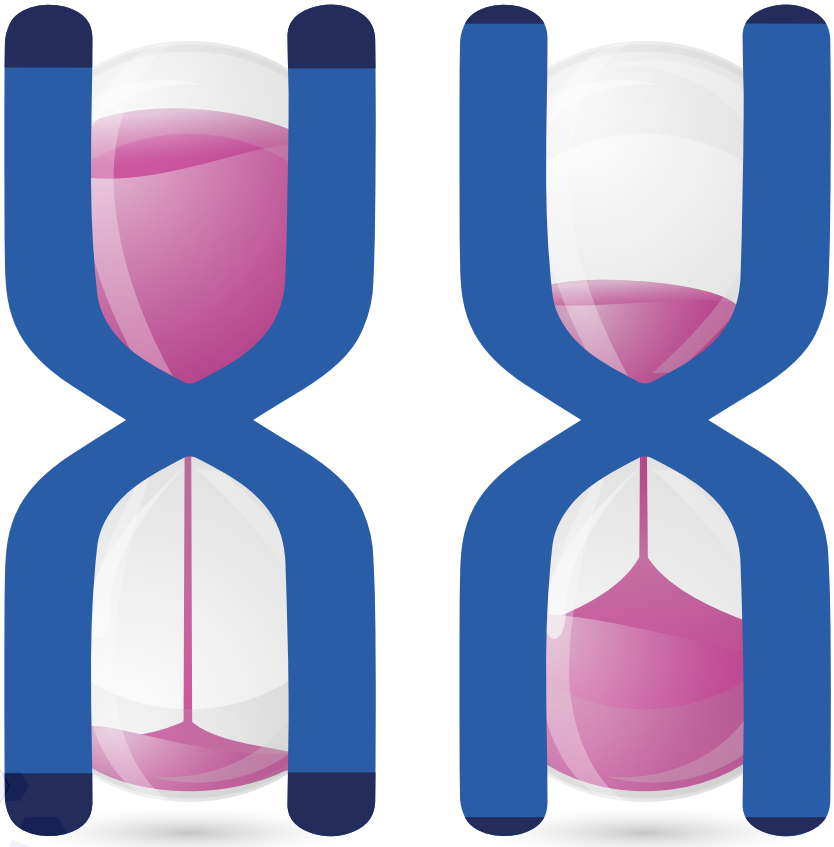
GENETIC AND METABOLIC STUDIES

OF AGING, DEPRESSION AND SLEEP IN THE GENERAL POPULATION



ASHLEY VAN DER SPEK

GENETIC AND METABOLIC STUDIES

OF AGING, DEPRESSION AND SLEEP IN THE GENERAL POPULATION

ASHLEY VAN DER SPEK

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 studies used in this thesis are gratefully acknowledged.

Printing of this thesis was financially supported by: Department of Epidemiology, Erasmus University Medical Center, Rotterdam, Erasmus University, Rotterdam and SkylineDx.

ISBN: 978-94-6423-430-5

Cover design & lay-out: Wendy Schoneveld || www.wenziD.nl
Printed by: ProefschriftMaken || www.proefschriftmaken.nl

Genetic and metabolic studies of aging, depression and sleep in the general population

Genetische en metabole studies naar veroudering, depressie
en slaap in de algemene bevolking

Proefschrift

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

Prof.dr. A.L. Bredenoord

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
vrijdag 29 oktober 2021 om 13:00 uur

door

Ashley van der Spek

geboren te Rotterdam

Erasmus University Rotterdam

Erasmus

Promotiecommissie

Promotor:	Prof.dr.ir. C.M. van Duijn
Overige leden:	Prof.dr. M.K. Ikram
	Prof.dr. P.E. Slagboom
	Prof.dr. A.G. Uitterlinden
Copromotor:	Dr. N. Amin

Table of Contents

Publications and manuscripts included in this thesis

Chapter 2.1
Ashley van der Spek, Sophie C. Warner, Linda Broer, Christopher P. Nelson, Dina Vojinovic, Shahzad Ahmad, Pascal P. Arp, Rutger W.W. Brouwer, Matthew Denniff, Mirjam C.G.N. van den Hout, Jeroen G.J. van Rooij, Robert Kraaij, Wilfred F.J. van IJcken, Nilesh J. Samani, M. Arfan Ikram, André G. Uitterlinden, Veryan Codd, Najaf Amin, Cornelia M. van Duijn**. Exome sequencing analysis identifies rare variants in *ATM* and *RPL8* that are associated with shorter telomere length.** *Front Genet*. 2020 Apr 30;11:337. doi:10.3389/fgene.2020.00337.

Chapter 2.2
Ashley van der Spek*, Linda Broer*, Harmen HM Draisma*, René Pool*, Eva Albrecht, Marian Beekman, Massimo Mangino, Mait Raag, Dale R. Nyholt, Harish K. Dharuri, Veryan Codd, Najaf Amin, Eco J.C. de Geus, Joris Deelen, Ayse Demirkan, Idil Yet, Krista Fischer, Toomas Haller, Anjali K. Henders, Aaron Isaacs, Sarah E. Medland, Grant W. Montgomery, Simon P. Mooijaart, Konstantin Strauch, H. Eka D. Suchiman, Anika A.M. Vaarhorst, Diana van Heemst, Rui Wang-Sattler, John B. Whitfield, Gonneke Willemsen, Margaret J. Wright, Nicholas G. Martin, Nilesh J. Samani, Andres Metspalu, P. Eline Slagboom, Tim D. Spector, Dorret I. Boomsma**, Cornelia M. van Duijn**, Christian Gieger**. **Metabolomics reveals a link between homocysteine and lipid metabolism and leukocyte telomere length: the ENGAGE consortium.** *Sci Rep*. 2019 Aug 12;9(1):11623. doi:10.1038/s41598-019-47282-6.
* These authors contributed equally to this work
** These senior authors contributed equally to this work

Chapter 2.3
Ashley van der Spek, Hata Comic, Jun Liu, René Pool, Mariska Bot, Marian Beekman, Sanzhima Garmaeva, Alexessander Couto Alves, Gonneke Willemsen, Gerard van Grootheest, Cisca Wijmenga, André Uitterlinden, M. Arfan Ikram, Marjo-Riitta Jarvelin, Alexandra Zhernakova, P. Eline Slagboom, Brenda W.J.H. Penninx, Dorret I. Boomsma, Najaf Amin, Cornelia M. van Duijn. **Fat metabolism is associated with telomere length in six population-based studies.** Accepted for publication in *Human Molecular Genetics*.

Chapter 2.4
Sara A. Galle, Ashley van der Spek, Madeleine L. Drent, Michael P. Brugts, Erik J.A. Scherder, Joseph A.M.J.L. Janssen, M. Arfan Ikram, Cornelia M. van Duijn. **Revisiting the role of insulin-like growth factor-I receptor stimulating activity and the apolipoprotein E in Alzheimer's disease.** *Front Aging Neurosci*. 2019 Feb 12;11:20. doi:10.3389/fnagi.2019.00020.

Chapter 3.1
Ashley van der Spek*, Isobel D. Stewart*, Brigitte Kühnel*, Maik Pietzner*, Tahani Alshehri*, Friederike Gauß, Pirro G. Hysi, Siamak Mahmoudian Dehkordi, Almut Heinken, Annemarie

Luik, Karl-Heinz Ladwig, Gabi Kastenmüller, Johannes Hertel, M. Arfan Ikram, Renée de Mutsert, Karsten Suhre, Christian Gieger, Konstantin Strauch, Henry Völzke, Thomas Meitinger, Antonia Flaquer, Melanie Waldenberger, Annette Peters, Ines Thiele, Rima Kaddurah-Daouk, Frits Rosendaal, Nicholas J. Wareham, Tim D. Spector, Sonja Kunze**, Hans Jörgen Grabe**, Dennis Mook-Kanamori**, Claudia Langenberg**, Cornelia M. van Duijn**, Najaf Amin** **Circulating metabolites related to diet, medication and the gut microbiome are associated with depression.** (Submitted)
* These authors contributed equally to this work
** These senior authors contributed equally to this work

Chapter 3.2
Najaf Amin, Karla V. Allebrandt, Ashley van der Spek, Bertram Müller-Myhsok, Karin Hek, Maris Teder-Laving, Caroline Hayward, Tõnu Esko, Josine G. van Mill, Hamdi Mbarek, Nathaniel F. Watson, Scott A. Melville, Fabiola Del Greco M., Enda M. Byrne, Edwin Oole, Ivana Kolcic, Ting-hsu Chen, Daniel S. Evans, Josef Coresh, Nicole Vogelzangs, Juha Karjalainen, Gonneke Willemsen, Sina A. Gharib, Lina Zgaga, Evelin Mihailov, Katie L. Stone, Harry Campbell, Rutger Brouwer, Ayse Demirkan, Aaron Isaacs, Zoran Dogas, Kristin D. Marciante, Susan Campbell, Fran Borovecki, Annemarie I. Luik, Man Li, Jouke Jan Hottenga, Jennifer E. Huffman, Mirjam C. G. N. van den Hout, Steven R. Cummings, Yurii S. Aulchenko, Philip R. Gehrman, Andre G. Uitterlinden, H. Erich Wichmann, Martina Müller-Nurasyid, Rudolf S. N. Fehrmann, Grant W. Montgomery, Albert Hofman, W.H. Linda Kao, Ben A. Oostra, Alan F. Wright, Jacqueline M. Vink, James F. Wilson, Peter P. Pramstaller, Andrew A. Hicks, Ozren Polasek, Naresh M. Punjabi, Susan Redline, Bruce M. Psaty, Andrew C. Heath, Martha Merrow, Gregory J. Tranah, Daniel J. Gottlieb, Dorret I. Boomsma, Nicholas G. Martin, Igor Rudan, Henning Tiemeier, Wilfred F. J. van Ijcken, Brenda W. Penninx, Andres Metspalu, Thomas Meitinger, Lude Franke, Till Roenneberg*, Cornelia M. van Duijn*. **Genetic variants in *RBFOX3* are associated with sleep latency.** *Eur J Hum Genet*. 2016 Oct;24(10):1488-95. doi:10.1038/ejhg.2016.31.
* These senior authors contributed equally to this work

Chapter 3.3
Ashley van der Spek, Annemarie I. Luik, Desana Kocevska, Chunyu Liu, Rutger Brouwer, Jeroen G.J. van Rooij, Mirjam C.G.N. van den Hout, Robert Kraaij, Albert Hofman, André G. Uitterlinden, Wilfred F.J. van IJcken, Daniel J. Gottlieb, Henning Tiemeier, Cornelia M. van Duijn, Najaf Amin. **Exome-wide meta-analysis identifies rare 3'-UTR variant in *ERCC1/CD3EAP* associated with symptoms of sleep apnea.** *Front Genet*. 2017 Oct 18;8:151. doi:10.3389/fgene.2017.00151.

CHAPTER 1

General introduction

Human life expectancy has increased significantly during the past 200 years[1] and is expected to increase even further (**Figure 1**). In 2015, there were up to 900 million individuals aged 60 years or over worldwide, and this number is expected to increase to 2 billion in 2050[2]. Consequently, a clear shift in the distribution of the population towards older age can be observed, the process known as population aging. Although this process started in high-income countries, this trend is also emerging in low- and middle-income countries nowadays[2]. However, the increase in lifespan was not accompanied by an increase in length of healthy life[3] as, on average, individuals spend 16-20% of their life in late-life morbidity[4].

Aging is the most important risk factor for the development of many common diseases such as cardiovascular disease, type 2 diabetes, neurodegenerative disorders like dementia and Alzheimer's disease, and various cancers[5]. Population aging is associated with many factors, including genetic, environmental and lifestyle factors[6]. Although considerable progress has been made in aging research, it is far from understood how aging leads to pathology and disentangling the underlying molecular mechanisms may lead to possibilities to intervene in the aging process[7].

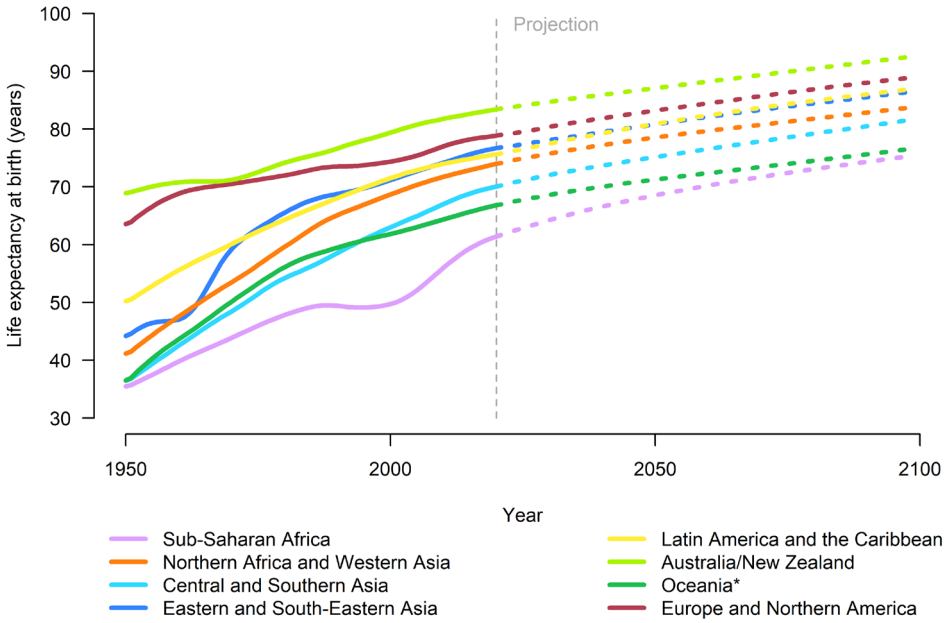


Figure 1. Estimated and projected life expectancy at birth by world region. The x-axis shows the time periods and the y-axis the life expectancy at birth in years. Source: World Population Prospects 2019: Highlights (United Nations publication ST/ESA/SER.A/423, https://population.un.org/wpp/Publications/Files/WPP2019_Highlights.pdf).

Therefore, in this thesis, I explored the role of genetic and metabolic factors in age-related outcomes including telomere length, the insulin/insulin-like growth factor signaling (IIS) pathway, late-onset depression, and sleep-related outcomes. Background information on these outcomes will be provided in this chapter, followed by the molecular approaches applied in this thesis, the study populations, and the aims of the studies included in this thesis.

Background information on age-related outcomes

Telomere length biology

Telomeres are located at the very ends of the chromosomes and consist of a stretch of several kilobases (5 – 15 kb) of tandemly repeated hexamers (5'-TTAGGG-3')[8]. Telomeres are important for genomic stability, which was recognized by Hermann Muller (Nobel Prize 1946) and Barbara McClintock (Nobel Prize 1983), who discovered that the ends of chromosomes had a different structure than the ends of a chromosomal break, and that loss of telomeres resulted in chromosome fusions, rearrangements and genomic instability[9,10]. To perform their function, telomeres interact with a protein complex called shelterin, which consists of six subunits (TRF1, TRF2, Rap1, TIN2, TPP1, POT1)[11]. Shelterin is involved in the generation of so-called t-loops of telomeric DNA and protects the telomeres from being recognized as "damaged DNA" by the DNA damage response machinery[11-13] (**Figure 2A**). In somatic cells, telomeres shorten during each cell division because of the inability of DNA polymerase to fully extend the telomere DNA at the 3' end of the DNA strand[14] (**Figure 2B**).

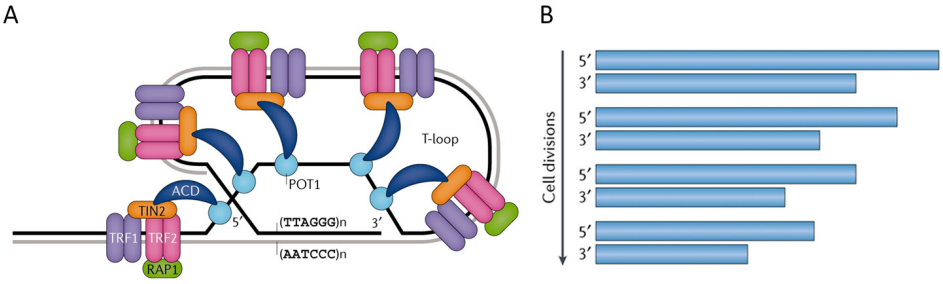


Figure 2. A) Protection of telomeres through formation of a T-loop binding with the Shelterin complex. B) Shortening of telomeres with each division in somatic cells. Figure adjusted from Shay & Wright[13].

This process is commonly referred to as the 'end replication problem'[16]. After telomeres reach a critical length, the cells are triggered to enter replicative senescence, which will eventually lead to cell death (apoptosis)[17,18]. Telomere length in germ cells is maintained by

the expression of telomerase, a ribonucleoprotein consisting of an RNA template (TERC), a reverse transcriptase or catalytic subunit (TERT) and associated proteins[19], which is normally inactive in somatic cells[20]. The discovery of the molecular structure of telomeres and the involvement of the telomerase enzyme led to a Nobel Prize in 2009 (Elizabeth H. Blackburn, Carol W. Greider, and Jack W. Szostak)[17].

Studies investigating the association of telomere length with aging showed that cells from multiple premature-aging syndromes have shorter telomere length, the so-called "telomeropathies" or telomere maintenance spectrum disorders. The primary telomeropathies are caused by mutations in genes involved in telomere maintenance, while secondary telomeropathies are mostly caused by mutations in genes involved in DNA repair or in telomere preservation[15]. Mean telomere length was found to be shorter in men than in women and in older age[21-23] and shorter telomere length has been associated with an increased risk of several age-related diseases, including coronary artery disease[21-24], diabetes[23,25,26] and dementia[27-29]. Some studies also found an association with mortality[29-35], although this association was inconsistent between studies[36-41]. Furthermore, both short and long telomeres have been associated with cancer[42-45]. Telomere length is affected by genetic, environmental, and lifestyle factors[46-48], and has been proposed as a marker of biological age[49]. As telomere length is highly variable between individuals, even between individuals of the same age, it can be hypothesized that it is largely explained by genetics[50].

Indeed, telomere length is highly heritable with heritability estimates (i.e., the proportion of observed variance in a trait that can be attributed to inherited genetic factors) varying between 44-86%[50-52]. Although most studies use blood samples to measure telomere length in leukocytes, mean leukocyte telomere length (LTL) and telomere length in other tissues are highly correlated within individuals[53,54]. Previous genome-wide association studies (GWAS) have identified multiple common genetic variants (minor allele frequency (MAF) > 5%) associated with LTL in Europeans[55-61] and South Asians[62,63]. The two most recent as well as largest GWAS studies of LTL have been published in Singaporean Chinese (N = 23,096)[63] and in Europeans (N = 78,592)[61]. The GWAS in Europeans[61] identified 20 variants at 17 genomic loci significantly associated with LTL ($P < 5x10^{-8}$) and an additional 32 loci at FDR < 0.05 (**Figure 3**).

In addition to the confirmation of associations between LTL and seven previously reported loci (*TERC*, *TERT*, *NAF1*, *OBFC1*, *DCAF4*, *ZNF208*, and *RTEL1*), 6 novel loci were found (*SENPJ*, *MOB1B*, *CARMIL1*, *PRRC2A*, *TERF2*, and *RFWD3*). Interestingly, multiple variants were located in loci reported by the Singaporean Chinese GWAS (*POT1*, *PARP1*, *ATM*, and *MPHOSPH6*). However, these variants explain only up to 5% of the telomere length heritability[61]. Advances in next-generation sequencing techniques have improved our ability to capture rare variants, which may explain part of the missing heritability. In this thesis, I aim to identify rare variants associated with LTL by performing a whole-exome sequencing analysis.

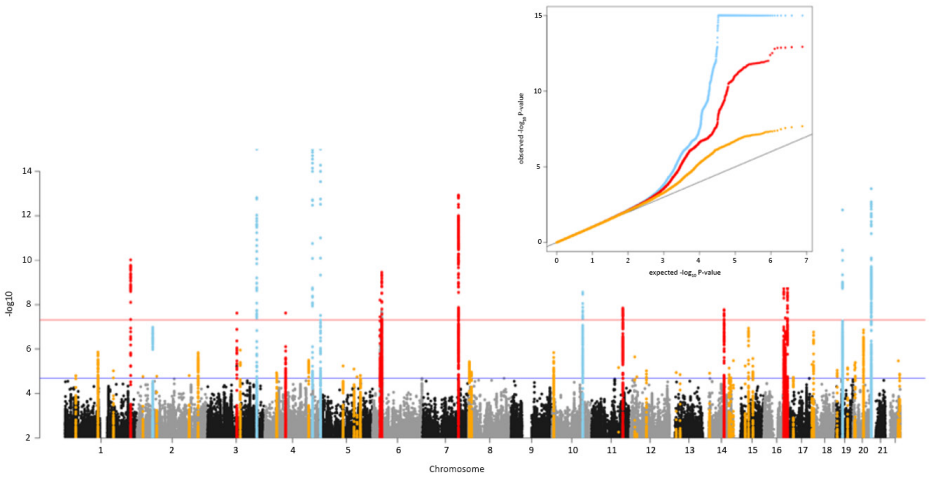


Figure 3. Manhattan plot and accompanying quantile-quantile plot of the European GWAS results[59]. Previously identified loci are depicted in blue, novel genome-wide significant loci in red ($P < 5x10^{-8}$), and identified loci at FDR < 0.05 in orange. Source: Li et al.[59]

Molecular mechanisms

The molecular mechanisms underlying the association of LTL with age-related diseases remain largely unknown. There has been increasing interest in the discovery of metabolomic biomarkers for aging and mortality as these may be valuable in clinical settings. Two large studies have investigated the association of metabolomic biomarkers with all-cause mortality using a Nuclear Magnetic Resonance (NMR) spectroscopy platform (Nightingale Health Ltd., Helsinki, Finland)[64,65]. The first study included data of 17,435 individuals and identified four biomarkers that predict the risk of all-cause mortality: alpha-1-acid glycoprotein, albumin, very-low-density lipoprotein particle size, and citrate[64]. The second study further increased the statistical power by increasing the sample size using a wide age range and thereby also increased the number of deaths[65]. This study included 44,168 individuals with metabolomics data measured in a standardized way on the same Nightingale Health platform and identified 14 metabolic biomarkers that were independently associated with all-cause mortality[65]: total lipids in chylomicrons and extremely large VLDL (XXL-VLDL-L), small high-density lipoprotein (S-HDL-L), the mean diameter for VLDL particles (VLDL-D), the ratio of polyunsaturated fatty acids to total fatty acids (PUFA/FA), histidine, leucine, valine, albumin, glucose, lactate, isoleucine, phenylalanine, acetoacetate, and glycoprotein acetyls. The research group further showed that combining these 14 markers in a risk model together with sex improved the prediction of mortality compared to a model based on conventional risk factors (i.e., sex, body mass index, systolic blood pressure, total cholesterol, HDL cholesterol, triglycerides, creatinine, smoking, alcohol, prevalent diabetes, prevalent cardiovascular disease, and prevalent cancer) for mortality[65]. These findings suggest that metabolic profiles may be used in clinical settings to improve treatment strategies.

Furthermore, multiple metabolites have also been associated with age-related outcomes, such as, among others, cardiovascular disease[66,67], type 2 diabetes[68-71], and dementia[72,73]. As telomere length is a potential biomarker of aging, identifying metabolomic biomarkers associated with telomere length is of interest and may provide important insights in the underlying biological mechanisms and the relation with age-related diseases. To date, only a few studies have investigated the association of metabolites with LTL[74-76]. The first study investigated metabolic biomarkers involved in aging and early development using data of 6,055 participants of the TwinsUK registry[74]. The authors found a set of 22 metabolites that were strongly correlated with age and age-related outcomes, but the individual metabolites were not significantly associated with LTL. The second study identified 19 metabolites associated with LTL in American-Indians, but the sample size was small (N = 423) and there was no replication[75]. The third study of 3,511 females from the Twins UK registry identified five significant metabolites, including two lysolipids, two gamma-glutamyl amino acids, and one xenobiotic. These findings suggest the involvement of lipid and fatty acid metabolism and oxidative stress in biological aging, reflected by LTL[76]. Therefore, larger metabolomics studies are needed to further investigate the metabolic pathways associated with LTL, as performed in the two studies included in this thesis.

The insulin/insulin-like growth factor signaling (IIS) pathway

Determinants of human lifespan, such as genetic, environmental and lifestyle factors, have been studied extensively in the past. Studies of human lifespan or longevity have used different phenotype definitions, either age at death or survival to an exceptional age, for example reaching 90 years or over[77], which makes the studies difficult to compare[78]. Based on the results of twin and pedigree studies, genetic studies have estimated the heritability of human lifespan to be around 27% in the general population[79]. However, a previous study in centenarians (i.e., individuals aged 100 years or over) showed that the influence of genetics appears to get stronger with older age[80]. Although the genetics of longevity has been studied by many candidate gene studies and GWASs, the most convincing loci that were reported in multiple independent studies are Forkhead Box O3 (*FOXO3*) and apolipoprotein E (*APOE*) on chromosomes 6 and 19[81-88].

