^{-3}$ (0.05 divided by nine metabolites and ratios). We noted a decrease in CA levels (P=9.12 x 10^{-4}) and an increase in ratios of GDCA:CA (P=1.63 x 10^{-3}) and TDCA:CA (P=1.72 x 10^{-3}) in MCI-converters (Figure 4 and Supplementary Table 5). Further survival analysis also revealed that levels of CA (hazard ratio [HR], 0.92;

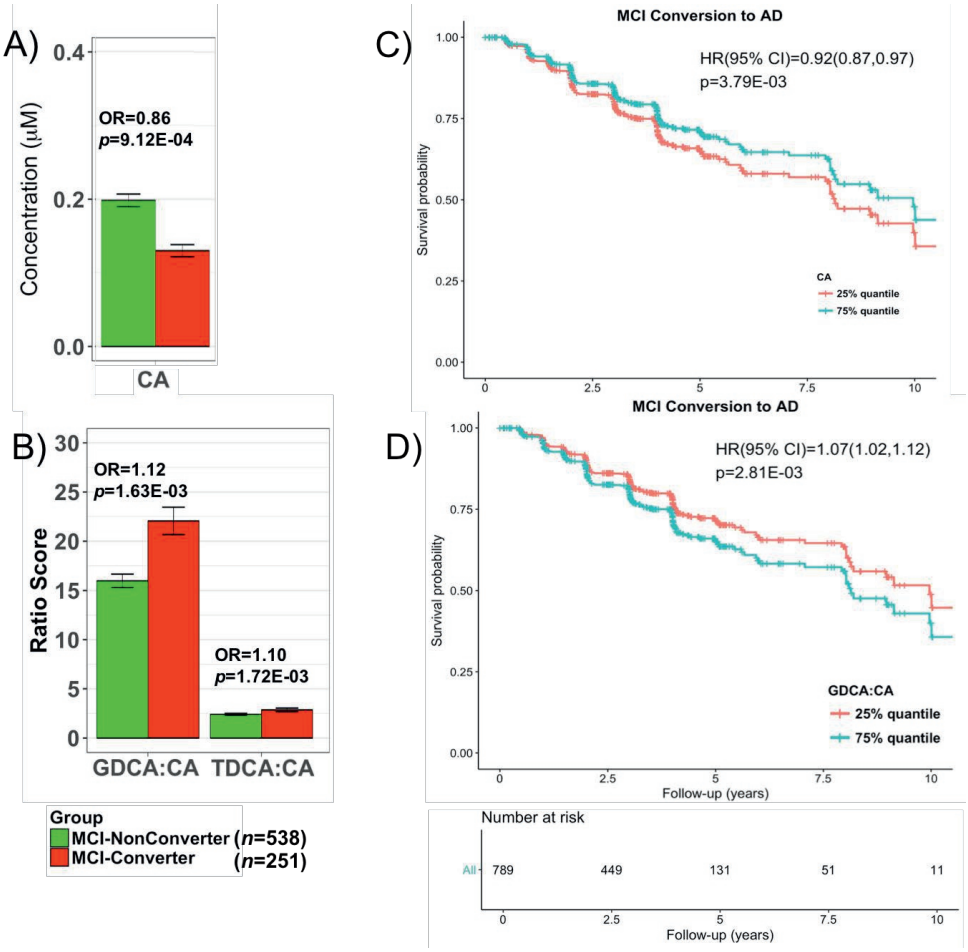


Figure 4: Comparison of bile levels in MCI subjects who convert and those who did not convert to AD dementia. (A and B) Lower levels of CA and higher levels of two secondary to primary ratios were significantly associated with higher odds of converting from MCI to AD. EMCI and LMCI patients that converted to AD dementia in 4 years after baseline were labeled as MCI-converter; 9 bile acids and ratios that were significantly dysregulated between CN to AD were assessed; *P*-values were estimated from logistic regression models and adjusted for age, sex, body mass index, and *APOE* ε4 status; the significance level was adjusted for multiple testing according to Bonferroni 0.05/9 = 5.56 x 10^{-3}. (C and D) Cox hazards model of the association of conversion from MCI to AD. Red line: 1st quantile, blue line: 3rd quantile. Analysis was conducted using quantitative values and stratification by quantiles was used only for graphical representation.

P =3.79 x 10^{-3}), GDCA:CA (HR, 1.07; P=2.81 x 10^{-3}), and TDCA:CA (HR, 1.06; P=3.19 x 10^{-3}) ratios predicted MCI progression (Figure 4).

Replication of association between cognition and DCA:CA ratio in serum and brain from ROS/MAP

To confirm the associations observed in ADNI, we used an independent cohort of older adults (ROS/MAP) with measures of BAs in serum and brain to replicate our findings. Because the sample sizes in ROS/MAP were smaller than ADNI and AD cases were strongly underrepresented (566 serum samples 11 of which were AD and 111 brains 27 of which were AD), we focused on replicating our key findings related to the association between cognition and the DCA:CA ratio (as proxy for BA processing by the gut microbiome). Here, we had to use global cognition scores where higher values indicate better cognition. Separate linear regression models were used for brain and serum samples. Pearson's correlation coefficient between serum DCA:CA and DCA:CA in 93 matching brain samples was 0.303 (P=0.003). In both serum and brain samples, higher levels of DCA:CA were associated with worse cognition (serum: β = -0.06; P = 0.011; brain: β = -0.21; P = 0.032), confirming our ADNI finding.

Genetic risk variants for AD in genes related to immune function are associated with bile acid levels

To further evaluate that altered BA profiles in AD are related to processes in the gut microbiome, we investigated if BA profiles were associated with immune-related AD risk genes which may contribute to differences in gut microbiome composition. Using the ADNI (n=817 with WGS data) and RS (n=488) cohorts, association of selected BAs in the primary BA pathway (CA, DCA, GDCA, and TDCA) as well as the DCA:CA ratio with the selected genetic risk variants in 9 candidate genes with immune-related functions were assessed. In addition, we included associations from a published large cohort-based study[45] to increase sample size. With the exception of rs983392 in *MS4A6A,* we found nominally significant associations for the candidate variants in all of these genes (Supplementary Table 10). Three associations were significant after Bonferroni-correction (P < 1.1 x 10^{-3}) in at least one of the studies: rs616338 (*ABI3*) and rs190982 (*MEF2C*) were significantly associated with the DCA:CA ratio, and rs11771145 (*EPHA1*) was significantly linked to both DCA and TDCA.

Genetic loci associated with DCA may influence susceptibility for AD

To follow up on the hypothesis that elevated DCA levels in AD that are linked to gut dysbiosis are relevant in the pathogenesis of AD, we collected (suggestive) significant genetic associations with DCA levels (P < 1.0 x 10^{-5}) from a previous study of genetic influences on blood metabolite levels in large population-based cohorts (n~7,800)[45]. We then annotated the resulting 13 loci with genetic trait associations, including AD associations from the IGAP study[2], and tried to replicate associations with DCA in ADNI (Supplementary Table 11). Two of the 13 genes, *CYP7A1* and *IMPA2*, also showed association with DCA levels in ADNI subjects. Notably, six of the 13 genes

have been previously linked via genetic studies to AD (*ABCA7*) or AD phenotypes, including cognitive decline and CSF protein levels (*LRRC7*, *CYCS*, *GPC6*, *FOXN3,* and *CNTNAP4*).

Discussion

In this study, we interrogated a possible role for BA end products of cholesterol metabolism and clearance in cognitive changes in AD. Using stored blood samples from ADNI studies we established that BA profile is significantly altered in AD patients. We noted a significant decrease in serum levels of a liver-derived primary BA (CA) and an increase in levels of a bacterially produced secondary BAs and their conjugated forms (DCA, GDCA and TDCA, GLCA) in AD patients compared to CN subjects (Table 4, Figure 1). Higher levels of secondary conjugated BAs (GDCA, GLCA, and TLCA) were significantly associated with worse cognitive function (ADAS-Cog13; Table 4). In a follow-up study, we illustrate that these changes are also correlated with changes in CSF markers of disease and brain imaging changes.

To inform about enzymatic activity changes in the liver and the gut, three types of metabolite ratios were evaluated to inform about mechanisms leading to the noted altered BA profile in AD. We found no shift in metabolism between primary and alternative pathways (Figure 3; no change in CA:CDCA); a significant change in production of secondary BAs via enzymatic activities in the gut microbiome (increased DCA:CA as well as GLCA:CDCA and TLCA:CDCA as proxies for LCA:CDCA) and no change in processes involved in glycine and taurine conjugation of secondary BAs in the liver (no change in GDCA:DCA or TDCA:DCA). The significant increase in ratios of secondary to primary BAs (e.g., DCA:CA; Figure 3), suggest altered activity of bacterial 7α-dehydroxylases leading to excess production of secondary BAs many of which are cytotoxic[34,60-62]. This indicates potential gut dysbiosis in AD patients possibly caused by enhanced colonization of the large and possibly the small intestine with anaerobic bacteria capable of CA and CDCA 7α-dehydroxylation. Increases in these ratios also significantly correlated with poorer cognition (Table 5). Together, these findings suggest that enzymatic steps in conversion of primary to secondary BAs in the gut might contribute to disease.

We also evaluated effects of BA levels on risk of progression to AD among 789 MCI patients. Lower levels of CA and higher ratio of secondary to primary BAs, GDCA:CA, and TDCA:CA were significantly associated with risk of developing AD dementia (Figure 4, Supplementary Table 6).

The increased production of bacterially produced DCA from CA modeled by ratio DCA:CA and its link to cognition was replicated in the independent ROS/MAP cohort. Association of the DCA:CA ratio with disease severity was evaluated separately in 566 serum and 111 brain samples. Because of the small number of AD patients (n=11, serum n = 27, brains), we used

global cognitive scores as an index of disease severity. Most of the BAs primary and bacterially produced secondary were found in the brain. Similar to ADNI findings, an increase in the DCA:CA ratio in both serum and brain were significantly associated with worse cognition. This finding suggests that downstream effects of the gut-directed dysregulation of primary versus secondary BAs are not limited to the periphery but also might affect metabolic homeostasis and/or signaling functions in the human brain.

Earlier smaller studies suggested differences in BA levels in AD[26,38-41]. For example, in a study of 495 plasma metabolites comparing MCI (*n*=58) and AD (*n* = 100) with those of cognitively normal controls (*n*=93), levels of DCA, LCA, and GLCA were significantly elevated in the disease state[41]. Mapstone and colleagues[39] identified increased levels of glycoursodeoxycholic acid (GUDCA) in subjects likely to develop amnestic MCI or AD within 2 to 3 years compared to control. In a small pilot study, Marksteiner and colleagues[38] reported increased levels of LCA, GDCA, and GLCA in AD (n=30) relative to MCI (n=20). We replicated these findings with the exception of LCA (Excluded during QC) and glyxoursodeoxycholic acid, which showed only a nonsignificant trend of upregulation in the AD group ($P = 0.054$). Marksteiner[38] did not report a significant increase in DCA or decrease in CA which we observed in the ADNI cohort. However, there is a trend in their data to suggest that DCA levels are increased in AD relative to CN. Our analyses build upon these pilot studies to include a large well-characterized cohort with rich clinical, neuroimaging, and genetics data. Our analyses include links to innate immunity-related genes which was not possible in smaller studies. In addition, we controlled for medication use which is known to significantly affect the gut microbiome and BAs. In our follow-up study, we explore the association of serum BAs with CSF and neuroimaging biomarkers of AD.

Composition and functional changes of the gut microbiome have been implicated in several diseases. Microbiome GWAS revealed that variants in many human genes involved in immunity and gut architecture are associated with an altered gut microbiome composition[63]. Although many factors such as diet can affect the microbial organisms residing in the gut, emerging data support the hypothesis that certain host genetic variants predispose an individual toward microbiome dysbiosis and this can be linked to disorders of metabolism and immunity such as type 2 diabetes mellitus, obesity, and autism[63].

Accumulating evidence links dysregulation of the immune system to AD pathology. In particular, genetic association studies in AD have robustly identified several genetic risk variants in immune-related genes[2,59]. Using the ADNI and RS cohorts, we investigated the association of BA profiles of CN subjects with genetic variants in nine AD-related and innate immunity genes. Eight genetic variants were associated with selected BA levels at nominal significance (Supplementary Table 10). Three of these associations were significant after Bonferroni-correction, with rs616338 (*ABI3*) and rs190982 (*MEF2C*) associated with the DCA:CA ratio, and rs11771145 (*EPHA1*) linked to both DCA and TDCA. The association of the BAs to AD genes suggests that these

immune-related genes may influence the risk of AD through BA metabolism or changes in the gut microbiome. Interestingly, both *ABI3* and *MEF2C* are thought to be involved in immune reactions to pro-inflammatory stimuli that are partially secreted by microbes[64,65]. The link to the DCA:CA ratio may thus mirror differences in gut microbiome composition due to altered immune response in AD, providing a mechanistic hypothesis for our findings. The function of *EPHA1* is not well understood but, it has been hypothesized that when activated, this receptor may affect the integrity of the blood-brain barrier (BBB)[66]. Its association with levels of DCA is intriguing as DCA is known to be cytotoxic and can disrupt the BBB and then enter the brain[28]. rs11771145 is associated with gene expression levels of *EPHA1*[55], and as DCA is not known to be produced by human metabolism, changed expression and activity of *EPHA1* may be related to DCA-mediated cytotoxic effects.

Using an established atlas of genetic influences on human blood metabolites[45], we further investigated a potential cytotoxic role of DCA. For almost half of the 13 identified loci, we found genetic evidence for involvement in AD-linked complex traits (Supplementary Table 11). In particular, *ABCA7* is an AD risk gene replicated in several genetic studies[67,68]. Five additional genes (*LRRC7*, *CYCS*, *GPC6*, *FOXN3,* and *CNTNAP4*) genetically influence AD phenotypes, including cognitive decline and CSF markers. While it remains speculative if and how these genes interact with DCA to contribute to AD risk, it is intriguing that we identified *ABCA7* by screening for associations with DCA levels. *ABCA7* is highly expressed in the brain, and functions in the efflux of lipids, including cholesterol, from cells. Because of the structural similarity of DCA and cholesterol, we hypothesize that *ABCA7* may be able to also transport this BA, reconciling metabolomics findings via a functional hypothesis to a risk gene for AD. The findings that BA levels are regulated by AD-related genes might provide new mechanistic insights.

There is growing support for strong connections between the intestinal environment, with its diverse microbial composition and activity, and the functions of the central nervous system. The "gut-brain metabolic axis" facilitates bidirectional chemical communication between the central and enteric nervous systems through mechanisms just starting to be defined[7-9]. Such a metabolic axis is thought to be involved in the regulation of multiple host metabolic pathways in which levels of hormones, neurotransmitters, amines, GABA, short-chain fatty acids (SCFA), lipid metabolites, and others are regulated by gut microbiome activity[12]. Changes in the composition of intestinal bacterial populations are associated with a wide array of neurological and neurodevelopmental disorders such as multiple sclerosis, autism, depression, schizophrenia, and Parkinson's disease[69-71]. In addition, increasing evidence suggests that liver disease may impact cognitive functions and contribute to AD[72].

Our findings suggest novel metabolic links in AD where BAs represent a component of the gut-liver-brain axis that relates to cognition. We hypothesize that interconnected immune and gut

microbiome dysregulation leads to increase in production of cytotoxic secondary BAs like DCA and its derivatives and these can modulate the BBB and build up in the

brain leading to impaired metabolic functions mediated by their receptors and targets. Such dysregulation includes cholesterol and glucose homeostasis.

It is of interest that BAs are ligands for nuclear receptors including FXR, LXRs among others and they act synergistically as metabolic sensors to regulate energy homeostasis [73,74] peripherally and might also propagate their effects to the brain. Interestingly, levels of four BAs produced by the gut microbiome and those we show to be significantly correlated with disease status and cognition (DCA, GLCA, TLCA, TDCA) are hydrophobic and cytotoxic[34,35,75,76]. Cell lines, animal models, and human studies suggest that levels of such BAs, particularly DCA, lead to a disruption of mitochondrial membranes resulting in increased reactive oxygen species, markers of inflammation, and apoptosis as well as decreases in cell viability and DNA synthesis[34,35,77]. Studies in rodents with deuterium-labeled DCA demonstrated that DCA crosses the BBB and increases its permeability[27,29]. Increased amounts of secondary BAs in blood may enter the brain through induced permeability of the BBB, affecting brain physiology and metabolism[28]. Several studies in human and animal brains also revealed that the full panel of BAs is found in the brain[24-27], but it is unclear whether this is due to transport from the periphery, from local synthesis, or both. The function of these BAs in the brain remains poorly defined with some support for them acting as neurosteroids[78].

BA levels and the gut microbiome influence each other, where bile salt hydrolase–rich bacteria readily modify the BA profile while, on the other hand, intestinal BAs control the growth and maintenance of commensal bacteria, maintain barrier integrity, and modulate the immune system[79-82]. Such changes might impact brain functions. Significant data support a role for cholesterol metabolism in the pathogenesis of AD including large genetic studies. Cholesterol homeostasis is regulated in part by the gut microbiome suggesting that cholesterol intermediates including those produced by gut might present as one gut-brain axis of communication that needs to be further investigated in human and animal studies.

Limitations

This is an observational study, the results of which may contain confounding biases. For example, diet, lifestyle, exposome, and other factors may contribute to changes in the gut. It remains unclear how these important factors are related to AD pathogenesis and whether the observed differences we note are causes or consequences of disease. Further studies of metabolic changes in normal aging are required to help define which aspects of BA metabolism might be related to disease versus normal aging. Fecal material was not collected in the ADNI cohorts or other large studies therefore precluding a direct analysis of microbiota changes across the trajectory of disease. Such studies have just been initiated. Use of medications was extensively evaluated as a

possible confound (Supplementary Methods and Tables 7–9), and our key findings remained after controlling for medication use but larger studies need to further evaluate the effect of these medications. Additional experimental studies are needed to more fully define the expression of BAs and their receptors in the brain and the mechanistic roles of BAs in the development of AD. The impact of BAs on FXR, TGR5, vitamin, and hormone receptors in the brain and the signaling pathways impacted are currently unclear. It is important to evaluate in other large community studies the generalizability of our findings. The genetic links need to be tested in large populations.

Longitudinal studies covering presymptomatic stages are needed to establish the influence of immune changes on gut microbiome composition and activity in AD patients and the impact of this on BAs and cholesterol homeostasis. Tracking earliest changes in BA and other gut-derived metabolites might provide insights into causality. Labeling studies are needed to evaluate if BAs cross the BBB and build up in brain with further elucidation of their signaling and regulatory functions centrally. However, we cannot exclude the possibility that changes in the brain during disease can also impact the gut and liver, and hence, some of our findings might be brain derived.

Conclusions

In summary, there is evidence of a relationship among the intestinal BA profile, gut microbial composition and/or activity, innate immunity, and genetic variants implicated in AD. When disrupted, BAs may contribute to cognitive changes, highlighting the importance of cholesterol clearance and its regulation in AD. Disorders in BA metabolism cause cholestatic liver diseases, dyslipidemia, fatty liver diseases, cardiovascular diseases, and diabetes, which are all associated with risk of cognitive decline, directly or indirectly. Our results lend support to this relationship in the context of AD and cohorts at risk for AD. Our evolving understanding of the gut microbiome's role in aging and in central nervous system diseases and their progression could open potential new hypotheses in the field, regardless of whether the role is ultimately found to be causative, consequence, or contributory. The role of the gut microbiome in AD needs to be further investigated along with the emerging links between central and peripheral metabolic failures that might contribute to brain health and disease during aging.

Acknowledgments

The Alzheimer's Disease Metabolomics Consortium (ADMC): A complete listing of ADMC investigators can be found at: https://sites.duke.edu/adnimetab/who-we-are/

The Alzheimer's Disease Neuroimaging Initiative (ADNI): Data used in the preparation of this article were obtained from the ADNI database (http://adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc .edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Funding for ADMC (Alzheimer's Disease Metabolomics Consortium, led by Rima Kaddurah-Daouk at Duke University) was provided by the National Institute on Aging grant R01AG046171, a component of the Accelerated Medicines Partnership for AD (AMP-AD) Target Discovery and Preclinical Validation Project (https://www.nia.nih.gov/research/dn/amp-ad-target-discovery-and-preclinical-validation-project) and the National Institute on Aging grant RF1 AG0151550, a component of the M²OVE-AD Consortium (Molecular Mechanisms of the Vascular Etiology of AD – Consortium https://www.nia.nih.gov/news/decoding-molecular-ties-between-vascular-disease-and-alzheimers).

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

The Religious Orders and the Rush Memory and Aging studies were supported by the National Institute on Aging grants P30AG10161, R01AG15819, R01AG17917, and U01AG46152.

The work of various Consortium Investigators is also supported by various NIA grants (U01AG024904-09S4, P50NS053488, R01AG19771, P30AG10133, P30AG10124, K01AG049050, R03AG054936), the National Library of Medicine (R01LM011360, R00LM011384, R01 LM0112535), and the National Institute of Biomedical Imaging and Bioengineering (R01EB022574). Additional support came from Helmholtz Zentrum, the Alzheimer's Association, the Indiana Clinical and Translational Science Institute, and the Indiana University-IU Health Strategic Neuroscience Research Initiative.