One of the genes associated with longevity, *FOXO3*, is part of the IIS pathway. *FOXO3* is the human homologue of *daf-16*, a lifespan increasing gene in *Caenorhabditis elegans*[89,90], which encodes a fork head transcription factor that targets genes involved in cellular stress response, metabolism and immune function[91]. The highly conserved IIS pathway is one of the most studied pathways in longevity and is involved in cell survival, cell growth and proliferation[92]. Down-regulation of this pathway has been associated with increased lifespan in multiple model organisms[93-98], although findings in humans are inconsistent[99-101]. In mammals, the IIS pathway involves a complex interplay of multiple ligands including insulin, insulin/insulin-like growth factor I (IGF-I) and IGF-II, and multiple tyrosine kinase receptors: the insulin receptor (IR), IGF-I receptor, and orphan IR related receptor (IRR)[102]. Following ligand binding, two different pathways are activated: PI3K-PKB/AKT and Ras-MAPK pathway, which regulate most of the metabolic and mitogenic effects of insulin/IGF-I signaling[102].

Alterations in IGF-I signaling have been reported in brains of patients with Alzheimer's disease[103], a disease for which age is the main driver. Higher levels of IGF-I receptor stimulating activity, a measure of IGF-I bioactivity[104], has been associated with a higher prevalence and incidence of Alzheimer's disease in the population-based Rotterdam Study[105]. However, these findings have not been replicated. As the Rotterdam Study is still ongoing, we tested the robustness of the association of IGF-I bioactivity with dementia by extending the follow-up period in this thesis. Additionally, we evaluated the possible interaction between IGF-I and *APOE* as *APOE*-ε4 is the most important genetic risk factor for Alzheimer's disease[106]. *APOE* is also the only locus that was significantly associated with longevity across multiple independent studies[83,86,88,107,108]. The most recent GWAS of longevity in Europeans showed that *APOE*-ε4 is associated with lower risk of surviving to old age, in this case to the 90th and 99th survival percentile based on life tables that were specific for each country, sex, and birth year[88]. The *APOE*-ε2 allele, which has been shown to reduce risk for the development of AD[109], showed an increased risk of surviving to old age[88].

Late onset depression

Depression is a common, complex disease characterized by sadness, loss of interest or pleasure, feelings of guilt, low self-worth, tiredness, poor concentration, and disturbed sleep or appetite[110]. The World Health Organization (WHO) has estimated that over 300 million people, which is more than 4% of the world's population, are affected by depression[110,111]. Depression is on average more common among females (5.1%) than males (3.6%) and the prevalence rates increase with age, where on average 7.6% of females and 4.9% of males aged 55-74 years suffer from depression (**Figure 4**).

Depression in older adults is often underdiagnosed as their symptoms do not meet the criteria for diagnosis. Without a doubt, life events at old age, such as those related to mortality of relatives and friends and morbidity are important to be taken into consideration. However, depression in older adults is also clinically relevant. Late onset depression is associated with cognitive impairment, physical disability and anxiety[112]. Multiple different treatments exist for late onset depression, including psychological treatments and antidepressant medication (e.g., selective serotonin reuptake inhibitors (SSRIs) and tricyclic antidepressants)[113]. Moreover, depression is a complex, multifactorial disease resulting from an interaction of social, psychological, physical[114], and biological factors (i.e., neuro-immunological, neuroendocrine, and genetic factors)[115].

The heritability of depression has been estimated to be around 40%. Multiple GWASs have been performed but there have been difficulties in the identification of common genetic variants associated with depression[116,117]. As depression is a heterogeneous disease, large sample sizes were needed to discover novel genetic variants[118]. Indeed, large GWASs including >300,000 individuals had enough power to detect single nucleotide polymorphisms (SNPs) associated with depression[119-121]. The latest meta-analysis combined data from the three largest GWASs performed until 2018 (Hyde et al.[119], Wray et al.[120], Howard et al.[121]).

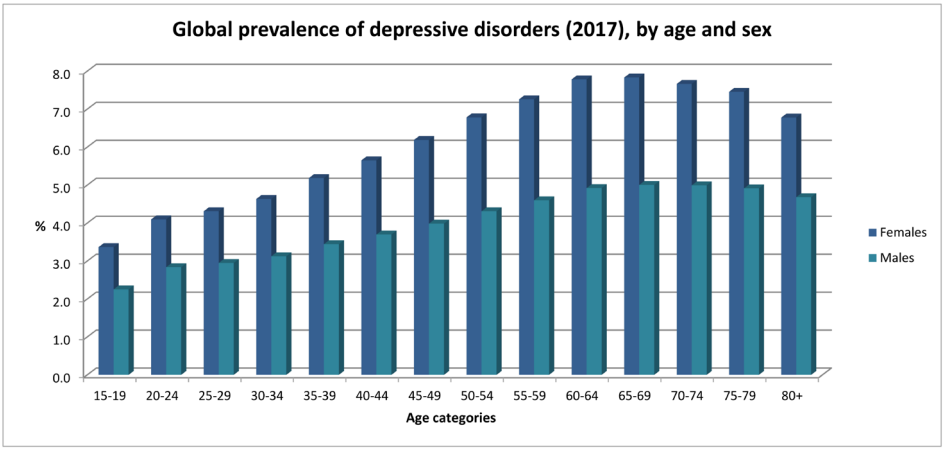
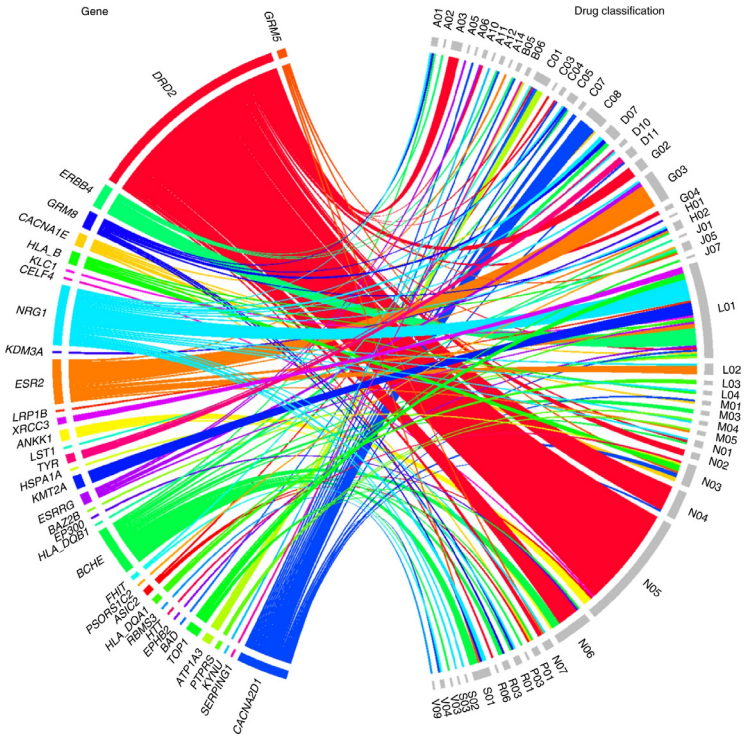


Figure 4. Global prevalence of depressive disorders in 2017. Source: Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2017 (GBD 2017) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2018. Available at http://ghdx.healthdata.org/gbd-results-tool.

The study included data from 807,553 individuals (N_{cases} = 246,363, $N_{controls}$ = 561,190) and identified 102 independent variants, 269 genes, and 15 pathways associated with depression including those involving synaptic structure and neurotransmission[122]. Evidence for the importance of prefrontal brain regions in depression was found and replicated in an independent dataset (N = 1,306,354). The authors further showed drug-gene interactions of 37 genes with 220 drugs belonging to 54 second-level Anatomical Therapeutic Chemical (ATC) classes (**Figure 5**).

These findings implicate that some of the genes may be targeted by the currently available pharmaceutical treatments. Most drug-gene interactions (N = 47) were observed between psycholeptics (ATC: N05, including antipsychotics and anxiolytics) and the dopamine receptor D2 (*DRD2*). Dopamine is a neurotransmitter in the brain and the dopaminergic system has previously been implicated in depression and is a target for antidepressant medication[122].

The findings of genetic studies provide novel knowledge and understanding of the underlying genetic architecture and provide information on the biological mechanisms involved. However, the pathophysiology of depression remains elusive. In order to provide an earlier and more precise diagnosis of depression and to improve treatment and potentially prevention, further research into the molecular mechanisms underlying depression is needed. Preclinical and clinical studies have provided evidence for changes in intracellular signaling, gene expression, neuronal function, and cellular architecture within the brain regions that control cognition, mood, and executive functions[123]. Given the strong link between depression and lipid dysregulations in the circulation[124], there is also increasing interest to investigate the relationship between depression and metabolic changes in blood.



ATC classifications: A01: stomatological preparations; A02: drugs for acid-related disorders; A03: drugs for functional gastrointestinal disorders; A05: bile and liver therapy; A06: drugs for constipation; A10: drugs used in diabetes; A11: vitamins; A12: mineral supplements; A14: anabolic agents for systemic use; B05: blood substitutes and perfusion solutions; B06: other hematological agents; C01: cardiac therapy; C03: diuretics; C04: peripheral vasodilators; C05: vasoprotectives; C07: beta-blocking agents; C08: calcium channel blockers; D07: corticosteroids, dermatological preparations; D10: anti-acne preparations; D11: other dermatological preparations; G02: other gynecologicals; G03: sex hormones and modulators of the genital system; G04: urologicals; H01: pituitary and hypothalamic hormones and analogs; H02: corticosteroids for systemic use; J01: antibacterials for systemic use; J05: antivirals for systemic use; J07: vaccines; L01: antineoplastic agents; L02: endocrine therapy; L03: immunostimulants; L04: immunosuppressants; M01: anti-inflammatory and antirheumatic products; M03: muscle relaxants; M04: antigout preparations; M05: drugs for treatment of bone diseases; N01: anesthetics; N02: analgesics; N03: anti-epileptics; N04: anti-Parkinson drugs; N05: psycholeptics; N06: psychoanaleptics; N07: other nervous system drugs; P01: antiprotozoals; P03: ectoparasiticides, including scabicides, insecticides, and repellents; R01: nasal preparations; R03: drugs for obstructive airway diseases; R06: antihistamines for systemic use; S01: ophthalmologicals; S02: otologicals; S03: ophthalmological and otological preparations; V03: all other therapeutic products; V04: diagnostic agents; V09: diagnostic radiopharmaceuticals.

Figure 5. Chord diagram representing the genes significantly associated with depression ($p < 2.80 \times 10^{-6}$) and the second level ATC classifications to which interacting drugs have been assigned. The lines are wider if more drugs interact with the gene and the most significant genes are located at the top of the diagram. Source: Howard et al.[110]

A powerful approach to understand metabolic changes is through the investigation of metabolites in blood, as these represent the direct downstream effects of cellular metabolism and are closely related to the disease[125]. A previous study has investigated the relationship between metabolites and depression using the Nightingale Health platform (Nightingale Health Ltd., Helsinki, Finland), as part of the Biobanking for Medical Research

Infrastructure of the Netherlands (BBMRI-NL) consortium. In this study, a specific lipid profile was associated with depression[126]. This profile showed that higher levels of apolipoprotein B, very-low-density lipoprotein cholesterol, diglycerides, triglycerides, total and monounsaturated fatty acids, fatty acid chain length, glycoprotein acetyls, isoleucine, tyrosine, and lower levels of apolipoprotein A1, high-density lipoprotein cholesterol, and acetate were associated with depression (**Figure 6**). These findings indicate that metabolomics studies may provide further insights into biological pathways implicated in depression pathophysiology, and the results may be used for the development of new therapies.

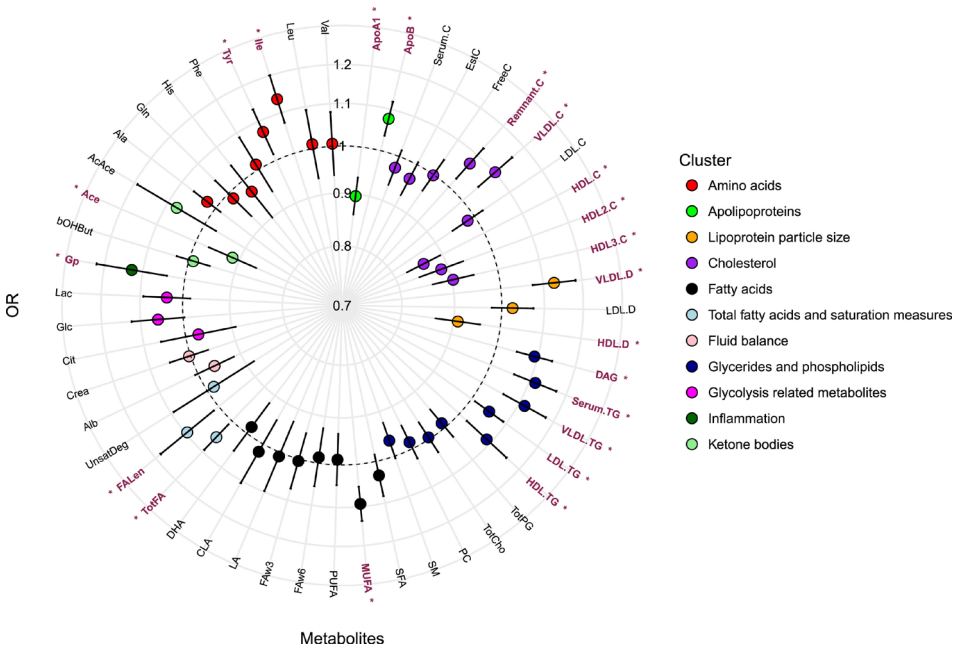


Figure 6. Polar plot showing pooled odds ratios (OR) and 95% confidence intervals for the association of the 51 lipids, fatty acids and various low-molecular-weight metabolites with depression. *Significant associations (FDR < 0.05). Dotted circle indicates an OR of 1.
Abbreviations: HDL2 = high-density lipoprotein (HDL) subfraction 2 (1.063–1.125 g/mL), HDL3 = high-density lipoprotein subfraction 3 (1.125–1.210 g/mL), AcAce = acetoacetate, Ace = acetate, Ala = alanine, Alb = albumin, ApoA1 = apolipoprotein A-I, ApoB = apolipoprotein B, bOHBut = 3-hydroxybutyrate, C = cholesterol, Cit = citrate, CLA = conjugated linoleic acids, Crea = creatinine, D = mean diameter, DAG = diglycerides, DHA = docosahexaenoic acid, Est = esterified, FA = fatty acids, FALen = estimated fatty acids chain length, FAw3 = ω-3 fatty acids, FAw6 = ω-6 fatty acids, Glc = glucose, Gln = glutamine, Gp = glycoprotein acetyls (mainly α1-acid glycoprotein), His = histidine, IDL = intermediate-density lipoprotein, Ile = isoleucine, LA = linoleic acid (18:2), Lac = lactate, Leu = leucine, LDL = low-density lipoprotein, MUFA = monounsaturated fatty acids (16:1, 18:1), PC = phosphatidylcholine and other cholines, Phe = phenylalanine, PUFA = polyunsaturated fatty acids, Remnant = non-HDL, non-LDL cholesterol, SFA = saturated fatty acids, SM = sphingomyelins, TG = triglycerides, TotCho = total cholines, TotFA = total fatty acids, TotPG = total phosphoglycerides, Tyr = tyrosine, UnsatDeg = estimated degree of unsaturation, Val = valine, VLDL = very-low-density lipoprotein. Source: Bot et al. 2020[126].

In this thesis, I aimed to study circulating metabolic determinants of depression. A recent study by the BBMRI-NL consortium also provided evidence for the major impact of the usage of drugs on the metabolism, where the research groups constructed a drug-metabolite atlas of associations of commonly prescribed drugs and clinically relevant metabolites[127]. Interestingly, this study showed that one of the 15 drugs that were associated with the largest number of metabolites were SSRIs[127]. Therefore, I also explored the effect of antidepressants on the association of metabolites with depression.

Sleep and associated outcomes and disorders in the elderly

Sleep is a complex and essential biological process that has been conserved across diverse animal species throughout evolution[128]. Although normal healthy sleep highly varies within and between adults[129-131], it has to consist of adequate duration, good quality, proper timing and regularity, whereas the absence of sleep leads to disturbances or disorders[132]. The ability to fall asleep and to stay asleep throughout the night has been shown to reduce with age[133,134]. Taking into account the classic sleep staging criteria[135], there are well-characterized changes in sleep stages of older adults compared to younger adults (**Figure 7**).

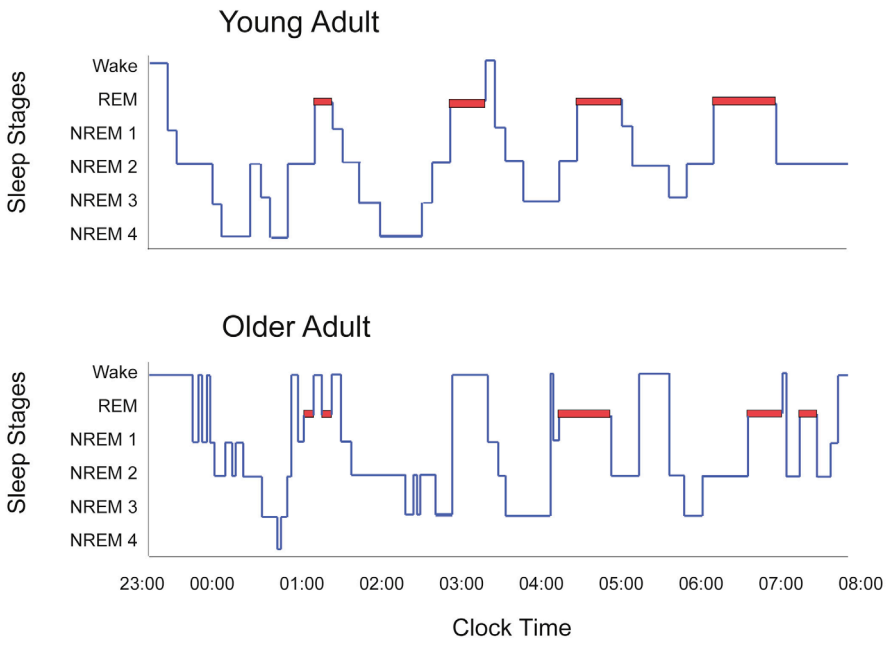


Figure 7. chematic overview of age-related changes in the different sleep stages. The top graph shows a 9 hour sleep period of a younger adult, while the lower graph shows this period in an older adult. Figure adjusted from Mander et al.[136]

For example, older adults show advanced sleep timing (i.e., earlier bedtimes and rise times), longer sleep latency (i.e., time it takes to fall asleep), shorter overall sleep duration, increased sleep fragmentation, and a reduced amount of deeper non-rapid eye movement (NREM) sleep[136]. An epidemiological study showed that more than half of the adults aged 65 years or older reported to have sleep-related complaints most of the time[137]. To obtain more insights in the molecular mechanisms underlying sleep, I performed genetic studies to explore the genetic architecture and biological pathways involved in two sleep-related outcomes, sleep latency and sleep apnea.

Sleep latency

The circadian rhythm is the natural process that regulates the sleep-wake cycle and is repeated every 24 hours[138]. This cycle is coordinated by a group of hypothalamic nerve cells, the suprachiasmatic nucleus (SCN), which functions as the brain pacemaker for the timing of the sleep-wake cycle[138-140]. Sleep latency is defined as the time of falling asleep (from being fully awake to being fully asleep). There are different methods to assess sleep latency including self-reports, actigraphy, and polysomnography. Subjectively reported sleep latency estimates of normal sleepers have been shown to correlate well with polysomnographic measures[141,142]. Sleep latency is different between males and females[143,144], and across ages[143]. Other risk factors for sleep latency are genetic factors as well as environmental and lifestyle factors such as consumption of stimulants[145], dietary intake[146], sedentary life[147] and diseases including depression[148]. Longer sleep latency is a measure of poor sleep quality[149] which can have serious health consequences, such as increased stress responsivity, pain, reduced quality of life, emotional distress and mood disorders, decreased performance and cognitive impairment. On the long run poor sleep quality can lead to the development of diseases[150]. While the environmental and lifestyle factors associated with sleep latency have been studied extensively, the genetic component received less attention, even though the heritability of sleep latency is moderate as it is estimated to be between 17-44%[151-153]. Thus, in this thesis, I aim to identify genetic determinants of sleep latency.

Sleep apnea

Obstructive sleep apnea (OSA) is a common sleep breathing disorder[154], which affects up to 38% of the general population[155] and is one of the most common causes of disturbed sleep. Two common symptoms of OSA are snoring and excessive daytime sleepiness[156]. Excessive daytime sleepiness can cause personal and job-related problems and can also cause an increased risk of traffic and work-related accidents[157-159]. If untreated, OSA is associated with serious health problems[157] such as hypertension[160,161], cardiovascular disease[162-164], stroke[165], type 2 diabetes[166-168], impaired cognitive function[169,170], depression[171], and increased mortality[172-174]. The main characteristic of OSA is the partial or complete obstruction of the upper airways during sleep, causing oxyhemoglobin desaturations and arousals from sleep leading to sleep fragmentation and decreased periods of slow wave and rapid eye movement (REM) sleep[158,159].

OSA is a complex disease that is influenced by genetic[175,176] and environmental factors, of which age, sex, and obesity are the most important risk factors[154,157,177-180]. The heritability of OSA has been estimated to be around 40%[176] and more than half of the genetic contribution to sleep apnea acts through mechanisms independent of obesity[181]. Previous candidate gene studies have investigated the association of multiple genes with breathing disorders, such as *ACE*[182-184], *TNF*[185-187], *APOE*-ε4[188,189], serotonin receptors and transporters genes: *5-HT2A*, *5-HT2C*, *5-HTT*[190-194], and adrenergic receptors genes: *ADRB2* and *ADRB3*[195,196]. However, the results have been conflicting or not validated[197]. Linkage analysis, a genetic method to identify the genomic location of the disease influencing genes, provided suggestive evidence for the linkage of OSA with two regions on chromosomes 2p16 and 19q13, independent of obesity[198]. Further, only a few GWASs or analysis with customized or targeted genotyping arrays have been performed, which found significant associations of loci in *GPR83*[199], *C6ORF183/CCD162P*[199], *PPARGC1B*[200], *PTGER3*[201], *PLEK*[201], and *LPAR1*[201] with OSA, albeit without replication. As the genetic determinants of OSA remain largely unknown, I aim to identify rare genetic variants associated with symptoms of sleep apnea.

Molecular approaches included in this thesis

Studying the alterations at the molecular level can improve our understanding of the biological mechanisms underlying diseases and associated outcomes (**Figure 8**). There are multiple "omics" approaches that involve various molecular levels. Genetic research focuses on the investigation of variation in the DNA base pair sequence in (parts of) genes between individuals, the transmission of this variation between generations, and how this variation behaves in populations. Genomics involves also down-stream effects of the genome (i.e., the complete DNA sequence, the methylation and transcription) and involves studying the relation between genes and their influence on organisms. DNA is transcribed into ribonucleic acid (RNA) and the process of transcription may be regulated by for instance methylation. RNA is subsequently translated into proteins (proteomics) but may also be involved in the regulation of other genes. Finally, proteins are catabolized into metabolites, which are studied in metabolomics studies. Thus, metabolites represent the final product of all the upstream processes in a hierarchy for which genomics forms the basis (**Figure 8**). Additionally, **Figure 8** shows that the omics layers are affected by environmental factors e.g., stress, life style factors and medication use and hence are dynamic and also tissue specific. In the projects included in this thesis, I have used genomics and metabolomics to study the association of genetic variants and metabolites with aging-related processes, pathways, and associated (disease) outcomes.

Genetics

All genetic information of an individual is stored in the human genome, which consists of four nucleotides: Adenine (A), Cytosine (C), Guanine (G), and Thymine (T). Together, they form base pairs, where A binds with T, and C with G. The human genome consists of approximately 3.3 billion base pairs and over 20,000 protein-coding genes. Humans have

different DNA variations and their genomes may differ at millions of places[203], of which the count increases as more and more genomes are being studied. There are multiple types of DNA variations, but the most observed in the general population are genetic variants such as single nucleotide variants (SNVs), which are single base pair differences between individuals of the same population. SNVs are called single nucleotide polymorphisms (SNPs) or common variants if they occur in >1% of the population, and rare variants if they occur in <1%. Other types of DNA variations also occur, however, this thesis focuses specifically on SNPs and rare variants.

Different techniques are used to study common and rare genetic variants. During the past 12 years, the focus of genetic studies has been on the identification of common genetic variants for various diseases in the population using GWAS. These studies aim at a better understanding of the biology of disease, which may lead to better diagnosis, treatment, or prevention[203]. To gather data for a GWAS, SNP arrays assess a dense set of SNPs across the genome. The idea is that these SNPs will tag the neighboring variation because of high correlation due to the physical proximity. Since these arrays do not assess the whole genome, i.e., > 3 billion base pairs, a whole lot of information is missing, specifically the information on the rare genetic variants that are poorly tagged by common variants. Part of this lost information, mostly on SNPs, can be recovered by imputation, a technique using the correlation between SNPs in the genome[204-206]. The genotypes of the unmeasured variants can be predicted by comparing them to fully known reference panels, such as the 1000 Genomes reference panel[203] or the Haplotype Reference Consortium (HRC) panel[207].