Role of the Funder/Sponsor: [Funders listed above] had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Contributions: The authors are grateful to Lisa Howerton for administrative support and the numerous ADNI study volunteers and their families.

Disclosures

J.B.T. reports investigator-initiated research support from Eli Lilly unrelated to the work reported here. S.S., S.K, and T.K. are employed by Biocrates Life Sciences AG. These authors have no other financial relationships relevant to this article to disclose. J.Q.T. may accrue revenue in the future on patents submitted by the University of Pennsylvania wherein he is a co-inventor and he received revenue from the sale of Avid to Eli Lily as a co-inventor on imaging-related patents submitted by the University of Pennsylvania. L.M.S. receives research funding from NIH (U01 AG024904; R01 MH 098260; R01 AG 046171; 1RF AG 051550) and MJFox Foundation for PD Research and is a consultant for Eli Lilly, Novartis, and Roche; he provides QC over-sight for the Roche Elecsys immunoassay as part of responsibilities for the ADNI3 study. A.J.S. reports investigator-initiated research support from Eli Lilly unrelated to the work reported here. He has received consulting fees and travel expenses from Eli Lilly and Siemens Healthcare and is a consultant to Arkley BioTek. He also receives support from Springer publishing as an editor in chief of Brain Imaging and Behavior. M.W.W. reports stock/stock options from Elan, Synarc, travel expenses from Novartis, Tohoku University, Fundacio Ace, Travel eDreams, MCI Group, NSAS, Danone Trading, ANT Congress, NeuroVigil, CHRU-Hopital Roger Salengro, Siemens, AstraZeneca, Geneva University Hospitals, Lilly, University of California, San Diego–ADNI, Paris University, Institut Catala de Neurociencies Aplicades, University of New Mexico School of Medicine, Ipsen, Clinical Trials on Alzheimer's Disease, Pfizer, AD PD meeting. P.M.D. has received research grants and advisory/speaking fees from several companies for other projects, and he owns stock in several companies. Full disclosures will be made through the IJCME

form. R.K.D. is inventor on key patents in the field of metabolomics including applications for Alzheimer disease. All other authors report no disclosures.

An email with links to the Authorship Form will be sent to authors for completion after manuscripts have been submitted

References

1. Alzheimer's, A. 2017 Alzheimer's disease facts and figures. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* **13**, 325-373 (2017).
2. Lambert, J.C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* **45**, 1452-8 (2013).
3. Jones, L. *et al.* Convergent genetic and expression data implicate immunity in Alzheimer's disease. *Alzheimer's & dementia: the journal of the Alzheimer's Association* **11**, 658-671 (2015).
4. Sampson, T.R. & Mazmanian, S.K. Control of brain development, function, and behavior by the microbiome. *Cell Host Microbe* **17**, 565-76 (2015).
5. Wu, S.-C., Cao, Z.-S., Chang, K.-M. & Juang, J.-L. Intestinal microbial dysbiosis aggravates the progression of Alzheimer's disease in Drosophila. *Nature Communications* **8**, 24 (2017).
6. Kumar, D.K. *et al.* Amyloid-beta peptide protects against microbial infection in mouse and worm models of Alzheimer's disease. *Sci Transl Med* **8**, 340ra72 (2016).
7. Ghaisas, S., Maher, J. & Kanthasamy, A. Gut microbiome in health and disease: linking the microbiome-gut-brain axis and environmental factors in the pathogenesis of systemic and neurodegenerative diseases. *Pharmacology & therapeutics* **158**, 52-62 (2016).
8. Yarandi, S.S., Peterson, D.A., Treisman, G.J., Moran, T.H. & Pasricha, P.J. Modulatory Effects of Gut Microbiota on the Central Nervous System: How Gut Could Play a Role in Neuropsychiatric Health and Diseases. *Journal of Neurogastroenterology and Motility* **22**, 201-212 (2016).
9. Tognini, P. Gut Microbiota: A Potential Regulator of Neurodevelopment. *Frontiers in Cellular Neuroscience* **11**, 25 (2017).
10. Tremlett, H., Bauer, K.C., Appel-Cresswell, S., Finlay, B.B. & Waubant, E. The gut microbiome in human neurological disease: A review. *Annals of Neurology* **81**, 369-382 (2017).
11. Holmes, E., Li, J.V., Marchesi, J.R. & Nicholson, J.K. Gut microbiota composition and activity in relation to host metabolic phenotype and disease risk. *Cell Metab* **16**, 559-64 (2012).
12. Nicholson, J.K. *et al.* Host-gut microbiota metabolic interactions. *Science* **336**, 1262-7 (2012).
13. Sampson, T.R. *et al.* Gut Microbiota Regulate Motor Deficits and Neuroinflammation in a Model of Parkinson's Disease. *Cell* **167**, 1469-1480 e12 (2016).
14. Minter, M.R. *et al.* Antibiotic-induced perturbations in microbial diversity during post-natal development alters amyloid pathology in an aged APPSWE/PS1DeltaE9 murine model of Alzheimer's disease. *Sci Rep* **7**, 10411 (2017).
15. Harach, T. *et al.* Reduction of Abeta amyloid pathology in APPPS1 transgenic mice in the absence of gut microbiota. *Sci Rep* **7**, 41802 (2017).
16. Cattaneo, A. *et al.* Association of brain amyloidosis with pro-inflammatory gut bacterial taxa and peripheral inflammation markers in cognitively impaired elderly. *Neurobiol Aging* **49**, 60-68 (2017).
17. Bester, J., Soma, P., Kell, D.B. & Pretorius, E. Viscoelastic and ultrastructural characteristics of whole blood and plasma in Alzheimer-type dementia, and the possible role of bacterial lipopolysaccharides (LPS). *Oncotarget* **6**, 35284-303 (2015).

18. Di Paolo, G. & Kim, T.W. Linking lipids to Alzheimer's disease: cholesterol and beyond. *Nat Rev Neurosci* **12**, 284-96 (2011).
19. Beecham, G.W. *et al.* Genome-Wide Association Meta-analysis of Neuropathologic Features of Alzheimer's Disease and Related Dementias. *PLOS Genetics* **10**, e1004606 (2014).
20. Ridlon, J.M., Kang, D.J. & Hylemon, P.B. Bile salt biotransformations by human intestinal bacteria. *J Lipid Res* **47**, 241-59 (2006).
21. Donova, M.V. [Transformation of steroids by actinobacteria: a review]. *Prikl Biokhim Mikrobiol* **43**, 5-18 (2007).
22. Islam, K.B.M.S. *et al.* Bile Acid Is a Host Factor That Regulates the Composition of the Cecal Microbiota in Rats. *Gastroenterology* **141**, 1773-1781 (2011).
23. Camilleri, M. & Gores, G.J. Therapeutic targeting of bile acids. *Am J Physiol Gastrointest Liver Physiol* **309**, G209-15 (2015).
24. Mano, N. *et al.* Presence of protein-bound unconjugated bile acids in the cytoplasmic fraction of rat brain. *J Lipid Res* **45**, 295-300 (2004).
25. Naqvi, S.H. *et al.* Detection of monohydroxy "bile" acids in the brains of guinea pigs afflicted with experimental allergic encephalomyelitis. *J Lipid Res* **10**, 115-20 (1969).
26. Pan, X. *et al.* Metabolomic Profiling of Bile Acids in Clinical and Experimental Samples of Alzheimer's Disease. *Metabolites* **7**(2017).
27. Higashi, T. *et al.* Unconjugated bile acids in rat brain: Analytical method based on LC/ESI-MS/MS with chemical derivatization and estimation of their origin by comparison to serum levels. *Steroids* **125**, 107-113 (2017).
28. Quinn, M. *et al.* Bile acids permeabilize the blood brain barrier after bile duct ligation in rats via Rac1-dependent mechanisms. *Digestive and Liver Disease* **46**, 527-534 (2014).
29. Lalic-Popovic, M. *et al.* Deoxycholic Acid as a Modifier of the Permeation of Gliclazide through the Blood Brain Barrier of a Rat. *J Diabetes Res* **2013**, 598603 (2013).
30. Cecilia, M.P.R. *et al.* Neuroprotection by a Bile Acid in an Acute Stroke Model in the Rat. *Journal of Cerebral Blood Flow & Metabolism* **22**, 463-471 (2002).
31. Dionisio, P.A. *et al.* Amyloid-beta pathology is attenuated by tauroursodeoxycholic acid treatment in APP/PS1 mice after disease onset. *Neurobiology of Aging* **36**, 228-240 (2015).
32. Nunes, A.F. *et al.* TUDCA, a bile acid, attenuates amyloid precursor protein processing and amyloid-beta deposition in APP/PS1 mice. *Mol Neurobiol* **45**, 440-54 (2012).
33. Paolini, M. *et al.* Ursodeoxycholic acid (UDCA) prevents DCA effects on male mouse liver via up-regulation of CYP [correction of CXP] and preservation of BSEP activities. *Hepatology* **36**, 305-14 (2002).
34. Ignacio Barrasa, J. *et al.* Deoxycholic and chenodeoxycholic bile acids induce apoptosis via oxidative stress in human colon adenocarcinoma cells. *Apoptosis* **16**, 1054-67 (2011).
35. Martinez-Diez, M.C., Serrano, M.A., Monte, M.J. & Marin, J.J. Comparison of the effects of bile acids on cell viability and DNA synthesis by rat hepatocytes in primary culture. *Biochim Biophys Acta* **1500**, 153-60 (2000).

36. Ramalho, R.M., Viana, R.J.S., Low, W.C., Steer, C.J. & Rodrigues, C.M.P. Bile acids and apoptosis modulation: an emerging role in experimental Alzheimer's disease. *Trends in Molecular Medicine* **14**, 54-62 (2008).
37. Schulz, S. *et al.* Progressive stages of mitochondrial destruction caused by cell toxic bile salts. *Biochimica et biophysica acta* **1828**, 2121-33 (2013).
38. Marksteiner, J., Blasko, I., Kemmler, G., Koal, T. & Humpel, C. Bile acid quantification of 20 plasma metabolites identifies lithocholic acid as a putative biomarker in Alzheimer's disease. *Metabolomics* **14**, 1 (2018).
39. Mapstone, M. *et al.* Plasma phospholipids identify antecedent memory impairment in older adults. *Nat Med* **20**, 415-8 (2014).
40. Greenberg, N., Grassano, A., Thambisetty, M., Lovestone, S. & Legido-Quigley, C. A proposed metabolic strategy for monitoring disease progression in Alzheimer's disease. *Electrophoresis* **30**, 1235-9 (2009).
41. Olazaran, J. *et al.* A blood-based, 7-metabolite signature for the early diagnosis of Alzheimer's disease. *J Alzheimers Dis* **45**, 1157-73 (2015).
42. Bennett, D.A., Schneider, J.A., Arvanitakis, Z. & Wilson, R.S. Overview and findings from the religious orders study. *Curr Alzheimer Res* **9**, 628-45 (2012).
43. Bennett, D.A. *et al.* Overview and findings from the rush Memory and Aging Project. *Curr Alzheimer Res* **9**, 646-63 (2012).
44. Adams, H.H. *et al.* Genetic risk of neurodegenerative diseases is associated with mild cognitive impairment and conversion to dementia. *Alzheimers Dement* **11**, 1277-85 (2015).
45. Shin, S.Y. *et al.* An atlas of genetic influences on human blood metabolites. *Nat Genet* **46**, 543-550 (2014).
46. Xie, G. *et al.* Alteration of bile acid metabolism in the rat induced by chronic ethanol consumption. *The FASEB Journal* **27**, 3583-3593 (2013).
47. St John-Williams, L. *et al.* Targeted metabolomics and medication classification data from participants in the ADNI1 cohort. *Sci Data* **4**, 170140 (2017).
48. Wilson, R. *et al.* Proneness to psychological distress is associated with risk of Alzheimer's disease. *Neurology* **61**, 1479-1485 (2003).
49. Wilson, R.S. *et al.* Participation in cognitively stimulating activities and risk of incident Alzheimer disease. *Jama* **287**, 742-748 (2002).
50. Wilson, R.S., Bienias, J.L., Evans, D.A. & Bennett, D.A. Religious Orders Study: overview and change in cognitive and motor speed. *Aging Neuropsychology and Cognition* **11**, 280-303 (2004).
51. Wilson, R.S. *et al.* Early and late life cognitive activity and cognitive systems in old age. *Journal of the International Neuropsychological Society* **11**, 400-407 (2005).
52. Saykin, A.J. *et al.* Genetic studies of quantitative MCI and AD phenotypes in ADNI: Progress, opportunities, and plans. *Alzheimer's & dementia: the journal of the Alzheimer's Association* **11**, 792-814 (2015).
53. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-83 (2016).

54. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-7 (2016).
55. Arnold, M., Raffler, J., Pfeufer, A., Suhre, K. & Kastenmuller, G. SNiPA: an interactive, genetic variant-centered annotation browser. *Bioinformatics* **31**, 1334-6 (2015).
56. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res* **45**, D896-D901 (2017).
57. Guerreiro, R., Bras, J. & Hardy, J. SnapShot: genetics of Alzheimer's disease. *Cell* **155**, 968-968 e1 (2013).
58. Villegas-Llerena, C., Phillips, A., Garcia-Reitboeck, P., Hardy, J. & Pocock, J.M. Microglial genes regulating neuroinflammation in the progression of Alzheimer's disease. *Curr Opin Neurobiol* **36**, 74-81 (2016).
59. Sims, R. *et al.* Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. *Nat Genet* **49**, 1373-1384 (2017).
60. Rolo, A.P., Palmeira, C.M. & Wallace, K.B. Mitochondrially mediated synergistic cell killing by bile acids. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1637**, 127-132 (2003).
61. Shivaram, K.N. *et al.* The effect of idebenone, a coenzyme Q analogue, on hydrophobic bile acid toxicity to isolated rat hepatocytes and hepatic mitochondria. *Free Radic Biol Med* **25**, 480-92 (1998).
62. Rodrigues, C.M., Fan, G., Ma, X., Kren, B.T. & Steer, C.J. A novel role for ursodeoxycholic acid in inhibiting apoptosis by modulating mitochondrial membrane perturbation. *J Clin Invest* **101**, 2790-9 (1998).
63. Hall, A.B., Tolonen, A.C. & Xavier, R.J. Human genetic variation and the gut microbiome in disease. *Nature Reviews Genetics* **18**, 690 (2017).
64. Han, J., Jiang, Y., Li, Z., Kravchenko, V.V. & Ulevitch, R.J. Activation of the transcription factor MEF2C by the MAP kinase p38 in inflammation. *Nature* **386**, 296-9 (1997).
65. Fairfax, B.P. *et al.* Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* **343**, 1246949 (2014).
66. Shoemark, D.K. & Allen, S.J. The microbiome and disease: reviewing the links between the oral microbiome, aging, and Alzheimer's disease. *J Alzheimers Dis* **43**, 725-38 (2015).
67. 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).
68. Le Guennec, K. *et al.* ABCA7 rare variants and Alzheimer disease risk. *Neurology* **86**, 2134-7 (2016).
69. Holmqvist, S. *et al.* Direct evidence of Parkinson pathology spread from the gastrointestinal tract to the brain in rats. *Acta Neuropathol* **128**, 805-20 (2014).
70. Jangi, S. *et al.* Alterations of the human gut microbiome in multiple sclerosis. *Nat Commun* **7**, 12015 (2016).
71. Dinan, T.G. & Cryan, J.F. Gut instincts: microbiota as a key regulator of brain development, ageing and neurodegeneration. *J Physiol* **595**, 489-503 (2017).

72. Kim, D.G. *et al.* Non-alcoholic fatty liver disease induces signs of Alzheimer's disease (AD) in wild-type mice and accelerates pathological signs of AD in an AD model. *J Neuroinflammation* **13**, 1 (2016).
73. Calkin, A.C. & Tontonoz, P. Transcriptional integration of metabolism by the nuclear sterol-activated receptors LXR and FXR. *Nat Rev Mol Cell Biol* **13**, 213-224 (2012).
74. Schaap, F.G., Trauner, M. & Jansen, P.L.M. Bile acid receptors as targets for drug development. *Nat Rev Gastroenterol Hepatol* **11**, 55-67 (2014).
75. Faubion, W.A. *et al.* Toxic bile salts induce rodent hepatocyte apoptosis via direct activation of Fas. *J Clin Invest* **103**, 137-45 (1999).
76. Perez, M.J. & Briz, O. Bile-acid-induced cell injury and protection. *World J Gastroenterol* **15**, 1677-89 (2009).
77. Sousa, T. *et al.* Deoxycholic acid modulates cell death signaling through changes in mitochondrial membrane properties. *Journal of Lipid Research* **56**, 2158-2171 (2015).
78. Keitel, V. *et al.* The bile acid receptor TGR5 (Gpbar-1) acts as a neurosteroid receptor in brain. *Glia* **58**, 1794-805 (2010).
79. Stenman, L.K., Holma, R., Eggert, A. & Korpela, R. A novel mechanism for gut barrier dysfunction by dietary fat: epithelial disruption by hydrophobic bile acids. *Am J Physiol Gastrointest Liver Physiol* **304**, 29 (2013).
80. Shapiro, H., Thaiss, C.A., Levy, M. & Elinav, E. The cross talk between microbiota and the immune system: metabolites take center stage. *Current Opinion in Immunology* **30**, 54-62 (2014).
81. Baptissart, M. *et al.* Bile acids: from digestion to cancers. *Biochimie* **95**, 504-17 (2013).
82. Allegretti, J.R. *et al.* Recurrent Clostridium difficile infection associates with distinct bile acid and microbiome profiles. *Aliment Pharmacol Ther* **43**, 1142-53 (2016).

Chapter 5.3

The Alzheimer's disease genetic risk variant in *ABCA7* is associated to the gut microbiome

Shahzad Ahmad*, Hata Karamujić-Čomić*, Djawad Radjabzadeh*, Bruno Bonnechere, Dina Vojinovic, Thomas Hankemier, M. Arfan Ikram, Andre G. Uitterlinden, Rima Kaddurah-Daouk, Robert Kraaij, Najaf Amin, Cornelia M. van Duijn

This chapter is in preparation

**First authors contributed equally to the work*

Supplementary information for this chapter is available at: https://drive.google.com/open?id=12nz0Otz2QCvHyqY8CDKI9iPAAUBlkvCx

Abstract

Studies of transgenic animals, based on the mutation in the amyloid precursor protein (*APP*)/ Presenilin-1 (*PS1*) genes, have suggested that Alzheimer's disease (AD) risk genes may influence the gut microbiome and subsequently determine the amyloid metabolism. Our study aims to examine whether also other AD implicated genetic variants influence the human gut microbiome in dementia-free individuals. The current study was performed in 1,221 subjects of the Rotterdam Study. The gut microbiome of the participants was profiled using Illumina MiSeq platform to sequence the V3-4 hypervariable region of the 16s rRNA gene. We evaluated the association of 28 AD genetic variants with 188 gut microbial taxa using linear regression analysis. Subsequently, we checked the association of microbial taxa associated with AD genes with cognition. We found that the AD-associated genetic variant in the *ABCA7* gene was associated with a lower abundance of phylum Firmicutes, class Clostridia, order Clostridiales, and with higher abundance of family Enterobacteriaceae and order Enterobacteriales. Moreover, a higher abundance of class Clostridia and order Clostridiales showed an association with higher cognition. We also found an association of lipid metabolism-related microbial pathways with the *ABCA7* gene. Data integration of earlier findings raises the hypothesis that high-density lipoprotein (HDL) subfractions may play a role in the cross-talk of the gut microbiome and the brain. Our study provides evidence for the role of gut microbiota in the AD-related molecular pathways in humans, which can be of great interest in future research.

Introduction

Accumulating evidence suggests that the gut microbiome is implicated in various psychiatric and neurodegenerative disorders, including Alzheimer's disease (AD)[1-3]. However, the question how the gut microbiome interacts with the brain is largely unanswered. It is likely that immune signals, lipid, and other metabolites in the blood circulation play a key role[4-7]. We have recently shown that the microbiome is associated to high-density lipoprotein (HDL) subfraction, amino acids and the set of glycoproteins (GP) implicated in the acute phase reaction that includes antichymotrypsin[8-10]. Gut microbiota may also influence the metabolism of bile acids, which act as signaling molecules and may modify the lipid metabolism[11]. Earlier, we have found bile acids, in particular cholic acid and deoxycholic acid (DCA), associated to AD[5]. Various animal model studies and human studies have shown the altered levels of gut microbiota in AD. At the phylum level, a decreased abundance of Firmicutes (families Clostridiaceae, Ruminococcaceae, Turicibacteraceae and genus *Clostridium sensu stricto*), Actinobacteria, and an increased abundance of Bacteroidetes were observed in AD patients[2,12]. Transgenic animal studies based on amyloid precursor protein (*APP*)/ Presenilin-1 (*PS-1*) mutations found also similar results related to phyla Firmicutes and Bacteroidetes[13]. Shifts in the gut microbial composition by introducing human *APP* and *PS-1* mutations in mice, which appear to change the amyloid-beta (Aβ) metabolism in the brain of the animal, can be reversed by introducing the gut microbiome of wild type mice into *APP* transgenic mice raised in a germ-free environment[13-15]. These pieces of evidence suggest an interaction of *APP* and *PS* -1 genes with the gut microbiome that alters cerebral Aβ metabolism. Up until now, these interactions have been studied only in transgenic animals, focusing only on these two genes.