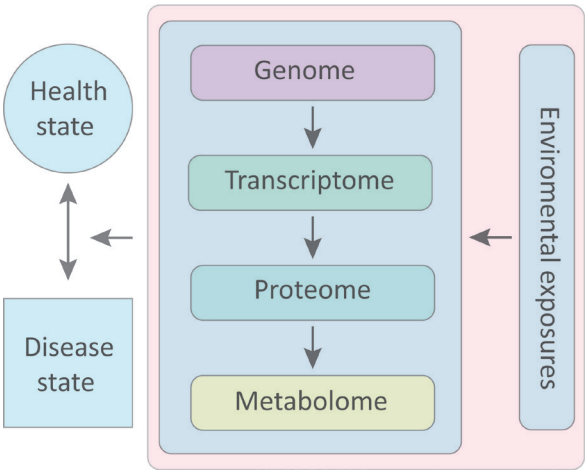


Figure 8. Interaction between genomics, transcriptomics, proteomics, and metabolomics as well as the impact of the environment, including microbiota and drugs, and how the end products can impact health and disease states. Figure adjusted from Kaddurah-Daouk & Krishnan[202]

A major advantage of GWAS is the hypothesis-free testing, i.e., no prior information about the trait or disease is assumed in the analysis. Additionally, combining GWAS results of multiple cohorts in a meta-analysis increases the power of the statistical analysis and allows replication of findings across studies, possibly leading to new discoveries as shown for multiple complex diseases[208,209].

However, the SNPs that are identified by GWAS often explain a relatively small proportion of the heritability. The "missing heritability"[210] might be explained by rare variants, which can have larger effects on the trait or disease[211]. Rare variants are not well covered by the SNP arrays and imputation panels. Therefore, these variants are better studied with next-generation sequencing techniques, including whole-exome sequencing and whole-genome sequencing[212]. Previous studies have shown that some complex diseases, such as autism spectrum disorders, are better explained by rare variants[213]. Rare variants have also been identified to be associated with extreme longevity[108], and with age-related diseases such as cardiovascular disease and Alzheimer's disease[214-216]. Family-based studies provide a good setting for the discovery of rare variants. The basic rationale is that variants which are rare in the general population, are more frequently found in families as they are inherited from one generation to another. I have used a family-based study to investigate the association of rare variants with telomere length and with symptoms of sleep apnea in **Chapters 2.1** and **3.3**, respectively.

Metabolomics

Metabolites are small molecules present in tissues and body fluids. All metabolites in an organism collectively constitute the metabolome. Metabolites are products of cellular metabolism and physiological processes that are influenced by disease mechanisms, environmental and lifestyle factors as well as pharmacological treatments[217]. As they are the direct quantifiable effects of biochemical activity, this provides opportunities to study their association with phenotypes or diseases[125]. Therefore, metabolomics is a powerful technique to find disease biomarkers that can have a diagnostic, prognostic or predictive role and also monitor the efficacy of a treatment[218,219].

Metabolites can be studied using a targeted or untargeted approach[220]. Targeted metabolomics has the advantage of studying specific pathways of interest, while the main advantage of untargeted metabolomics is that as many metabolites as possible are quantified without bias[220]. Metabolic profiles have been identified for multiple age-related diseases, such as type 2 diabetes[69], Alzheimer's disease[72,221], and cardiovascular disease[222], but also for all-cause mortality[64,65]. A recent study identified a metabolic profile associated with all-cause mortality, which predicted long term mortality risk (5-10 years) more accurately than conventional risk factors for mortality[65]. These findings prove that metabolomics is a promising field in aging research and metabolic profiles associated with age-related outcomes should be explored further in epidemiological studies with large sample sizes.

In this thesis, I have used both targeted as well as untargeted metabolomics. I studied the association of metabolites, using a targeted approach, with telomere length in **Chapter 2**. In **Chapter 2.2**, metabolites were measured with the AbsoluteIDQTM p150 kit from BIOCRATES Life Sciences AG (Innsbruck, Austria), while in **Chapter 2.3** metabolites were measured with the Nightingale Health Ltd. (Helsinki, Finland) platform. In **Chapter 3.1**, I have used an untargeted metabolomics approach to study the association of metabolites with depression using the Metabolon Inc. (Durham, USA) platform.

Study populations

The primary cohorts used in the studies described in this thesis include the family-based Erasmus Rucphen Family (ERF) study, and the population-based Rotterdam Study (RS). The ERF study consists of approximately 3000 inhabitants of a genetically isolated community in the southwest of the Netherlands, and was ascertained as part of the Genetic Research in Isolated Populations (GRIP) program[223,224]. The community was founded in the middle of the 18th century and has remained in isolation with minimal immigration[223,224]. All study participants are descendants (or their spouses) of 22 founder couples that lived in this area in the 19th century[223,224]. All participants were 18 years or older when included in the study and baseline data was collected between 2002 and 2005.

The RS is a prospective cohort study that aims to investigate the determinants of diseases that occur in older adults, with the focus on cardiovascular, endocrine, hepatic, neurological, ophthalmic, psychiatric, dermatological, otolaryngological, locomotor, and respiratory diseases[225]. The first RS cohort (RS-I) started in 1990 and comprised 7,983 participants of 55 years or older, living in the Ommoord area in the city of Rotterdam. The RS was extended with a second (RS-II) and third (RS-III) cohort in 2000 and 2006, respectively, which included new participants of 45 years or older. As of 2008, 14,926 participants were included. However, the RS is now expanding, as a fourth cohort was started in 2016, including individuals aged 40 years or over (RS-IV). All participants had examinations at baseline, which included examinations at the research facility and through an at home interview. The examinations were repeated every 3-4 years for characteristics that could change over time. Both the ERF study and the RS were approved by the Medical Ethics Committee of the Erasmus MC and by the Ministry of Health, Welfare and Sport of the Netherlands. All participants provided written informed consent to participate in the study and all investigations were carried out in accordance with the Declaration of Helsinki.

In addition to ERF and RS, most studies described in this thesis are a result of collaboration with cohorts that are a part of various consortia, namely, the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE), the European Network for Genetic and Genomic Epidemiology (ENGAGE) consortium, and the BBMRI-NL consortium.

Scope of this thesis

The overall aim of this thesis is to improve the understanding of the molecular mechanisms involved in aging and longevity and associated outcomes by using multiple omics approach and data from family- and population-based cohorts.

In the studies included in **Chapter 2** of this thesis, I have investigated biological processes and pathways influencing age-related outcomes, including multiple studies investigating telomere length biology and the role of the IIS pathway in dementia and Alzheimer's disease. **Chapter 2.1** describes a whole-exome sequencing analysis aiming to identify rare genetic variants that are associated with telomere length. **Chapter 2.2** searches for metabolites associated with telomere length by performing a meta-analysis including seven different cohorts that are part of the ENGAGE consortium. A second study investigating the association of metabolites with telomere length was performed in **Chapter 2.3** using a different platform for metabolite quantification and data from six Dutch studies which are part of the BBMRI-NL consortium**. Chapter 2.4** focuses on investigating the robustness of the association of IGF-I bioactivity with dementia and Alzheimer's disease by extending the follow-up period in the Rotterdam Study and focuses on studying the possible effect modification by the *APOE* gene. **Chapter 3** focuses on genetic and metabolic determinants of late onset depression and sleep-related outcomes and diseases. **Chapter 3.1** describes a large meta-analysis, including data from five cohorts, aiming to identify metabolites associated with depression. Further, the aim of **Chapter 3.2** was to identify common genetic variants associated with sleep latency in a GWAS using data from seven cohorts. The association of rare genetic variants with sleep apnea, an age-related sleep breathing disorder, was studied in **Chapter 3.3. Chapter 4** includes the general discussion of all research included in this thesis and provides suggestions for future research. **Chapter 5** provides a summary of the main findings of this thesis.

References

1 Oeppen, J. & Vaupel, J. W. Demography. Broken limits to life expectancy. *Science* **296**, 1029-1031, doi:10.1126/science.1069675 (2002).
2 Organization, W. H. *Ageing and health*, <https://www.who.int/news-room/fact-sheets/detail/ageing-and-health> (2018).
3 Crimmins, E. M. Lifespan and Healthspan: Past, Present, and Promise. *Gerontologist* **55**, 901-911, doi:10.1093/geront/gnv130 (2015).
4 Jagger, C. *et al.* Inequalities in healthy life years in the 25 countries of the European Union in 2005: a cross-national meta-regression analysis. *Lancet* **372**, 2124-2131, doi:10.1016/S0140-6736(08)61594-9 (2008).
5 Niccoli, T. & Partridge, L. Ageing as a risk factor for disease. *Curr Biol* **22**, R741-752, doi:10.1016/j.cub.2012.07.024 (2012).
6 Partridge, L., Deelen, J. & Slagboom, P. E. Facing up to the global challenges of ageing. *Nature* **561**, 45-56, doi:10.1038/s41586-018-0457-8 (2018).
7 Campisi, J. *et al.* From discoveries in ageing research to therapeutics for healthy ageing. *Nature* **571**, 183-192, doi:10.1038/s41586-019-1365-2 (2019).
8 Blackburn, E. H. Structure and function of telomeres. *Nature* **350**, 569-573, doi:10.1038/350569a0 (1991).
9 Muller, H. The remaking of chromosomes. *The Collecting Net*, 181-198 (1938).
10 McClintock, B. The Fusion of Broken Ends of Chromosomes Following Nuclear Fusion. *Proc Natl Acad Sci U S A* **28**, 458-463, doi:10.1073/pnas.28.11.458 (1942).
11 de Lange, T. Shelterin-Mediated Telomere Protection. *Annu Rev Genet* **52**, 223-247, doi:10.1146/annurev-genet-032918-021921 (2018).
12 de Lange, T. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev* **19**, 2100-2110, doi:10.1101/gad.1346005 (2005).
13 Griffith, J. D. *et al.* Mammalian telomeres end in a large duplex loop. *Cell* **97**, 503-514, doi:10.1016/s0092-8674(00)80760-6 (1999).
14 Blackburn, E. H. Switching and signaling at the telomere. *Cell* **106**, 661-673, doi:10.1016/S0092-8674(01)00492-5 (2001).
15 Shay, J. W. & Wright, W. E. Telomeres and telomerase: three decades of progress. *Nat Rev Genet* **20**, 299-309, doi:10.1038/s41576-019-0099-1 (2019).
16 Watson, J. D. Origin of concatemeric T7 DNA. *Nat New Biol* **239**, 197-201 (1972).
17 Blackburn, E. H., Greider, C. W. & Szostak, J. W. Telomeres and telomerase: the path from maize, Tetrahymena and yeast to human cancer and aging. *Nat Med* **12**, 1133-1138, doi:10.1038/nm1006-1133 (2006).
18 Allsopp, R. C. *et al.* Telomere length predicts replicative capacity of human fibroblasts. *Proc Natl Acad Sci U S A* **89**, 10114-10118 (1992).
19 Blackburn, E. H. & Collins, K. Telomerase: an RNP enzyme synthesizes DNA. *Cold Spring Harb Perspect Biol* **3**, doi:10.1101/cshperspect.a003558 (2011).
20 Harley, C. B. Telomere loss: mitotic clock or genetic time bomb? *Mutat Res* **256**, 271-282, doi:10.1016/0921-8734(91)90018-7 (1991).
21 Brouilette, S., Singh, R. K., Thompson, J. R., Goodall, A. H. & Samani, N. J. White cell telomere length and risk of premature myocardial infarction. *Arterioscler Thromb Vasc Biol* **23**, 842-846, doi:10.1161/01.ATV.0000067426.96344.32 (2003).
22 Brouilette, S. W. *et al.* Telomere length, risk of coronary heart disease, and statin treatment in the West of Scotland Primary Prevention Study: a nested case-control study. *Lancet* **369**, 107-114, doi:10.1016/S0140-6736(07)60071-3 (2007).
23 Fitzpatrick, A. L. *et al.* Leukocyte telomere length and cardiovascular disease in the cardiovascular health study. *Am J Epidemiol* **165**, 14-21, doi:10.1093/aje/kwj346 (2007).
24 Benetos, A. *et al.* Short telomeres are associated with increased carotid atherosclerosis in hypertensive subjects. *Hypertension* **43**, 182-185, doi:10.1161/01.HYP.0000113081.42868.f4 (2004).
25 Jeanclos, E. *et al.* Shortened telomere length in white blood cells of patients with IDDM. *Diabetes* **47**, 482-486 (1998).
26 Zhao, J., Miao, K., Wang, H., Ding, H. & Wang, D. W. Association between telomere length and type 2 diabetes mellitus: a meta-analysis. *PLoS One* **8**, e79993, doi:10.1371/journal.pone.0079993 (2013).

27 Martin-Ruiz, C. *et al.* Telomere length predicts poststroke mortality, dementia, and cognitive decline. *Ann Neurol* **60**, 174-180, doi:10.1002/ana.20869 (2006).

28 Grodstein, F. *et al.* Shorter telomeres may mark early risk of dementia: preliminary analysis of 62 participants from the nurses' health study. *PLoS One* **3**, e1590, doi:10.1371/journal.pone.0001590 (2008).

29 Honig, L. S., Kang, M. S., Schupf, N., Lee, J. H. & Mayeux, R. Association of shorter leukocyte telomere repeat length with dementia and mortality. *Arch Neurol* **69**, 1332-1339, doi:10.1001/archneurol.2012.1541 (2012).

30 Cawthon, R. M., Smith, K. R., O'Brien, E., Sivatchenko, A. & Kerber, R. A. Association between telomere length in blood and mortality in people aged 60 years or older. *Lancet* **361**, 393-395, doi:10.1016/S0140-6736(03)12384-7 (2003).

31 Kimura, M. *et al.* Telomere length and mortality: a study of leukocytes in elderly Danish twins. *Am J Epidemiol* **167**, 799-806, doi:10.1093/aje/kwm380 (2008).

32 Fitzpatrick, A. L. *et al.* Leukocyte telomere length and mortality in the Cardiovascular Health Study. *J Gerontol A Biol Sci Med Sci* **66**, 421-429, doi:10.1093/gerona/glq224 (2011).

33 Deelen, J. *et al.* Leukocyte telomere length associates with prospective mortality independent of immune-related parameters and known genetic markers. *Int J Epidemiol* **43**, 878-886, doi:10.1093/ije/dyt267 (2014).

34 Rode, L., Nordestgaard, B. G. & Bojesen, S. E. Peripheral blood leukocyte telomere length and mortality among 64,637 individuals from the general population. *J Natl Cancer Inst* **107**, djv074, doi:10.1093/jnci/djv074 (2015).

35 Marioni, R. E. *et al.* The epigenetic clock and telomere length are independently associated with chronological age and mortality. *Int J Epidemiol*, doi:10.1093/ije/dyw041 (2016).

36 Martin-Ruiz, C. M., Gussekloo, J., van Heemst, D., von Zglinicki, T. & Westendorp, R. G. Telomere length in white blood cells is not associated with morbidity or mortality in the oldest old: a population-based study. *Aging Cell* **4**, 287-290, doi:10.1111/j.1474-9726.2005.00171.x (2005).

37 Harris, S. E. *et al.* The association between telomere length, physical health, cognitive ageing, and mortality in non-demented older people. *Neurosci Lett* **406**, 260-264, doi:10.1016/j.neulet.2006.07.055 (2006).

38 Njajou, O. T. *et al.* Association between telomere length, specific causes of death, and years of healthy life in health, aging, and body composition, a population-based cohort study. *J Gerontol A Biol Sci Med Sci* **64**, 860-864, doi:10.1093/gerona/glp061 (2009).

39 Houben, J. M., Giltay, E. J., Rius-Ottenheim, N., Hageman, G. J. & Kromhout, D. Telomere length and mortality in elderly men: the Zutphen Elderly Study. *J Gerontol A Biol Sci Med Sci* **66**, 38-44, doi:10.1093/gerona/glq164 (2011).

40 Strandberg, T. E. *et al.* Association of telomere length in older men with mortality and midlife body mass index and smoking. *J Gerontol A Biol Sci Med Sci* **66**, 815-820, doi:10.1093/gerona/glr064 (2011).

41 Needham, B. L. *et al.* Leukocyte telomere length and mortality in the National Health and Nutrition Examination Survey, 1999-2002. *Epidemiology* **26**, 528-535, doi:10.1097/EDE.0000000000000299 (2015).

42 Hastie, N. D. *et al.* Telomere reduction in human colorectal carcinoma and with ageing. *Nature* **346**, 866-868, doi:10.1038/346866a0 (1990).

43 Willeit, P., Willeit, J., Kloss-Brandstatter, A., Kronenberg, F. & Kiechl, S. Fifteen-year follow-up of association between telomere length and incident cancer and cancer mortality. *JAMA* **306**, 42-44, doi:10.1001/jama.2011.901 (2011).

44 Willeit, P. *et al.* Telomere length and risk of incident cancer and cancer mortality. *JAMA* **304**, 69-75, doi:10.1001/jama.2010.897 (2010).

45 Bojesen, S. E. *et al.* Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat Genet* **45**, 371-384, 384e371-372, doi:10.1038/ng.2566 (2013).

46 Bekaert, S. *et al.* Telomere length and cardiovascular risk factors in a middle-aged population free of overt cardiovascular disease. *Aging Cell* **6**, 639-647, doi:10.1111/j.1474-9726.2007.00321.x (2007).

47 Lin, J., Epel, E. & Blackburn, E. Telomeres and lifestyle factors: roles in cellular aging. *Mutat Res* **730**, 85-89, doi:10.1016/j.mrfmmm.2011.08.003 (2012).

48 Epel, E. S. *et al.* Accelerated telomere shortening in response to life stress. *Proc Natl Acad Sci U S A* **101**, 17312-17315, doi:10.1073/pnas.0407162101 (2004).

49 Samani, N. J. & van der Harst, P. Biological ageing and cardiovascular disease. *Heart* **94**, 537-539, doi:10.1136/hrt.2007.136010 (2008).

50 Slagboom, P. E., Droog, S. & Boomsma, D. I. Genetic determination of telomere size in humans: a twin study of three age groups. *Am J Hum Genet* **55**, 876-882 (1994).

51 Broer, L. *et al.* Meta-analysis of telomere length in 19,713 subjects reveals high heritability, stronger maternal inheritance and a paternal age effect. *Eur J Hum Genet* **21**, 1163-1168, doi:10.1038/ejhg.2012.303 (2013).

52 Vasa-Nicotera, M. *et al.* Mapping of a major locus that determines telomere length in humans. *Am J Hum Genet* **76**, 147-151, doi:10.1086/426734 (2005).

53 Wilson, W. R. *et al.* Blood leucocyte telomere DNA content predicts vascular telomere DNA content in humans with and without vascular disease. *Eur Heart J* **29**, 2689-2694, doi:10.1093/eurheartj/ehn386 (2008).

54 Okuda, K. *et al.* Telomere length in the newborn. *Pediatr Res* **52**, 377-381, doi:10.1203/00006450-200209000-00012 (2002).

55 Codd, V. *et al.* Common variants near TERC are associated with mean telomere length. *Nat Genet* **42**, 197-199, doi:10.1038/ng.532 (2010).

56 Codd, V. *et al.* Identification of seven loci affecting mean telomere length and their association with disease. *Nat Genet* **45**, 422-427, 427e421-422, doi:10.1038/ng.2528 (2013).

57 Levy, D. *et al.* Genome-wide association identifies OBFC1 as a locus involved in human leukocyte telomere biology. *Proc Natl Acad Sci U S A* **107**, 9293-9298, doi:10.1073/pnas.0911494107 (2010).

58 Mangino, M. *et al.* Genome-wide meta-analysis points to CTC1 and ZNF676 as genes regulating telomere homeostasis in humans. *Hum Mol Genet* **21**, 5385-5394, doi:10.1093/hmg/dds382 (2012).

59 Mangino, M. *et al.* DCAF4, a novel gene associated with leucocyte telomere length. *J Med Genet* **52**, 157-162, doi:10.1136/jmedgenet-2014-102681 (2015).

60 Pooley, K. A. *et al.* A genome-wide association scan (GWAS) for mean telomere length within the COGS project: identified loci show little association with hormone-related cancer risk. *Hum Mol Genet* **22**, 5056-5064, doi:10.1093/hmg/ddt355 (2013).

61 Li, C. *et al.* Genome-wide Association Analysis in Humans Links Nucleotide Metabolism to Leukocyte Telomere Length. *Am J Hum Genet* **106**, 389-404, doi:10.1016/j.ajhg.2020.02.006 (2020).

62 Delgado, D. A. *et al.* Genome-wide association study of telomere length among South Asians identifies a second RTEL1 association signal. *J Med Genet* **55**, 64-71, doi:10.1136/jmedgenet-2017-104922 (2018).

63 Dorajoo, R. *et al.* Loci for human leukocyte telomere length in the Singaporean Chinese population and trans-ethnic genetic studies. *Nat Commun* **10**, 2491, doi:10.1038/s41467-019-10443-2 (2019).

64 Fischer, K. *et al.* Biomarker profiling by nuclear magnetic resonance spectroscopy for the prediction of all-cause mortality: an observational study of 17,345 persons. *PLoS Med* **11**, e1001606, doi:10.1371/journal.pmed.1001606 (2014).

65 Deelen, J. *et al.* A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals. *Nature Communications* **10**, 3346, doi:10.1038/s41467-019-11311-9 (2019).

66 Matsumoto, T., Kobayashi, T. & Kamata, K. Role of lysophosphatidylcholine (LPC) in atherosclerosis. *Curr Med Chem* **14**, 3209-3220 (2007).

67 Vojinovic, D. *et al.* Metabolic profiling of intra- and extracranial carotid artery atherosclerosis. *Atherosclerosis* **272**, 60-65, doi:10.1016/j.atherosclerosis.2018.03.015 (2018).

68 Wang-Sattler, R. *et al.* Novel biomarkers for pre-diabetes identified by metabolomics. *Mol Syst Biol* **8**, 615, doi:10.1038/msb.2012.43 (2012).

69 Liu, J. *et al.* Metabolomics based markers predict type 2 diabetes in a 14-year follow-up study. *Metabolomics* **13**, 104, doi:10.1007/s11306-017-1239-2 (2017).

70 Suhre, K. *et al.* Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. *PLoS One* **5**, e13953, doi:10.1371/journal.pone.0013953 (2010).

71 Floegel, A. *et al.* Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. *Diabetes* **62**, 639-648, doi:10.2337/db12-0495 (2013).

72 van der Lee, S. J. *et al.* Circulating metabolites and general cognitive ability and dementia: Evidence from 11 cohort studies. *Alzheimers Dement* **14**, 707-722, doi:10.1016/j.jalz.2017.11.012 (2018).

73 Oresic, M. *et al.* Metabolome in progression to Alzheimer's disease. *Transl Psychiatry* **1**, e57, doi:10.1038/tp.2011.55 (2011).

74 Menni, C. *et al.* Metabolomic markers reveal novel pathways of ageing and early development in human populations. *Int J Epidemiol* **42**, 1111-1119, doi:10.1093/ije/dyt094 (2013).

75 Zhao, J. *et al.* Metabolic profiles of biological aging in American Indians: the Strong Heart Family Study. *Aging (Albany NY)* **6**, 176-186, doi:10.18632/aging.100644 (2014).

76 Zierer, J. *et al.* Metabolomics profiling reveals novel markers for leukocyte telomere length. *Aging (Albany NY)* **8**, 77-94, doi:10.18632/aging.100874 (2016).

77 Murabito, J. M., Yuan, R. & Lunetta, K. L. The search for longevity and healthy aging genes: insights from epidemiological studies and samples of long-lived individuals. *J Gerontol A Biol Sci Med Sci* **67**, 470-479, doi:10.1093/gerona/gls089 (2012).

78 Sebastiani, P. *et al.* Limitations and risks of meta-analyses of longevity studies. *Mech Ageing Dev* **165**, 139-146, doi:10.1016/j.mad.2017.01.008 (2017).

79 van den Berg, N., Beekman, M., Smith, K. R., Janssens, A. & Slagboom, P. E. Historical demography and longevity genetics: Back to the future. *Ageing Res Rev* **38**, 28-39, doi:10.1016/j.arr.2017.06.005 (2017).

80 Sebastiani, P. & Perls, T. T. The genetics of extreme longevity: lessons from the new England centenarian study. *Front Genet* **3**, 277, doi:10.3389/fgene.2012.00277 (2012).

81 Soerensen, M. *et al.* Replication of an association of variation in the FOXO3A gene with human longevity using both case-control and longitudinal data. *Aging Cell* **9**, 1010-1017, doi:10.1111/j.1474-9726.2010.00627.x (2010).

82 Willcox, B. J. *et al.* FOXO3A genotype is strongly associated with human longevity. *Proc Natl Acad Sci U S A* **105**, 13987-13992, doi:10.1073/pnas.0801030105 (2008).

83 Broer, L. *et al.* GWAS of longevity in CHARGE consortium confirms APOE and FOXO3 candidacy. *J Gerontol A Biol Sci Med Sci* **70**, 110-118, doi:10.1093/gerona/glu166 (2015).

84 Flachsbart, F. *et al.* Identification and characterization of two functional variants in the human longevity gene FOXO3. *Nat Commun* **8**, 2063, doi:10.1038/s41467-017-02183-y (2017).

85 Grossi, V. *et al.* The longevity SNP rs2802292 uncovered: HSF1 activates stress-dependent expression of FOXO3 through an intronic enhancer. *Nucleic Acids Res* **46**, 5587-5600, doi:10.1093/nar/gky331 (2018).