AD is a highly heritable disorder[16], and recent genome-wide association studies have identified genes that are highly expressed in microglial cells in brain[17,18] and circulating macrophages in the blood, resulting in the working hypothesis that these cells play a vital role in the pathogenesis of AD[19]. The gut microbiome can directly or indirectly influence the function of microglia and macrophage[20] through the immune or metabolite signals in the blood[21]. The findings of clinical, pathological and epidemiological studies corroborate the recent experimental studies[22,23].

In this study, we have addressed the question of whether the human gut microbiome is associated with other, more common genetic variants implicated in AD in humans. Finally, we evaluated whether the identified gut microbial taxa are associated with cognitive function and circulating metabolites that have been associated with AD and cognition.

Study populations

Rotterdam Study

The Rotterdam Study is a prospective population-based study located in the Ommoord district of Rotterdam, the Netherlands[24]. At the baseline of the study in 1990, the cohort was initiated including 7,983 subjects ≥ 55 years of age (RS-I). From 2000 to 2001. The baseline cohort was expanded with the addition of 3,011 participants aged 55 years or older (RS-II). The cohort was further extended, with the inclusion of 3,932 participants age 45 years or older during 2006-2008 (RS-III). All study participants were extensively interviewed and physically examined at baseline and after every 3 to 4 years. The study has been approved by the Medical Ethical Committee of Erasmus Medical Center and by the Ministry of Health, Welfare and Sport of the Netherlands. Written Informed consent was obtained from each study participant to participate and to collect information from their treating physicians. In the current study, we focused participants from the second follow-up of the RS-III cohort for which gut microbiota and genetic information was available (N = 1,221).

Genotyping and imputations

Blood from the Rotterdam Study participants was collected during the baseline visit. DNA was extracted from blood and genotyping was performed with the 550K, 550K duo, or 610K Illumina arrays. In genotyping quality control, genetic variants exclusion criteria include: call rate < 95%, Hardy-Weinberg equilibrium $P < 1.0x10^{-6}$ and Minor Allele Frequency (MAF) < 1%. Exclusion criteria include excess autosomal heterozygosity, call rate < 97.5%, ethnic outliers, and duplicates or family relationships. Further, genetic variants were imputed with the Haplotype Reference Consortium (HRC) reference panel (version 1.0)[25], using the Michigan imputation server[26]. The server uses SHAPEIT2 (v2.r790)[27] to phase the genotype data and performs imputation with Minimac 3 software[28].

Gut microbiota profiling

In the Rotterdam Study, stool samples were collected at home using a Commode Specimen Collection System (Covidien, Mansfield, MA). Around 1 gram of stool was transferred to a 25x76mm feces collection tube and sent to the Erasmus MC via regular mail. Information about the time and date of defecation, current or recent use of antibiotics, current use of probiotics and recent travel activities were recorded using a short questionnaire. Samples were stored at -20^{0}C upon arrival at the center and we excluded samples that took more than 3 days to arrive at the Erasmus MC. For the remaining samples, time in the mail (in days) was used as a technical covariate[29].

To perform DNA isolation, each participant's frozen stool sample was thawed for 10 minutes at room temperature and a sample aliquot (~300 mg) was homogenized using 0.1mm silica beads (MP Biomedicals, LLC, Bio Connect Life Sciences BV, Huissen, The Netherlands). DNA isolation

from samples was performed using the Arrow stool DNA kit according to the manufacturer's defined protocol (Arrow Stool DNA; Isogen Life Science, de Meern, The Netherlands).

We used 309F-806R primer pair and dual indexing to amplify the V3 and V4 hypervariable regions of the 16s rRNA gene whose details are described elsewhere[30]. Amplicons were normalized using the SequalPrep Normalization Plate kit (Thermo Fischer Scientific, Waltham, MA, USA) and pooled. Amplicon pools were purified prior to sequencing (Agencourt AMPure XP, Beckman Coulter Life Science, Indianapolis, IN, USA) and size and quantity was assessed (LabChip GX, PerkinElmer Inc., Groningen, The Netherlands). A control library was added to ~10% of each amplicon pool as a positive control (PhiX Control v3 library, Illumina Inc., San Diego, CA, USA). The hypervariable V3 and V4 regions of the 16s rRNA gene were sequenced in paired-end mode (2x300bp) using the MiSeq platform (Illumina Inc.) with an average depth of 50,000 paired reads per sample[29].

In quality control, low-quality and chimeric reads were excluded. In the sample quality control, samples with less than 10,000 reads, and from participants who reported the use of antibiotics in the 6 months before sample collection were excluded. The reads passing quality control (~93%) were normalized using random 10,000 read subsampling (rarefication).

The ribosomal database project (RDP) classifier (2.12)[31] and SILVA 16S database (release 128)[32] were employed to infer taxonomic hierarchy (domain, phylum, class, order, family, and genus) of microbial communities (Binning posterior cutoff = 0.8)[33]. Absolute abundance was calculated for each taxonomic level prior to any additional QC (domain, phylum, class, order, family and genus). Further, to obtain information about the functional metagenome of gut bacteria, we used the PICRUSt (phylogenetic investigation of communities by reconstruction of unobserved states)(v.1.1.0) tool[34]. HUMAnN2 (the HMP Unified Metabolic Analysis Network 2) (v0.99)[35] was used to identify the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. Gut microbiota composition is available for a total of 1,427 Rotterdam Study participants, of which 1,221 individuals also have genotype information available.

Assessment of cognitive function

Participants of the Rotterdam Study also were tested with a battery of cognitive tests at the same follow-up round as the microbiome sampling. In the current study, we used a general cognitive measure based on principal component analysis of different cognitive tests and the first principal component was used as measure of general cognitive function (g-factor)[36]. Detailed methods for the calculation of general cognitive function are provided elsewhere[37].

Statistical analysis

Single Nucleotide Polymorphism selection

In our study, we selected 28 single nucleotide polymorphisms (SNPs) associated with AD at a genome-wide significance level, which include 24 common, and four relatively rare genetic variants (MAF<0.01) of *TREM2, PLCG2* and *ABI3* genes. We included common genetic variants based on the largest genome-wide association study (GWAS) of AD reported in 2019 by *Kunkle et al.*[17] and rare genetic variants from other studies[18,38]. All the selected genetic variants were well imputed ($R^2 > 0.5$) except a variant rs9271058 from *HLA-DRB1* gene ($R^2 < 0.3$) which was excluded from the analysis. A detailed list of SNPs used in the current study is provided in Supplementary Table 1.

Association of SNPs with gut microbiota

Taxa having a size less than 0.005% of the total datasets and present in less than 3% of the total sample size were excluded and we performed a natural logarithmic transformation of each microbial taxa abundance values prior to the association analysis. To test the association of each SNP with different microbial taxa, we performed linear regression analysis using microbial taxa abundance as the outcome and risk allele dosage of SNP as predictor. Association analysis was adjusted for age, sex, body mass index, smoking, alcohol intake and medications (lipid-lowering, proton pump inhibitor, metformin, antibiotics) as well as technical covariates including time in mail (TIM) and DNA isolation batch (Batch). Multiple testing correction was performed using false discovery rate (*FDR* < $5x10^{-2}$) by *Benjamini and Hochberg* method[39]. All the analyses were performed using R version 3.5.2 (2018-12-20) (https://www.r-project.org/).

Additional analysis

For microbial taxa associated with AD genetic variants, we examined their association with g-factor. We performed linear regression analysis using g-factor as outcome while adjusting for the same covariates as in our primary outcomes of interest. Additional lookups were performed for the microbial taxa identified in this study to assess their relationship with metabolites (cholesterol sub-fractions including small high-density lipoproteins (S-HDLs), medium high-density lipoproteins (M-HDLs), very high-density lipoproteins (VLDLs) and amino acids) in blood circulation that have been previously identified in other studies to be associated with cognitive function or AD[37,40]. Information about the association between metabolites and microbial taxa was extracted from a previous study performed in the Rotterdam Study and the Life Lines-DEEP cohorts[33]. With this, we aimed to discover which of the AD/cognition related metabolites are influenced by the taxa we found in our study.

Results

The results of the association analysis between the genetic variants implicated in AD and gut microbial taxa are provided in Supplementary Table 2. Table 1 shows the genetic variants which passed the multiple testing correction (*FDR* < $5x10^{-2}$). Our study showed that a genetic variant in *ABCA7* (rs3752246-G) is significantly associated with a lower abundance of phylum Firmicutes (β = -0.043, SE = 0.012, *P* = $3.05x10^{-4}$) and more specifically the class Clostridia (β = -0.045, SE = 0.013, *P* = $7.04x10^{-4}$) and order Clostridiales (β = -0.045, SE = 0.013, *P* = $7.04x10^{-4}$). Further, the genetic variant was associated with a higher abundance of order Enterobacteriales (β = 0.429, SE = 0.127, *P* = $7.38x10^{-4}$) and the family Enterobacteriaceae (β = 0.429, SE = 0.127, *P* = $7.38x10^{-4}$). At genus level, microbial taxa of order Clostridiales have showed lower abundance in *ABCA7-G* allele carriers including genus *LachnospiraceaeND3007 group id 11317 (P* = $2.29x10^{-3}$*), LachnospiraceaeFCS020group id 11314 (P* = $3.45x10^{-3}$*), LachnospiraceaeUCG004 id 11324 (P* = $3.96x10^{-3}$*), LachnospiraceaeNK4A136 group id 11319 (P* = $1.31x10^{-2}$*), Eubacteriumrectalegroup id 14374 (P* = $1.96x10^{-2}$*), Faecalibacterium id 2057 (P* = $4.86x10^{-2}$*), Marvinbryantia id 2005 (P* = $4.87x10^{-2}$*), Roseburia id 2012 (P* = $1.52x10^{-2}$*)* and *RuminococcaceaeUCG013 id 11370* (*P* = $1.73x10^{-2}$) (Supplementary Table 2). However, associations at this taxonomic level did not remain significant after multiple testing correction. At the genus level, a higher abundance of the genus *Escherichia/Shigella* id 3504 (*P* < $3.29x10^{-3}$) of order Enterobacteriales was seen in *ABCA7* gene risk allele carriers.

Table 1: Association of gut microbiota with Alzheimer's disease-associated genes

SNP (Gene)	Microbial taxa	β	SE	*P-value*	*FDR*
rs3752246 (ABCA7)	Phylum Firmicutes.id.1672	-0.043	0.012	$3.05x10^{-4}$	$2.78x10^{-2}$
	Class Clostridia.id.1859	-0.045	0.013	$7.04x10^{-4}$	$2.78x10^{-2}$
	Order Clostridiales.id.1863	-0.045	0.013	$7.04x10^{-4}$	$2.78x10^{-2}$
	Family Enterobacteriaceae.id.3469	0.429	0.127	$7.38x10^{-4}$	$2.78x10^{-2}$
	Order Enterobacteriales.id.3468	0.429	0.127	$7.38x10^{-4}$	$2.78x10^{-2}$

Abbreviations: SNP, single nucleotide polymorphism; β, regression coefficient; SE, standard error; FDR, false discovery rate

Note: Genetic variants associated with gut microbiota at *False discovery rate (FDR) < 0.05*

In the analysis of microbial functional pathways (Supplementary Table 3), rs3752246 in the *ABCA7* gene showed an association with various pathways including tryptophan, linoleic metabolism pathway, lipopolysaccharide biosynthesis, lipoic acid metabolism, ubiquinone and other terpenoid-quinone biosynthesis and caprolactam degradation pathways (*P* < 0.01).

We subsequently associated the abundance of these microbial taxa with cognitive function (see Table 2). We found that the class Clostridia (β = 0.236, SE = 0.083, *P* = $4.44x10^{-3}$) and order Clostridiales (β = 0.236, SE = 0.083, *P* = $4.46x10^{-3}$) showed a significant positive association with

Table 2: Association of microbiota with cognition

Microbial taxa	β	SE	*P-value*
Phylum Firmicutes.id.1672	0.179	0.093	$5.48x10^{-2}$
Class Clostridia.id.1859	0.236	0.083	$4.44x10^{-3}$
Order Clostridiales.id.1863	0.236	0.083	$4.46x10^{-3}$
Family Enterobacteriaceae.id.3469	-0.013	0.009	$1.26x10^{-1}$
Order Enterobacteriales.id.3468	-0.013	0.009	$1.26x10^{-1}$

Abbreviations: β, regression coefficient; SE, standard error

cognitive function (Bonferroni adjusted threshold for significance: 0.05/5=0.01). The phylum Firmicutes also showed a marginal positive association with cognition (β = 0.176, SE = 0.093, *P* = $5.48x10^{-2}$). The Enterobacteria did not show a significant association with cognitive function at the family or order level.

Next, we integrated the findings of the current study with those of previous studies assessing the relationships between AD/cognition and metabolome and a study assessing the microbiome and metabolome (Figures 1 and 2). Higher abundance of class and order of Clostridia (β = 0.059, SE = 0.015, *P* = $7.47x10^{-5}$) and phylum Firmicutes (β = 0.084, SE = 0.018, *P* = $3.46x10^{-6}$) were

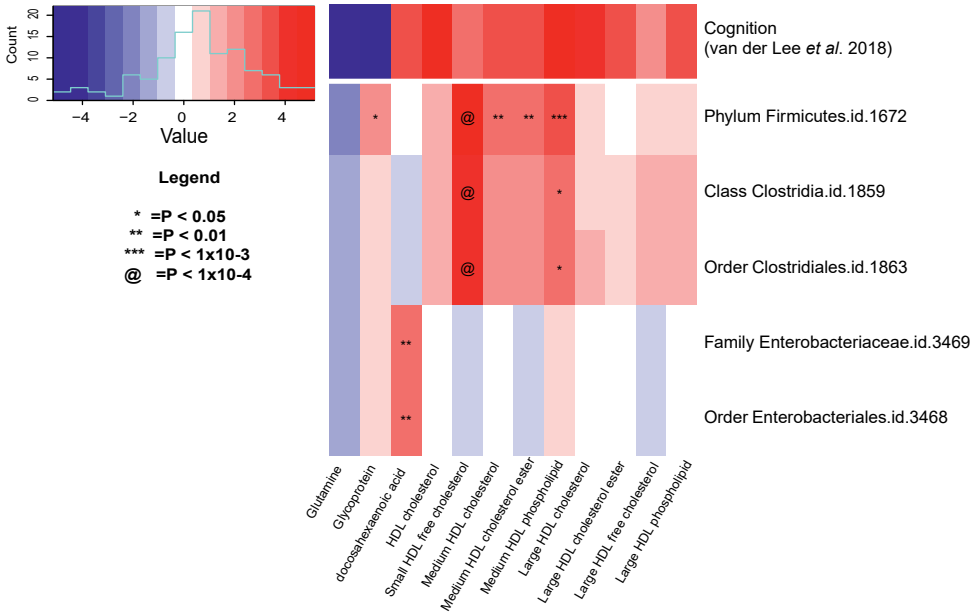


Figure 1: Integration of significant microbial taxa with levels of circulating metabolites associated with cognition. Abbreviations: HDL, high-density lipoprotein; VLDL, very large density lipoprotein; LDL, large density lipoprotein.

strongly associated with higher levels ($P < 10^{-4}$) of small high-density lipoprotein (HDL) free cholesterol (Figure 1). It is of note that higher levels of small and medium HDL are associated with higher cognition in our earlier study[37]. There is a cluster involving medium and large HDL-particles. A higher abundance of Phylum Firmicutes showed an association with higher levels of medium HDL subfractions including medium HDL cholesterol (β = 0.076, SE = 0.027, $P = 3.46\text{x}10^{-6}$), medium HDL cholesterol ester (β = 0.074, SE = 0.027, $P = 4.68\text{x}10^{-4}$) and medium HDL phospholipids (β = 0.101, SE = 0.027, $P = 6.60\text{x}10^{-4}$). The order and family Enterobacteria were associated to docosahexaenoic acid (DHA) (β = 0.015, SE = 0.005, $P = 4.02\text{x}10^{-3}$) while the phylum of Firmutes also showed a marginal significant association with glycoprotein levels (β = 0.041, SE = 0.019, $P = 3.31\text{x}10^{-2}$) as well as to valine (β = 0.065, SE = 0.031, $P = 3.33\text{x}10^{-2}$) and leucine (β = 0.059, SE = 0.024, $P = 1.67\text{x}10^{-2}$). Moreover, the order and class Clostridia showed evidence of association with medium HDL phospholipid (β = 0.056, SE = 0.022, $P = 1.26\text{x}10^{-2}$). A very similar pattern was seen for AD (Figure 2).

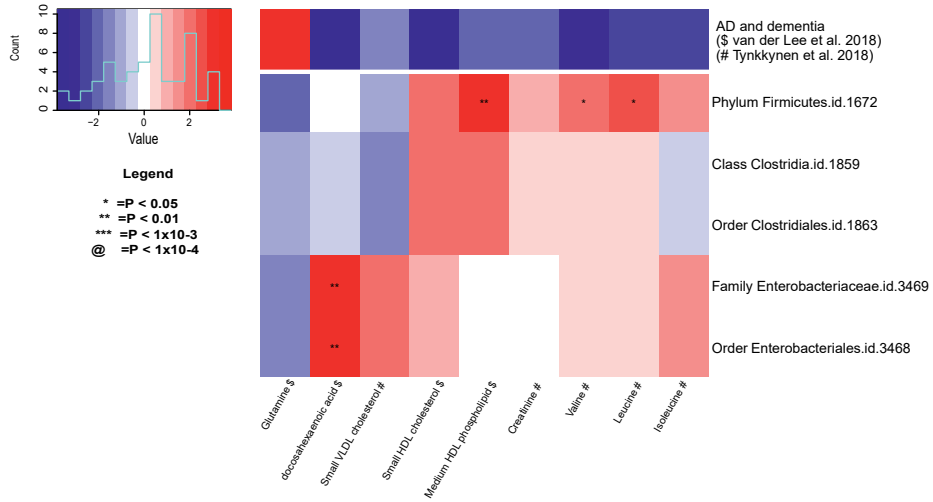


Figure 2: Integration of significant microbial taxa with levels of circulating metabolites associated with Alzheimer's disease and dementia. Abbreviations: AD (Alzheimer's disease), HDL, high-density lipoprotein; VLDL, very large density lipoprotein.

Discussion

In this study, we found that a genetic variant in or close by the *ABCA7* gene that increases the risk of AD showed a significant association with a lower abundance of phylum Firmicutes, its class Clostridia and order Clostridiales and a higher abundance of order Enterobacteriales and its family of Enterobacteriaceae. There is marginal significant evidence for association of the microbial taxa at the genus level of the order Clostridials, class Clostridia and order Enterobacteriales. The higher abundance of the class Clostridia and order Clostridiales also showed an association

with cognition. The pathway and metabolite-based analysis showed that the microbial taxa are associated with various lipid metabolism-related pathways. Integration of these findings with the metabolites that are associated with cognitive function and AD suggests a role of lipids, in particular, medium HDL and DHA metabolism in the cross-talk of the gut microbiome and the brain.

To our knowledge, there are no earlier studies on the role of AD-related genes in the gut microbiome. However, our findings corroborate with those of AD patients and in transgenic animals[41,42]. The *ABCA7* variant (which increases the risk of AD) in our study was associated with a lower abundance of phylum Firmicutes, which is in line with an earlier study that showed a decreased abundance of Firmicutes in AD patients compared to controls[2,12]. A decreased abundance of Firmicutes was also demonstrated in *APP/PS1* transgenic mice[14]. Firmicutes consists of both Gram-positive and Gram-negative bacteria and is considered a common bacterial phylum, which accounts for >51% of the whole human gut microbiota[22]. We found that a lower abundance of class Clostridia and order Clostridiales was observed in carriers of the *ABCA7* risk allele. In line with our findings, the class Clostridia, family Clostridiaceae and genus Clostridium were found to be less abundant in AD patients compared to controls in an earlier study[2]. In the present study, we show that Clostridia are associated with higher cognition. At the upper taxonomic level, both the class Clostridia and order Closteridiales belong to the phylum Firmicutes[43], which showed very similar relation to *ABCA7* and cognition in our study. Similar to phylum Firmicutes, class Clostridia and order Clostridiales, we observed decreased levels of the *Faecalibacterium, Eubacterium rectal*, *Lachnospiraceae* and *Roseburia* genera in *ABCA7* carriers which showed that our observed lower abundance in upper microbial taxonomic is driven by the cumulative effect of several genera belonging to this phylum.

We also observed an association of the *ABCA7* genetic risk with higher abundance of the order Enterobacteriales and family Enterobacteriaceae, which belong to phylum Proteobacteria (P = $2.30x10^{-3}$, FDR = $5.42x10^{-2}$). Although we did not find an association of these microbial taxa with cognitive function, an increased abundance of Proteobacteria is reported earlier in AD patients compared to controls[44]. A high abundance of Enterobacteriaceae has been found in other neurodegenerative diseases including Parkinson's disease[1], asking for further research. Moreover, a study investigating the impact of inflammation on gut microbiota has found that an increased abundance of Enterobacteriaceae is associated with inflammation of the human host[45]. Similarly, a higher abundance of the proinflammatory bacterial genus *Escherichia/Shigella* has been found correlated with the proinflammatory state in amyloid positive individuals[46], and we also found increased abundance of this genus in *ABCA7-G* allele carriers.

Interestingly, among all tested AD-associated genetic loci, only the *ABCA7* genetic variant showed association with various microbial taxa after multiple testing. The *ABCA7* gene is a member of the ATP-binding cassette (ABC) transporter and involved in cholesterol transport and highly

expressed in brain[41,42]. The ABCA7 protein is known to interact with APOA1[47], which is an important determinant of HDL subfractions[48] and these sub-fractions have been associated with AD and cognitive function. As we reported earlier, we also found that variants in the *ABCA7* gene are associated with altered blood levels of deoxycholic acid (DCA), a secondary bile acid. DCA is associated with gut microbiota[49] as well as to AD and cognition[5]. It is also known that secondary bile acids may absorb into bloodstream and can modulate lipid metabolism through G-proteins coupled receptors such as FXR[11]. In addition to cholesterol transport, the *ABCA7* gene is also implicated in amyloid-β clearance[50] and is expressed in microglia[51], which suggests that it may also be implicated in microglial proinflammatory response during old age[52]. Thus, we suggest that hyperactivated inflammatory response may drive the abundance of order Enterobacteriales and family Enterobacteriaceae.

The weakness of our study is that we do not have a single dataset on the human genome, gut microbiome, metabolome and AD. Instead, we integrated data using genetic variant analysis and gut microbiota profiling data from the participants of the Rotterdam Study, and connected findings of cognition and AD within various population-based studies. Thus, our study does not yield conclusive answers on the question whether the human *ABCA7* gene changes the gut microbiome, which leads to a change in the human metabolome and risk of AD. This hypothesis can be tested in experimental studies (e.g., transgenic mice) or in future epidemiological follow-up studies on the role of the microbiome in AD. At present, human studies to exclude pleiotropic effects are lacking, i.e., the genetic variants in *ABCA7* determine amyloid-beta clearance in the brain, but this process is independent of the effect of the *ABCA7* genetic variant on bile acid and HDL subfractions levels in blood. Such a pleiotropic mechanism may also explain the findings of earlier studies about the microbial taxa in AD

One of the major strengths of our study is the population-based design in a large epidemiological study. This allows us to account for multiple covariates, which can confound the association between genetic variants and gut-microbiota. This includes various medications, alcohol consumption, smoking, and obesity. We also have adjusted for these covariates when determining the association between the microbiota and the metabolites in our earlier study[33].

In conclusion, we found that a genetic variant in the *ABCA7* gene is associated with a lower abundance of phylum Firmicutes and a higher abundance of Proteobacteria, which are also associated with cognitive function and AD. The taxa are also associated with the HDL lipid fractions and DHA that have been associated with cognitive function and AD. The findings of this data integration study form the basis of a testable hypothesis for future animal or human studies.

Acknowledgment

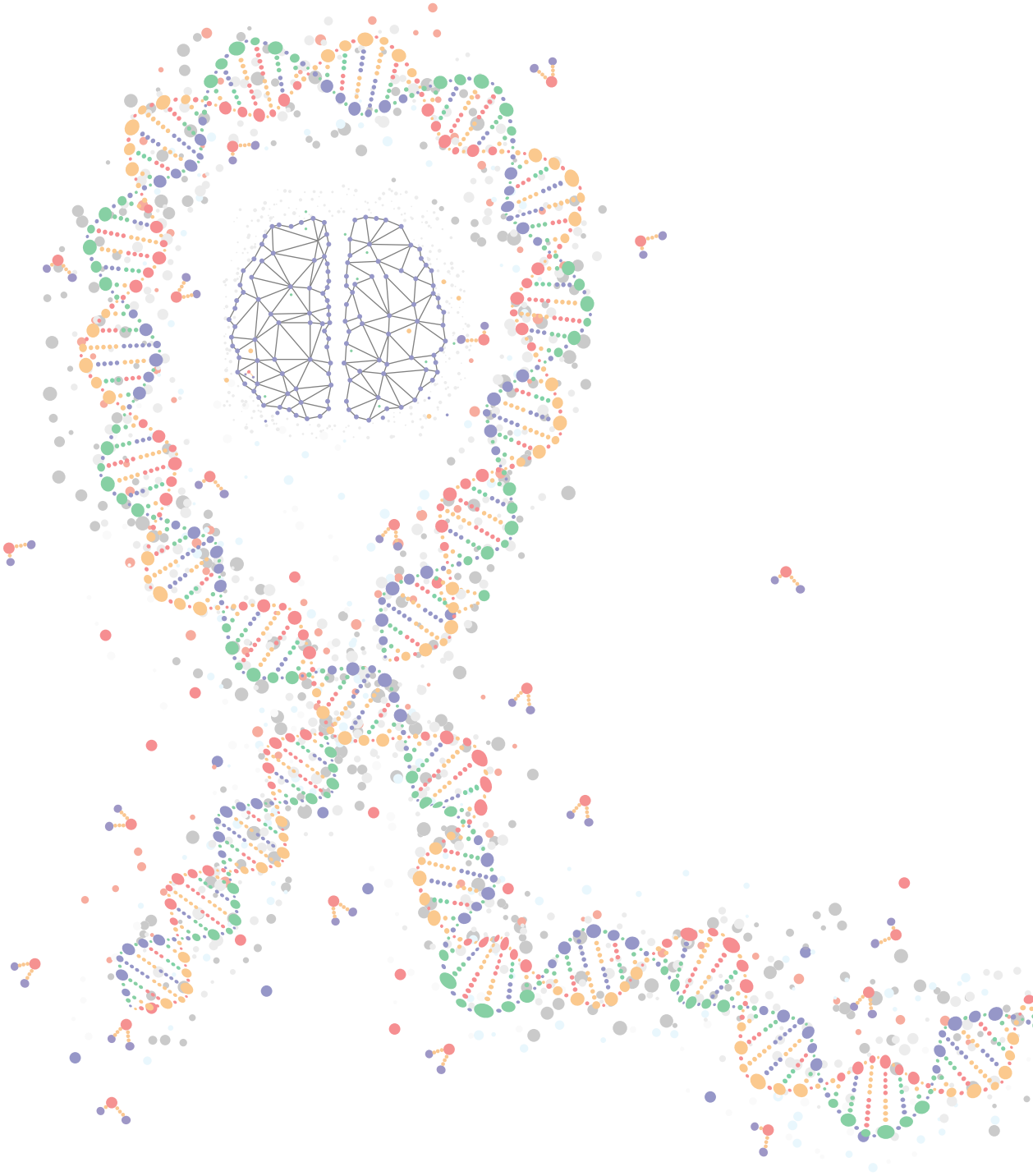
The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. This work has received funding from RO1 (Grant number Project No. 1RF1AG058942-01), Memorabel program (project number 733050814), AMP-AD project (U01-AG061359 NIA) and the Common mechanisms and pathways in Stroke and Alzheimer's disease (CoSTREAM) project (www.costream.eu, grant agreement No 667375). Djawad Radjabzadeh was funded by an Erasmus MC mRACE grant "Profiling of the human gut microbiome". The generation and management of stool microbiome data for the Rotterdam Study (RSIII-2) were executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. We thank Nahid El Faquir and Jolande Verkroost-van Heemst for their help in sample collection and registration, and Pelle van der Wal, Kamal Arabe, Hedayat Razawy and Karan Singh Asra for their help in DNA isolation and sequencing. Furthermore, we thank drs. Jeroen Raes and Jun Wang (KU Leuven, Belgium) for their guidance in 16S rRNA profiling and dataset generation. The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists.

References

1. Scheperjans, F. *et al.* Gut microbiota are related to Parkinson's disease and clinical phenotype. *Mov Disord* **30**, 350-8 (2015).
2. Vogt, N.M. *et al.* Gut microbiome alterations in Alzheimer's disease. *Scientific reports* **7**, 13537 (2017).
3. Valles-Colomer, M. *et al.* The neuroactive potential of the human gut microbiota in quality of life and depression. *Nature microbiology* **4**, 623 (2019).
4. Le Chatelier, E. *et al.* Richness of human gut microbiome correlates with metabolic markers. *Nature* **500**, 541 (2013).
5. MahmoudianDehkordi, S. *et al.* Altered bile acid profile associates with cognitive impairment in Alzheimer's disease-An emerging role for gut microbiome. *Alzheimers Dement* **15**, 76-92 (2019).
6. Sochocka, M. *et al.* The gut microbiome alterations and inflammation-driven pathogenesis of Alzheimer's disease—a critical review. *Molecular neurobiology* **56**, 1841-1851 (2019).
7. Foster, J.A. & McVey Neufeld, K.A. Gut-brain axis: how the microbiome influences anxiety and depression. *Trends Neurosci* **36**, 305-12 (2013).
8. Vojinovic, D. *et al.* Relationship between gut microbiota and circulating metabolites in population-based cohorts. *Nature Communications* **10**, 1-7 (2019).
9. Rothschild, D. *et al.* Environment dominates over host genetics in shaping human gut microbiota. *Nature* **555**, 210-215 (2018).
10. Zhernakova, A. *et al.* Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science* **352**, 565-569 (2016).
11. Fu, J. *et al.* The gut microbiome contributes to a substantial proportion of the variation in blood lipids. *Circulation research* **117**, 817-824 (2015).
12. Liu, P. *et al.* Altered microbiomes distinguish Alzheimer's disease from amnestic mild cognitive impairment and health in a Chinese cohort. *Brain, behavior, and immunity* (2019).
13. Harach, T. *et al.* Reduction of Abeta amyloid pathology in APPPS1 transgenic mice in the absence of gut microbiota. *Scientific reports* **7**, 41802 (2017).
14. Harach, T. *et al.* Reduction of Alzheimer's disease beta-amyloid pathology in the absence of gut microbiota. *arXiv preprint arXiv:1509.02273* (2015).
15. Bäuerl, C., Collado, M.C., Diaz Cuevas, A., Viña, J. & Pérez Martínez, G. Shifts in gut microbiota composition in an APP/PSS 1 transgenic mouse model of Alzheimer's disease during lifespan. *Letters in applied microbiology* **66**, 464-471 (2018).
16. Pedersen, N.L., Gatz, M., Berg, S. & Johansson, B. How heritable is Alzheimer's disease late in life? Findings from Swedish twins. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* **55**, 180-185 (2004).
17. Kunkle, B.W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet* **51**, 414-430 (2019).
18. Sims, R. *et al.* Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. *Nat Genet* **49**, 1373-1384 (2017).

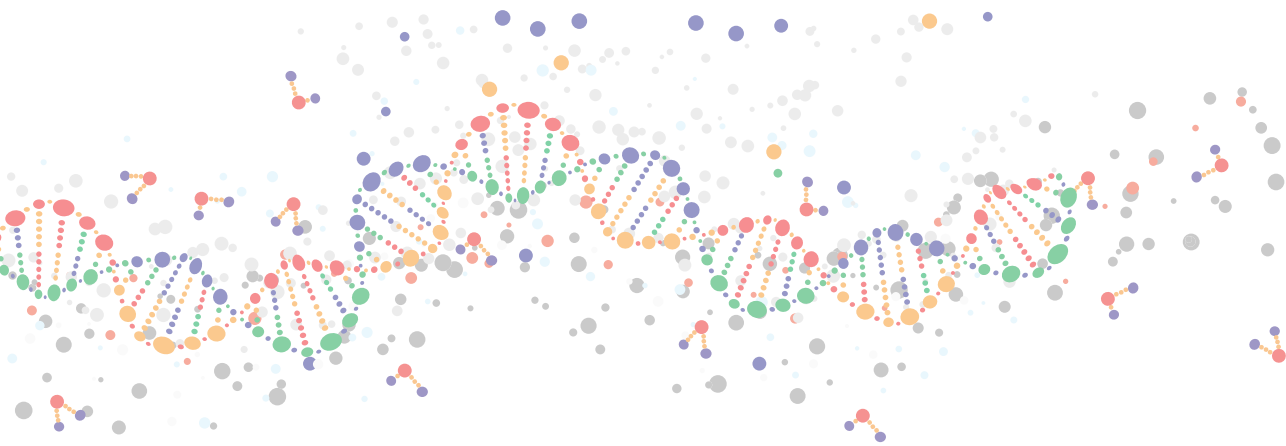
19. McQuade, A. & Blurton-Jones, M. Microglia in Alzheimer's Disease: Exploring How Genetics and Phenotype Influence Risk. *J Mol Biol* **431**, 1805-1817 (2019).
20. Katsumoto, A., Takeuchi, H., Takahashi, K. & Tanaka, F. Microglia in Alzheimer's Disease: Risk Factors and Inflammation. *Front Neurol* **9**, 978 (2018).
21. Abdel-Haq, R., Schlachetzki, J.C.M., Glass, C.K. & Mazmanian, S.K. Microbiome-microglia connections via the gut-brain axis. *J Exp Med* **216**, 41-59 (2019).
22. Mancuso, C. & Santangelo, R. Alzheimer's disease and gut microbiota modifications: the long way between preclinical studies and clinical evidence. *Pharmacological research* **129**, 329-336 (2018).
23. Duncan, S.H., Louis, P. & Flint, H.J. Cultivable bacterial diversity from the human colon. *Lett Appl Microbiol* **44**, 343-50 (2007).
24. Hofman, A. *et al.* The Rotterdam Study: 2016 objectives and design update. *Eur J Epidemiol* **30**, 661-708 (2015).
25. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-83 (2016).
26. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-7 (2016).
27. Delaneau, O., Marchini, J., Genomes Project, C. & Genomes Project, C. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat Commun* **5**, 3934 (2014).
28. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G.R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* **44**, 955-9 (2012).
29. Boer, C.G. *et al.* Intestinal microbiome composition and its relation to joint pain and inflammation. *Nature communications* **10**, 1-9 (2019).
30. Fadrosh, D.W. *et al.* An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome* **2**, 6 (2014).
31. Cole, J.R. *et al.* The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic acids research* **37**, D141-D145 (2008).
32. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research* **41**, D590-D596 (2012).
33. Vojinovic, D. *et al.* Relationship between gut microbiota and circulating metabolites in population-based cohorts. *bioRxiv*, 617431 (2019).
34. Langille, M.G.I. *et al.* Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature biotechnology* **31**, 814 (2013).
35. Abubucker, S. *et al.* Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS computational biology* **8**, e1002358 (2012).
36. Hoogendam, Y.Y., Hofman, A., van der Geest, J.N., van der Lugt, A. & Ikram, M.A. Patterns of cognitive function in aging: the Rotterdam Study. *Eur J Epidemiol* **29**, 133-40 (2014).
37. van der Lee, S.J. *et al.* Circulating metabolites and general cognitive ability and dementia: Evidence from 11 cohort studies. *Alzheimer's & Dementia* **14**, 707-722 (2018).

38. Jonsson, T. *et al.* Variant of TREM2 associated with the risk of Alzheimer's disease. *New England Journal of Medicine* **368**, 107-116 (2013).
39. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300 (1995).
40. Tynkkynen, J. *et al.* Association of branched-chain amino acids and other circulating metabolites with risk of incident dementia and Alzheimer's disease: A prospective study in eight cohorts. *Alzheimer's & Dementia* **14**, 723-733 (2018).
41. Hollingworth, P. *et al.* Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nature genetics* **43**, 429 (2011).
42. Le Guennec, K. *et al.* ABCA7 rare variants and Alzheimer disease risk. *Neurology* **86**, 2134-2137 (2016).
43. Rinninella, E. *et al.* What is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases. *Microorganisms* **7**(2019).
44. Li, Z., Zhu, H., Zhang, L. & Qin, C. The intestinal microbiome and Alzheimer's disease: A review. *Animal models and experimental medicine* **1**, 180-188 (2018).
45. Lupp, C. *et al.* Host-mediated inflammation disrupts the intestinal microbiota and promotes the overgrowth of Enterobacteriaceae. *Cell Host Microbe* **2**, 119-29 (2007).
46. Cattaneo, A. *et al.* Association of brain amyloidosis with pro-inflammatory gut bacterial taxa and peripheral inflammation markers in cognitively impaired elderly. *Neurobiology of aging* **49**, 60-68 (2017).
47. Carter, C. Alzheimer's disease: APP, gamma secretase, APOE, CLU, CR1, PICALM, ABCA7, BIN1, CD2AP, CD33, EPHA1, and MS4A2, and their relationships with herpes simplex, C. pneumoniae, other suspect pathogens, and the immune system. *International Journal of Alzheimer's Disease* **2011**(2011).
48. Cohen, J.C. *et al.* Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* **305**, 869-872 (2004).
49. Kakiyama, G. *et al.* Modulation of the fecal bile acid profile by gut microbiota in cirrhosis. *Journal of hepatology* **58**, 949-955 (2013).
50. Satoh, K., Abe-Dohmae, S., Yokoyama, S., St George-Hyslop, P. & Fraser, P.E. ATP-binding cassette transporter A7 (ABCA7) loss of function alters Alzheimer amyloid processing. *Journal of Biological Chemistry* **290**, 24152-24165 (2015).
51. Karch, C.M. & Goate, A.M. Alzheimer's disease risk genes and mechanisms of disease pathogenesis. *Biological psychiatry* **77**, 43-51 (2015).
52. Efthymiou, A.G. & Goate, A.M. Late onset Alzheimer's disease genetics implicates microglial pathways in disease risk. *Molecular neurodegeneration* **12**, 43 (2017).



Chapter 6

General Discussion



In this thesis, I aim to identify the genetic, metabolomic, and proteomic determinants of Alzheimer's disease (AD) risk and pathophysiology using integrated multi-omics and epidemiological approaches. Furthermore, I investigate the interaction of lifestyle factors with genetic risk factors in AD. The general discussion that follows will highlight the main findings of this thesis, their possible implications in understanding the AD pathophysiology, challenges, and opportunities for future research.

Genetics of Alzheimer's disease

In Chapter 2 of this thesis, I performed linkage analysis coupled with whole-genome sequencing to identify rare genetic variants. Genetic linkage analysis is a powerful approach to identify rare genetic variants, particularly when coupled with deep genomic sequencing[1]. The study described in this chapter was performed as part of the Alzheimer's disease Sequencing Project (ADSP) and is based on multi-generational families from the Genetic Research in Isolated Populations (GRIP), a highly inbred Dutch isolate[2]. Shared genetic and environmental factors in isolated family-based studies increase the power to detect rare variants[3]. I identified a genomic region, 5p14.3 (logarithm of odds (LOD) score = 3.3) linked to late-onset AD, and a haplotype region spanning over 9.1 cM shared by ten individuals contributing to the linkage signal at 5p14.3 region. With whole-genome sequencing of the affected family members, I identified a number of candidate single nucleotide polymorphisms (SNPs) in the *CDH18* gene, which may explain the identified genetic linkage region. *CDH18* encodes a type II cadherin protein, which is highly expressed in brain[4] and is involved in neuronal development and synaptic function[5].

In a separate study, A large GWAS of short-term verbal memory from The Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) has identified a genome-wide significant variant ($P = 5x10^{-8}$) in the *CDH1*8 gene (Jari *et al.* 2019 [unpublished]). It is of note that loss of short-term memory is considered one of the early clinical symptoms of AD[6] and is associated with an increased risk of AD[7,8], thus relevant to our finding. Further evidence from earlier studies have shown the association of various members of the cadherin gene family in neuropsychiatric disorders[9,10], CSF levels of Aβ-42[11], cognition, and education attainment[12,13,14,15]. Taken together, findings from our study and those of others highlight the importance of a genomic region, i.e., 5p14.3, in the central nervous system, but its precise role in AD is yet to be explored. Due to the highly variable frequencies of rare variants, population specificity, and unstable genomic region, I did not find a specific variant associated with AD in our identified genomic region based on the statistical analysis. Future functional studies, in cellular models or transgenic animals, aims at deciphering the role of various cadherin genes in AD pathology are needed to validate our findings, and to identify the pathways contributing to our observation.

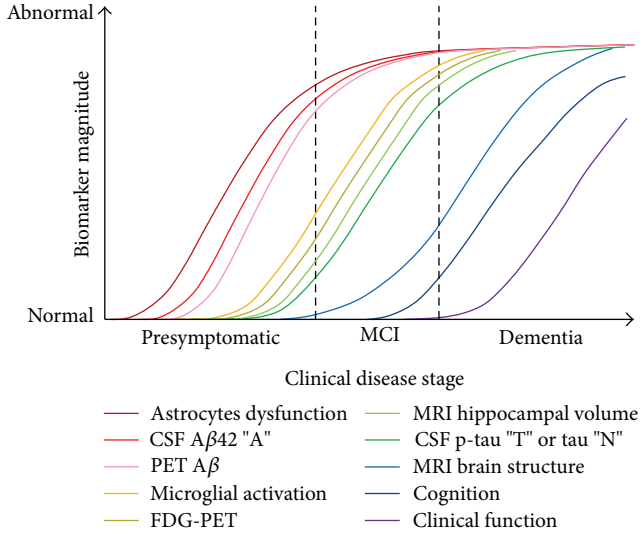


Figure 1: Biomarkers of AD trajectory over time, presenting three stages of AD, presymptomatic stage, MCI and eventually AD dementia. Reprinted from Leclerc *et al.*[27] Abbreviations: MCI, mild cognitive impairment; MRI, magnetic resonance imaging; CSF, cerebrospinal fluid.