86 Deelen, J. *et al.* Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. *Hum Mol Genet* **23**, 4420-4432, doi:10.1093/hmg/ddu139 (2014).

87 Soerensen, M. *et al.* Evidence from case-control and longitudinal studies supports associations of genetic variation in APOE, CETP, and IL6 with human longevity. *Age (Dordr)* **35**, 487-500, doi:10.1007/s11357-011-9373-7 (2013).

88 Deelen, J. *et al.* A meta-analysis of genome-wide association studies identifies multiple longevity genes. *Nat Commun* **10**, 3669, doi:10.1038/s41467-019-11558-2 (2019).

89 Lin, K., Dorman, J. B., Rodan, A. & Kenyon, C. daf-16: An HNF-3/forkhead family member that can function to double the life-span of Caenorhabditis elegans. *Science* **278**, 1319-1322, doi:10.1126/science.278.5341.1319 (1997).

90 Ogg, S. *et al.* The Fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in C. elegans. *Nature* **389**, 994-999, doi:10.1038/40194 (1997).

91 Murphy, C. T. *et al.* Genes that act downstream of DAF-16 to influence the lifespan of Caenorhabditis elegans. *Nature* **424**, 277-283, doi:10.1038/nature01789 (2003).

92 Laviola, L., Natalicchio, A. & Giorgino, F. The IGF-I signaling pathway. *Curr Pharm Des* **13**, 663-669, doi:10.2174/138161207780249146 (2007).

93 Kenyon, C., Chang, J., Gensch, E., Rudner, A. & Tabtiang, R. A C. elegans mutant that lives twice as long as wild type. *Nature* **366**, 461-464, doi:10.1038/366461a0 (1993).

94 Bohni, R. *et al.* Autonomous control of cell and organ size by CHICO, a Drosophila homolog of vertebrate IRS1-4. *Cell* **97**, 865-875, doi:10.1016/s0092-8674(00)80799-0 (1999).

95 Kimura, K. D., Tissenbaum, H. A., Liu, Y. & Ruvkun, G. daf-2, an insulin receptor-like gene that regulates longevity and diapause in Caenorhabditis elegans. *Science* **277**, 942-946, doi:10.1126/science.277.5328.942 (1997).

96 Coschigano, K. T., Clemmons, D., Bellush, L. L. & Kopchick, J. J. Assessment of growth parameters and life span of GHR/BP gene-disrupted mice. *Endocrinology* **141**, 2608-2613, doi:10.1210/endo.141.7.7586 (2000).

97 Clancy, D. J. *et al.* Extension of life-span by loss of CHICO, a Drosophila insulin receptor substrate protein. *Science* **292**, 104-106, doi:10.1126/science.1057991 (2001).

98 Libina, N., Berman, J. R. & Kenyon, C. Tissue-specific activities of C. elegans DAF-16 in the regulation of lifespan. *Cell* **115**, 489-502, doi:10.1016/s0092-8674(03)00889-4 (2003).

99 Vitale, G., Barbieri, M., Kamenetskaya, M. & Paolisso, G. GH/IGF-I/insulin system in centenarians. *Mech Ageing Dev* **165**, 107-114, doi:10.1016/j.mad.2016.12.001 (2017).

100 Junnila, R. K., List, E. O., Berryman, D. E., Murrey, J. W. & Kopchick, J. J. The GH/IGF-1 axis in ageing and longevity. *Nat Rev Endocrinol* **9**, 366-376, doi:10.1038/nrendo.2013.67 (2013).

101 Reddy, S. S. K. & Chaiban, J. T. The Endocrinology of Aging: A Key to Longevity "Great Expectations". *Endocr Pract* **23**, 1107-1116, doi:10.4158/EP171793.RA (2017).

102 Taniguchi, C. M., Emanuelli, B. & Kahn, C. R. Critical nodes in signalling pathways: insights into insulin action. *Nat Rev Mol Cell Biol* **7**, 85-96, doi:10.1038/nrm1837 (2006).

103 Moloney, A. M. *et al.* Defects in IGF-1 receptor, insulin receptor and IRS-1/2 in Alzheimer's disease indicate possible resistance to IGF-1 and insulin signalling. *Neurobiol Aging* **31**, 224-243, doi:10.1016/j.neurobiolaging.2008.04.002 (2010).

104 Brugts, M. P. *et al.* Igf-I bioactivity in an elderly population: relation to insulin sensitivity, insulin levels, and the metabolic syndrome. *Diabetes* **59**, 505-508, doi:10.2337/db09-0583 (2010).

105 de Bruijn, R. F. *et al.* Insulin-like growth factor-I receptor stimulating activity is associated with dementia. *J Alzheimers Dis* **42**, 137-142, doi:10.3233/JAD-140186 (2014).

106 Farrer, L. A. *et al.* Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *JAMA* **278**, 1349-1356 (1997).

107 Deelen, J. *et al.* Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited. *Aging Cell* **10**, 686-698, doi:10.1111/j.1474-9726.2011.00705.x (2011).

108 Sebastiani, P. *et al.* Four Genome-Wide Association Studies Identify New Extreme Longevity Variants. *J Gerontol A Biol Sci Med Sci* **72**, 1453-1464, doi:10.1093/gerona/glx027 (2017).

109 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, doi:10.1016/S1474-4422(18)30053-X (2018).

110 Organization, G. W. H. *Depression and Other Common Mental Disorders: Global Health Estimates.* . (Licence: CC BY-NC-SA 3.0 IGO, 2017).

111 https://www.who.int/news-room/fact-sheets/detail/depression, W. H. O.

112 Beekman, A. T. *et al.* Anxiety and depression in later life: Co-occurrence and communality of risk factors. *Am J Psychiatry* **157**, 89-95, doi:10.1176/ajp.157.1.89 (2000).

113 Rodda, J., Walker, Z. & Carter, J. Depression in older adults. *BMJ* **343**, d5219, doi:10.1136/bmj.d5219 (2011).

114 Beekman, A. T. *et al.* Major and minor depression in later life: a study of prevalence and risk factors. *J Affect Disord* **36**, 65-75, doi:0165-0327(95)00061-5 [pii] (1995).

115 Tiemeier, H. Biological risk factors for late life depression. *Eur J Epidemiol* **18**, 745-750 (2003).

116 Major Depressive Disorder Working Group of the Psychiatric, G. C. *et al.* A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry* **18**, 497-511, doi:10.1038/mp.2012.21 (2013).

117 Hek, K. *et al.* A genome-wide association study of depressive symptoms. *Biol Psychiatry* **73**, 667-678 (2013).

118 Levinson, D. F. *et al.* Genetic studies of major depressive disorder: why are there no genome-wide association study findings and what can we do about it? *Biol Psychiatry* **76**, 510-512, doi:10.1016/j.biopsych.2014.07.029 (2014).

119 Hyde, C. L. *et al.* Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat Genet* **48**, 1031-1036, doi:10.1038/ng.3623 (2016).

120 Wray, N. R. & Sullivan, P. F. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *bioRxiv*, doi:10.1101/167577 (2017).

121 Howard, D. M. *et al.* Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. *Nat Commun* **9**, 1470, doi:10.1038/s41467-018-03819-3 (2018).

122 Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci* **22**, 343-352, doi:10.1038/s41593-018-0326-7 (2019).

123 Duric, V. & Duman, R. S. Depression and treatment response: dynamic interplay of signaling pathways and altered neural processes. *Cell Mol Life Sci* **70**, 39-53, doi:10.1007/s00018-012-1020-7 (2013).

124 Bot, M. *et al.* Metabolomics Profile in Depression: A Pooled Analysis of 230 Metabolic Markers in 5283 Cases With Depression and 10,145 Controls. *Biol Psychiatry*, doi:10.1016/j.biopsych.2019.08.016 (2019).

125 Patti, G. J., Yanes, O. & Siuzdak, G. Innovation: Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol* **13**, 263-269, doi:10.1038/nrm3314 (2012).

126 Bot, M. *et al.* Metabolomics Profile in Depression: A Pooled Analysis of 230 Metabolic Markers in 5283 Cases With Depression and 10,145 Controls. *Biol Psychiatry* **87**, 409-418, doi:10.1016/j.biopsych.2019.08.016 (2020).

127 Liu, J. *et al.* Integration of epidemiologic, pharmacologic, genetic and gut microbiome data in a drug-metabolite atlas. *Nat Med* **26**, 110-117, doi:10.1038/s41591-019-0722-x (2020).

128 Rechtschaffen, A. Current perspectives on the function of sleep. *Perspect Biol Med* **41**, 359-390 (1998).

129 Knutson, K. L., Rathouz, P. J., Yan, L. L., Liu, K. & Lauderdale, D. S. Intra-individual daily and yearly variability in actigraphically recorded sleep measures: the CARDIA study. *Sleep* **30**, 793-796 (2007).

130 Van Dongen, H. P., Vitellaro, K. M. & Dinges, D. F. Individual differences in adult human sleep and wakefulness: Leitmotif for a research agenda. *Sleep* **28**, 479-496 (2005).

131 Mezick, E. J. *et al.* Intra-individual variability in sleep duration and fragmentation: associations with stress. *Psychoneuroendocrinology* **34**, 1346-1354, doi:10.1016/j.psyneuen.2009.04.005 (2009).

132 Consensus Conference, P. *et al.* Recommended Amount of Sleep for a Healthy Adult: A Joint Consensus Statement of the American Academy of Sleep Medicine and Sleep Research Society. *J Clin Sleep Med* **11**, 591-592, doi:10.5664/jcsm.4758 (2015).

133 Reid, K. J. *et al.* Sleep: a marker of physical and mental health in the elderly. *Am J Geriatr Psychiatry* **14**, 860-866, doi:10.1097/01.JGP.0000206164.56404.ba (2006).

134 Landolt, H. P. & Borbely, A. A. Age-dependent changes in sleep EEG topography. *Clin Neurophysiol* **112**, 369-377, doi:10.1016/s1388-2457(00)00542-3 (2001).

135 Rechtschaffen, A., Kales, A. *A Manual of Standardized Terminology, Techniques and Scoring System of Sleep Stages in Human Subjects*. (Los Angeles: UCLA Brain Information Services, 1968).

136 Mander, B. A., Winer, J. R. & Walker, M. P. Sleep and Human Aging. *Neuron* **94**, 19-36, doi:10.1016/j.neuron.2017.02.004 (2017).

137 Foley, D. J. *et al.* Sleep complaints among elderly persons: an epidemiologic study of three communities. *Sleep* **18**, 425-432, doi:10.1093/sleep/18.6.425 (1995).

138 Hastings, M. The brain, circadian rhythms, and clock genes. *BMJ* **317**, 1704-1707, doi:10.1136/bmj.317.7174.1704 (1998).

139 Hastings, M. H. Central clocking. *Trends Neurosci* **20**, 459-464, doi:S0166-2236(97)01087-4 [pii] (1997).

140 Moore, R. Y. Suprachiasmatic nucleus in sleep-wake regulation. *Sleep Med* **8 Suppl 3**, 27-33, doi:10.1016/j.sleep.2007.10.003 (2007).

141 Baekeland, F. & Hoy, P. Reported vs recorded sleep characteristics. *Arch Gen Psychiatry* **24**, 548-551 (1971).

142 Hoch, C. C. *et al.* Empirical note: self-report versus recorded sleep in healthy seniors. *Psychophysiology* **24**, 293-299 (1987).

143 Geisler, P. *et al.* The influence of age and sex on sleep latency in the MSLT-30--a normative study. *Sleep* **29**, 687-692, doi:10.1093/sleep/29.5.687 (2006).

144 Krishnan, V. & Collop, N. A. Gender differences in sleep disorders. *Curr Opin Pulm Med* **12**, 383-389 (2006).

145 Bonnet, M. H. & Webb, W. B. The return to sleep. *Biol Psychol* **8**, 225-233 (1979).

146 Afaghi, A., O'Connor, H. & Chow, C. M. High-glycemic-index carbohydrate meals shorten sleep onset. *Am J Clin Nutr* **85**, 426-430, doi:85/2/426 [pii] (2007).

147 Nixon, G. M. *et al.* Falling asleep: the determinants of sleep latency. *Arch Dis Child* **94**, 686-689, doi:10.1136/adc.2009.157453 (2009).

148 Gupta, R., Dahiya, S. & Bhatia, M. S. Effect of depression on sleep: Qualitative or quantitative? *Indian J Psychiatry* **51**, 117-121 (2009).

149 Buysse, D. J., Reynolds, C. F., 3rd, Monk, T. H., Berman, S. R. & Kupfer, D. J. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res* **28**, 193-213, doi:0165-1781(89)90047-4 [pii] (1989).

150 Medic, G., Wille, M. & Hemels, M. E. Short- and long-term health consequences of sleep disruption. *Nat Sci Sleep* **9**, 151-161, doi:10.2147/NSS.S134864 (2017).

151 Heath, A. C., Kendler, K. S., Eaves, L. J. & Martin, N. G. Evidence for genetic influences on sleep disturbance and sleep pattern in twins. *Sleep* **13**, 318-335 (1990).

152 Klei, L. *et al.* Heritability of morningness-eveningness and self-report sleep measures in a family-based sample of 521 hutterites. *Chronobiol Int* **22**, 1041-1054, doi:10.1080/07420520500397959 (2005).

153 Partinen, M., Kaprio, J., Koskenvuo, M., Putkonen, P. & Langinvainio, H. Genetic and environmental determination of human sleep. *Sleep* **6**, 179-185, doi:10.1093/sleep/6.3.179 (1983).

154 Peppard, P. E. *et al.* Increased prevalence of sleep-disordered breathing in adults. *Am J Epidemiol* **177**, 1006-1014, doi:10.1093/aje/kws342 (2013).
155 Senaratna, C. V. *et al.* Prevalence of obstructive sleep apnea in the general population: A systematic review. *Sleep Med Rev*, doi:10.1016/j.smrv.2016.07.002 (2016).
156 Gottlieb, D. J. *et al.* Relation of sleepiness to respiratory disturbance index: the Sleep Heart Health Study. *Am J Respir Crit Care Med* **159**, 502-507, doi:10.1164/ajrccm.159.2.9804051 (1999).
157 Young, T., Peppard, P. E. & Gottlieb, D. J. Epidemiology of obstructive sleep apnea: a population health perspective. *Am J Respir Crit Care Med* **165**, 1217-1239 (2002).
158 American Academy of Sleep Medicine. *International classification of sleep disorders*. (American Academy of Sleep Medicine, 2014).
159 McNicholas, W. T. Diagnosis of obstructive sleep apnea in adults. *Proc Am Thorac Soc* **5**, 154-160, doi:10.1513/pats.200708-118MG (2008).
160 Pedrosa, R. P. *et al.* Obstructive sleep apnea: the most common secondary cause of hypertension associated with resistant hypertension. *Hypertension* **58**, 811-817, doi:10.1161/HYPERTENSIONAHA.111.179788 (2011).
161 Peppard, P. E., Young, T., Palta, M. & Skatrud, J. Prospective study of the association between sleep-disordered breathing and hypertension. *N Engl J Med* **342**, 1378-1384, doi:10.1056/NEJM200005113421901 (2000).
162 Gottlieb, D. J. *et al.* Prospective study of obstructive sleep apnea and incident coronary heart disease and heart failure: the sleep heart health study. *Circulation* **122**, 352-360, doi:10.1161/CIRCULATIONAHA.109.901801 (2010).
163 Shamsuzzaman, A. S., Gersh, B. J. & Somers, V. K. Obstructive sleep apnea: implications for cardiac and vascular disease. *JAMA* **290**, 1906-1914, doi:10.1001/jama.290.14.1906 (2003).
164 Marin, J. M., Carrizo, S. J., Vicente, E. & Agusti, A. G. Long-term cardiovascular outcomes in men with obstructive sleep apnoea-hypopnoea with or without treatment with continuous positive airway pressure: an observational study. *Lancet* **365**, 1046-1053, doi:10.1016/S0140-6736(05)71141-7 (2005).
165 Yaggi, H. K. *et al.* Obstructive sleep apnea as a risk factor for stroke and death. *N Engl J Med* **353**, 2034-2041, doi:10.1056/NEJMoa043104 (2005).
166 Kendzerska, T., Gershon, A. S., Hawker, G., Tomlinson, G. & Leung, R. S. Obstructive sleep apnea and incident diabetes. A historical cohort study. *Am J Respir Crit Care Med* **190**, 218-225, doi:10.1164/rccm.201312-2209OC (2014).
167 Shaw, J. E. *et al.* Sleep-disordered breathing and type 2 diabetes: a report from the International Diabetes Federation Taskforce on Epidemiology and Prevention. *Diabetes Res Clin Pract* **81**, 2-12, doi:10.1016/j.diabres.2008.04.025 (2008).
168 Aurora, R. N. & Punjabi, N. M. Obstructive sleep apnoea and type 2 diabetes mellitus: a bidirectional association. *Lancet Respir Med* **1**, 329-338, doi:10.1016/S2213-2600(13)70039-0 (2013).
169 Kim, H. C. *et al.* Sleep-disordered breathing and neuropsychological deficits. A population-based study. *Am J Respir Crit Care Med* **156**, 1813-1819, doi:10.1164/ajrccm.156.6.9610026 (1997).
170 Yaffe, K. *et al.* Sleep-disordered breathing, hypoxia, and risk of mild cognitive impairment and dementia in older women. *JAMA* **306**, 613-619, doi:10.1001/jama.2011.1115 (2011).
171 Peppard, P. E., Szklo-Coxe, M., Hla, K. M. & Young, T. Longitudinal association of sleep-related breathing disorder and depression. *Arch Intern Med* **166**, 1709-1715, doi:10.1001/archinte.166.16.1709 (2006).
172 Punjabi, N. M. *et al.* Sleep-disordered breathing and mortality: a prospective cohort study. *PLoS Med* **6**, e1000132, doi:10.1371/journal.pmed.1000132 (2009).
173 Young, T. *et al.* Sleep disordered breathing and mortality: eighteen-year follow-up of the Wisconsin sleep cohort. *Sleep* **31**, 1071-1078 (2008).
174 Marshall, N. S. *et al.* Sleep apnea as an independent risk factor for all-cause mortality: the Busselton Health Study. *Sleep* **31**, 1079-1085 (2008).
175 Redline, S. & Tishler, P. V. The genetics of sleep apnea. *Sleep Med Rev* **4**, 583-602, doi:10.1053/smrv.2000.0120 (2000).
176 Redline, S. *et al.* The familial aggregation of obstructive sleep apnea. *Am J Respir Crit Care Med* **151**, 682-687, doi:10.1164/ajrccm.151.3.7881656 (1995).
177 Redline, S., Kump, K., Tishler, P. V., Browner, I. & Ferrette, V. Gender differences in sleep disordered breathing in a community-based sample. *Am J Respir Crit Care Med* **149**, 722-726, doi:10.1164/ajrccm.149.3.8118642 (1994).

178 Bixler, E. O. *et al.* Prevalence of sleep-disordered breathing in women: effects of gender. *Am J Respir Crit Care Med* **163**, 608-613, doi:10.1164/ajrccm.163.3.9911064 (2001).
179 Young, T., Skatrud, J. & Peppard, P. E. Risk factors for obstructive sleep apnea in adults. *Jama-Journal of the American Medical Association* **291**, 2013-2016, doi:DOI 10.1001/jama.291.16.2013 (2004).
180 Young, T. *et al.* Predictors of sleep-disordered breathing in community-dwelling adults: the Sleep Heart Health Study. *Arch Intern Med* **162**, 893-900, doi:ioi10232 [pii] (2002).
181 Patel, S. R., Larkin, E. K. & Redline, S. Shared genetic basis for obstructive sleep apnea and adiposity measures. *Int J Obes (Lond)* **32**, 795-800, doi:10.1038/sj.ijo.0803803 (2008).
182 Lin, L., Finn, L., Zhang, J., Young, T. & Mignot, E. Angiotensin-converting enzyme, sleep-disordered breathing, and hypertension. *Am J Respir Crit Care Med* **170**, 1349-1353, doi:10.1164/rccm.200405-616OC (2004).
183 Bostrom, K. B. *et al.* Interaction between the angiotensin-converting enzyme gene insertion/deletion polymorphism and obstructive sleep apnoea as a mechanism for hypertension. *J Hypertens* **25**, 779-783, doi:10.1097/HJH.0b013e328017f6d5 (2007).
184 Patel, S. R., Larkin, E. K., Mignot, E., Lin, L. & Redline, S. The association of angiotensin converting enzyme (ACE) polymorphisms with sleep apnea and hypertension. *Sleep* **30**, 531-533 (2007).
185 Riha, R. L. *et al.* Tumour necrosis factor-alpha (-308) gene polymorphism in obstructive sleep apnoea-hypopnoea syndrome. *Eur Respir J* **26**, 673-678, doi:10.1183/09031936.05.00130804 (2005).
186 Bhushan, B., Guleria, R., Misra, A., Luthra, K. & Vikram, N. K. TNF-alpha gene polymorphism and TNF-alpha levels in obese Asian Indians with obstructive sleep apnea. *Respir Med* **103**, 386-392, doi:10.1016/j.rmed.2008.10.001 (2009).
187 Popko, K. *et al.* Frequency of distribution of inflammatory cytokines IL-1, IL-6 and TNF-alpha gene polymorphism in patients with obstructive sleep apnea. *J Physiol Pharmacol* **59 Suppl 6**, 607-614 (2008).
188 Kadotani, H. *et al.* Association between apolipoprotein E epsilon4 and sleep-disordered breathing in adults. *JAMA* **285**, 2888-2890, doi:jbr10003 [pii] (2001).
189 Gottlieb, D. J. *et al.* APOE epsilon4 is associated with obstructive sleep apnea/hypopnea: the Sleep Heart Health Study. *Neurology* **63**, 664-668, doi:63/4/664 [pii] (2004).
190 Bayazit, Y. A. *et al.* Association of the -1438G/A polymorphism of the 5-HT2A receptor gene with obstructive sleep apnea syndrome. *ORL J Otorhinolaryngol Relat Spec* **68**, 123-128, doi:10.1159/000091216 (2006).
191 Ylmaz, M. *et al.* Association of serotonin transporter gene polymorphism with obstructive sleep apnea syndrome. *Laryngoscope* **115**, 832-836 (2005).
192 Sakai, K. *et al.* Serotonin-2A and 2C receptor gene polymorphisms in Japanese patients with obstructive sleep apnea. *Intern Med* **44**, 928-933, doi:JST.JSTAGE/internalmedicine/44.928 [pii] (2005).
193 Qin, B., Sun, Z., Liang, Y., Yang, Z. & Zhong, R. The association of 5-HT2A, 5-HTT, and LEPR polymorphisms with obstructive sleep apnea syndrome: a systematic review and meta-analysis. *PLoS One* **9**, e95856 (2014).
194 Larkin, E. K. *et al.* A candidate gene study of obstructive sleep apnea in European Americans and African Americans. *Am J Respir Crit Care Med* **182**, 947-953, doi:10.1164/rccm.201002-0192OC (2010).
195 Grote, L., Kraiczi, H. & Hedner, J. Reduced alpha- and beta(2)-adrenergic vascular response in patients with obstructive sleep apnea. *Am J Respir Crit Care Med* **162**, 1480-1487, doi:10.1164/ajrccm.162.4.9912028 (2000).
196 Mills, P. J. *et al.* Beta 2-adrenergic receptor characteristics in sleep apnea patients. *Sleep* **18**, 39-42 (1995).
197 Sleiman, P. & Hakonarson, H. Genetic underpinnings of obstructive sleep apnea: are we making progress? *Sleep* **34**, 1449-1452, doi:10.5665/sleep.1366 (2011).
198 Palmer, L. J. *et al.* A whole-genome scan for obstructive sleep apnea and obesity. *Am J Hum Genet* **72**, 340-350, doi:10.1086/346064 (2003).
199 Cade, B. E. *et al.* Genetic Associations with Obstructive Sleep Apnea Traits in Hispanic/Latino Americans. *Am. J. Respir. Crit. Care Med.* **194**, 886-897, doi:10.1164/rccm.201512-2431OC (2016).
200 Kripke, D. F. *et al.* Genetic variants associated with sleep disorders. *Sleep Med* **16**, 217-224, doi:10.1016/j.sleep.2014.11.003 (2015).
201 Patel, S. R. *et al.* Association of genetic loci with sleep apnea in European Americans and African-Americans: the Candidate Gene Association Resource (CARe). *PLoS One* **7**, e48836 (2012).
202 Kaddurah-Daouk, R. & Krishnan, K. R. Metabolomics: a global biochemical approach to the study of central nervous system diseases. *Neuropsychopharmacology* **34**, 173-186, doi:10.1038/npp.2008.174 (2009).

203 Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).

204 Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* **34**, 816-834, doi:10.1002/gepi.20533 (2010).

205 Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* **81**, 1084-1097, doi:10.1086/521987 (2007).

206 Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-913, doi:10.1038/ng2088 (2007).

207 Iglesias, A. I. *et al.* Haplotype reference consortium panel: Practical implications of imputations with large reference panels. *Hum Mutat* **38**, 1025-1032, doi:10.1002/humu.23247 (2017).

208 International Schizophrenia, C. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748-752, doi:10.1038/nature08185 (2009).

209 Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427, doi:10.1038/nature13595 (2014).

210 Maher, B. Personal genomes: The case of the missing heritability. *Nature* **456**, 18-21, doi:10.1038/456018a (2008).

211 Gibson, G. Rare and common variants: twenty arguments. *Nat Rev Genet* **13**, 135-145, doi:10.1038/nrg3118 (2012).

212 Metzker, M. L. Sequencing technologies - the next generation. *Nat Rev Genet* **11**, 31-46, doi:10.1038/nrg2626 (2010).

213 Griswold, A. J. *et al.* Targeted massively parallel sequencing of autism spectrum disorder-associated genes in a case control cohort reveals rare loss-of-function risk variants. *Mol Autism* **6**, 43, doi:10.1186/s13229-015-0034-z (2015).

214 Raghavan, N. S. *et al.* Whole-exome sequencing in 20,197 persons for rare variants in Alzheimer's disease. *Ann Clin Transl Neurol* **5**, 832-842, doi:10.1002/acn3.582 (2018).

215 Guerreiro, R., Bras, J., Hardy, J. & Singleton, A. Next generation sequencing techniques in neurological diseases: redefining clinical and molecular associations. *Hum Mol Genet* **23**, R47-53, doi:10.1093/hmg/ddu203 (2014).

216 Guerreiro, R. *et al.* TREM2 variants in Alzheimer's disease. *N Engl J Med* **368**, 117-127, doi:10.1056/NEJMoa1211851 (2013).

217 Menni, C., Zierer, J., Valdes, A. M. & Spector, T. D. Mixing omics: combining genetics and metabolomics to study rheumatic diseases. *Nat Rev Rheumatol* **13**, 174-181, doi:10.1038/nrrheum.2017.5 (2017).

218 Tektonidou, M. G. & Ward, M. M. Validity of clinical associations of biomarkers in translational research studies: the case of systemic autoimmune diseases. *Arthritis Res Ther* **12**, R179, doi:10.1186/ar3143 (2010).

219 Kaddurah-Daouk, R., Kristal, B. S. & Weinshilboum, R. M. Metabolomics: a global biochemical approach to drug response and disease. *Annu Rev Pharmacol Toxicol* **48**, 653-683, doi:10.1146/annurev.pharmtox.48.113006.094715 (2008).

220 Bingol, K. Recent Advances in Targeted and Untargeted Metabolomics by NMR and MS/NMR Methods. *High Throughput* **7**, doi:10.3390/ht7020009 (2018).

221 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. *Alzheimers Dement* **14**, 723-733, doi:10.1016/j.jalz.2018.01.003 (2018).

222 Wurtz, P. *et al.* Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation* **131**, 774-785, doi:10.1161/CIRCULATIONAHA.114.013116 (2015).

223 Pardo, L. M., MacKay, I., Oostra, B., van Duijn, C. M. & Aulchenko, Y. S. The effect of genetic drift in a young genetically isolated population. *Annals of Human Genetics* **69**, 288-295, doi:DOI 10.1046/j.1529-8817.2005.00162.x (2005).

224 Aulchenko, Y. S. *et al.* Linkage disequilibrium in young genetically isolated Dutch population. *Eur J Hum Genet* **12**, 527-534, doi:10.1038/sj.ejhg.5201188 (2004).

225 Ikram, M. A. *et al.* The Rotterdam Study: 2018 update on objectives, design and main results. *Eur J Epidemiol* **32**, 807-850, doi:10.1007/s10654-017-0321-4 (2017).

CHAPTER 2

Telomere length and related outcomes

CHAPTER 2.1

Exome sequencing analysis identifies rare variants in *ATM* and *RPL8* that are associated with shorter telomere length

Ashley van der Spek, Sophie C. Warner, Linda Broer, Christopher P. Nelson, Dina Vojinovic, Shahzad Ahmad, Pascal P. Arp, Rutger W.W. Brouwer, Matthew Denniff, Mirjam C.G.N. van den Hout, Jeroen G.J. van Rooij, Robert Kraaij, Wilfred F.J. van IJcken, Nilesh J. Samani, M. Arfan Ikram, André G. Uitterlinden, Veryan Codd, Najaf Amin, Cornelia M. van Duijn

This chapter was published in *Frontiers in Genetics*. 2020 Apr 30;11:337.

The supplemental information for this paper is available at https://doi.org/10.3389/fgene.2020.00337.

Abstract

Telomeres are important for maintaining genomic stability. Telomere length has been associated with aging, disease, and mortality and is highly heritable (~82%). In this study, we aimed to identify rare genetic variants associated with telomere length using whole-exome sequence data. We studied 1,303 participants of the Erasmus Rucphen Family (ERF) study, 1,259 of the Rotterdam Study (RS), and 674 of the British Heart Foundation Family Heart Study (BHF-FHS). We conducted two analyses, first we analyzed the family-based ERF study and used the RS and BHF-FHS for replication. Second, we combined the summary data of the three studies in a meta-analysis. Telomere length was measured by quantitative polymerase chain reaction in blood. We identified nine rare variants significantly associated with telomere length (p-value < 1.42×10^{-7}, minor allele frequency of 0.2-0.5%) in the ERF study. Eight of these variants (in *C11orf65*, *ACAT1*, *NPAT*, *ATM*, *KDELC2* and *EXPH5*) were located on chromosome 11q22.3 that contains *ATM*, a gene involved in telomere maintenance. Although we were unable to replicate the variants in the RS and BHF-FHS (p-value ≥ 0.21), segregation analysis showed that all variants segregate with shorter telomere length in a family. In the meta-analysis of all studies, a nominally significant association with LTL was observed with a rare variant in *RPL8* (p-value = 1.48×10^{-6}), which has previously been associated with age. Additionally, a novel rare variant in the known *RTEL1* locus showed suggestive evidence for association (p-value = 1.18×10^{-4}) with LTL. To conclude, we identified novel rare variants associated with telomere length. Larger samples sizes are needed to confirm these findings and to identify additional variants.

Introduction

Telomeres are DNA structures located at the ends of chromosomes and consist of tandem hexanucleotide sequence repeats (TTAGGG)[1]. They are important for maintaining genomic stability by preventing DNA degradation and chromosomal fusions[2]. Telomeres are shortened with each cell division due to the inability of DNA polymerase to fully extend the 3' end of the DNA strand during replication. When the telomeres reach a critical length, this leads to cellular senescence and ultimately to cell death, making them regulators of the replicative capacity of a cell and markers of biological age[3,4].

Shorter leukocyte telomere length (LTL) has been associated with several age-related diseases including cardiovascular diseases[5-10], cancer[11-14] and dementia[15-17]. LTL has also been associated with mortality[15,18-23]. However, this association has been inconsistent[24-29]. LTL is highly heritable with heritability estimates ranging from 34% to 82%[30-34]. Previously, genome-wide association studies (GWASs) in European ancestry studies have identified common variants associated with LTL located in multiple genes, including: *TERC*[35-38], *TERT*[36,38], *NAF1*[36], *OBFC1*[36-39], *RTEL1*[36], *CTC1*[37], *ZNF676*[37], *ZNF208*[36], *ACYP2*[36], *DCAF4*[40], and *PXK*[38]. However, these variants explain < 5% of the heritability.

Up until now, no systemic whole exome or genome screen for rare variants has been published, despite the fact that these may explain part of the heritability[41]. Rare variants are not well captured by microarrays used for GWAS and remain difficult to impute, despite the recent improvements in imputation panels[42]. Next generation sequencing technologies, such as whole-exome sequencing (WES), are better suited to study rare variants. In this study, we present a dual analysis. First, we conducted a genome-wide WES analysis of LTL in 1,303 individuals of the Dutch Erasmus Rucphen Family (ERF) study to search for rare genetic variants associated with LTL. The advantage of a family-based study is that the segregation of rare variants can be studied. We performed a replication analysis in the Rotterdam Study (RS) and the British Heart Foundation Family Heart Study (BHF-FHS). Next, we pooled the data together and conducted a meta-analysis of the association results of all three cohorts.

Results

Descriptive statistics of the family-based and population-based studies are provided in **Table 1**. Mean age at LTL measurement was 49 years (SD = 15.0) in the ERF study and 61% of the study participants were female. The RS participants were older ($\bar{x}_{age}$ = 75 years, SD = 7.7) and 57% of the participants were female, while mean age in the BHF-FHS was 58 years (SD = 8.2) and most study participants were male (26% female). Mean LTL values, measured in each participant using a quantitative polymerase chain reaction (qPCR) based technique as the ratio of telomere repeat length to copy number of the single copy gene *36B4*, were

higher in the ERF study ($\bar{x}_{LTL}$ = 1.85, SD = 0.35) than in the RS ($\bar{x}_{LTL}$ = 0.94. SD = 0.18) and BHF-FHS ($\bar{x}_{LTL}$ = 1.37, SD = 0.22). After adjusting LTL values for age and sex, mean LTL was comparable between studies (**Table 1**).

Table 1. Descriptive statistics of the study populations.

	ERF	RS	BHF-FHS
N	1303	1257	674
Mean age (SD)	48.9 (15.0)	74.5 (7.7)	58.0 (8.2)
Age range	18.2 – 95.7	55.0 – 105.8	36.0 – 81.0
% female	60.5	57.0	25.8
Mean LTL (SD)	1.85 (0.35)	0.94 (0.18)	1.37 (0.22)
LTL range	0.77 – 3.17	0.31 – 1.79	0.69 – 2.14
Adjusted mean LTL (SD)*	$8.85x10^{-18}$ (0.32)	$1.37x10^{-17}$ (0.18)	$1.11x10^{-10}$ (0.21)
Adjusted LTL range*	-1.15 – 1.08	-0.71 – 0.89	-0.63 – 0.71

Abbreviations: ERF = Erasmus Rucphen Family study; RS = Rotterdam Study; BHF-FHS = British Heart Foundation Family Heart Study; N = Number of variants; SD = Standard Deviation, LTL = Leukocyte Telomere Length. * LTL values were adjusted for age and sex and information on residuals is shown.

The Manhattan plot and the distribution of the test statistic (quantile-quantile (QQ) plot, λ = 1.04) of the WES analysis in the ERF study are presented in **Figures 1**, **2**, respectively. We observed significant association of nine rare variants (Minor Allele Frequency (MAF) between 0.2% and 0.5%) with LTL as shown in **Table 2**. The significance threshold (*p*-value < $1.42x10^{-7}$) was adjusted for multiple testing using Bonferroni correction based on the number of variants in the analysis (0.05/353,075 variants). Each variant was negatively associated with LTL and the estimated effects of the minor allele of these variants were large (-2.18 < standardized β < -1.34), suggesting a significant decrease in LTL for each minor allele.

The top eight variants are located in a dense region on chromosome 11q22.3 (position range: 108004687 – 108384666, **Figure 3**) and appear to be a part of a linked haplotype that spans the *C11orf65, ACAT1, NPAT, ATM, KDELC2*, and *EXPH5* genes. A haplotype can describe a pair of genes inherited together from one parent on one chromosome, or it can describe all of the genes on a chromosome that were inherited together from a single parent. This haplotype segregates with shorter LTL in a family (**Supplementary Figure 1**), where it is carried by 14 individuals, 11 of whom were related within 4 generations according to the pedigree data (**Figure 4**). Further, the genetic kinship estimates show that the other three individuals are also related within 3-4 generations. These 8 variants are in strong linkage disequilibrium (pairwise LD: r^2 between 0.93 and 1.00, D′ = 1) and show very similar *p*-values. The top variant rs185270276 is located in an intron of the *C11orf65* gene (MAF = 0.5%, β = -1.34, SE = 0.25, *p-value* = $7.99x10^{-8}$). The next six variants significantly associated with LTL (MAF = 0.5%, β = -1.38, SE = 0.26, *p-value* = $1.21x10^{-7}$) are located in the *ACAT1*, *NPAT*, *ATM*/

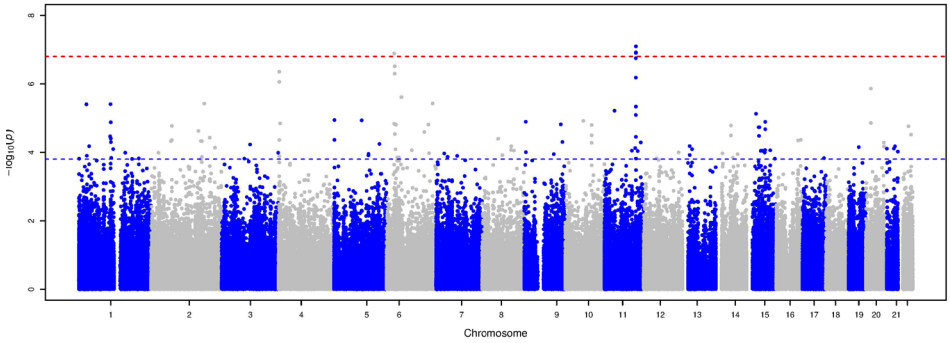


Figure 1. Manhattan plot of the association analysis with LTL in the ERF study.
This plot shows –log10 transformed *p*-values (y-axis) for all variants present in the association analysis according to their position on each chromosome (x-axis). The red dashed line represents the Bonferroni corrected *p*-value threshold for significance (*p*-value < 1.42×10^{-7}), while the blue dashed line represents the *p*-value threshold for suggestive significance (*p*-value < 1.42×10^{-4}).

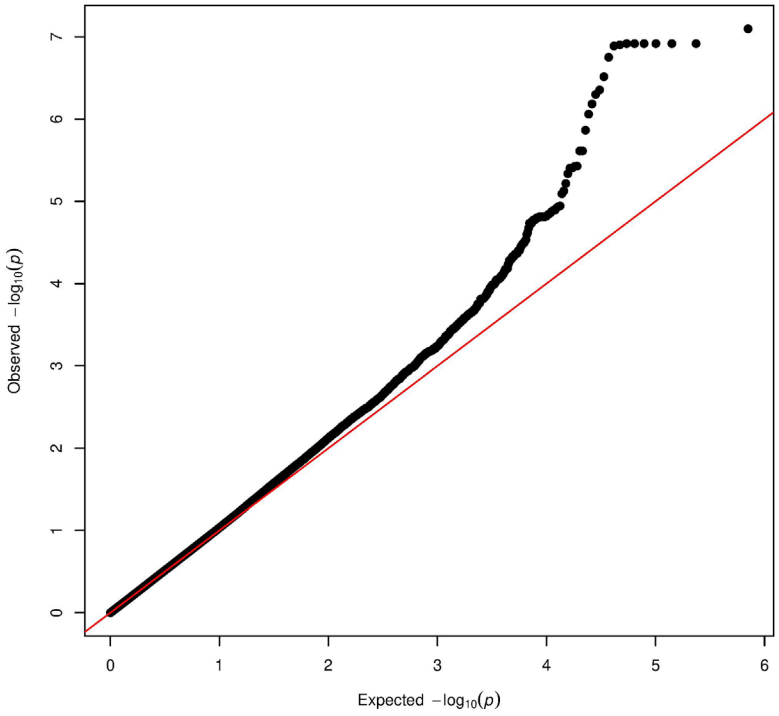


Figure 2. Quantile-quantile plot of the association analysis with LTL in the ERF study.
The QQ-plot shows the observed test statistics (y-axis) plotted against the expected values of the test statistics (x-axis) (X^2-distribution). The red line shows the distribution under the null hypothesis.

Table 2. Significant variants from the association analysis in the ERF study.

rsID	Gene	Chr	Position*	MAF	REF/ ALT	GVS function**	PolyPhen score**	PhastCons score**	CADD score**	β	SE	P-value
rs185270276	*C11orf65*	11	108263828	0.005	T/C	intron	unknown	0	6.51	-1.34	0.25	$7.99x10^{-8}$
rs12365364	*ACAT1*	11	108004687	0.005	G/A	intron	unknown	0	3.28	-1.38	0.26	$1.21x10^{-7}$
rs79119325	*NPAT*	11	108032614	0.005	C/T	missense	1	0.998	18.13	-1.38	0.26	$1.21x10^{-7}$
rs3092910	*ATM,C11orf65*	11	108180917	0.005	T/C	intron, synonymous	unknown	0.997	9.87	-1.38	0.26	$1.21x10^{-7}$
rs3218711	*ATM,C11orf65*	11	108236264	0.005	C/G	3-prime-UTR, intron	unknown	0.002	5.56	-1.38	0.26	$1.21x10^{-7}$
rs11212668	*KDELC2*	11	108352576	0.005	T/C	intron	unknown	0	1.44	-1.38	0.26	$1.21x10^{-7}$
rs12146512	*EXPH5*	11	108384666	0.005	T/C	missense	0.624	0.011	4.47	-1.38	0.26	$1.21x10^{-7}$
rs2234993	*ATM*	11	108129599	0.005	C/G	intron	unknown	0	4.25	-1.37	0.26	$1.25x10^{-7}$
rs144114619	*BTN3A1*	6	26408145	0.002	T/A	missense	1	0.002	12.24	-2.18	0.41	$1.29x10^{-7}$

Abbreviations: Chr = Chromosome; MAF = Minor Allele Frequency; REF = Reference allele; ALT = Alternative allele; CADD = Combined Annotation Dependent Depletion; β = effect of the minor allele; SE = Standard Error.
* Position according to Hg19
** Seattleseq Annotation Database 138

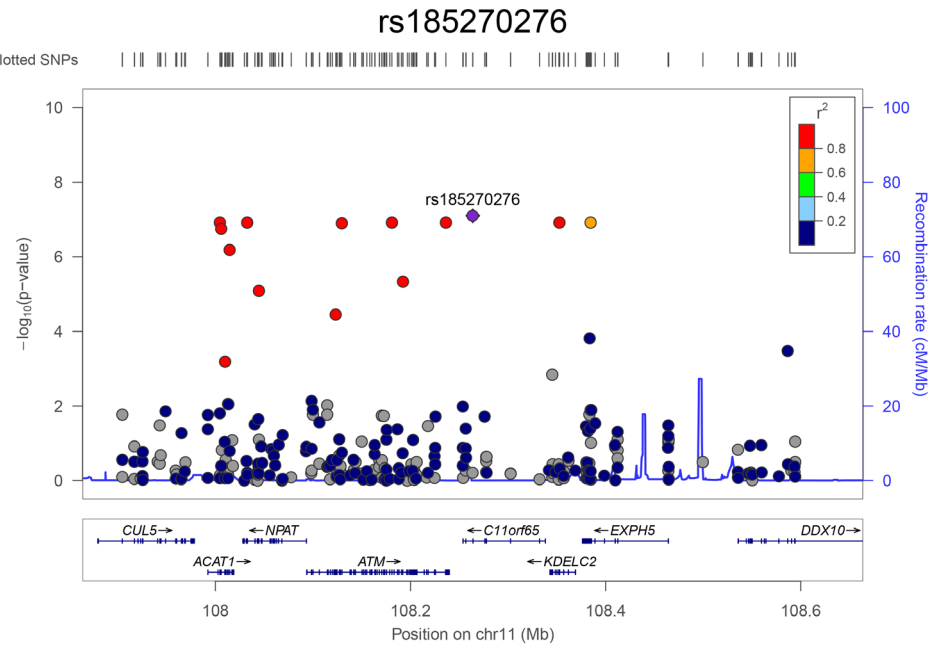


Figure 3. Regional association plot for the top hits on chromosome 11.
The plot was constructed using LocusZoom (http://locuszoom.org/). The –log10 transformed p-values are plotted on the y-axis. The x-axis shows the position of the variants (dots) on chromosome 11 and the genes in this region. The most significant variant (rs185270276) is shown in purple and the color of the dots indicates the extend of linkage disequilibrium between the variant and the top variant.

C11orf65, *KDELC2*, and *EXPH5* genes. Two of these six variants are missense variants: rs79119325 (*NPAT*, PolyPhen = 1, CADD score = 18.1) and rs12146512 (*EXPH5*, PolyPhen = 0.624, CADD score = 4.5). The eighth significant variant, rs2234993, is located within an intron of *ATM* (MAF = 0.5%, β = -1.37, SE = 0.26, *p-value* = 1.25×10^{-7}). The ninth significant variant, rs144114619 (MAF = 0.2%, *p-value* = 1.29×10^{-7}), is a missense variant located on chromosome six in the *BTN3A1* gene, which is predicted to be damaging (PolyPhen = 1, CADD score = 12.2) and has the largest effect size (β = -2.18, SE = 0.41). There were six carriers of this variant in the ERF population. Interestingly, four of these carriers are also carriers of the rare variants in the chromosome 11q22.3 region (**Supplementary Figure 2**).

For replication analysis, we used WES data from two independent cohorts of European ancestry, the RS and the BHF-FHS. Results of the replication analysis are shown in **Table 3**, together with the results of the meta-analysis of summary statistics from all three cohorts for these variants. Six out of nine rare variants significantly associated with LTL in the ERF study (located in *ACAT1*, *NPAT*, *ATM/C11orf65*, *EXPH5*, and *BTN3A1*) were present in the RS and/or the BHF-FHS but were not significantly (*p-value* ≥ 0.21) associated with LTL (*p-value* < 0.025, 0.05/2 independent tests). The direction of effect of most variants was similar in

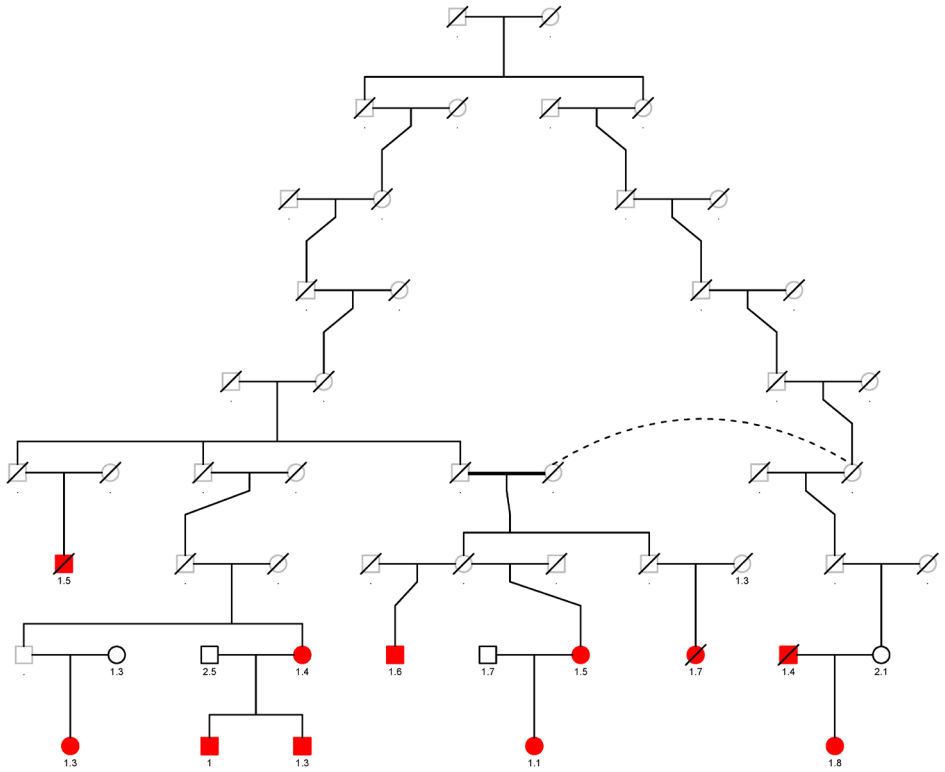


Figure 4. Segregation plot of the rare variants on chromosome 11q22.3 in the ERF study.
The carriers of the rare variants located on chromosome 11q22.3 are depicted in red. The T/S ratio is added below the individuals of whom whole-exome sequencing data was available. Squares are males and rounds are females. Deceased individuals are denoted by a line through that specific individual. One person is twice in the pedigree, this is shown with the dotted line.

the ERF study and the RS, while the BHF-FHS showed opposite direction of effect for most variants. Three variants, located in *C11orf65*, *KDELC2*, and *ATM*, were not present in the RS and BHF-FHS data. As these are unique to an isolated population, we were unable to confirm or reject their association with LTL in the replication cohorts.