Pathways implicated in Alzheimer's disease and lifestyle factors

A primary goal of genetic knowledge of a disease is to gain insight into novel pathways implicated in the etiology. Integrating genome-wide association study (GWAS) findings in AD with differential gene expression showed the role of eight biological pathways in AD[16] and, more recently also of Aβ processing[17]. In chapter 3.1, I showed that *endocytosis* pathway-based risk score associates to AD, mild cognitive impairment (MCI) and progression from normal (dementia-free) to AD. This observation is in line with earlier studies reporting the *endocytosis* pathway as the earliest manifestation of AD[18-20]. *The endocytosis* pathway is involved in neuronal uptake of cholesterol as well as excretion of secretory vesicles, thus crucial for synaptic function and axonal elongation[21]. Evidence from other studies also suggests a role of *endocytosis* dysfunction in Aβ aggregation[22], impaired synaptic function, and release of neurotransmitters[23]. Findings from this chapter indicate that the *endocytosis* pathway is shared between MCI and AD, thus may serve as a target for early AD risk assessment (Figure 1). I also showed that the increased genetic risk burden of risk variants in *clathrin/AP2 adaptor complex pathways* and *immune response* may play a role in early AD pathogenesis through white matter pathology. Among two genes clustered in *clathrin/AP2 adaptor complex* pathway (*PICALM, CLU*), *PICALM* is mainly expressed in the blood vessel walls and brain tissues[24] and is also expressed in the white matter, whereas *CLU* plays a role in the efflux of free insoluble amyloid-beta (Aβ) peptide through blood-brain barrier[25]. Moreover, increased plasma levels of *CLU* were found to be associated with decreased integrity of white matter at a young age[26]. Thus, the combined genetic risk of these two genes (*PICALM,*

CLU) may explain changes in white matter hyperintensities at MRI. Interestingly, a combined genetic risk score based on all genome-wide significant variants did not show association with any of the early features of AD pathophysiology, which suggest that aggregating multiple variants into a single score may dilute meaningful biological relevance of specific sets of genes belonging to a particular pathway. Findings from this study advances our understanding about the role of AD implicated pathways in early phase of AD and this also open evenues for research in pathway based genetic risk stratification of AD patients. Future studies are needed to explore the role of pathways in AD based on recent meta-analysis of AD GWAS[17] which first time showed the enrichment of genetic variants related to the Aβ processing pathway.

Chapter 3.2 of this thesis addressed the question of whether adherence to a healthy lifestyle can attenuate or even offset the genetic risk of dementia contributed by *APOE* in the Rotterdam study, a prospective population-based study. To study the healthy lifestyle pattern, a composite profile was constructed based on six lifestyle factors including abstaining from smoking, avoiding social isolation, absence of depression, absence of diabetes, regular physical activity and healthy diet including limited alcohol consumption. Findings from this chapter showed that modifiable risk factors lower the risk of dementia in individuals with low (*APOE* ε2ε3, ε2ε2 genotypes) and intermediate (*APOE* ε3ε3) genetic risk of dementia but not in the high genetic risk group (*APOE* ε2ε4, ε3ε4 or ε4ε4 genotype carriers). If replicated in an independent cohort, the clinical implication of our findings may be that adopting a healthy lifestyle could potentially prevent or postpone the onset of dementia in individuals at low and intermediate genetic risk. However, in contrast to our findings, a recent study based on the UK Biobank data showed that the association of lifestyle and genetic risk factors with AD risk is independent and additive in nature[28]. There may be several explanations of why our findings did not replicate. Although the UK Biobank-based study used a similar set of lifestyle factors, the study lacks deep dementia phenotyping in the cohort. Secondly, the use of polygenic risk score based on non-genome-wide significant loci by the UK Biobank study to account for the genetic component may explain the differences in both investigations. However, more extensive studies on the interaction of lifestyle factors with genetic factors in AD are needed to overcome the conflicting results.

Proteomics and Metabolomics of Alzheimer's disease

An increasing number of studies are focusing on identifying proteomic and metabolic determinants of AD during the preclinical and prodromal phase of the disease (Figure 1). In chapter 4.1, I studied the association of proteins profiled in blood with AD in the overall population and stratified by *APOE*. In this study, brain-related proteins were profiled in pre-symptomatic subjects from a prospective population-based study, the Rotterdam Study. In this chapter, I reported the association of CDH6 and HAGH protein levels in blood with AD in *APOE* ε4 carriers. The findings of this study were replicated in an independent Swedish BioFINDER cohort.

Furthermore, a genome-wide association study of protein levels in blood identified a genetic variant (rs111283466) that is also associated with the expression levels of *CDH6* transcripts. In an additional analysis, I showed a significant positive correlation of CDH6 levels in CSF with p-tau and total-tau in Amsterdam Dementia Cohort (ADC), which suggests that CDH6 levels in CSF may act as a biomarker of neurodegeneration. CDH6 is a cell surface glycoprotein that is highly expressed in blood and is involved in synaptic functions[29-31], synaptogenesis[32], TGF-B signaling[33], neural crest differentiation[34], presenilin-mediated signaling, and integrity of blood-brain barrier[35]. CDH6 protein-coding gene is part of a larger cadherin (CDH) gene cluster including *CDH8*, *CDH9, CDH10, CDH12, CDH13, CDH15, and CDH18.* It is of note that most of the cadherin genes including *CDH6, CDH9, CDH10, CDH12,* and *CDH18* are located in chromosome 5p14.3 region which I identified to be linked with AD in the genetic linkage analysis reported in Chapter 2. Positive associations of CSF levels of CDH6 protein with p-tau and total-tau levels in the ADC cohort suggest the role of CDH6 in the molecular process of neuronal and axonal cell injury, which begins years before the symptomatic phase of AD[36,37]. I also find elevated levels of HAGH protein in *APOE* ε4 carriers. The HAGH protein (glyoxalase-2) is involved in the glyoxalase system along with glyoxalase-1, a key player is oxidative stress control[38,39]. The glyoxalase system plays a role in detoxification of cytotoxic metabolite, methylglyoxal[40]. In plasma, higher levels of methylglyoxal are observed in hyperglycemia, which aids in the production of reactive oxygen species (ROS) and causes oxidative stress. Moreover, methylglyoxal is also a precursor of glycation end products (AGEs) and are implicated in neurodegeneration and AD[41,42]. Moreover, AGEs and glyoxalase system is involved in the regulation of amyloid precursor protein (*APP*) expression[43,44]. Findings from this chapter reiterate observations from earlier studies that suggest the role of oxidative stress during the prodromal stage of AD[45-47]. Additionally, findings from this study suggest that the oxidative stress-related protein marker (glyoxalase-2 or HAGH) can be observed in the blood of the presymptomatic phase of AD carrying *APOE* ε4 allele.

In chapter 4.2, I set out to investigate the role of signaling lipids in AD pathophysiology and MCI to AD progression. Among signaling lipids, I focused on lysophosphatidic acids (LPAs) which are bioactive phospholipids and known to regulate several key physiological processes including brain immune response[48], myelination[49], synaptic transmission[50] as well as having a role in endothelial cells and neurovascular function[51]. In this study, I observed the significant positive association of five LPAs (C16:0, C16:1, C22:4, C22:6, and isomer-LPA C22:5) to all three AD biomarkers, i.e., Aβ-42, p-tau and total tau levels in CSF. LPA C14:0 and C20:1 showed association to Aβ-42 and alkyl-LPA C18:1, LPA C20:1 to tau pathology biomarkers (p-tau and total tau) in CSF. The positive association between various LPA and Aβ-42 levels in CSF is concordant with an earlier finding, which suggested a role of LPAs in the upregulation of β-secretase (BACE1)[52], a key enzyme involved in Aβ production. Accumulation of Aβ is an early event in the pathogenesis of AD[53], which may lead to synaptic failure, tau pathology, and neuronal death (see Chapter 1)[54]. In the *APOE* stratified analysis, LPA C16:1 associated to Aβ-42 levels in both *APOE* ε4 and *APOE* ε*33* carriers but LPA C16:0 showed significant

association only in the *APOE* ε3 stratum, which implies that the *APOE* genotype may influence the association of LPAs with Aβ-42 levels. I also found that decreased CSF levels of LPA C16:0 and C16:1 were associated with MCI to AD progression. Inverse relation of LPA levels to MCI to AD conversion may be due to their correlation with CSF Aβ-42 levels, which is associated with a higher risk for MCI to AD progression[55]. In line with these findings, an earlier study also showed decreased levels of LPA C16:0 in frontal cortex tissue of Lewy body disease patients with AD compared to controls[56]. The fact that the association of LPAs to MCI to AD progression loses significance and the regression coefficient reducing to half after adjusting for Aβ-42 levels suggests the role of LPA 16:0 and 16:1 is dependent on Aβ-42. This study suggests a relationship between LPAs and Aβ metabolism; thus, future research should focus on investigating the nature of this relationship. Our study provides evidence for the role of an important class of signaling lipids in the pathophysiology of AD, but there is an open question whether these signalling lipids triggers the biological cascade of events leading to AD pathology or are products of neurodegeneration.

Gut-Liver-Brain axis

The study described in chapter 5.1 of this thesis presents the association of liver function biomarkers with AD, cognition, neuroimaging, and CSF biomarkers of AD in participants of the Alzheimer's Disease Neuroimaging Initiative (ADNI) study. This study has provided evidence for the association between lower levels of alanine aminotransferase (ALT) enzyme and elevated liver enzyme ratio (aspartate aminotransferase [AST] to ALT ratio) with AD and reduced cognitive performance, which is concordant with findings in an earlier study[57]. Moreover, an increased AST to ALT ratio and lower levels of ALT were associated with increased Aβ deposition, and reduced brain glucose metabolism and atrophy. The observation that the increased AST to ALT ratio associated with reduced brain glucose metabolism may explain the disturbed energy metabolism in AD and cognition[58,59] and brain hypometabolism during the prodromal phase of AD[60,61]. Findings from this chapter are concordant with a large number of studies that suggested the association of disturbed liver enzymes with altered levels of branched-chain amino acids, ether-linked phosphatidylcholines[62], which relate to AD[63-66] and its pathophysiology[67]. AST and ALT enzymes also play a crucial role in gluconeogenesis in liver and neurotransmitter (glutamate) production in brain[68]. In neuronal synapses, glutamate acts as a neurotransmitter and plays a role in long-term potentiation[69]. Earlier studies have shown a positive correlation between plasma levels of liver enzymes and glutamate in blood[58,70], which indicates that reduced levels of ALT enzyme may downregulate glutamate levels. It is of note that plasma levels of glutamate correlate positively with CSF glutamate levels[71], and lower levels of glutamate in plasma[72] and brain tissues are reported in AD patients compared to controls[73-76]. This study relates liver function biomarkers in blood to various features of AD pathophysiology (Figure 2). One of the limitations of our study is that we cannot account for medication use in ADNI, which can potentially confound our results. Our study implicates the role of liver enzymes in the connection between

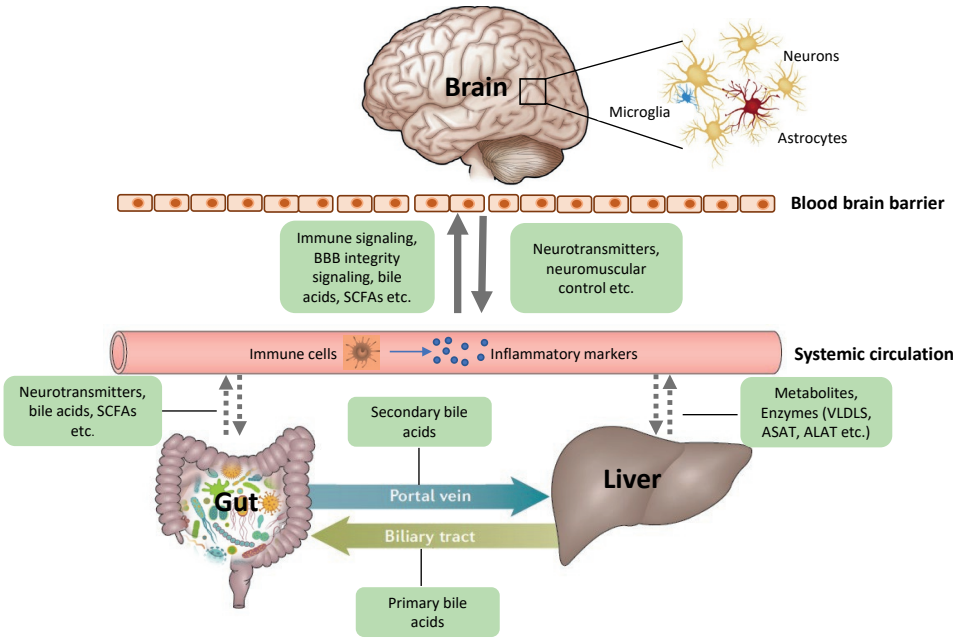


Figure 2: Gut-Liver-Brain axis. Abbreviations: BBB, blood-brain barrier; SCFA, Short-chain fatty acids; ASAT, Aspartate aminotransferase; ALAT, Alanine aminotransferase. Source: Adapted from Tripathi *et al*[96].

peripheral circulation and pathophysiological biomarkers of neurodegeneration. Our findings also raise questions about whether medications and lifestyle factors influence the link between peripheral circulation and brain. In future studies, other measures of liver function and non-alcoholic fatty liver disease should be investigated in addition to the effects of medication. Furthermore, Mendelian randomization may help to determine the causal relationships between the associations of liver function biomarkers and AD.

Chapter 5.2 describes the results from an association study of plasma levels of bile acids with AD diagnosis, cognition, and AD-associated genetic variants. Bile acids are products of cholesterol metabolism in the liver and further metabolized by gut microbiota[77,78]. Bile acids play a role in various signaling and regulatory functions and may be relevant in the gut-liver-brain axis[79] (Figure 2). In AD patients, we observed significantly lower serum levels of primary bile acid (cholic acid CA) and increased levels of the microbial product, secondary bile acid, deoxycholic acid (DCA), and its glycine and taurine conjugated forms. Moreover, an increased ratio of DCA:CA was associated with cognitive decline which was replicated in serum and brain samples in the Rush Religious Orders and Memory and Aging Project. Analyzing the MCI to AD progression analysis showed that lower levels of CA and a higher ratio of secondary to primary BAs, GDCA:CA, and TDCA:CA were significantly associated with the risk of developing AD. Findings from

this chapter suggest a unique metabolic link of the gut-liver-brain axis and AD. Bile acids can act as ligands for the nuclear transcription factor FXR, which is involved in energy homeostasis pathways[80,81]. It is of note that gut microbial derived bile acids (DCA, GLCA, TLCA, TDCA) which showed association with AD and cognition are hydrophobic and cytotoxic in nature[82-85]. DCA can disrupt mitochondrial membranes to produce reactive oxygen species and may thus influence cell viability via apoptotic and inflammatory pathways[82,83,86]. Moreover, studies have shown that DCA can cross the BBB, affecting brain physiology and metabolism[87-89]. As AD is a genetic disease, we studied the association of AD implicated immune-related genetic variants with bile acid levels. We found that *ABI3* and *MEF2C* showed association with the DCA:CA ratio, and *EPHA1* with DCA and TDCA levels. These findings are compatible with the hypothesis that immune-related genes may influence the risk of AD in part through the bile acid metabolism or changes in the gut microbiome. Additional lookups in the atlas of genetic influences on human blood metabolites[90], showed that six genes (*ABCA7*, *LRRC7*, *CYCS*, *GPC6*, *FOXN3,* and *CNTNAP4*) are related to cognition and CSF AD biomarkers as well as levels of DCA levels in serum. Of the identified genes, *ABCA7* is among highly replicated risk loci for AD[91,92], which is highly expressed in the brain and plays a role in cholesterol transport[93,94] and amyloid-β clearance[95]. Overall, this study contributed to the understanding of the complex interconnection between gut microbiota, liver function, and genetics of AD, where bile acids as the product of liver and gut microbiota may act as a biochemical signal in the gut-brain metabolic axis. Furthermore, altered levels of bile acids highlight the importance of cholesterol clearance and its regulation in AD. Bile acids are the product of cholesterol metabolism, which highlights the need to investigate the relationship between bile acids and cholesterol sub-fractions in circulation. Future studies should focus on investigating the relationship between bile acids and gut microbiota and on identifying the microbial taxa whose abundance may derive the bile acids related to AD pathology.

In chapter 5.3, I further explored the connection between gut microbiota and genetics of AD and their relation with metabolites implicated in AD and cognition. I found an association of the risk (G) allele of *ABCA7* genetic variant with a lower abundance of phylum Firmicutes, class Clostridia, order Clostridiales and, a higher abundance of order Enterobacteriales and family Enterobacteriaceae. The findings related to Firmicutes, Clostridia, and Clostridiales are in line with earlier studies, which showed the decreased abundance of Firmicutes in AD patients compared to controls[97] and in *APP* transgenic mice[98]. Additionally, I found evidence of an association between the higher abundance of class Clostridia and order Clostridiales with higher cognition levels. I also observed an association of the *ABCA7* genetic risk with higher abundance of order Enterobacteriales and family Enterobacteriaceae, which belong to phylum Proteobacteria (P = 2.30x10^{-3}, FDR = 5.42x10^{-2}). Although we did not find an association of these microbial taxa with cognitive function, an increased abundance of Proteobacteria is known earlier in AD patients[99]. An increased abundance of Enterobacteriaceae has also been found in Parkinson's disease[100]. In our study, integration of findings with earlier studies on the

association of cholesterol sub-fractions and gut microbiota[101], suggest that HDL sub-fractions may play a role in the cross-talk between the gut microbiome and the brain. Of the 28 tested AD genetic risk variants, only the variant in the *ABCA7* gene showed association with gut microbial taxa. The *ABCA7* gene is highly expressed in the brain and microglia[102] and plays a role in cholesterol transport[93,94] and amyloid-β clearance[95]. This gene is also implicated in microglial proinflammatory response in elderly individuals[103]. One possible mechanism of the *ABCA7* gene association with gut microbial taxa may be the hyper-activated inflammatory response, which may result in an increased abundance of order Enterobacteriales and family Enterobacteriaceae. *ABCA7* gene may also influence gut microbiota through bile acid metabolism (Chapter 5.2). The variants in the *ABCA7* gene may alter blood levels of deoxycholic acid (DCA), a secondary bile acid. DCA is associated with gut microbiota[77] as well as to AD and cognition[104]. It is also known that secondary bile acids may absorb into bloodstream and can modulate lipid metabolism through G-proteins coupled receptors such as FXR[105]. Functional studies are required to validate these findings. Transgenic animal studies, aiming to unravel the role of *ABCA7* genetic variant in modulating gut microbiota, can also help to understand the mechanism. The integration of identified microbial taxa with measures of AD neuropathology and other endophenotypes of AD is of interest in future studies.

Future research

In this section, I describe the future directions of the AD research in multi-omics, which may open new avenues for understanding the risk and pathophysiology of AD. One of the large whole-exome and whole-genome sequencing in AD gene discovery efforts, ADSP discovered few rare variants due to inadequate sample size (AD cases = 2,778 and control = 7,262), which warrants the need for exome sequencing studies in a larger sample such as in the UK Biobank[106]. Recent developments in the discovery of rare genetic variants of AD with large effect estimates[107-109] may improve our understanding about the AD genetic risk. Rare genetic variants may also help to decipher functional pathways in disease pathophysiology. Future genetic research should utilize long read sequencing techniques to overcome the problems encountered in sequence depth, which are otherwise difficult to characterize using traditional next-generation sequencing[110]. At present, this is feasible at small scale studies. Moreover, research is needed to assess the functional aspects of genetic variants on AD pathology. Discovery of more disease-relevant genetic variants among known genetic loci of the disease may guide drug target studies[111]. A recent study showed that most of the pleiotropic SNPs are located in intergenic regions, while the majority of the lead or causal genetic variants are located in the coding or flanking region of genes[112]. Future genomic studies should also focus on even other types of genomic fingerprints (short repeats, other sequencing methods)[113,114], which may explain unexplained linkage findings as in Chapter 2 but also may elucidate causal variants in GWAS.

One of the several challenges in AD research is to assess the functional impact of known genetic variants with small effect sizes on the molecular pathways implicated in AD pathophysiology[115,116]. Future research should expand beyond the highly researched amyloid hypothesis. Pathway-specific genetic risk scores, as developed in Chapter 3.1, open opportunities to evaluate the pathway-specific genetic risk score interaction with lifestyle factors such as education, alcohol intake, physical activity in longitudinal cohorts, which would help to identify the relevant pathways interacting with lifestyle factors in AD risk. It is of note that prior biological pathway information, which was used in our study, was based on the integration of gene expression data with available AD GWAS information published in 2013[16,117]. A new pathway enrichment analysis based on a larger GWAS has expanded pathways including amyloid precursor protein metabolism and tau binding protein pathways that were not included in the analysis of Chapter 3.1[17], which warrants the need for a new evaluation. GWAS based on Haplotype Reference Consortium (HRC) and TOPMed imputation panels such as European Alzheimer's Disease DNA BioBank (EADB) is uncovering more insight into biological pathways implicated in AD. Complex systems biology approaches to integrate multi-omics layers with genetic risk variants of AD would lead to more complex molecular pathways relevant to AD and may help to identify disease targets for drug research.