Finally, to increase the statistical power, we performed an inverse-variance weighted meta-analysis of the association results from all three cohorts using METAL software. Variants were included in the meta-analysis if they were present in at least two out of three cohorts and had a minimum minor allele count of five in one or more cohorts, resulting in a multiple testing corrected significance threshold of 3.02×10^{-7} (0.05/165,311 variants). The top results of the meta-analysis (*p-value* < 3.02×10^{-4}) are available in **Table 4** and in **Supplementary Table 1**, which also contains cohort specific information. The Manhattan and QQ-plots are provided in **Supplementary Figures 3** and **4**, respectively. The λ of 0.97 suggests the power has been low. Although there were no variants genome-wide significantly associated with

Table 3. Replication results of the association analysis.

rsID	Gene	ERF (N = 1,303)				RS (N = 1,257)				BHF-FHS (N = 674)				Meta-analysis (N = 3,234)			
		MAF	β	SE	P-value	MAF	β	SE	P-value	MAF	β	SE	P-value	Direction*	β	SE	P-value
rs185270276	*C11orf65*	0.005	-1.34	0.25	$7.99x10^{-8}$	--	--	--	--	--	--	--	--	--	--	--	--
rs12365364	*ACAT1*	0.005	-1.38	0.26	$1.21x10^{-7}$	0.006	-0.20	0.25	0.43	0.010	0.30	0.26	0.25	--+	-0.40	0.15	$7.84x10^{-3}$
rs79119325	*NPAT*	0.005	-1.38	0.26	$1.21x10^{-7}$	0.006	-0.24	0.28	0.39	0.010	0.30	0.26	0.25	--+	-0.43	0.16	$5.91x10^{-3}$
rs3092910	*ATM,C11orf65*	0.005	-1.38	0.26	$1.21x10^{-7}$	0.007	-0.18	0.27	0.50	0.010	0.30	0.26	0.25	--+	-0.41	0.16	$8.67x10^{-3}$
rs3218711	*ATM,C11orf65*	0.005	-1.38	0.26	$1.21x10^{-7}$	0.005	-0.24	0.28	0.39	0.010	0.30	0.26	0.25	--+	-0.43	0.16	$5.93x10^{-3}$
rs11212668	*KDELC2*	0.005	-1.38	0.26	$1.21x10^{-7}$	--	--	--	--	--	--	--	--	--	--	--	--
rs12146512	*EXPH5*	0.005	-1.38	0.26	$1.21x10^{-7}$	--	--	--	--	0.016	0.27	0.21	0.21	-?+	-0.38	0.17	$2.61x10^{-2}$
rs2234993	*ATM*	0.005	-1.37	0.26	$1.25x10^{-7}$	--	--	--	--	--	--	--	--	--	--	--	--
rs144114619	*BTN3A1*	0.002	-2.18	0.41	$1.29x10^{-7}$	0.004	0.008	0.29	0.98	0.003	-0.43	0.48	0.38	-+-	-0.63	0.22	$3.73x10^{-3}$

Abbreviations: ERF = Erasmus Rucphen Family study; RS = Rotterdam Study; BHF-FHS = British Heart Foundation Family Heart Study; N = sample size; MAF = Minor Allele Frequency; β = Effect of the minor allele; SE = Standard Error.
* Order of cohorts in direction column: ERF – RS – BHF-FHS; Direction of effect represented by - (negative association) + (positive association) or ? (not available).

LTL in the meta-analysis after adjusting for multiple testing, many of the top findings show a consistent effect across cohorts. The variant most significantly associated with LTL was a highly conserved synonymous variant (PhastCons score = 0.999) located in the *RPL8* gene on chromosome 8 (*p-value* = 1.48×10^{-6}), which is predicted to be deleterious (CADD score = 15.11). Additionally, we used the meta-analysis results to perform a look-up of variants in loci identified by previous European ancestry GWASs (**Supplementary Table 2**). There were 153 variants present in these loci and we found suggestive evidence ($p\text{-value}_{meta} < 3.02 \times 10^{-4}$) for a positive association of a rare variant with LTL (rs181080831, β = 0.74, SE = 0.19, *p-value* = 1.18×10^{-4}) in the known *RTEL1* locus.

Discussion

In the family-based ERF study, we identified nine rare variants (MAF between 0.2% - 0.5%) associated with LTL by performing a WES association analysis. The eight most significantly associated variants are located in a region on chromosome 11q22.3 and segregate together with shorter LTL in one family. This region contains the *ATM* locus that has previously been shown to be involved in telomere maintenance and genomic stability and is thus an obvious candidate gene. In the meta-analysis of discovery and replication cohorts, we identified another rare missense variant in the *RPL8* gene strongly associated with LTL. Although we were not able to replicate either of the associations, both *ATM* and *RPL8* have been previously found to be strong predictors of telomere length (*ATM*) and chronological age (*ATM* and *RPL8*).

Interestingly, we identified three unique rare variants in the *ATM* (Ataxia Telangiectasia Mutated) gene associated with LTL in the ERF study. *ATM* is the homolog of the *Tel1* gene in yeast[43] and has been implicated in important telomere maintenance processes[44-46]. The ATM protein kinase is a master controller of cell cycle checkpoint signaling pathways required for cell response to DNA damage and for genomic stability (https://www.ncbi.nlm.nih.gov/gene/?term=472). Additionally, ATM kinase is necessary for telomere elongation[44,46]. *ATM* is involved in the genetic disorder ataxia telangiectasia (AT), which is characterized by cerebellar ataxia, oculocutaneous telangiectasia, immunodeficiency, and a predisposition to cancer[47]. Cells of AT patients also show telomeric fusions and have accelerated telomere shortening with increasing age[48]. Furthermore, *ATM* was significantly associated with chronological age in a meta-analysis of gene expression profiles, showing lower transcript abundance in older individuals[49]. Although genetic variants in *ATM* have been associated with various cancers[50-53], only one genetic variant in *ATM*, rs227080, was genome-wide significantly associated with LTL in a Singaporean Chinese population[54]. This variant was not significantly associated with LTL in the ERF study, the Rotterdam Study or the BHF-FHS, implicating it was not driving the association observed in the current study.

Table 4. Suggestive findings of the meta-analysis (*p-value* ≤ 3.02×10^{-4}).

rsID	Gene	Chr	Postion*	REF/ALT	GVS function**	PolyPhen score**	PhastCons score**	CADD score**	β	SE	P-value	Direction‡
8:146017422	*RPL8*	8	146017422	G/A	synonymous	unknown	0.999	15.11	1.93	0.40	1.48×10^{-6}	+++
rs77919685	*LATS2*	13	21563311	G/A	missense	0.002	0	7.001	0.44	0.09	2.23×10^{-6}	+++
rs56041036	*ZFPM1*	16	88599023	A/G	synonymous-near-splice	unknown	0.998	10.77	-0.20	0.04	7.44×10^{-6}	---
rs199779997	*MGA*	15	42058958	A/C	missense	0.001	0.866	9.45	1.77	0.40	8.60×10^{-6}	+++
rs4895944	*VNN2*	6	133070995	G/T	missense, non-coding-exon	0.038	0.003	16.69	-1.29	0.29	8.99×10^{-6}	--?
rs7735563	*RAPGEF6*	5	130764936	T/C	intron	unknown	1	19.08	-0.77	0.18	1.53×10^{-5}	---
rs189691392	*SF3B5*	6	144416667	G/A	5-prime-UTR	unknown	0	7.119	-1.82	0.42	1.77×10^{-5}	-?-
rs138765444	*ZKSCAN4*	6	28219377	A/G	synonymous, intron	unknown	0.036	9.324	-1.27	0.30	2.59×10^{-5}	---
rs1783091	*none*	21	33964605	T/C	intergenic	unknown	0.004	0.154	0.22	0.05	2.67×10^{-5}	+?+
rs2170177	*DOCK3*	3	51349887	C/A	intron	unknown	0	0.044	-0.36	0.09	5.06×10^{-5}	---
rs5930	*LDLR*	19	11224265	A/G	synonymous	unknown	0.001	0.579	-0.10	0.03	5.73×10^{-5}	---
rs55648406	*TUB*	11	8060566	G/A	missense, intron	0.086	1	15.76	-0.62	0.16	5.85×10^{-5}	---
rs140456008	*SLC35G2*	3	136574420	A/G	missense	0.941	1	10.84	1.89	0.47	6.67×10^{-5}	++?
rs139380413	*COL8A1*	3	99513830	G/A	missense	0.071	0.966	11.23	-1.26	0.32	7.88×10^{-5}	---
rs11656725	*LRRC48*	17	17900726	C/T	intron	unknown	0	0.471	0.38	0.10	8.25×10^{-5}	+?+
rs7193541	*RFWD3*	16	74664743	T/C	missense	0.008	0.485	9.1	-0.10	0.03	8.28×10^{-5}	---
rs17222435	*SLC28A1*	15	85488335	C/T	intron	unknown	0	3.598	0.69	0.18	9.07×10^{-5}	+++
rs117223521	*none*	8	38964715	T/C	upstream-gene	unknown	0	1.194	-1.36	0.35	1.01×10^{-4}	-?-
rs11700220	*MTG2*	20	60770931	A/G	missense	1	1	21.6	0.47	0.12	1.03×10^{-4}	+++
rs187466877	*GPN1*	2	27862872	A/G	intron	unknown	0	1.47	0.85	0.22	1.05×10^{-4}	+?+
rs56188826	*MARK1*	1	220791870	C/T	synonymous	unknown	0.123	6.414	0.41	0.11	1.14×10^{-4}	+?+
rs1872592	*PIF1*	15	65113493	G/A	intron	unknown	0	0.005	-0.10	0.02	1.15×10^{-4}	---
rs73056605	*CLEC4C*	12	7894056	G/A	missense	0.037	0	0.005	0.11	0.03	1.18×10^{-4}	+++

Table 4. Continued

rsID	Gene	Chr	Postion*	REF/ALT	GVS function**	PolyPhen score**	PhastCons score**	CADD score**	β	SE	P-value	Direction‡
rs181080831	*RTEL1, RTEL1-TNFRSF6B*	20	62326874	G/A	intron, non-coding-exon, synonymous	unknown	0	4.04	0.74	0.19	$1.18x10^{-4}$	+++
rs13014800	*CENPA*	2	27015118	G/A	intron	unknown	0.163	11.67	-0.13	0.03	$1.23x10^{-4}$	-?-
rs143463783	*TRIM27*	6	28889741	G/A	synonymous	unknown	1	9.216	-1.22	0.32	$1.24x10^{-4}$	-?-
rs374215951	*PIEZO1*	16	88788318	G/A	synonymous	unknown	0.21	0.893	-2.57	0.67	$1.31x10^{-4}$	-?-
rs181215404	*EPPK1*	8	144941659	G/A	synonymous	unknown	0.011	4.835	0.94	0.25	$1.32x10^{-4}$	+?+
rs377359525	*FTCD*	21	47572869	A/G	missense	1	1	16.44	1.68	0.44	$1.33x10^{-4}$	+++
rs10936599	*MYNN*	3	169492101	C/T	synonymous, non-coding-exon, 5-prime-UTR	unknown	1	10.1	-0.11	0.03	$1.38x10^{-4}$	---
rs74730846	*STXBP5L*	3	120924764	C/T	intron-near-splice	unknown	0.629	5.818	-0.19	0.05	$1.39x10^{-4}$	-?-
rs137853096	*HSD17B4*	5	118788316	G/A	missense, 5-prime-UTR	1	1	24	-0.73	0.19	$1.67x10^{-4}$	--?
rs41284136	*IFIT3*	10	91087805	G/C	5-prime-UTR	unknown	0	7.796	0.40	0.11	$1.76x10^{-4}$	+++
rs58106741	*SLC4A1AP*	2	27886820	G/T	synonymous	unknown	0	5.939	0.70	0.19	$1.80x10^{-4}$	+?+
rs58068845	*UTP6*	17	30200363	G/A	intron	unknown	0	4.223	0.35	0.09	$1.90x10^{-4}$	+++
rs200602887	*GREB1*	2	11751072	G/C	synonymous	unknown	0.986	10.5	-1.02	0.27	$2.01x10^{-4}$	---
rs141180155	*LRP2*	2	170127559	G/A	synonymous, intron	unknown	0	13.06	-0.36	0.10	$2.08x10^{-4}$	---
rs7837242	*LONRF1*	8	12600622	C/T	intron	unknown	0.001	4.941	-0.18	0.05	$2.08x10^{-4}$	-?-
rs115018606	*C2orf16*	2	27799773	A/C	missense	0.972	0.002	5.869	0.41	0.11	$2.12x10^{-4}$	+?+
rs151309008	*REXO2*	11	114310345	C/T	missense	0.437	1	17.76	-0.64	0.17	$2.12x10^{-4}$	---
rs143759519	*PYGL*	14	51382637	G/A	missense	1	0.975	34	0.55	0.15	$2.13x10^{-4}$	+++
rs10936600	*LRRC34*	3	169514585	A/T	missense	1	0.001	12.07	-0.11	0.03	$2.14x10^{-4}$	---
rs117178504	*DYNC2H1*	11	103153788	C/A	synonymous	unknown	0.996	8.615	0.38	0.10	$2.18x10^{-4}$	+++
rs367644268	*COA5*	2	99224742	C/T	intron	unknown	0.015	8.626	-1.42	0.39	$2.28x10^{-4}$	--?
rs146979490	*GPN1*	2	27864089	A/G	intron	unknown	0.025	11.52	0.51	0.14	$2.43x10^{-4}$	+++

rs146033252	*MTA3, OXER1*	2	42991127	G/A	missense, intron	0.084	0.023	8.404	0.71	0.19	$2.48x10^{-4}$	+++
rs141280036	*PADI4*	1	17634718	A/G	missense	0.992	0.881	13.72	0.79	0.22	$2.51x10^{-4}$	+++
rs79400176	*DZANK1*	20	18414309	C/T	missense	0.129	0.994	2.801	-0.36	0.10	$2.55x10^{-4}$	---
rs116604207	*RBM5*	3	50147061	G/A	synonymous, non-coding-exon	unknown	0.453	11.12	0.79	0.22	$2.62x10^{-4}$	+++
rs41307740	*LBR*	1	225601614	C/A	intron	unknown	0.004	5.956	-0.65	0.18	$2.76x10^{-4}$	-?-
rs369623673	*ELAC2*	17	12899160	C/T	intron	unknown	0	6	0.93	0.26	$2.77x10^{-4}$	+?+
rs9997727	*C4orf50*	4	5969113	G/A	intron	unknown	0	0.613	-0.10	0.03	$2.85x10^{-4}$	---
rs150538926	*PDZD2*	5	32037369	C/T	synonymous	unknown	0	0.482	0.40	0.11	$2.87x10^{-4}$	+++
rs55868421	*TRIM5*	11	5688948	A/G	intron-near-splice, intron	unknown	0	5.512	0.44	0.12	$2.94x10^{-4}$	+++
rs323895	*ACY1, ABHD14A-ACY1*	3	52021316	A/G	intron	unknown	0	1.818	-0.36	0.10	$3.00x10^{-4}$	---
rs7188880	*RFWD3*	16	74664810	A/T	synonymous	unknown	1	10.78	-0.09	0.02	$3.01x10^{-4}$	---

Abbreviations: Chr = Chromosome; REF = Reference allele; ALT = Alternative allele; CADD = Combined Annotation Dependent Depletion; β = effect of the minor allele; SE = Standard Error.
* Position according to Hg19
** Seattleseq Annotation Database 138
‡ Order of cohorts in direction column: ERF, RS, BHF-FHS; Direction of effect represented by - (negative association) + (positive association) or ? (not available)

Segregation analysis of the variants within the chromosome 11q22.3 region in the ERF study showed that the variants segregate with shorter LTL in one family. We did not detect a specific disease that segregates in this family, which may be explained by the relatively young age of most of the carriers. In the replication analysis, we were not able to confirm the association of the nine variants associated with LTL in the RS and BHF-FHS, due to lack of association (six variants) or absence of the variant in the replication cohorts (three variants). However, it is possible this signal comes from variants that are family-specific and thus may not be transferable to the general population. Likewise, it may also be that the whole haplotype has an effect on telomere length, a haplotype that is most likely unique to the isolated ERF population. At this locus, the *ATM* gene is currently the most likely causative gene. Lastly, we performed a meta-analysis of the three cohorts. The top variant of the meta-analysis is located in *RPL8* (Ribosomal Protein L8). This is interesting as *RPL8*, together with six other *RPL* genes, was negatively associated with chronological age[49]. In this transcriptomic study, Peters et al. identified 1,497 genes whose expression level changes associated with age; this gene list includes *RPL8* and *ATM*. We additionally found suggestive evidence for an association of a novel rare variant in the known *RTEL1* locus. It would be interesting to perform a WES or whole-genome sequencing meta-analysis with larger sample size to increase the statistical power.

The advantage of our study is that we used a family-based setting that allowed us to show segregation of the variants located on chromosome 11q22.3 in a family. The ERF study population has shown a low immigration rate and a high level of inbreeding, which has increased the frequency of many rare alleles[55]. Another advantage is that all three studies used the standardized qPCR method to quantify LTL. However, there are also several limitations to our study. The first limitation is that our findings in the ERF study are not easy to generalize to the general population as these findings may be family-specific. However, we identified genes that are known to be related to telomere processes and thus are plausible candidate genes. The second limitation of our study may be that we used blood measurements of telomere length instead of tissue specific measurements, although previous studies have shown that mean telomere length in blood and other tissues are highly correlated[56,57] and, therefore, we expect this did not influence our findings. The third limitation is the difference in LTL distribution between the three studies, which may be explained by the different age distributions in the studies. Mean age is lowest in the ERF study (49 years), highest in the RS (75 years) and mean age in the BHF-FHS is in between (58 years). As LTL decreases significantly with age[3,30,58] and is associated with mortality[15,18-23], the variation in LTL becomes less with the aging populations. Mean LTL values were comparable between studies after adjusting LTL values for age and sex. To further standardize across the three cohorts, a z-transformation of the LTL values was performed resulting in comparable distributions with mean of zero and standard deviation of one. Nevertheless, these differences between studies, together with the small sample size of the BHF-FHS cohort, could potentially explain the lack of replication. The fourth limitation is that we were unable to calculate the effect of the variants on telomere length in base pairs

because of the quantification method of LTL in our study together with the z-transformation that was applied to the LTL values to standardize across the cohorts.

To conclude, this first study using WES data to search for rare genetic variants associated with LTL has identified interesting variants and genes associated with shorter LTL. Eight out of nine rare variants associated with LTL are located on chromosome 11q22.3 and all variants segregate within an ERF family. As we were not able to replicate findings, future studies should further investigate this region and the other genes identified in this study to confirm their involvement in telomere length regulation.

Materials and Methods

Study populations

Our discovery population consisted of participants from the family-based Erasmus Rucphen Family (ERF) study. The ERF study comprises approximately 3,000 inhabitants of a recent genetically isolated community in the Southwest of the Netherlands, studied as part of the Genetic Research in Isolated Population program[72]. All ERF participants are descendants of 22 founder couples who had at least six children baptized in the community church in the 18th century, or their spouses. Baseline data collection, including blood withdrawal, took place between 2002 and 2005. As the ERF population shows a low rate of immigration and a high level of inbreeding, the frequency of several rare alleles is increased in this population[72]. The ERF study was approved by the Medical Ethics Committee of the Erasmus Medical Center (MC), Rotterdam, The Netherlands. All participants provided written informed consents and all investigations were carried out in accordance with the Declaration of Helsinki.

The replication cohorts included participants from the Rotterdam Study (RS) and the British Heart Foundation Family Heart Study (BHF-FHS). The RS is a prospective cohort study ongoing since 1990 in the well-defined Ommoord district in Rotterdam, The Netherlands. The original RS cohort (RS-I) included 7,983 individuals of 55 years of age or over. At baseline, participants were interviewed at home and had an extensive set of examinations, which were repeated every 3-4 years[75]. The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC and by the Dutch Ministry of Health, Welfare and Sport. The Rotterdam Study has been entered into the Netherlands National Trial Register (NTR; www.trialregister.nl) and into the WHO International Clinical Trials Registry Platform (ICTRP; www.who.int/ictrp/network/primary/en/) under shared catalogue number NTR6831. All participants provided written informed consent to participate in the study and to have their information obtained from treating physicians.

The British Heart Foundation Family Heart study recruited families with at least two siblings diagnosed with premature (<66yrs) coronary artery disease (CAD) within the United Kingdom between 1998 and 2003. Full details are provided elsewhere[76,77].

Telomere length

Mean LTL values in the ERF study and the BHF-FHS study were measured using a qPCR method in all samples[78]. The measurements were performed in Leicester, United Kingdom, and details of the measurements were previously described[35,36]. In summary, mean LTL was measured in leukocytes and expressed as the ratio (T/S ratio) of telomere repeat length (T) to the copy number of a single-copy gene, *36B4* (S). Samples were quantified relative to a calibrator sample used on each run (DNA from the K562 cell line)[35].

In the RS, mean LTL values were also measured using a qPCR assay based on the method described by Cawthon[78] with minor modifications. For each sample the telomere and *36B4* assay were run in separate wells but in the same 384 wells PCR plate. Each reaction contained 5 ng DNA, 1 uM of each of the telomere primers (tel1b-forward: GGTTTGTTTGGGTTTGGGTTTGGGTTTGGGTTTGGGTT, tel2b-reverse: GGCTTGCCTTACCCTTACCCTTACCCTTACCCTTACCCT) or 250 nM of the *36B4* primers (36B4u-forward: CAGCAAGTGGGAAGGTGTAATCC, 36B4d-reverse: CCCATTCTATCATCAACGGGTACAA) and 1x Quantifast SYBR green PCR Mastermix (Qiagen). The reactions for both assays were performed in duplicate for each sample in a 7900HT machine (Applied Biosystems). Ct values and PCR efficiencies were calculated per plate using the MINER algorithm[79]. Duplicate Ct values that had a Coefficient of Variance (CV) of more than 1% were excluded from further analysis. Using the average Ct value per sample and the average PCR efficiency per plate the samples were quantified using the formula Q = 1/(1 + PCR eff)^Ct. The relative telomere length was calculated by dividing the Q of the telomere assay by the Q of the *36B4* assay. To validate the assay 96 random samples were run twice and the CV of that experiment was 4.5%.

Exome sequencing

The exomes of 1,336 ERF participants were sequenced at the Erasmus Center for Biomics of the department of Cell Biology, Erasmus MC, The Netherlands. The exomes of a randomly selected subset of 2,628 individuals from the RS-I population were sequenced at the Human Genotyping facility of the Department of Internal Medicine, Erasmus MC, The Netherlands. Details of the methods and quality control for ERF and RS are described elsewhere[80,81]. In total, 1,303 ERF participants and 1,257 RS participants had both exome sequence and LTL data available and were included in this analysis. A subset of the BHF-FHS, comprising of 674 unrelated individuals of Caucasian ancestry who had previously undergone exome sequencing as part of the Leicester Myocardial Infarction Study[82] and had LTL data available[35] were included in this analysis.

Statistical analyses

For each individual cohort quantitative trait association analysis was performed using Rare Variant tests (RVtests) software, which supports the analysis of related individuals[83]. Association analysis was performed using a score test, assuming an additive model, suitable for analysis with related and unrelated individuals. We applied a z-transformation of LTL

values for the three cohorts separately to standardize values across cohorts. All analyses were adjusted for age, sex and batch effects (if needed). Furthermore, we adjusted for familial relationships in ERF using the kinship matrix estimated from the genotyped data, while in the RS we corrected for the first four principal components as the fourth principal component was significantly associated with LTL. Only variants with a minor allele count ≥ 5 were included.

In the ERF study, we calculated the pairwise LD (r^2 and D') between the top eight variants on chromosome 11 that were significantly associated with LTL using the –ld command of PLINK 1.9 software[68] (www.cog-genomics.org/plink/1.9/). Additionally, we performed an inverse-variance weighted meta-analysis using METAL software[69]. In the analysis using data of the family-based ERF study, we corrected the significance thresholds for multiple testing using Bonferroni correction, resulting in a significance threshold of 1.42×10^{-7} (0.05/353,075). In the replication analysis, the multiple testing corrected *p*-value threshold was 0.025 (0.05/2 independent tests). In the meta-analysis, we adjusted for the number of variants tested, resulting in a significance threshold of 3.02×10^{-7} (0.05/165,311).