In this thesis, we identified the role of lysophosphatidic acids (LPAs) in AD pathophysiology. LPAs are bioactive signaling phospholipids whose receptors are highly expressed on astrocytes, and also on neurons[118]. LPAs are known to be involved in neuronal cell survival and outgrowth[119,120], but there is a need to explore their precise molecular mechanism in AD pathophysiology. One way to unravel their mechanism of action is through using the multi-omics approach in neuronal and astrocyte cell cultures and their co-cultures to identify the influence of LPAs on cell metabolism, expression of the gene, and various proteins involved in AD pathogenesis.

Metabolomics and proteomics of preclinical and predementia stages of AD have provided useful insight into the molecular mechanism of AD and yielded biomarkers of disease risk and progression (Aβ42/40, p-tau, and total-tau). However, since metabolomics and proteomics are more sensitive to deviation in analysis approaches, there is a need to harmonize and standardize the global efforts to address study power issues[121]. Variability in study design, selection, storage and processing of samples, instrumental analysis, data analysis and most importantly, the inclusion of potential confounders such as medication use can lead to inconsistent and conflicting results[122]. Future research should focus on standardizing workflow in large metabolomics investigations in order to improve the generalizability of findings and their ultimate use in clinical diagnosis[123]. Moreover, metabolomics studies in AD and dementia should be extended and focus on the interactions between the metabolome and proteomics with sex, lifestyle, and genetic risk factors. Variation in AD risk across genetic risk strata[124] may also be explained by variation in circulating metabolites. Thus, efforts should focus on integrated genetic and metabolomics information to identifying personalized signatures (unique metabolic fingerprint) of disease. There are growing

opportunities to incorporate advance machine learning and network analysis approach in large studies to address the challenges of data integration, which will improve the discovery of biomarkers of early disease assessment, allow patient stratification and lead to discovery of novel drug targets.

Conclusions

In the thesis, I study genetics, metabolomics, and proteomics to elucidate the pathophysiology of AD and the interaction of lifestyle factors with genetics, which determine the risk of AD. The research described in this thesis contributes to our understanding of the pathophysiology and molecular mechanism of AD in the prodromal and predementia phase of AD. In particular, the findings in the area of the gut-liver-brain axis may shed new light on potential new modifiable risk factors in AD.

References

1. Ott, J., Wang, J. & Leal, S.M. Genetic linkage analysis in the age of whole-genome sequencing. *Nat Rev Genet* **16**, 275-84 (2015).
2. Liu, F. *et al.* A genomewide screen for late-onset Alzheimer disease in a genetically isolated Dutch population. *The American Journal of Human Genetics* **81**, 17-31 (2007).
3. Zeggini, E. Using genetically isolated populations to understand the genomic basis of disease. *Genome Medicine* **6**, 83 (2014).
4. Thal, D.R., Rüb, U., Orantes, M. & Braak, H. Phases of Aβ-deposition in the human brain and its relevance for the development of AD. *Neurology* **58**, 1791-1800 (2002).
5. Seong, E., Yuan, L. & Arikkath, J. Cadherins and catenins in dendrite and synapse morphogenesis. *Cell adhesion & migration* **9**, 202-213 (2015).
6. Peters, F. *et al.* The neural correlates of verbal short-term memory in Alzheimer's disease: an fMRI study. *Brain* **132**, 1833-46 (2009).
7. Sullivan, E.V., Corkin, S. & Growdon, J.H. Verbal and nonverbal short-term memory in patients with Alzheimer's disease and in healthy elderly subjects. *Developmental Neuropsychology* **2**, 387-400 (1986).
8. Peters, F. *et al.* The neural correlates of verbal short-term memory in Alzheimer's disease: an fMRI study. *Brain* **132**, 1833-1846 (2009).
9. Wang, K. *et al.* Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* **459**, 528-33 (2009).
10. Singh, S.M., Castellani, C. & O'Reilly, R. Autism meets schizophrenia via cadherin pathway. *Schizophr Res* **116**, 293-4 (2010).
11. Deming, Y. *et al.* Genome-wide association study identifies four novel loci associated with Alzheimer's endophenotypes and disease modifiers. *Acta Neuropathol* **133**, 839-856 (2017).
12. Davies, G. *et al.* Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nat Commun* **9**, 2098 (2018).
13. Kichaev, G. *et al.* Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *Am J Hum Genet* **104**, 65-75 (2019).
14. Lee, J.J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* **50**, 1112-1121 (2018).
15. Savage, J.E. *et al.* Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat Genet* **50**, 912-919 (2018).
16. International Genomics of Alzheimer's Disease, C. Convergent genetic and expression data implicate immunity in Alzheimer's disease. *Alzheimers Dement* **11**, 658-71 (2015).
17. Kunkle, B.W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet* **51**, 414-430 (2019).
18. Funderburk, S.F., Marcellino, B.K. & Yue, Z. Cell "self-eating" (autophagy) mechanism in Alzheimer's disease. *Mt Sinai J Med* **77**, 59-68 (2010).

19. Ginsberg, S.D. *et al.* Upregulation of select rab GTPases in cholinergic basal forebrain neurons in mild cognitive impairment and Alzheimer's disease. *J Chem Neuroanat* **42**, 102-10 (2011).
20. Cataldo, A.M. *et al.* Endocytic pathway abnormalities precede amyloid beta deposition in sporadic Alzheimer's disease and Down syndrome: differential effects of APOE genotype and presenilin mutations. *Am J Pathol* **157**, 277-86 (2000).
21. Vance, J.E., Campenot, R.B. & Vance, D.E. The synthesis and transport of lipids for axonal growth and nerve regeneration. *Biochim Biophys Acta* **1486**, 84-96 (2000).
22. Maxfield, F.R. Role of endosomes and lysosomes in human disease. *Cold Spring Harb Perspect Biol* **6**, a016931 (2014).
23. Woodruff, G. *et al.* Defective Transcytosis of APP and Lipoproteins in Human iPSC-Derived Neurons with Familial Alzheimer's Disease Mutations. *Cell Reports* **17**, 759-773 (2016).
24. Baig, S. *et al.* Distribution and expression of picalm in Alzheimer disease. *J Neuropathol Exp Neurol* **69**, 1071-7 (2010).
25. Zlokovic, B.V. Neurovascular pathways to neurodegeneration in Alzheimer's disease and other disorders. *Nat Rev Neurosci* **12**, 723-38 (2011).
26. Braskie, M.N. *et al.* Common Alzheimer's disease risk variant within the CLU gene affects white matter microstructure in young adults. *J Neurosci* **31**, 6764-70 (2011).
27. Leclerc, B. & Abulrob, A. Perspectives in molecular imaging using staging biomarkers and immunotherapies in Alzheimer's disease. *ScientificWorldJournal* **2013**, 589308 (2013).
28. Lourida, I. *et al.* Association of lifestyle and genetic risk with incidence of dementia. *Jama* **322**, 430-437 (2019).
29. Fannon, A.M. & Colman, D.R. A model for central synaptic junctional complex formation based on the differential adhesive specificities of the cadherins. *Neuron* **17**, 423-34 (1996).
30. Obst-Pernberg, K. & Redies, C. Cadherins and synaptic specificity. *J Neurosci Res* **58**, 130-8 (1999).
31. Yamagata, M., Duan, X. & Sanes, J.R. Cadherins Interact With Synaptic Organizers to Promote Synaptic Differentiation. *Front Mol Neurosci* **11**, 142 (2018).
32. Arikkath, J. & Reichardt, L.F. Cadherins and catenins at synapses: roles in synaptogenesis and synaptic plasticity. *Trends Neurosci* **31**, 487-94 (2008).
33. Sancisi, V. *et al.* Cadherin 6 is a new RUNX2 target in TGF-beta signalling pathway. *PLoS One* **8**, e75489 (2013).
34. Clay, M.R. & Halloran, M.C. Cadherin 6 promotes neural crest cell detachment via F-actin regulation and influences active Rho distribution during epithelial-to-mesenchymal transition. *Development* **141**, 2506-15 (2014).
35. Montagne, A. *et al.* Blood-brain barrier breakdown in the aging human hippocampus. *Neuron* **85**, 296-302 (2015).
36. Sutphen, C.L. *et al.* Longitudinal decreases in multiple cerebrospinal fluid biomarkers of neuronal injury in symptomatic late onset Alzheimer's disease. *Alzheimers Dement* **14**, 869-879 (2018).
37. Hesse, C. *et al.* Transient increase in total tau but not phospho-tau in human cerebrospinal fluid after acute stroke. *Neuroscience letters* **297**, 187-190 (2001).

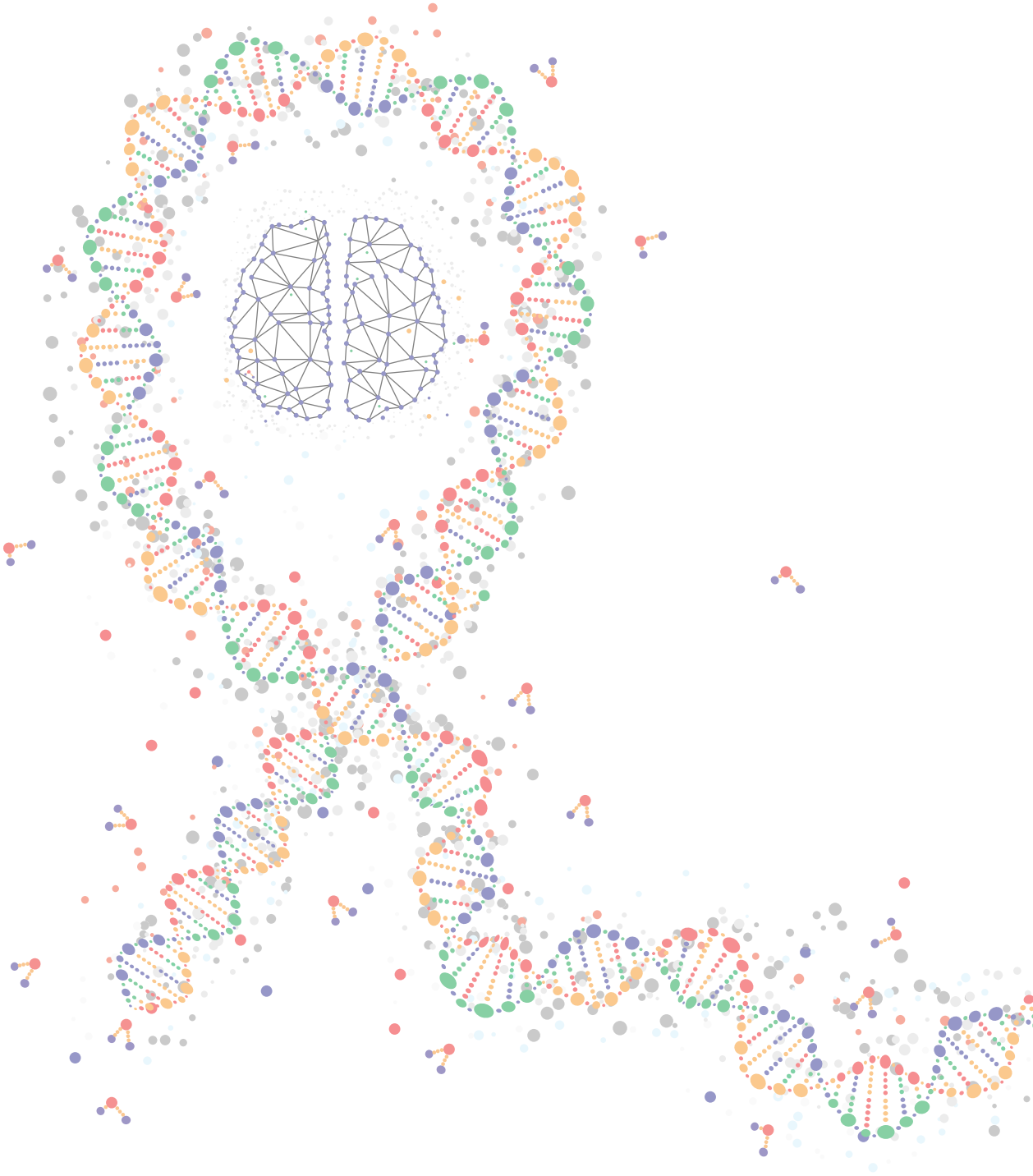
38. Chen, F. *et al.* Role for glyoxalase I in Alzheimer's disease. *Proc Natl Acad Sci U S A* **101**, 7687-92 (2004).
39. Thornalley, P.J. The glyoxalase system: new developments towards functional characterization of a metabolic pathway fundamental to biological life. *Biochem J* **269**, 1-11 (1990).
40. Xu, Y. & Chen, X. Glyoxalase II, a detoxifying enzyme of glycolysis byproduct methylglyoxal and a target of p63 and p73, is a pro-survival factor of the p53 family. *J Biol Chem* **281**, 26702-13 (2006).
41. Desai, K.M. *et al.* Oxidative stress and aging: is methylglyoxal the hidden enemy? *Can J Physiol Pharmacol* **88**, 273-84 (2010).
42. Takeuchi, M. *et al.* Involvement of advanced glycation end-products (AGEs) in Alzheimer's disease. *Curr Alzheimer Res* **1**, 39-46 (2004).
43. Ko, S.Y., Lin, Y.P., Lin, Y.S. & Chang, S.S. Advanced glycation end products enhance amyloid precursor protein expression by inducing reactive oxygen species. *Free Radic Biol Med* **49**, 474-80 (2010).
44. Wang, J.Z., Wang, Z.H. & Tian, Q. Tau hyperphosphorylation induces apoptotic escape and triggers neurodegeneration in Alzheimer's disease. *Neurosci Bull* **30**, 359-66 (2014).
45. Shea, T.B., Rogers, E., Ashline, D., Ortiz, D. & Sheu, M.S. Apolipoprotein E deficiency promotes increased oxidative stress and compensatory increases in antioxidants in brain tissue. *Free Radic Biol Med* **33**, 1115-20 (2002).
46. Baldeiras, I. *et al.* Peripheral oxidative damage in mild cognitive impairment and mild Alzheimer's disease. *J Alzheimers Dis* **15**, 117-28 (2008).
47. Guidi, I. *et al.* Oxidative imbalance in patients with mild cognitive impairment and Alzheimer's disease. *Neurobiol Aging* **27**, 262-9 (2006).
48. Schilling, T., Stock, C., Schwab, A. & Eder, C. Functional importance of Ca2+-activated K+ channels for lysophosphatidic acid-induced microglial migration. *European Journal of Neuroscience* **19**, 1469-1474 (2004).
49. Anliker, B. *et al.* Lysophosphatidic acid (LPA) and its receptor, LPA1, influence embryonic schwann cell migration, myelination, and cell-to-axon segregation. *Glia* **61**, 2009-2022 (2013).
50. Armati, P.J. & Mathey, E.K. An update on Schwann cell biology—immunomodulation, neural regulation and other surprises. *Journal of the neurological sciences* **333**, 68-72 (2013).
51. Yukiura, H. *et al.* Autotaxin regulates vascular development via multiple lysophosphatidic acid (LPA) receptors in zebrafish. *Journal of Biological Chemistry* **286**, 43972-43983 (2011).
52. Shi, J., Dong, Y., Cui, M.-Z. & Xu, X. Lysophosphatidic acid induces increased BACE1 expression and Aβ formation. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* **1832**, 29-38 (2013).
53. Walsh, D., Klyubin, I., Fadeeva, J., Rowan, M. & Selkoe, D. Amyloid-β oligomers: their production, toxicity and therapeutic inhibition. (Portland Press Limited, 2002).
54. Hardy, J. & Selkoe, D.J. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *science* **297**, 353-356 (2002).
55. Hampel, H. *et al.* Value of CSF β-amyloid 1–42 and tau as predictors of Alzheimer's disease in patients with mild cognitive impairment. *Molecular psychiatry* **9**, 705 (2004).

56. Wood, P.L., Tippireddy, S., Feriante, J. & Woltjer, R.L. Augmented frontal cortex diacylglycerol levels in Parkinson's disease and Lewy Body Disease. *PloS one* **13**, e0191815 (2018).
57. Giambattistelli, F. *et al.* Effects of hemochromatosis and transferrin gene mutations on iron dyshomeostasis, liver dysfunction and on the risk of Alzheimer's disease. *Neurobiol Aging* **33**, 1633-41 (2012).
58. Sookoian, S. *et al.* Serum aminotransferases in nonalcoholic fatty liver disease are a signature of liver metabolic perturbations at the amino acid and Krebs cycle level, 2. *The American journal of clinical nutrition* **103**, 422-434 (2016).
59. Kapogiannis, D. & Mattson, M.P. Disrupted energy metabolism and neuronal circuit dysfunction in cognitive impairment and Alzheimer's disease. *Lancet Neurol* **10**, 187-98 (2011).
60. Mosconi, L. *et al.* Hypometabolism exceeds atrophy in presymptomatic early-onset familial Alzheimer's disease. *J Nucl Med* **47**, 1778-86 (2006).
61. Drzezga, A. *et al.* Cerebral metabolic changes accompanying conversion of mild cognitive impairment into Alzheimer's disease: a PET follow-up study. *Eur J Nucl Med Mol Imaging* **30**, 1104-13 (2003).
62. Pietzner, M. *et al.* Hepatic steatosis associates with adverse molecular signatures in subjects without diabetes. *J Clin Endocrinol Metab* (2018).
63. Toledo, J.B. *et al.* Metabolic network failures in Alzheimer's disease: A biochemical road map. *Alzheimers Dement* **13**, 965-984 (2017).
64. Varma, V.R. *et al.* Brain and blood metabolite signatures of pathology and progression in Alzheimer disease: A targeted metabolomics study. *PLoS Med* **15**, e1002482 (2018).
65. Kaddurah-Daouk, R. *et al.* Alterations in metabolic pathways and networks in mild cognitive impairment and early Alzheimer's disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* **9**, P571 (2013).
66. Kaddurah-Daouk, R. *et al.* Metabolomic changes in autopsy-confirmed Alzheimer's disease. *Alzheimers Dement* **7**, 309-17 (2011).
67. Botros, M. & Sikaris, K.A. The de ritis ratio: the test of time. *Clin Biochem Rev* **34**, 117-30 (2013).
68. Rui, L. Energy metabolism in the liver. *Compr Physiol* **4**, 177-97 (2014).
69. Francis, P.T. Glutamatergic systems in Alzheimer's disease. *Int J Geriatr Psychiatry* **18**, S15-21 (2003).
70. Kamada, Y. *et al.* Impact of plasma transaminase levels on the peripheral blood glutamate levels and memory functions in healthy subjects. *BBA Clin* **5**, 101-7 (2016).
71. Alfredsson, G., Wiesel, F.A. & Tylec, A. Relationships between glutamate and monoamine metabolites in cerebrospinal fluid and serum in healthy volunteers. *Biol Psychiatry* **23**, 689-97 (1988).
72. Wang, G. *et al.* Plasma metabolite profiles of Alzheimer's disease and mild cognitive impairment. *J Proteome Res* **13**, 2649-58 (2014).
73. Hayasaka, S., Phan, K.L., Liberzon, I., Worsley, K.J. & Nichols, T.E. Nonstationary cluster-size inference with random field and permutation methods. *Neuroimage* **22**, 676-87 (2004).

74. Lowe, S.L., Bowen, D.M., Francis, P.T. & Neary, D. Ante mortem cerebral amino acid concentrations indicate selective degeneration of glutamate-enriched neurons in Alzheimer's disease. *Neuroscience* **38**, 571-7 (1990).
75. Fayed, N., Modrego, P.J., Rojas-Salinas, G. & Aguilar, K. Brain glutamate levels are decreased in Alzheimer's disease: a magnetic resonance spectroscopy study. *Am J Alzheimers Dis Other Demen* **26**, 450-6 (2011).
76. Procter, A.W. *et al.* Evidence of glutamatergic denervation and possible abnormal metabolism in Alzheimer's disease. *J Neurochem* **50**, 790-802 (1988).
77. Kakiyama, G. *et al.* Modulation of the fecal bile acid profile by gut microbiota in cirrhosis. *Journal of hepatology* **58**, 949-955 (2013).
78. Frimmer, M. & Ziegler, K. The transport of bile acids in liver cells. *Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes* **947**, 75-99 (1988).
79. Chiang, J.Y. Bile acid metabolism and signaling. *Compr Physiol* **3**, 1191-212 (2013).
80. Calkin, A.C. & Tontonoz, P. Transcriptional integration of metabolism by the nuclear sterol-activated receptors LXR and FXR. *Nat Rev Mol Cell Biol* **13**, 213-224 (2012).
81. Schaap, F.G., Trauner, M. & Jansen, P.L.M. Bile acid receptors as targets for drug development. *Nat Rev Gastroenterol Hepatol* **11**, 55-67 (2014).
82. Martinez-Diez, M.C., Serrano, M.A., Monte, M.J. & Marin, J.J. Comparison of the effects of bile acids on cell viability and DNA synthesis by rat hepatocytes in primary culture. *Biochim Biophys Acta* **1500**, 153-60 (2000).
83. Ignacio Barrasa, J. *et al.* Deoxycholic and chenodeoxycholic bile acids induce apoptosis via oxidative stress in human colon adenocarcinoma cells. *Apoptosis* **16**, 1054-67 (2011).
84. Faubion, W.A. *et al.* Toxic bile salts induce rodent hepatocyte apoptosis via direct activation of Fas. *J Clin Invest* **103**, 137-45 (1999).
85. Perez, M.J. & Briz, O. Bile-acid-induced cell injury and protection. *World J Gastroenterol* **15**, 1677-89 (2009).
86. Sousa, T. *et al.* Deoxycholic acid modulates cell death signaling through changes in mitochondrial membrane properties. *Journal of Lipid Research* **56**, 2158-2171 (2015).
87. Quinn, M. *et al.* Bile acids permeabilize the blood brain barrier after bile duct ligation in rats via Rac1-dependent mechanisms. *Digestive and Liver Disease* **46**, 527-534 (2014).
88. Higashi, T. *et al.* Unconjugated bile acids in rat brain: Analytical method based on LC/ESI-MS/MS with chemical derivatization and estimation of their origin by comparison to serum levels. *Steroids* **125**, 107-113 (2017).
89. Lalic-Popovic, M. *et al.* Deoxycholic Acid as a Modifier of the Permeation of Gliclazide through the Blood Brain Barrier of a Rat. *J Diabetes Res* **2013**, 598603 (2013).
90. Shin, S.Y. *et al.* An atlas of genetic influences on human blood metabolites. *Nat Genet* **46**, 543-550 (2014).
91. 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).

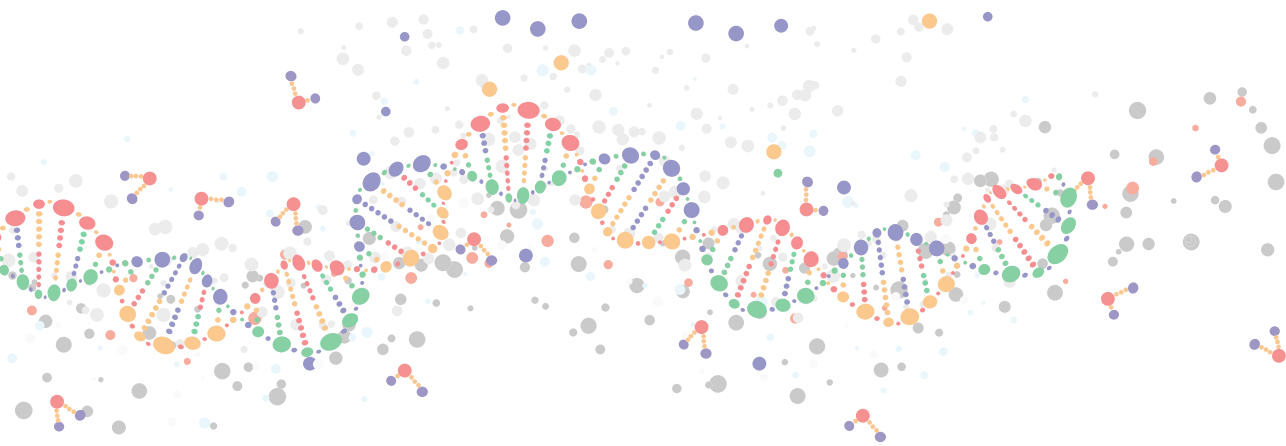
92. Le Guennec, K. *et al.* ABCA7 rare variants and Alzheimer disease risk. *Neurology* **86**, 2134-7 (2016).
93. Hollingworth, P. *et al.* Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nature genetics* **43**, 429 (2011).
94. Le Guennec, K. *et al.* ABCA7 rare variants and Alzheimer disease risk. *Neurology* **86**, 2134-2137 (2016).
95. Satoh, K., Abe-Dohmae, S., Yokoyama, S., St George-Hyslop, P. & Fraser, P.E. ATP-binding cassette transporter A7 (ABCA7) loss of function alters Alzheimer amyloid processing. *Journal of Biological Chemistry* **290**, 24152-24165 (2015).
96. Tripathi, A. *et al.* The gut-liver axis and the intersection with the microbiome. *Nat Rev Gastroenterol Hepatol* **15**, 397-411 (2018).
97. Vogt, N.M. *et al.* Gut microbiome alterations in Alzheimer's disease. *Scientific reports* **7**, 13537 (2017).
98. Harach, T. *et al.* Reduction of Alzheimer's disease beta-amyloid pathology in the absence of gut microbiota. *arXiv preprint arXiv:1509.02273* (2015).
99. Li, Z., Zhu, H., Zhang, L. & Qin, C. The intestinal microbiome and Alzheimer's disease: A review. *Animal models and experimental medicine* **1**, 180-188 (2018).
100. Scheperjans, F. *et al.* Gut microbiota are related to Parkinson's disease and clinical phenotype. *Mov Disord* **30**, 350-8 (2015).
101. Vojinovic, D. *et al.* Relationship between gut microbiota and circulating metabolites in population-based cohorts. *bioRxiv*, 617431 (2019).
102. Karch, C.M. & Goate, A.M. Alzheimer's disease risk genes and mechanisms of disease pathogenesis. *Biological psychiatry* **77**, 43-51 (2015).
103. Efthymiou, A.G. & Goate, A.M. Late onset Alzheimer's disease genetics implicates microglial pathways in disease risk. *Molecular neurodegeneration* **12**, 43 (2017).
104. MahmoudianDehkordi, S. *et al.* Altered bile acid profile associates with cognitive impairment in Alzheimer's disease-An emerging role for gut microbiome. *Alzheimers Dement* **15**, 76-92 (2019).
105. Fu, J. *et al.* The gut microbiome contributes to a substantial proportion of the variation in blood lipids. *Circulation research* **117**, 817-824 (2015).
106. Bis, J.C. *et al.* Whole exome sequencing study identifies novel rare and common Alzheimer's-associated variants involved in immune response and transcriptional regulation. *Molecular psychiatry*, 1 (2018).
107. Sims, R. *et al.* Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. *Nature genetics* **49**, 1373 (2017).
108. Guerreiro, R. *et al.* TREM2 variants in Alzheimer's disease. *N Engl J Med* **368**, 117-27 (2013).
109. Cruchaga, C. *et al.* Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. *Nature* **505**, 550-4 (2014).
110. Mantere, T., Kersten, S. & Hoischen, A. Long-Read Sequencing Emerging in Medical Genetics. *Front Genet* **10**, 426 (2019).

111. Watanabe, K., Taskesen, E., Van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nature communications* **8**, 1826 (2017).
112. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nature genetics* **51**, 1339-1348 (2019).
113. Bertram, L. & Tanzi, R.E. Alzheimer disease risk genes: 29 and counting. *Nature Reviews Neurology* **15**, 191 (2019).
114. Zuk, O., Hechter, E., Sunyaev, S.R. & Lander, E.S. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences* **109**, 1193-1198 (2012).
115. Katsumata, Y., Nelson, P.T., Estus, S., Fardo, D.W. & Initiative, A.s.D.N. Translating Alzheimer's disease–associated polymorphisms into functional candidates: a survey of IGAP genes and SNPs. *Neurobiology of aging* **74**, 135-146 (2019).
116. Rossin, E.J. *et al.* Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS genetics* **7**, e1001273 (2011).
117. 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).
118. Mutoh, T., Rivera, R. & Chun, J. Insights into the pharmacological relevance of lysophospholipid receptors. *British journal of pharmacology* **165**, 829-844 (2012).
119. Spohr, T.C., Dezonne, R.S., Rehen, S.K. & Gomes, F.C. LPA-primed astrocytes induce axonal outgrowth of cortical progenitors by activating PKA signaling pathways and modulating extracellular matrix proteins. *Front Cell Neurosci* **8**, 296 (2014).
120. TC, E.S., Dezonne, R.S., Rehen, S.K. & Gomes, F.C. Astrocytes treated by lysophosphatidic acid induce axonal outgrowth of cortical progenitors through extracellular matrix protein and epidermal growth factor signaling pathway. *J Neurochem* **119**, 113-23 (2011).
121. Yi, L., Liu, W., Wang, Z., Ren, D. & Peng, W. Characterizing Alzheimer's disease through metabolomics and investigating anti-Alzheimer's disease effects of natural products. *Annals of the New York Academy of Sciences* **1398**, 130-141 (2017).
122. Jiang, Y. *et al.* Metabolomics in the development and progression of dementia: A systematic review. *Frontiers in Neuroscience* **13**, 343 (2019).
123. Lindon, J.C. *et al.* Summary recommendations for standardization and reporting of metabolic analyses. *Nature biotechnology* **23**, 833 (2005).
124. van der Lee, S.J. *et al.* The effect of APOE and other common genetic variants on the onset of Alzheimer's disease and dementia: a community-based cohort study. *The Lancet Neurology* **17**, 434-444 (2018).



Chapter 7

Summary / Samenvatting



Summary

Alzheimer's disease (AD) is a complex neurodegenerative disorder that accounts for more than 70% of worldwide dementia cases. With no treatment available, the rising AD prevalence in aging populations is posing a serious economic and health challenge. Advances in multi-omics technology in recent decade enables us to disentangle the molecular mechanism underlying AD pathophysiology to guide drug and preventive treatment research. I used several layers of omics in highly characterized epidemiological cohorts to understand the pathophysiology, molecular mechanism and the risk of AD dementia.

Chapter 2 focused on identifying rare genetic variants in Erasmus Rucphen Family Study using linkage analysis and whole genome-sequencing approach under the Alzheimer's disease sequencing project (ADSP). Our study identified a genomic region 5p14.3 linked with AD. Functional bioinformatics analysis of the genetic variants in the AD linked region identified few rare genetic variants in *CDH18* and *CDH12* genes shared by the individuals contributing to the linkage signal at 5p14.3 regions.

Chapter 3 studied the role of AD implicated biological pathways in the prodromal phase of AD and the interaction of genetic risk factors with modifiable risk factors in AD. In Chapter 3.1, I identified the association of *endocytosis* pathway-based genetic risk score with mild cognitive impairment (MCI). Further, this study identified a relevance between the *immune response* and *clathrin/AP2 adaptor complex* pathways with white matter lesions at magnetic resonance imaging in predementia individuals. In Chapter 3.2, we studied the interaction of lifestyle factors with *APOE ε4* and polygenic risk score in the Rotterdam Study. We reported that the healthy lifestyle factors may lower the risk of dementia in individuals with low and intermediate genetic risk of dementia, while healthy lifestyle did not offset the dementia risk in high genetic risk group.

In Chapter 4, I studied the role of proteins and metabolites in AD risk and pathophysiology. Chapter 4.1, I studied the association of plasma levels of brain-related proteins with AD in overall sample and stratified by *APOE* genotype in the Rotterdam Study. I discovered the association of higher plasma levels of CDH6 and HAGH proteins with AD in *APOE ε4* carriers. The findings of this study were also replicated in the BioFINDER study. We further showed significant positive correlation between CDH6 protein levels and both phosphorylated tau and total tau in cerebrospinal fluid (CSF) in the Amsterdam Dementia Cohort. In Chapter 4.2, I studied the association of CSF and plasma levels of signaling lipids (lysophosphatidic acids (LPAs)) with CSF biomarkers of AD pathology including amyloid-beta 42, phosphorylated tau and total tau.

In Chapter 5, I studied the genetic and metabolic determinants of gut-liver-brain axis. In Chapter 5.1, we studied the association of liver function biomarkers with cognition, AD and imaging biomarkers of brain atrophy in The Alzheimer's Disease Neuroimaging Initiative (ADNI). In

this chapter, we identified that lower levels of alanine aminotransferase (ALT) enzyme and elevated liver enzyme ratio (aspartate aminotransferase (AST):ALT) were associated with AD and lower cognitive measure. Moreover, lower levels of ALT and increased ratio (AST:ALT) were associated with increased Aβ deposition, and reduced brain glucose metabolism and atrophy. In the next Chapter 5.2, we showed significant association of lower levels of primary bile acids (cholic acid) and increased levels of secondary bile acids (deoxycholic acid and its glycine and taurine conjugates) with AD while lower levels of cholic acid also showed association with faster progression from MCI to AD. In the next step (Chapter 5.3), I further investigated the association of AD-associated genetic variants with the abundance of gut microbiota and integrated the identified microbiota with metabolomics. I identified association of risk allele of a genetic variant of *ABCA7* gene with lower abundance of phylum Firmicutes, class Clostridia, order Clostridiales and, higher abundance of order Enterobacteriales and family Enterobacteriaceae.

Samenvatting

De ziekte van Alzheimer is een complexe neurodegeneratieve aandoening, meer dan 70% van alle wereldwijde gevallen van dementie betreft de ziekte van Alzheimer. Omdat er geen behandeling beschikbaar is, vormt de stijgende prevalentie van de ziekte van Alzheimer in de steeds ouder wordende populatie een ernstige economische- en gezondheidsuitdaging. Vooruitgang in multi-omics-technologie in het afgelopen decennium stelt ons in staat om het moleculaire mechanisme achter de pathofysiologie van de ziekte van Alzheimer te ontrafelen om onderzoek naar geneesmiddelen en preventieve behandelingen te begeleiden. Ik gebruikte verschillende lagen omics in zeer gekarakteriseerde epidemiologische cohorten om de pathofysiologie, het moleculaire mechanisme en het risico op ziekte van Alzheimer te begrijpen.

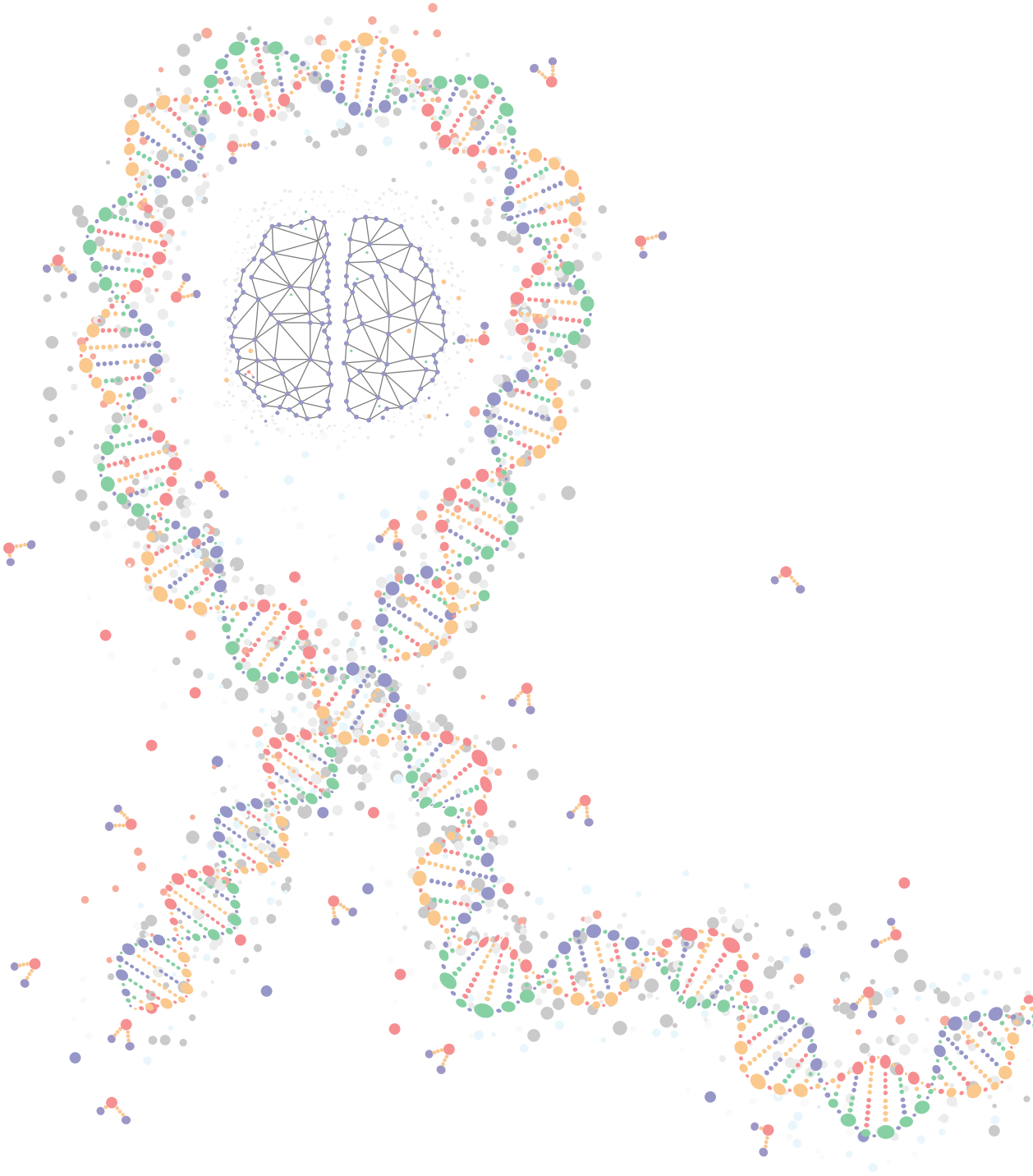
Hoofdstuk 2 richtte zich op het ontdekken van zeldzame genetische varianten in de Erasmus Rucphen Familie Studie met behulp van linkage-analyse en een whole genoom-sequencing-aanpak in het kader van het Alzheimer's disease sequencing project (ADSP). Onze studie ontdekte een genomisch gebied, 5p14.3, gekoppeld aan de ziekte van Alzheimer. Met functionele bioinformatica-analyse van de genetische varianten in het ziekte van Alzheimer gekoppelde gebied, ontdekte we enkele zeldzame genetische varianten in *CDH18*- en *CDH12*-genen die worden gedeeld door de individuen die bijdragen aan het linkage signaal op 5p14.3-regio's.

Hoofdstuk 3 bestudeerde de rol van de bij de ziekte van Alzheimer betrokken biologische paden in de prodromale fase van de ziekte en de interactie van genetische risicofactoren met aanpasbare risicofactoren bij de ziekte van Alzheimer. In hoofdstuk 3.1 ontdekte ik de associatie van op endocytose pathway gebaseerde genetische risicoscore met milde cognitieve stoornis (MCI). Verder toonde deze studie een relevantie tussen de immuunrespons en clathrin/AP2-adapter complexe paden met witte stoflaesies bij MRI-beeldvorming bij pre-dementie individuen. In

hoofdstuk 3.2 hebben we de interactie van leefstijlfactoren met *APOE ε4* en polygene risicoscore onderzocht in de Rotterdam Studie. We hebben gevonden dat gezonde levensstijlfactoren het risico op dementie kunnen verlagen bij personen met een laag en gemiddeld genetisch risico op dementie, terwijl een gezonde levensstijl het dementierisico in de hoge genetische risicogroep niet compenseerde.

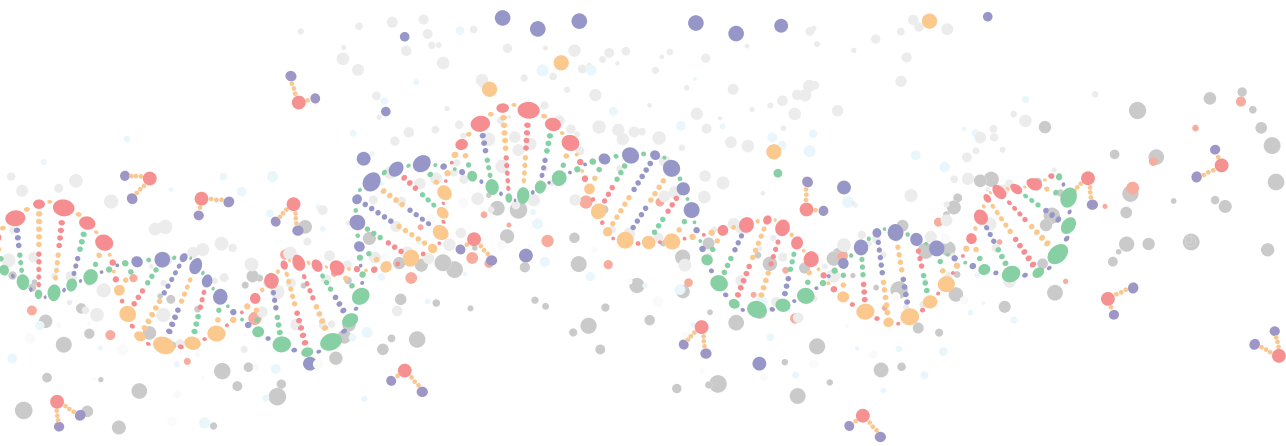
In hoofdstuk 4 heb ik de rol van eiwitten en metabolieten op het risico op de ziekte van Alzheimer en pathofysiologie bestudeerd. In hoofdstuk 4.1 bestudeerde ik de associatie van plasmaconcentraties van hersen gerelateerde eiwitten met de ziekte van Alzheimer in het totale sample en gestratificeerd naar *APO*E-genotype in de Rotterdam Studie. Ik ontdekte de associatie van hogere plasmaspiegels van CDH6- en HAGH-eiwitten met de ziekte van Alzheimer in *APOE ε4*-dragers. De bevindingen van deze studie werden ook gerepliceerd in de BioFINDER-studie. We toonden verder een significante positieve correlatie tussen CDH6-eiwitlevels en zowel gefosforyleerde tau als totale tau in hersenvocht (CSF) in het Amsterdam Dementia Cohort. In hoofdstuk 4.2 bestudeerde ik de associatie van CSF en plasmaspiegels van signaallipiden (lysofosfatidinezuren (LPA's)) met CSF biomarkers van ziekte van Alzheimer pathologie waaronder amyloïde beta-42, gefosforyleerde tau en totale tau.