References

1. Blackburn, E. H. & Gall, J. G. A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. *J Mol Biol* **120**, 33-53, doi:10.1016/0022-2836(78)90294-2 (1978).
2 Blackburn, E. H. Telomere states and cell fates. *Nature* **408**, 53-56, doi:10.1038/35040500 (2000).
3 Lindsey, J., McGill, N. I., Lindsey, L. A., Green, D. K. & Cooke, H. J. In vivo loss of telomeric repeats with age in humans. *Mutat Res* **256**, 45-48, doi:10.1016/0921-8734(91)90032-7 (1991).
4 Allsopp, R. C. *et al.* Telomere length predicts replicative capacity of human fibroblasts. *Proc Natl Acad Sci U S A* **89**, 10114-10118 (1992).
5 Brouilette, S., Singh, R. K., Thompson, J. R., Goodall, A. H. & Samani, N. J. White cell telomere length and risk of premature myocardial infarction. *Arterioscler Thromb Vasc Biol* **23**, 842-846, doi:10.1161/01.ATV.0000067426.96344.32 (2003).
6 Brouilette, S. W. *et al.* Telomere length, risk of coronary heart disease, and statin treatment in the West of Scotland Primary Prevention Study: a nested case-control study. *Lancet* **369**, 107-114, doi:10.1016/S0140-6736(07)60071-3 (2007).
7 Fitzpatrick, A. L. *et al.* Leukocyte telomere length and cardiovascular disease in the cardiovascular health study. *Am J Epidemiol* **165**, 14-21, doi:10.1093/aje/kwj346 (2007).
8 Benetos, A. *et al.* Short telomeres are associated with increased carotid atherosclerosis in hypertensive subjects. *Hypertension* **43**, 182-185, doi:10.1161/01.HYP.0000113081.42868.f4 (2004).
9 Willeit, P. *et al.* Cellular aging reflected by leukocyte telomere length predicts advanced atherosclerosis and cardiovascular disease risk. *Arterioscler Thromb Vasc Biol* **30**, 1649-1656, doi: 10.1161/ATVBAHA.110.205492 (2010).
10 Haycock, P. C. *et al.* Leucocyte telomere length and risk of cardiovascular disease: systematic review and meta-analysis. *BMJ* **349**, g4227 (2014).
11 Hastie, N. D. *et al.* Telomere reduction in human colorectal carcinoma and with ageing. *Nature* **346**, 866-868, doi:10.1038/346866a0 (1990).
12 Willeit, P. *et al.* Telomere length and risk of incident cancer and cancer mortality. *JAMA* **304**, 69-75, doi:10.1001/jama.2010.897 (2010).
13 Willeit, P., Willeit, J., Kloss-Brandstatter, A., Kronenberg, F. & Kiechl, S. Fifteen-year follow-up of association between telomere length and incident cancer and cancer mortality. *JAMA* **306**, 42-44, doi:10.1001/jama.2011.901 (2011).
14 Bojesen, S. E. *et al.* Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat Genet* **45**, 371-384, 384e371-372, doi:10.1038/ng.2566 (2013).
15 Honig, L. S., Kang, M. S., Schupf, N., Lee, J. H. & Mayeux, R. Association of shorter leukocyte telomere repeat length with dementia and mortality. *Arch Neurol* **69**, 1332-1339, doi:10.1001/archneurol.2012.1541 (2012).
16 Martin-Ruiz, C. *et al.* Telomere length predicts poststroke mortality, dementia, and cognitive decline. *Ann Neurol* **60**, 174-180, doi:10.1002/ana.20869 (2006).
17 Grodstein, F. *et al.* Shorter telomeres may mark early risk of dementia: preliminary analysis of 62 participants from the nurses' health study. *PLoS One* **3**, e1590, doi:10.1371/journal.pone.0001590 (2008).
18 Marioni, R. E. *et al.* The epigenetic clock and telomere length are independently associated with chronological age and mortality. *Int J Epidemiol*, doi:10.1093/ije/dyw041 (2016).
19 Cawthon, R. M., Smith, K. R., O'Brien, E., Sivatchenko, A. & Kerber, R. A. Association between telomere length in blood and mortality in people aged 60 years or older. *Lancet* **361**, 393-395, doi:10.1016/S0140-6736(03)12384-7 (2003).
20 Kimura, M. *et al.* Telomere length and mortality: a study of leukocytes in elderly Danish twins. *Am J Epidemiol* **167**, 799-806, doi:10.1093/aje/kwm380 (2008).
21 Fitzpatrick, A. L. *et al.* Leukocyte telomere length and mortality in the Cardiovascular Health Study. *J Gerontol A Biol Sci Med Sci* **66**, 421-429, doi:10.1093/gerona/glq224 (2011).
22 Rode, L., Nordestgaard, B. G. & Bojesen, S. E. Peripheral blood leukocyte telomere length and mortality among 64,637 individuals from the general population. *J Natl Cancer Inst* **107**, djv074, doi:10.1093/jnci/djv074 (2015).
23 Deelen, J. *et al.* Leukocyte telomere length associates with prospective mortality independent of immune-related parameters and known genetic markers. *Int J Epidemiol* **43**, 878-886, doi:10.1093/ije/dyt267 (2014).
24 Needham, B. L. *et al.* Leukocyte telomere length and mortality in the National Health and Nutrition Examination Survey, 1999-2002. *Epidemiology* **26**, 528-535, doi:10.1097/EDE.0000000000000299 (2015).

25 Martin-Ruiz, C. M., Gussekloo, J., van Heemst, D., von Zglinicki, T. & Westendorp, R. G. Telomere length in white blood cells is not associated with morbidity or mortality in the oldest old: a population-based study. *Aging Cell* **4**, 287-290, doi:10.1111/j.1474-9726.2005.00171.x (2005).

26 Harris, S. E. *et al.* The association between telomere length, physical health, cognitive ageing, and mortality in non-demented older people. *Neurosci Lett* **406**, 260-264, doi:10.1016/j.neulet.2006.07.055 (2006).

27 Strandberg, T. E. *et al.* Association of telomere length in older men with mortality and midlife body mass index and smoking. *J Gerontol A Biol Sci Med Sci* **66**, 815-820, doi:10.1093/gerona/glr064 (2011).

28 Houben, J. M., Giltay, E. J., Rius-Ottenheim, N., Hageman, G. J. & Kromhout, D. Telomere length and mortality in elderly men: the Zutphen Elderly Study. *J Gerontol A Biol Sci Med Sci* **66**, 38-44, doi:10.1093/gerona/glq164 (2011).

29 Njajou, O. T. *et al.* Association between telomere length, specific causes of death, and years of healthy life in health, aging, and body composition, a population-based cohort study. *J Gerontol A Biol Sci Med Sci* **64**, 860-864, doi:10.1093/gerona/glp061 (2009).

30 Slagboom, P. E., Droog, S. & Boomsma, D. I. Genetic determination of telomere size in humans: a twin study of three age groups. *Am J Hum Genet* **55**, 876-882 (1994).

31 Bischoff, C. *et al.* The heritability of telomere length among the elderly and oldest-old. *Twin Res Hum Genet* **8**, 433-439, doi:10.1375/183242705774310141 (2005).

32 Andrew, T. *et al.* Mapping genetic loci that determine leukocyte telomere length in a large sample of unselected female sibling pairs. *Am J Hum Genet* **78**, 480-486, doi: 10.1086/500052 (2006).

33 Vasa-Nicotera, M. *et al.* Mapping of a major locus that determines telomere length in humans. *Am J Hum Genet* **76**, 147-151, doi:10.1086/426734 (2005).

34 Broer, L. *et al.* Meta-analysis of telomere length in 19,713 subjects reveals high heritability, stronger maternal inheritance and a paternal age effect. *Eur J Hum Genet* **21**, 1163-1168, doi:10.1038/ejhg.2012.303 (2013).

35 Codd, V. *et al.* Common variants near TERC are associated with mean telomere length. *Nat Genet* **42**, 197-199, doi:10.1038/ng.532 (2010).

36 Codd, V. *et al.* Identification of seven loci affecting mean telomere length and their association with disease. *Nat Genet* **45**, 422-427, 427e421-422, doi:10.1038/ng.2528 (2013).

37 Mangino, M. *et al.* Genome-wide meta-analysis points to CTC1 and ZNF676 as genes regulating telomere homeostasis in humans. *Hum Mol Genet* **21**, 5385-5394, doi:10.1093/hmg/dds382 (2012).

38 Pooley, K. A. *et al.* A genome-wide association scan (GWAS) for mean telomere length within the COGS project: identified loci show little association with hormone-related cancer risk. *Hum Mol Genet* **22**, 5056-5064, doi:10.1093/hmg/ddt355 (2013).

39 Levy, D. *et al.* Genome-wide association identifies OBFC1 as a locus involved in human leukocyte telomere biology. *Proc Natl Acad Sci U S A* **107**, 9293-9298, doi:10.1073/pnas.0911494107 (2010).

40 Mangino, M. *et al.* DCAF4, a novel gene associated with leucocyte telomere length. *J Med Genet* **52**, 157-162, doi:10.1136/jmedgenet-2014-102681 (2015).

41 Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747-753, doi: 10.1038/nature08494 (2009).

42 McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-1283, doi: 10.1038/ng.3643 (2016).

43 Greenwell, P. W. *et al.* TEL1, a gene involved in controlling telomere length in S. cerevisiae, is homologous to the human ataxia telangiectasia gene. *Cell* **82**, 823-829, doi:0092-8674(95)90479-4 [pii] (1995).

44 Tong, A. S. *et al.* ATM and ATR Signaling Regulate the Recruitment of Human Telomerase to Telomeres. *Cell Rep* **13**, 1633-1646, doi:10.1016/j.celrep.2015.10.041 (2015).

45 Pandita, T. K. ATM function and telomere stability. *Oncogene* **21**, 611-618, doi:10.1038/sj.onc.1205060 (2002).

46 Lee, S. S., Bohrson, C., Pike, A. M., Wheelan, S. J. & Greider, C. W. ATM Kinase Is Required for Telomere Elongation in Mouse and Human Cells. *Cell Rep* **13**, 1623-1632, doi:10.1016/j.celrep.2015.10.035 (2015).

47 Savitsky, K. *et al.* A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science* **268**, 1749-1753 (1995).

48 Metcalfe, J. A. *et al.* Accelerated telomere shortening in ataxia telangiectasia. *Nat Genet* **13**, 350-353, doi:10.1038/ng0796-350 (1996).

49 Peters, M. J. *et al.* The transcriptional landscape of age in human peripheral blood. *Nat Commun* **6**, 8570, doi:10.1038/ncomms9570 (2015).

50 Barrett, J. H. *et al.* Genome-wide association study identifies three new melanoma susceptibility loci. *Nat Genet* **43**, 1108-1113, doi:10.1038/ng.959 (2011).

51 Helgason, H. *et al.* Loss-of-function variants in ATM confer risk of gastric cancer. *Nat Genet* **47**, 906-910, doi:10.1038/ng.3342 (2015).

52 Scelo, G. *et al.* Genome-wide association study identifies multiple risk loci for renal cell carcinoma. *Nat Commun* **8**, 15724, doi:10.1038/ncomms15724 (2017).

53 Ransohoff, K. J. *et al.* Two-stage genome-wide association study identifies a novel susceptibility locus associated with melanoma. *Oncotarget* **8**, 17586-17592, doi:10.18632/oncotarget.15230 (2017).

54 Dorajoo, R. *et al.* Loci for human leukocyte telomere length in the Singaporean Chinese population and trans-ethnic genetic studies. *Nat Commun* **10**, 2491, doi:10.1038/s41467-019-10443-2 (2019).

55 Pardo, L. M., MacKay, I., Oostra, B., van Duijn, C. M. & Aulchenko, Y. S. The effect of genetic drift in a young genetically isolated population. *Annals of Human Genetics* **69**, 288-295, doi:DOI 10.1046/j.1529-8817.2005.00162.x (2005).

56 Okuda, K. *et al.* Telomere length in the newborn. *Pediatr Res* **52**, 377-381, doi:10.1203/00006450-200209000-00012 (2002).

57 Wilson, W. R. *et al.* Blood leucocyte telomere DNA content predicts vascular telomere DNA content in humans with and without vascular disease. *Eur Heart J* **29**, 2689-2694, doi:10.1093/eurheartj/ehn386 (2008).

58 Blackburn, E. H. Switching and signaling at the telomere. *Cell* **106**, 661-673, doi:10.1016/S0092-8674(01)00492-5 (2001).

59 Ikram, M. A. *et al.* The Rotterdam Study: 2018 update on objectives, design and main results. *Eur J Epidemiol* **32**, 807-850, doi:10.1007/s10654-017-0321-4 (2017).

60 Samani, N. J. *et al.* A genomewide linkage study of 1,933 families affected by premature coronary artery disease: The British Heart Foundation (BHF) Family Heart Study. *Am J Hum Genet* **77**, 1011-1020, doi:10.1086/498653 (2005).

61 Samani, N. J. *et al.* Genomewide association analysis of coronary artery disease. *N Engl J Med* **357**, 443-453, doi:10.1056/NEJMoa072366 (2007).

62 Cawthon, R. M. Telomere measurement by quantitative PCR. *Nucleic Acids Res* **30**, e47 (2002).

63 Zhao, S. & Fernald, R. D. Comprehensive algorithm for quantitative real-time polymerase chain reaction. *J Comput Biol* **12**, 1047-1064, doi:10.1089/cmb.2005.12.1047 (2005).

64 Amin, N. *et al.* Exome-sequencing in a large population-based study reveals a rare Asn396Ser variant in the LIPG gene associated with depressive symptoms. *Mol Psychiatry* **22**, 537-543, doi:10.1038/mp.2016.101 (2016).

65 van Rooij, J. G. J. *et al.* Population-specific genetic variation in large sequencing data sets: why more data is still better. *Eur J Hum Genet* **25**, 1173-1175, doi:10.1038/ejhg.2017.110 (2017).

66 Khera, A. V. *et al.* Association of Rare and Common Variation in the Lipoprotein Lipase Gene With Coronary Artery Disease. *JAMA* **317**, 937-946, doi:10.1001/jama.2017.0972 (2017).

67 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**, 1423-1426, doi:10.1093/bioinformatics/btw079 (2016).

68 Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7, doi:10.1186/s13742-015-0047-8 (2015).

69 Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191, doi:10.1093/bioinformatics/btq340 (2010).

CHAPTER 2.2

Metabolomics reveals a link between homocysteine and lipid metabolism and leukocyte telomere length: the ENGAGE consortium

Ashley van der Spek*, Linda Broer*, Harmen H.M. Draisma*, René Pool*, Eva Albrecht, Marian Beekman, Massimo Mangino, Mait Raag, Dale R. Nyholt, Harish K. Dharuri, Veryan Codd, Najaf Amin, Eco J.C. de Geus, Joris Deelen, Ayse Demirkan, Idil Yet, Krista Fischer, Toomas Haller, Anjali K. Henders, Aaron Isaacs, Sarah E. Medland, Grant W. Montgomery, Simon P. Mooijaart, Konstantin Strauch, H. Eka D. Suchiman, Anika A.M. Vaarhorst, Diana van Heemst, Rui Wang-Sattler, John B. Whitfield, Gonneke Willemsen, Margaret J. Wright, Nicholas G. Martin, Nilesh J. Samani, Andres Metspalu, P. Eline Slagboom, Tim D. Spector, Dorret I. Boomsma#, Cornelia M. van Duijn#, Christian Gieger#

* These authors contributed equally to this manuscript

# These senior authors contributed equally to this manuscript

This chapter was published in *Scientific Reports*. 2019 Aug 12;9(1):11623.

The supplemental information for this paper is available at https://doi.org/10.1038/s41598-019-47282-6.

Abstract

Telomere shortening has been associated with multiple age-related diseases such as cardiovascular disease, diabetes, and dementia. However, the biological mechanisms responsible for these associations remain largely unknown. In order to gain insight into the metabolic processes driving the association of leukocyte telomere length (LTL) with age-related diseases, we investigated the association between LTL and serum metabolite levels in 7,853 individuals from seven independent cohorts. LTL was determined by quantitative polymerase chain reaction and the levels of 131 serum metabolites were measured with mass spectrometry in biological samples from the same blood draw. With partial correlation analysis, we identified six metabolites that were significantly associated with LTL after adjustment for multiple testing: lysophosphatidylcholine acyl C17:0 (lysoPC a C17:0, *p*-value = 7.1×10^{-6}), methionine (*p*-value = 9.2×10^{-5}), tyrosine (*p*-value = 2.1×10^{-4}), phosphatidylcholine diacyl C32:1 (PC aa C32:1, *p*-value = 2.4×10^{-4}), hydroxypropionylcarnitine (C3-OH, *p*-value = 2.6×10^{-4}), and phosphatidylcholine acyl-alkyl C38:4 (PC ae C38:4, *p*-value = 9.0×10^{-4}). Pathway analysis showed that the three phosphatidylcholines and methionine are involved in homocysteine metabolism and we found supporting evidence for an association of lipid metabolism with LTL. In conclusion, we found longer LTL associated with higher levels of lysoPC a C17:0 and PC ae C38:4, and with lower levels of methionine, tyrosine, PC aa C32:1, and C3-OH. These metabolites have been implicated in inflammation, oxidative stress, homocysteine metabolism, and in cardiovascular disease and diabetes, two major drivers of morbidity and mortality.

Introduction

Telomeres are located at the ends of chromosomes and protect against spontaneous DNA damage, thus preserving genomic integrity[1,2]. The progressive shortening of telomere length with each subsequent cell division underlies the so-called mitotic clock, i.e. the limited replicative capacity of a cell[3]. Replicative senescence and subsequent cell death occur when the mean telomere length reaches a critical value and telomere length is therefore seen as a marker for biological age[4-6]. Short leukocyte telomere length (LTL) has been associated with age[5,7-9] and multiple age-related diseases such as cardiovascular disease[10-15], diabetes[10,16,17] and dementia[18,19]. Short LTL has also been associated with mortality[20-27], although not all studies support this association[28-33]. However, the biological mechanisms underlying the associations of LTL with age-related diseases and mortality are still largely unknown.

Longevity in humans has previously been associated with various metabolic traits in the elderly, including traits related to blood pressure and lipids, suggesting that changes at the metabolic level are key features in longevity[34-36]. Metabolic profiles have been associated with age and various age-related diseases, such as type 2 diabetes, atherosclerosis, cancer, and Alzheimer's Disease[37-44]. Only a few studies investigated the association of metabolic markers with LTL, all using untargeted metabolomics[45-47]. One study focused on metabolic markers involved in aging and early development in 6,055 individuals included in the TwinsUK registry[45]. Although a combined set of 22 metabolites was strongly correlated with age and age-related traits, the individual metabolites were not significantly associated with LTL[45]. Another study identified 19 metabolites associated with LTL in a small sample of American Indians (n = 423)[46]. The most recent study was conducted in 3,511 females from the TwinsUK registry, reporting associations of five metabolites with LTL. These include gamma-glutamyltyrosine, gamma-glutamylphenylalanine, 1-stearoylglycerophosphoinositol, 1-palmitoylglycerophosphoinositol, and 4-vinylphenol sulfate[47].

In the current study, we used a standardized targeted metabolomics approach to investigate the association between LTL and serum metabolites of key biochemical pathways in the largest sample so far consisting of 7,853 individuals from seven independent population-based cohorts from Europe and Australia. We further aimed to disentangle the metabolic pathways that are represented by the metabolites significantly associated with LTL.

Methods

Cohort descriptions and measurements of LTL and metabolites

The cohorts included in this study are the Cooperative Health Research in the Region of Augsburg (KORA) study, the Netherlands Twin Register (NTR), the Estonian Genome Center University of Tartu (EGCUT) study, the TwinsUK cohort, the Erasmus Rucphen Family (ERF) study, the Leiden Longevity Study (LLS), and the Queensland Institute of Medical Research

(QIMR) study, all part of the ENGAGE consortium. Details on the individual cohorts as well as details on the LTL quantitative polymerase chain reaction measurements and the metabolites as measured with the Absolute*IDQ*™ p150 kit (BIOCRATES Life Sciences AG, Innsbruck, Austria) are provided in the **Supplementary Materials**. In summary, both LTL and metabolite concentrations were measured in the same laboratories according to a common protocol, using blood samples taken at the same time point. To ensure good data quality, each metabolite had to meet three criteria for inclusion in the study: 1) coefficient of variation (CV) not exceeding 25%; 2) less than 5% missing values; 3) median of metabolite concentrations above the limit of detection. This quality control was performed per cohort. **Supplementary Table 1** describes the reasons for exclusion of metabolites from the analysis for each cohort. The metabolites measured include hexoses (H1), amino acids (AA), acyl-carnitines (AC), sphingomyelins (SMs), diacyl phosphatidylcholines (PC aa), acyl-alkyl-phosphatidylcholines (PC ae) and lysophosphatidylcholines (lysoPC).

Written informed consent was obtained from all study participants. The study protocol was approved by the medical ethics boards of the Helmholtz Zentrum München, VUmc Amsterdam, University of Tartu, St. Thomas' Hospital London, Erasmus MC Rotterdam, LUMC Leiden, and Queensland Institute of Medical Research and all investigations were carried out in accordance with the Declaration of Helsinki.

Statistical analysis

To standardize LTL measurements across cohorts we Z-transformed the LTL values. Metabolite concentration values were natural log-transformed to attain a better approximation of the normal distribution. We performed partial correlation analysis per cohort, adjusting for age and sex, and if necessary for family relationships (model 1). In the extended model (model 2) we additionally adjusted for body mass index (BMI). We performed a sensitivity analysis to test the robustness of the results and repeated the meta-analysis excluding studies with low sample size or high mean LTL values.

A multiple testing-corrected statistical significance threshold for association of metabolite concentrations with LTL was defined at the meta-analysis level. We accounted for multiple testing by Bonferroni correction based on the effective number of independent variables (VeffLi) in the metabolite concentration data[48] (http://gump.qimr.edu.au/general/daleN/matSpDlite/). The VeffLi value was determined using the correlation matrix of the quality controlled and log-transformed metabolomics data in the ERF and NTR cohorts, yielding a VeffLi (representing the number of independent metabolites) of 46 for both cohorts. This resulted in a Bonferroni corrected significance threshold of p-value < 1.1×10^{-3} (= 0.05/46).

Pathway analysis of the associated metabolites

For interpretation of the observed associations, we followed two bioinformatics approaches in parallel. First, we employed a bioinformatics pipeline based on a workflow management software tool called "Taverna" (http://www.taverna.org.uk)[49] to determine if the metabolites associated with LTL share a network space within two reaction steps.

We took as input all possible pairs of significantly associated metabolites. In this pipeline, all the reactions within a radius of two steps in the reaction space of the first metabolite were obtained from the KEGG database[50]. The second metabolite is searched against the substrates and the products of the reactions obtained in the previous step. The presence of the second metabolite in any of the reaction steps is an indication that the two metabolites participate in reactions within two steps of each other. In the final step, the path between the two metabolites is returned to the user. In order to prevent non-specific connections, an intermediate step filters out hub metabolites such as ATP, ADP, and NADP. Next, we used the function "heatmap.2" included in the R package "gplots" (https://CRAN.R-project.org/package=gplots) to prepare a heat map of the correlation of the individual metabolites with LTL in models 1 and 2. For this analysis, default functions for clustering were used (distance measure: euclidean distance). We also derived a correlation matrix for the significantly associated metabolites in ERF and visualized this in a correlogram using the R package "corrplot"[51].

Results

General characteristics of the study populations are provided in **Table 1**. The study covers a wide age range, with the mean age of the participants ranging from 18.4 to 62.9 years. Most studies had approximately equal numbers of males and females, except for NTR (33% female) and TwinsUK (only females). Mean LTL ranged from 1.44 (LLS) to 3.58 (TwinsUK). BMI was on average between 25.2 and 27.6 kg/m^2, but was unavailable at the time of metabolite and LTL assessment in the QIMR study.

Out of the 131 metabolites that passed quality control, 27 showed at least nominally significant correlation (p-value < 0.05) with LTL when adjusting for age and sex in model 1 (**Table 2**). Six metabolites surpassed the multiple-testing corrected significance threshold (p-value < 1.1×10^{-3}). Five of these metabolites were consistently associated with LTL in the same direction in at least five out of seven studies: lysophosphatidylcholine acyl C17:0 (lysoPC a C17:0, r = 0.05, p-value = 7.1×10^{-6}) and phosphatidylcholine acyl-alkyl C38:4 (PC ae C38:4, r = 0.04, p-value = 9.0×10^{-4}) were positively associated with LTL, while methionine (Met, r = -0.04, p-value = 9.2×10^{-5}), tyrosine (Tyr, r = -0.04, p-value = 2.1×10^{-4}), and phosphatidylcholine diacyl C32:1 (PC aa C32:1, r = -0.04, p-value = 2.4×10^{-4}) were negatively associated with LTL. Although hydroxypropionylcarnitine (C3-OH, r = -0.10, p-value = 2.6×10^{-4}) was also found negatively associated with LTL, this effect was only based on two out of seven studies. Additionally adjusting for BMI in model 2 had limited effect on the correlation coefficients of the six significant metabolites in model 1 and all metabolites except PC ae C38:4 remained significantly associated with LTL (**Table 2**). The summary statistics for all metabolite-LTL correlations for both models can be found in **Supplementary Table 2**. Study-specific results for the age- and sex-adjusted model are provided in **Supplementary Table 3**.

Table 1. General characteristics of study populations.

				LTL			Age			BMI		
	n	n_BMI	% female	mean	SD	95% CI	mean	SD	95% CI	mean	SD	95% CI
KORA	3003	2988	51.8	1.85	0.33	1.84 - 1.86	56.08	13.25	55.61 - 56.55	27.61	4.80	27.44 - 27.78
NTR	1314	1307	33.3	2.54	0.47	2.51 - 2.57	50.60	14.13	49.84 - 51.36	25.97	3.80	25.76 - 26.18
EGCUT	1084	1084	50.3	1.90	0.30	1.88 - 1.92	37.78	15.70	36.85 - 38.71	25.16	4.56	24.89 - 25.43
TwinsUK	810	810	100.0	3.58	0.64	3.54 - 3.62	53.72	10.76	53.10 - 54.34	26.44	5.35	26.13 - 26.75
ERF	806	806	53.7	1.79	0.37	1.76 - 1.82	47.76	13.97	46.80 - 48.72	27.17	4.81	26.84 - 27.50
LLS	643	643	50.1	1.44	0.27	1.42 - 1.46	62.91	6.61	62.40 - 63.42	26.65	4.01	26.34 - 26.96
QIMR	193	0	48.2	3.43	0.56	3.35 - 3.51	18.44	12.65	16.66 - 20.22	NA	NA	NA

Abbreviations: n = number of participants with data available on metabolites, telomere length, age, and sex; n_BMI = number of participants with data available on metabolites, telomere length, age, sex, and BMI; LTL = leukocyte telomere length; SD = standard deviation. LTL as a ratio of telomere repeat length to copy number of the single copy gene *36B4*; Age in years; BMI in kg/m^2.

Table 2. Partial correlation meta-analysis results of LTL and metabolites (*p*-value in model 1 < 0.05).