In hoofdstuk 5 heb ik de genetische en metabole determinanten van de darm-lever-hersenen as bestudeerd. In hoofdstuk 5.1 bestudeerden we de associatie van biomarkers van leverfunctie met cognitie, ziekte van Alzheimer en beeldvormende biomarkers van hersenatrofie in The Alzheimer's Disease Neuroimaging Initiative (ADNI). In dit hoofdstuk hebben we een rol vastgesteld dat lagere levels van alanine aminotransferase (ALAT)-enzym en verhoogde leverenzymverhouding (aspartaataminotransferase (ASAT): ALAT) met ziekte van Alzheimer en lagere cognitie. Bovendien werden lagere ALAT-waarden en een verhoogde ratio (ASAT: ALAT) geassocieerd met verhoogde Aβ-afzetting en verminderd hersenglucosemetabolisme en atrofie. In de volgende studie, in hooftstuk 5.2, toonden we een significante associatie tussen lagere levels van primaire galzuren en verhoogde levels van secundaire galzuren (deoxycholzuur en de glycine- en taurine-conjugaten) en de ziekte van Alzheimer. Terwijl lagere levels van galzuur ook geassocieerd waren met snellere progressie van MCI naar ziekte van Alzheimer. In de volgende stap (hoofdstuk 5.3) heb ik de associatie onderzocht van de ziekte van Alzheimer geassocieerde genetische varianten met de overvloed aan darmmicrobiota en ik heb de ontdekte microbiota geïntegreerd met metabolomics. Ik ontdekte een associatie van het risico-allel van een genetische variant van het *ABCA7*-gen met een lagere overvloed aan phylum Firmicutes, klasse Clostridia, orde Clostridiales en hogere overvloed aan orde Enterobacteriales en familie Enterobacteriaceae.



Chapter 8

Appendix



Chapter 8.1

Acknowledgements

Alhamdulillah, I am writing this section to pay my gratitude to mentors, teachers, supervisors, friends, and family who directly or indirectly paved my way to successfully finish my Ph.D trajactory. In this acknowledgment, first of all, i would like to thank all study participants and people behind the data collection, management, and processing of the Rotterdam Study, GRIP, Erasmus Ruchphen Family Study and other cohorts.

My deepest gratitude goes to my promoter Professor Cornelia M. Van Duijn and co-promoter Najaf Amin. Your guidance, advice, and criticism helped me to accomplish my Ph.D. journey. Someone said about Cornelia that you are like an oracle, where one can find a solution for their problems , which is very true for you. Your ambitious nature and leadership skills are inspiring. You were approachable and available for the scientific discussions. Overall, in these years, you helped me to grow as a researcher and develop perseverance. I was anxious when you moved to Oxford, but your e-supervision worked well. Dear Najaf, you have a unique personality and teaching style. You were always available for help and 'soft' statistical comments. Your valuable tips helped me during writing my manuscripts and presentations. Another thing which I appreciate is your way of writing scripts on paper for new PhD students and crash genetic courses. Now you are finally settled on a new job, and I wish you best of luck for your scientific career.

I would like to extend my gratitude to reading committee members Professor Thomas Hankemeier, Professor M. Arfan Ikram and Professor Fernando Rivadeneira. Dear Thomas, we have collaborated and worked on several projects. I am also thankful that you provided me an opportunity to work with you in Leiden. Your broad vision will help me to shape my personality and future goals. Dear Arfan, I must thank you for being a part of my Ph.D. thesis committee and your overall active role during my PhD. You emerged as a young leader and, indeed, an inspiration for me and many other young researchers. Dear Fernando, thank you very much for being part of my thesis reading committee on such a short notice. I appreciate your feedback on my thesis. I always enjoyed working with you as a teaching assistant in summer Genetic Epidemiology courses, and I am looking forward to collaborate with you in future. I would like to extend my gratitude to members of plenary committee, Professor Wiesje van der Flier and Professor Kristel Sleegers. Thank you for being part of the plenary committee and your time to read my thesis.

Ph.D. journey was not easy without the support of my colleagues and friends from the Genetic-Epi unit. I would like to thank Adriana, Ashley, Ayse, Dina, Elisa, Hata, Ivana, Jun liu, Maaike, and Sven. Jun liu, we came together and attended our first summer course together. Your enthusiasm for science, diabetics, and big projects is something that no one can ignore. It was extremely pleasant sharing office with you. Dear Ashley, you are a nice friend and a colleague. I highly appreciate your help in finalizing my thesis. Your critical vision also helped me to improve my presentations. I will miss our conversations about culture, work, and various work-related issues. Dear Hata, thank you for your help and support during my Ph.D. You were available for

support, discussions especially on very short notices. We worked together on multiple projects and travelled to most of the Alzheimer's events. Your are intelligent and determined. Ivana, thank you for your discussions and comments on my presentations. In addition to genetics, cooking and presentations, I recently learned about your entrepreneurial skills. Adriana, thank you for your constructive discussions and moral support. Dear Dina, it was my pleasure sharing office with you for three years. Thank you for the discussions and help on various projects. Dear Sven, it is indeed hard to describe you in few lines. You were first senior PhD from whom i inherited several projects. You started helping me right from finding a house in Rotterdam till free transportation of the furniture in your car. Travel to CHARGE meeting was "WOW" experience, starting from a short visit to Washington DC till almost missing the flight. I would like to thank Ayse Demirkan for useful scientific discussions and starting several metabolomics projects in genetic-epi group, which is now my main area of interest. My heartfelt thanks to Genetic-Epi lab members; Andrea, Andy, Bernadette and Jeannette. Dear Jeannettee, thank you very much for your welcoming and polite attitude towards my urgent requests. You were always there to solve mysteries related to study samples. As a most senior member of our group, I highly appreciate your work for research. I would like to thank all my colleagues from NeuroEpi group and Internal medicine department. It was great experience working with Silvan, Frank, Tom, Lana, Gennady, Hieab, Sander, Natalie, Mohsen Ghanbari and Oscar. I would like to thank friends and colleagues from other departments for their support and discussions, including Annemarie, Katerina, Georgina, Nisreen, Fasa, Desiree, Pooja, Ilda, Irma, Anirudh, Milly, Djawad and Robert.

I would like to extend my gratitude to Maaike, Gabrielle, Sholange, and Andrea. Their tremendous support throughout the PhD was a great help. Dear Maaike, I am grateful that you always face my requests with smiling face. I am also grateful for your moral support during whole period of my PhD. I am thankful to Nela in Oxford university for her continuous help for the completion of PhD thesis submission and evaluation process. Dear Nano and Frank , I appreciate yours day to day support which was just a step away. Without your support, finishing whole PhD trajectory would not be efficient and smooth.

During my PhD, i had an opportunity to work in several international collaborations and consortia, which truely helped me to shape my personality as a researcher, and taught me the benefits of collaborations. It was valuable experience working with members from ADSP, ADMC, IGAP, CHARGE, CoSTREAM, IMI-ADAPTED, ADES, BBMRI, and EADB consortia. Specially, I would like to thank for support and valuable advise from Professor Sudha Sheshedri and Professor Ellen Wijsman on ADSP and CHARGE projects. I would like thank other members of the consortium, particularly Claudia, Amanda, Elizabeth and Rafael for useful discussions and sharing knowledge. In ADMC consortium, It is a pleasant experience working with the team of Professor Rima Kaddurah-Daouk. I am thankful for your help and collaboration. I appreciate your encouraging and supportive attitude towards ideas from young researchers. Your words and advice are great support for my future career. I would like to thank Mathias, Gabi, Alexandra,

Siamak, Kwangsik, Gregory and Elisa for your help in all collaborative projects. I would like to thank members of IMI-ADAPTED consortium including Agustin Ruiz, Adelina, Alfredo Ramirez, Maria Saez, Margot H.M Bakker and Bridget. I always learn something new from ADAPTED consortium meetings.

I would like to thank my colleagues from the Leiden Academic Centre for Drug Research, Amy Harmen, Michael, Naama, Aga, and Kevin for friendly atmosphere and moral support during submission of my PhD thesis. Dear Amy thank you very much for providing support for proofreading the part of my thesis. Your feedback was useful for me.

There is a saying, 'When the roots are deep, there is no reason to fear the wind'. There are several teachers, mentors and supervisors from my earlier studies, who played a pivotal role to deepen my roots as a researcher. To achieve my ultimate goal , coming to Europe and joining prestigious institute for doctoral research, several people played their role. First, I would like to pay an attribute to my high secondary school teacher Professor Arshad Batti. His kindness, support and teaching ignited in me interest in biological sciences. Moreover, I would like to thank professor Zahoor and Javed Iqbal for their unconditional support. My teachers from Government College University Lahore, my alma-matter, gave a new orientation to my goals. I would like to thank Ashraf Mirza, Mohsin Javed, Farah Shakoori, and Shabzada faisal Khursheed, who played a crucial role in my life. I would like to thank Manfred Schmidt, Fronza Rafael, my research supervisors of master studies from Germany. They made me capable to accept the challenges of this PhD. I would also like to acknowledge Professor Steffen Neuman and his team, for help and support during my first internship in bioinformatics. Thanks to Punjab Education Endowment Fund and their team who provided me an opportunity to fullfil my dreams.

I would like to pay my gratitude to my family members. I would like to thank Shahnawaz, Shafiq, Shahid, Shazia, Sabia, Shanaz, Misbah, Nida, Guria and my father and mother in law for their support and prayers. Dear Shahnawaz and Shafiq due to your support and ecouragment, I was able to achieve higher education. My sister Shazia, although you were just one grade ahead of me but your excellence in studies always inspired me to do my best. Special thank to my father in law, Shahzada Mumtaz, who kept a fatherly eye during my PhD trajectory and used to ask questions about progress of my manuscripts. In addition to his support to govern my matters in Pakistan, he alway encouraged and pushed me to excellence. Countless prayers and *nawafal* from my mother in law have eased several difficulties. I would like to thank all my friends who supported me during these years, including Ali Fawad, Anas, Ehtesham, Imtiaz, Khizar, Rizwan, Rashad, Saleem, Shahid, Shoib, Shahzaman, Faheem, Hammad, Mubashar, and Zia Amjad. They were always there to listen and provide me their moral support.

From the beginning to end, PhD journey would not have been possible without the support of my wife, Noreen Mumtaz, because she held me and all of this together. Many of my friends

used to convince me that the idea of getting married during PhD would not be easy, but I always contradicted. Now, I reached to a firm conclusion that your presence helped me for having less grey hairs during this time. I highly appreciate your patience for managing wedding events and moving to Netherlands in a short notice of just two weeks. Incredibly, you also maintained equally well your social and PhD life in Netherlands. I wish you best of luck for defending your PhD next year (*InshaAllah*). Now we have our sweet daughter Minha, who is like a pleasant breeze in our life and she smiles like you.

I am short of words for paying my acknowledgment to my mother Ramzano Noor and my father, Noor Muhammad, whose countless prayers and efforts made it possible where I am today. They devoted their whole life to support my education in their challenging times. My words cannot pay them the reward. I am lucky for having both of you as my parents.

Shahzad Ahmad
Rotterdam

Chapter 8.2

PhD Portfolio

PhD Portfolio

Name PhD student: **Shahzad Ahmad**

Erasmus MC department: **Epidemiology**

Research school: **Netherlands Institute of Health Sciences (NIHES)**

PhD Period: **August 2014 – September 2019**

Promotor: **Prof. C.M. van Duijn**

Co-promotor: **Dr. N. Amin**

Education	Year	ECTS*
Master of Science in Health Sciences, Genetic epidemiology		
Core Courses		
Study Design	2016	4.3
Biostatistical Methods I: Basic Principles	2015	5.7
Biostatistical Methods II: Classical Regression Models	2015	4.3
Genetic-Epidemiologic Research Methods	2014	5.1
SNP's and Human Diseases	2014	1.4
Advance short courses		
Linux for Scientist	2014	0.6
Advances in Genome-Wide Association Studies	2015	1.4
Family-based Genetic Analysis	2015	1.4
Psychiatric Epidemiology	2015	1.1
Principals of Epidemiologic Data analysis	2015	0.7
An introduction to the next-generation sequencing data	2015	1.4
Courses for the quantitative Researchers	2017	1.4
Erasmus Summer Programme		
Principles of Research in Medicine	2014	0.7
Genome-Wide Association Analysis	2014	1.4
Principles of Genetic Epidemiology	2014	0.7
Genomics in Molecular Medicine	2014	1.4
Advances in Genomics Research	2015	0.4
Human Epigenomics	2017	0.7

(Inter) National conferences, meetings and presentations	Year	ECTS*
AAIC, Los Angeles, USA – (poster)	2019	1.2
ADAPTED consortium meeting Madrid, Spain	2019	0.6
Co-STREAM consortium meeting Stockholm, Sweden – (oral)	2018	1
ADAPTED consortium meeting Seville, Spain – (oral)	2018	1.2
AAIC, Chicago, USA (oral, poster and a media talk)	2018	1.2
ADES meeting, Amsterdam, The Netherlands	2018	0.6
Society for CSF analysis and clinical neurochemistry meeting, Amsterdam, The Netherlands – (poster)	2018	0.9
ADAPTED consortium meeting, Edinburg, UK	2018	0.6
CHARGE consortium meeting, Rotterdam, The Netherlands – (oral)	2018	0.9
ADES and EADB meeting, Barcelona, Spain	2018	0.9
Health RI conference, Utrecht, The Netherlands (poster)	2017	0.9
Co-STREAM consortium meeting, Cambridge, UK (oral)	2017	1
AAIC, London, UK (poster)	2017	1.2
EABD meeting, Rotterdam, The Netherlands	2017	0.6
CHARGE consortium meeting, Charlottesville, USA	2016	0.9
AAIC, Toronto, Canada (oral)	2016	1.2
Co-STREAM consortium meeting, Rotterdam, The Netherlands	2015	0.6
EADB meeting, Lille, France	2015	0.6
Seminars and workshops		
Seminars at the Department of Epidemiology	2014-2019	3
Genetic epidemiology working group meetings	2014-2019	2
BBMRI-OMICS: Introduction and Hands on application, Leiden, The Netherlands	2017	0.6
Applying data science to accelerate precision medicine, Uppsala Sweden	2018	0.6
Teaching		
Genetic Analysis of Family-Based Studies (GE05) – Coordinator	2019	1.5
An Introduction to the analysis of the Next Generation Sequencing Data (GE13) -Lecturing	2017-2019	3
Genome-wide Association Analysis (ESP29) – Teaching Assistant	2015-2018	3
Principles of Genetic Epidemiology (ESP43) – Teaching Assistant	2017-2018	3
Supervising Master's thesis		
María Fernanda Vinueza Veloz, Topic: Genetic risk for Alzheimer's disease and child brain white-matter microstructure	2016-2017	3
Others		
Genetic Epidemiology biweekly working group meeting organization	2015-2018	3
Reviewing assignments		1
· Alzheimer's & Dementia: The Journal of the Alzheimer's Association		
· Neurobiology of Aging		

*ECTS (European Credit Transfer System) equals a workload of 28 hours

List of Publications

Ahmad S, Bannister C, van der Lee SJ, Vojinovic D, Adams HHH, Ramirez A, et al. Disentangling the biological pathways involved in early features of Alzheimer's disease in the Rotterdam Study. Alzheimers Dement. 2018;14(7):848-57.

Ahmad S, Milan MDC, Hansson O, Demirkan A, Agustin R, Saez ME, et al. CDH6 and HAGH protein levels in plasma associate with Alzheimer's disease in *APOE* epsilon4 carriers. Sci Rep. 2020;10(1):8233.

Nho K*, Kueider-Paisley A*, Ahmad S*, MahmoudianDehkordi S, Arnold M, Risacher SL, et al. Association of Altered Liver Enzymes With Alzheimer Disease Diagnosis, Cognition, Neuroimaging Measures, and Cerebrospinal Fluid Biomarkers. JAMA Netw Open. 2019;2(7):e197978.

Licher S, Ahmad S, Karamujic-Comic H, Voortman T, Leening MJG, Ikram MA, et al. Genetic predisposition, modifiable-risk-factor profile and long-term dementia risk in the general population. Nat Med. 2019;25(9):1364-9.

MahmoudianDehkordi S*, Arnold M*, Nho K*, Ahmad S, Jia W, Xie G, et al. Altered bile acid profile associates with cognitive impairment in Alzheimer's disease-An emerging role for gut microbiome. Alzheimers Dement. 2019;15(1):76-92.

Karamujić-Čomić H, Ahmad S, Lysen TS, Heshmatollah A, Roshchupkin GV, Vernooij MW, et al. Prion protein codon 129 polymorphism in mild cognitive impairment and dementia: the Rotterdam Study. Brain Communications. 2020;2(1):fcaa030.

van der Spek A, Warner SC, Broer L, Nelson CP, Vojinovic D, Ahmad S, et al. Exome Sequencing Analysis Identifies Rare Variants in ATM and RPL8 That Are Associated With Shorter Telomere Length. Front Genet. 2020;11:337.

Vojinovic D, Brison N, Ahmad S, Noens I, Pappa I, Karssen LC, et al. Variants in TTC25 affect autistic trait in patients with autism spectrum disorder and general population. Eur J Hum Genet. 2017;25(8):982-7.

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

Rueda-Ochoa OL, Rojas LZ, Ahmad S, van Duijn CM, Ikram MA, Deckers JW, et al. Impact of cumulative SBP and serious adverse events on efficacy of intensive blood pressure treatment: a randomized clinical trial. J Hypertens. 2019;37(5):1058-69.

Nho K, Kueider-Paisley A, MahmoudianDehkordi S, Arnold M, Risacher SL, Louie G, et al. Altered bile acid profile in mild cognitive impairment and Alzheimer's disease: Relationship to neuroimaging and CSF biomarkers. Alzheimers Dement. 2019;15(2):232-44.

Colijn JM, den Hollander AI, Demirkan A, Cougnard-Gregoire A, Verzijden T, Kersten E, et al. Increased High-Density Lipoprotein Levels Associated with Age-Related Macular Degeneration: Evidence from the EYE-RISK and European Eye Epidemiology Consortia. Ophthalmology. 2019;126(3):393-406.

Vojinovic D, Adams HH, Jian X, Yang Q, Smith AV, Bis JC, et al. Genome-wide association study of 23,500 individuals identifies 7 loci associated with brain ventricular volume. Nat Commun. 2018;9(1):3945.

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

Hofer E, Roshchupkin GV, Adams H, Knol M, Lin H, Li S, et al. Genetic Determinants of Cortical Structure (Thickness, Surface Area and Volumes) among Disease Free Adults in the CHARGE Consortium. bioRxiv. 2018:409649.

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

Leeuw FA*, Karamujić-Čomić H*, Tijms B.M., Peeters CFW, Kester MI, Scheltens P, Ahmad S, et al. Circulating metabolites are associated with brain atrophy and white matter hyperintensities. (Accepted for publication in the Journal of Alzheimer's and Dementia)

Ahmad S, Rafael A. Nafikov Rafael A, Blue E, Navas Patrick A, Karamujić-Čomić H, et al. Variation in cadherin genes in the 5p14.3 genomic region is implicated in Alzheimer's disease in an extended ADSP pedigree. (Submitted for publication)

Ahmad S, Orellana A, Kohler I, Frölich L, Rojas I, Gil S, et al. Association of lysophosphatidic acids with cerebrospinal fluid biomarkers and progression to Alzheimer's disease. (Under review for publication in Alzheimer's research and Therapy)

Ahmad S*, Karamujić-Čomić H*, Radjabzadeh D*, Bonnechere B, Vojinovic D, Hankemier T, et al. The Alzheimer's disease genetic risk variant in ABCA7 is associated to the gut microbiome. (In preparation)

**shared first authors*

Chapter 8.3

About the author

Shahzad Ahmad was born on 3rd of February 1988 in Kasur, Pakistan. After completing his higher secondary school education from Government College Kot Radha Kishan, Kasur (2005), he started his bachelors in Microbiology at Government College University, Lahore, Pakistan. After finishing his bachelor studies, he started working in a pharmaceutical company as a microbiologist. In 2011, he received a foreign study scholarship from Punjab Education Endowment Fund Pakistan for master in Life Science Informatics from the Bonn-Aachen International Center for Information technology, Bonn, Germany. During his master studies in Germany, he performed an internship at the Leipzig institute for plant biochemistry. He carried out his master thesis research at the Department of Translational Oncology, German Cancer Research Center (DKFZ) Heidelberg, Germany, during 2013-2014. In the same years, he also worked as a student research assistant in the same department. In 2014, he started his doctoral degree in the department of Epidemiology under the supervision of Professor Cornelia M. Van Duijn. During his PhD, he was also enrolled in master in Health Sciences (Genetic Epidemiology) from the Netherlands Institute for Health Sciences, Rotterdam. During his Ph.D, he worked in international consortia, including the Alzheimer's disease sequencing project (ADSP), the Alzheimer's Disease Metabolomics Consortium (ADMC), the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) and the International Genomics of Alzheimer's Project (IGAP). He had presented his research work at multiple international conferences and meetings. In August 2019, he started working as a postdoctoral researcher at the Leiden Academic Centre for Drug Research (LACDR). Since April 2020, he moved back to Erasmus MC, Rotterdam and is working as a joint postdoctoral researcher in the department of Epidemiology, Erasmus MC, Rotterdam and the LACDR, Leiden.