	Model 1: age + sex					Model 2: age + sex + BMI					
Metabolite	**n**	**direction***	**r**	**p-value**	**FDR**	**n**	**direction***	**r**	**p-value**	**FDR**	**Metabolite full name**
lysoPC a C17:0	7853	++-++++	0.05	**$7.1x10^{-6}$**	$9.3x10^{-4}$	7638	++-+++	0.04	**$4.7x10^{-4}$**	$6.9x10^{-3}$	Lysophosphatidylcholine acyl C17:0
Met	7852	------+	-0.04	**$9.2x10^{-5}$**	$6.0x10^{-3}$	7637	------	-0.05	**$7.5x10^{-5}$**	$9.3x10^{-4}$	Methionine
Tyr	7047	----?-+	-0.04	**$2.1x10^{-4}$**	$6.9x10^{-3}$	6832	----?-	-0.04	**$8.9x10^{-4}$**	$6.9x10^{-3}$	Tyrosine
PC aa C32:1	7851	----+-+	-0.04	**$2.4x10^{-4}$**	$6.9x10^{-3}$	7636	----++	-0.04	**$3.4x10^{-4}$**	$6.9x10^{-3}$	Phosphatidylcholine diacyl C32:1
C3-OH	1449	????--?	-0.10	**$2.6x10^{-4}$**	$6.9x10^{-3}$	1449	????--	-0.10	**$2.7x10^{-4}$**	$6.0x10^{-3}$	Hydroxypropionylcarnitine
PC ae C38:4	7853	+-+-+++	0.04	**$9.0x10^{-4}$**	$2.0x10^{-2}$	7638	+-+-++	0.03	$4.7x10^{-3}$	$3.1x10^{-2}$	Phosphatidylcholine acyl-alkyl C38:4
PC ae C40:3	7853	+++++++	0.04	$1.6x10^{-3}$	$3.0x10^{-2}$	7638	++++++	0.03	$8.5x10^{-3}$	$3.8x10^{-2}$	Phosphatidylcholine acyl-alkyl C40:3
PC ae C40:5	7853	+-+++++	0.04	$1.9x10^{-3}$	$3.1x10^{-2}$	7638	+-++++	0.03	$1.9x10^{-2}$	$1.0x10^{-1}$	Phosphatidylcholine acyl-alkyl C40:5
SM C20:2	7853	+-+-+++	0.03	$2.4x10^{-3}$	$3.5x10^{-2}$	7638	+-+-++	0.03	$2.7x10^{-3}$	$2.0x10^{-2}$	Sphingomyeline C20:2
C9	5262	+??-++?	0.04	$2.9x10^{-3}$	$3.8x10^{-2}$	5247	+??-++	0.04	$1.1x10^{-2}$	$7.5x10^{-2}$	Nonaylcarnitine
PC ae C40:4	7853	+0+++++	0.03	$4.1x10^{-3}$	$4.9x10^{-2}$	7638	+-++++	0.03	$2.8x10^{-2}$	$1.1x10^{-1}$	Phosphatidylcholine acyl-alkyl C40:4
PC aa C38:3	7852	----+++	-0.03	$7.0x10^{-3}$	$7.5x10^{-2}$	7637	----++	-0.02	$3.5x10^{-2}$	$1.3x10^{-1}$	Phosphatidylcholine diacyl C38:3
PC ae C36:1	7852	+-+-+++	0.03	$7.4x10^{-3}$	$7.5x10^{-2}$	7637	+-+-++	0.02	$5.8x10^{-2}$	$2.1x10^{-1}$	Phosphatidylcholine acyl-alkyl C36:1
PC aa C36:1	7850	----+++	-0.03	$9.3x10^{-3}$	$8.3x10^{-2}$	7635	----++	-0.03	$4.0x10^{-3}$	$3.0x10^{-2}$	Phosphatidylcholine diacyl C36:1
PC ae C40:6	7853	+-+-+++	0.03	$9.5x10^{-3}$	$8.3x10^{-2}$	7638	+-+-++	0.02	$8.5x10^{-2}$	$2.4x10^{-1}$	Phosphatidylcholine acyl-alkyl C40:6
SM (OH) C16:1	7047	+-+-?++	0.03	$1.1x10^{-2}$	$9.4x10^{-2}$	6832	+-+-?+	0.02	$5.3x10^{-2}$	$1.6x10^{-1}$	Hydroxysphingomyeline C16:1
C2	7853	--+---+	-0.03	$1.4x10^{-2}$	$1.0x10^{-1}$	7638	--+---	-0.03	$1.5x10^{-2}$	$8.3x10^{-2}$	Acetylcarnitine
PC ae C36:2	7853	++--+++	0.03	$1.4x10^{-2}$	$1.0x10^{-1}$	7638	+---++	0.02	$1.3x10^{-1}$	$2.4x10^{-1}$	Phosphatidylcholine acyl-alkyl C36:2
PC ae C38:3	7852	+++-+++	0.03	$1.7x10^{-2}$	$1.1x10^{-1}$	7637	+++-++	0.02	$3.6x10^{-2}$	$1.3x10^{-1}$	Phosphatidylcholine acyl-alkyl C38:3
PC aa C42:0	7853	+++++++	0.03	$2.0x10^{-2}$	$1.3x10^{-1}$	7638	+++-++	0.02	$8.9x10^{-2}$	$2.4x10^{-1}$	Phosphatidylcholine diacyl C42:0
PC aa C32:0	7853	----+++	-0.03	$2.1x10^{-2}$	$1.3x10^{-1}$	7638	----++	-0.03	$5.9x10^{-3}$	$3.5x10^{-2}$	Phosphatidylcholine diacyl C32:0
PC aa C40:5	7849	----+++	-0.03	$2.2x10^{-2}$	$1.3x10^{-1}$	7634	----++	-0.03	$1.7x10^{-2}$	$8.3x10^{-2}$	Phosphatidylcholine diacyl C40:5
PC aa C38:1	836	?????++	0.08	$2.6x10^{-2}$	$1.5x10^{-1}$	643	?????+	0.05	$1.7x10^{-1}$	$2.4x10^{-1}$	Phosphatidylcholine diacyl C38:1

Table 2. *Continued*

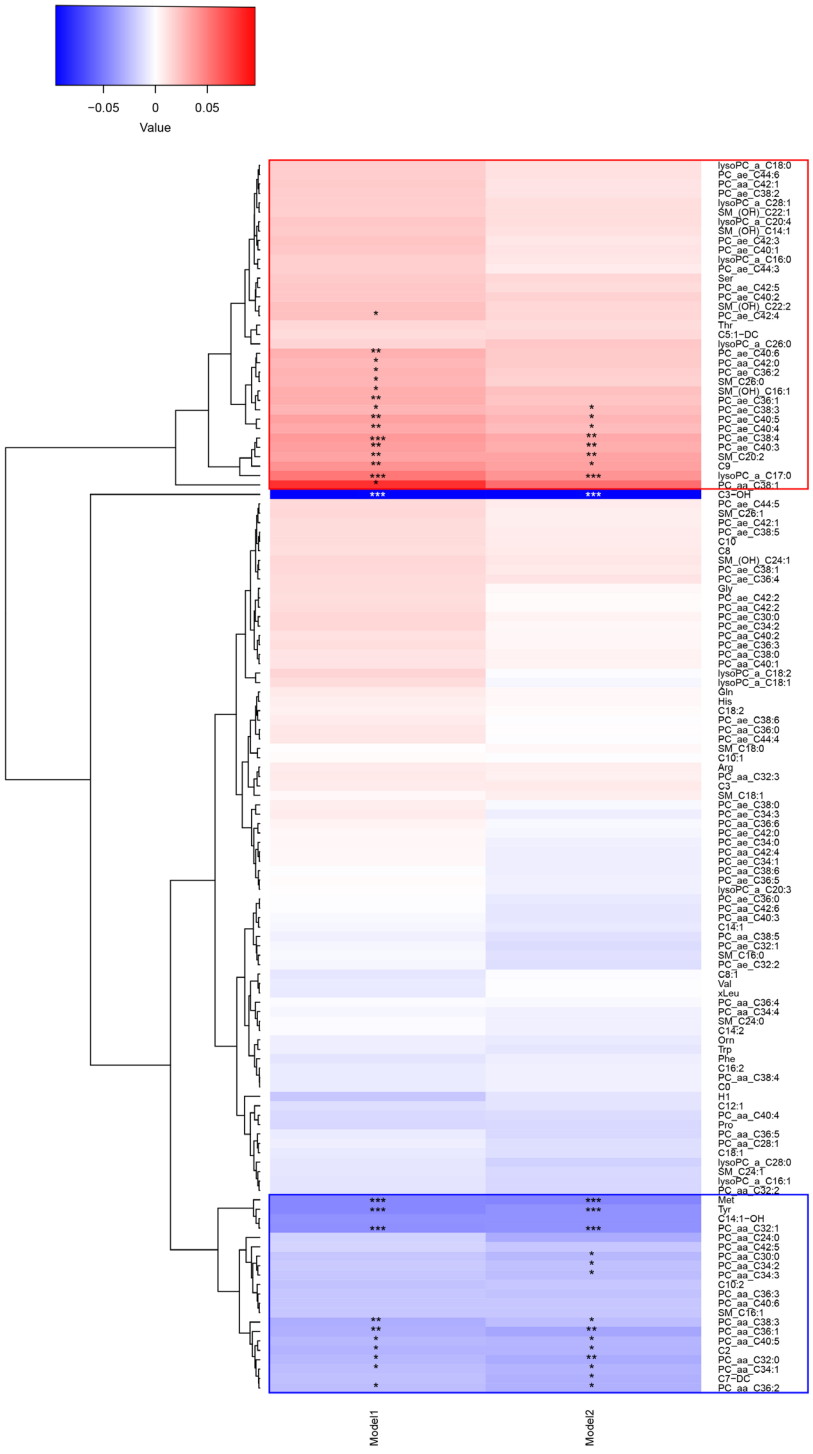
	Model 1: age + sex					Model 2: age + sex + BMI					
Metabolite	**n**	**direction***	**r**	**p-value**	**FDR**	**n**	**direction***	**r**	**p-value**	**FDR**	**Metabolite full name**
PC aa C36:2	7853	-----++	-0.02	$3.0x10^{-2}$	$1.6x10^{-1}$	7638	-----+	-0.03	$1.0x10^{-2}$	$4.9x10^{-2}$	Phosphatidylcholine diacyl C36:2
PC aa C34:1	7852	----+++	-0.02	$3.1x10^{-2}$	$1.6x10^{-1}$	7637	----++	-0.03	$1.3x10^{-2}$	$7.5x10^{-2}$	Phosphatidylcholine diacyl C34:1
PC ae C42:4	7853	+-+++++	0.02	$3.5x10^{-2}$	$1.8x10^{-1}$	7638	+-++++	0.01	$2.1x10^{-1}$	$2.5x10^{-1}$	Phosphatidylcholine acyl-alkyl C42:4
SM C26:0	5454	+??-+++	0.03	$4.3x10^{-2}$	$2.1x10^{-1}$	5246	+??-++	0.02	$2.3x10^{-1}$	$2.6x10^{-1}$	Sphingomyeline C26:0

Abbreviations: n = number of participants, r = correlation coefficient, BMI = body mass index, FDR = false discovery rate.
*Order of cohorts in direction column: KORA, NTR, EGCUT, TwinsUK, ERF, LLS, QIMR; Direction of effect represented by - (negative correlation) + (positive correlation) or ? (not included) for each study. Bold *p*-values: associations surpassing significance threshold (*p*-value < 0.001).

We next conducted a sensitivity analysis to determine whether the analyses were driven by a single cohort. We removed two cohorts from the analysis: the TwinsUK cohort, which had a high mean LTL ($\bar{x}_{LTL}$= 3.58) and the QIMR cohort, which had a small sample size (N = 193) and was on average younger than the other cohorts. After excluding the TwinsUK cohort from the meta-analysis, all metabolites remained significantly associated with LTL, except for PC aa C32:1 (*p*-value = $1.1x10^{-3}$) (**Supplementary Table 4**). All metabolites, except for PC ae C38:4 (*p*-value = $1.8x10^{-3}$), remained significantly associated with LTL after excluding the QIMR cohort from the meta-analysis (**Supplementary Table 5**).

To explore to which extent the various metabolites cluster, we constructed a heat map based on the correlation of each individual metabolite with LTL in both model 1 and model 2 (**Figure 1**). The heat map shows two clusters of which one (hereafter referred to as "cluster 1") includes lysoPC a C17:0, PC ae C38:4, and a series of PC ae metabolites positively associated with LTL, while the second cluster (hereafter referred to as "cluster 2") includes methionine, tyrosine, PC aa C32:1, and a series of PC aa metabolites negatively associated with LTL. **Figure 1** further shows that C3-OH is relatively dissimilar from all other metabolites. A correlogram of the six metabolites associated with LTL after correction for multiple testing is presented in **Figure 2** and shows a positive correlation of methionine with the three PC metabolites. LysoPC a C17:0 and PC ae C38:4 (cluster 1) are most strongly correlated in **Figure 2**. Methionine and tyrosine are highly correlated with each other and both amino acids are correlated to PC aa C32.1 (**Figure 2**).

Pathway analysis using the Taverna workflow showed that phosphatidylcholines (lysoPC a C17:0, PC aa C32:1, and PC ae C38:4) and methionine are involved in homocysteine metabolism. Homocysteine is the intermediate product in the conversion of the amino acid methionine to cysteine, a precursor of the antioxidant glutathione (**Figure 3**). Briefly, PC is a precursor of choline which is oxidized to betaine. Betaine is used to convert homocysteine to methionine. Methionine is first converted to S-adenosylmethionine followed by demethylation to S-adenosylhomocysteine (SAH). Next, hydrolysis of SAH forms homocysteine, which can either be re-methylated into methionine (transmethylation cycle) or metabolized to cysteine (transsulfuration pathway) as shown in **Figure 3**.


−0.05
0
0.05
Value
lysoPC_a_C18:0
PC_ae_C44:6
PC_aa_C42:1
PC_ae_C38:2
lysoPC_a_C28:1
SM_(OH)_C22:1
lysoPC_a_C20:4
SM_(OH)_C14:1
PC_ae_C42:3
PC_ae_C40:1
lysoPC_a_C16:0
PC_ae_C44:3
Ser
PC_ae_C42:5
PC_ae_C40:2
SM_(OH)_C22:2
PC_ae_C42:4
Thr
C5:1−DC
lysoPC_a_C26:0
PC_ae_C40:6
PC_aa_C42:0
PC_ae_C36:2
SM_C26:0
SM_(OH)_C16:1
PC_ae_C36:1
PC_ae_C38:3
PC_ae_C40:5
PC_ae_C40:4
PC_ae_C38:4
PC_ae_C40:3
SM_C20:2
C9
lysoPC_a_C17:0
PC_aa_C36:1
C3−OH
PC_ae_C44:5
SM_C28:1
PC_ae_C42:1
PC_ae_C38:5
C10
C8
SM_(OH)_C24:1
PC_ae_C38:1
PC_ae_C36:4
Gly
PC_ae_C42:2
PC_aa_C42:2
PC_ae_C30:0
PC_ae_C34:2
PC_aa_C40:2
PC_ae_C36:3
PC_aa_C38:0
PC_aa_C40:1
lysoPC_a_C18:2
lysoPC_a_C18:1
Gln
His
C18:2
PC_ae_C38:6
PC_aa_C36:0
PC_ae_C44:4
SM_C18:0
C10:1
Arg
PC_aa_C32:3
C3
SM_C18:1
PC_ae_C38:0
PC_ae_C34:3
PC_aa_C36:6
PC_ae_C42:0
PC_ae_C34:0
PC_aa_C42:4
PC_ae_C34:1
PC_aa_C38:6
PC_ae_C36:5
lysoPC_a_C20:3
PC_ae_C36:0
PC_aa_C42:6
PC_aa_C40:3
C14:1
PC_aa_C38:5
PC_ae_C32:1
SM_C16:0
PC_ae_C32:2
C8:1
Val
xLeu
PC_aa_C36:4
PC_aa_C34:4
SM_C24:0
C14:2
Orn
Trp
Phe
C16:2
PC_aa_C38:4
C0
H1
C12:1
PC_aa_C40:4
Pro
PC_aa_C36:5
PC_aa_C28:1
C18:1
lysoPC_a_C28:0
SM_C24:1
lysoPC_a_C16:1
PC_aa_C32:2
Met
Tyr
C14:1−OH
PC_aa_C32:1
PC_aa_C24:0
PC_aa_C42:5
PC_aa_C30:0
PC_aa_C34:2
PC_aa_C34:3
C10:2
PC_aa_C36:3
PC_aa_C40:6
SM_C18:1
PC_aa_C38:3
PC_aa_C36:1
PC_aa_C40:5
C2
PC_aa_C32:0
PC_aa_C34:1
C7−DC
PC_aa_C36:2
Model1
Model2

◀ **Figure 1.** Heat map showing the results of the cluster analysis of metabolite correlations with LTL.
The two statistical models used in our analysis are shown on the horizontal axis and all individual metabolites are depicted on the vertical axis. In model 1 we investigated the association of the metabolites with LTL adjusting for age and sex and in model 2 we additionally adjusted for BMI. A blue color indicates a negative partial correlation point estimate, while a red color indicates a positive partial correlation point estimate. Cluster 1 is shown by a blue rectangle and cluster 2 by a red rectangle. The stars represent the significance: * = *p*-value < 0.05; ** = *p*-value < 0.01; *** = *p*-value < 0.001.

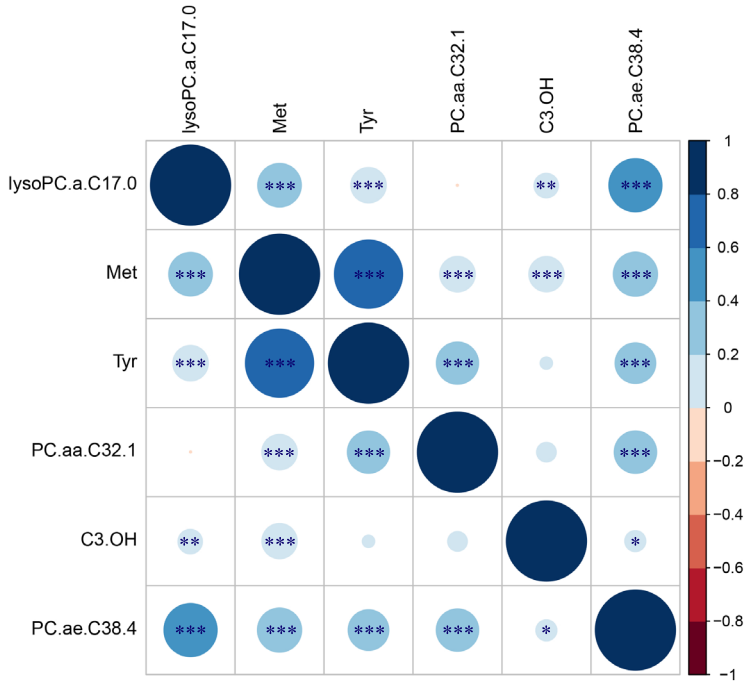


Figure 2. Correlogram of the six metabolites associated with LTL after correction for multiple testing in the first model using ERF data. Positive correlations are displayed in blue and negative correlations in red. Color intensity and the size of the circle are proportional to the correlation coefficients, with larger circles indicating higher correlation point estimates. Abbreviations: lysoPC.a.C17.0 = lysophosphatidylcholine acyl C17:0; Met = Methionine; Tyr = Tyrosine; PC.aa.C32.1 = phosphatidylcholine diacyl C32:1; C3.OH = hydroxypropionylcarnitine; PC.ae.C38.4 = phosphatidylcholine acyl-alkyl C38:4. * = *p*-value < 0.05; ** = *p*-value < 0.01; *** = *p*-value < 0.001.

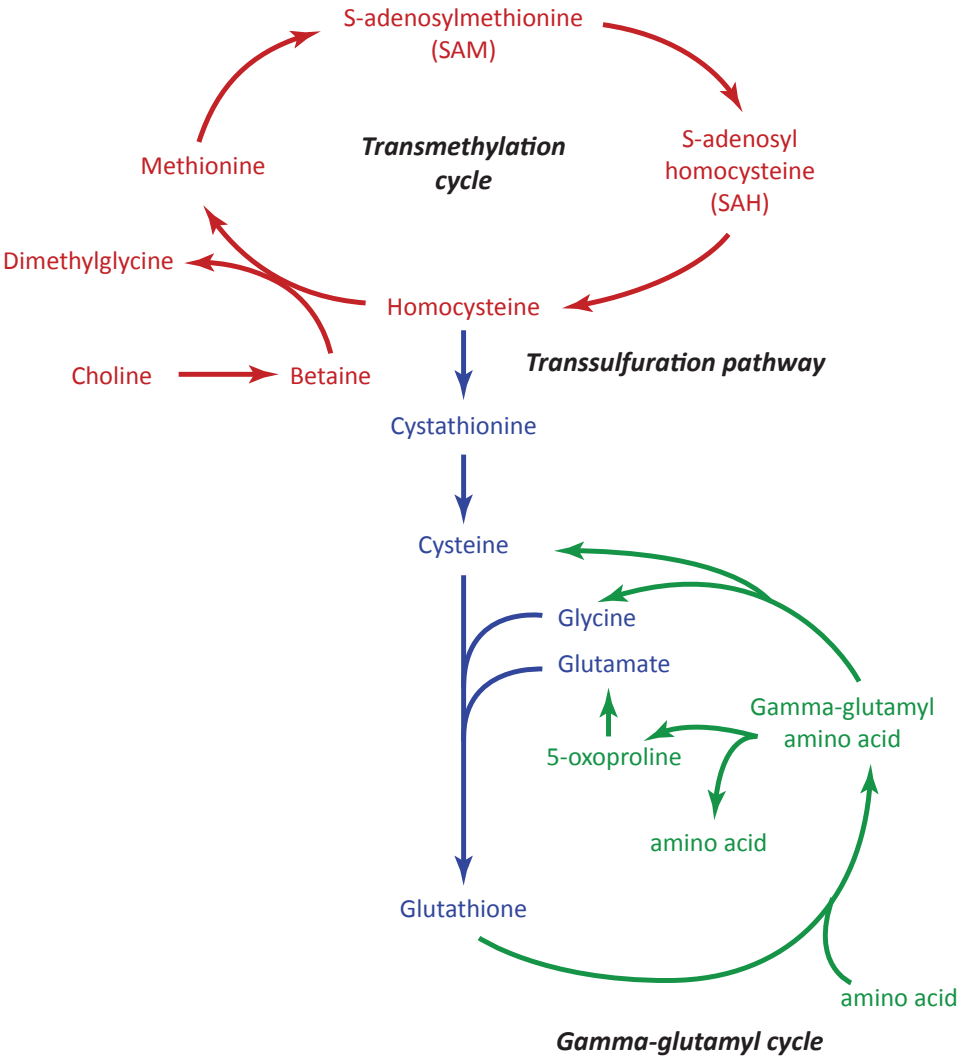


Figure 3. Methionine metabolic pathways. Indicated in red are the metabolites of the transmethylation pathway, in blue those of the transsulfuration pathway, and in green those involved in the gamma-glutamyl cycle. Modified figure from Dash et al.[112]

Discussion

When adjusting for false positive findings due to multiple testing, this study identified significant associations between LTL and six metabolites, which form two extended clusters. These metabolites include three phosphatidylcholines (lysoPC a C17:0, PC aa C32:1, PC ae C38:4), two amino acids (methionine, tyrosine), and one acylcarnitine (C3-OH). Longer LTL was associated with higher levels of lysoPC a C17:0 and PC ae C38:4, and with lower levels of methionine, tyrosine, PC aa C32:1, and C3-OH. Pathway analysis highlighted a key role of the homocysteine metabolism.

A problem when comparing our findings to those in earlier studies is that the metabolites are platform-specific and thus we cannot replicate directly previous findings [43,44]. However, the PCs significantly associated with LTL in our study belong to the same metabolite family of glycerophospholipids[52] as two metabolites significantly associated with LTL in the study of Zierer et al.[47]: 1-stearoylglycerophosphoinositol and 1-palmitoylglycero-phosphoinositol. These metabolites are involved in fatty acid metabolism and particularly membrane composition in biological aging. The small study (N = 423) performed in American Indians from the Strong Family Heart Study also found associations of glycerophospholipids (e.g., glycerophosphoethanolamines, glycero-phosphocholine, and glycerophosphoglycerol) with LTL[46].

The metabolite most significantly associated with LTL in our study was lysoPC a C17:0. LysoPCs are formed through hydrolysis of PCs by phospholipase A2[53] and have pro-atherogenic and pro-inflammatory effects through impairment of endothelium-dependent vascular relaxation[54], monocyte recruitment and macrophage proliferation[55,56], and increased expression of adhesion molecules[57]. Previously, negative associations of lysoPC a C17:0 levels with high-sensitivity C-reactive protein (hsCRP), interleukin-6, insulin, and myocardial infarction have been found[58,59]. These results indicate the involvement of inflammation[58,59]. Inflammation and cardiovascular disease are related to telomere shortening[60,61] and our study brings to surface lysoPC a C17:0 as a novel key player.

We further identified a negative association of the two highly correlated amino acids methionine and tyrosine with LTL. Methionine is an essential amino acid and involved in multiple important biological processes necessary for normal growth and development in mammals, including protein synthesis, methylation, the transsulfuration pathway, and homocysteine metabolism[62]. Previous studies have shown that a methionine-restricted diet increased lifespan in rodents[63-67]. Tyrosine is a non-essential amino acid and a precursor for several catecholamines, including dopamine, as well as thyroid hormones (T_3 and T_4)[68,69]. Low thyroid hormone levels have been associated with increased lifespan in multiple animal models[70-72] and in humans[73-75]. Moreover, a role of tyrosine as developmental regulator and modulator of longevity has been described in *Caenorhabditis elegans*[76]. Tyrosine can also form a dipeptide with gamma-glutamate called gamma-glutamyltyrosine (http://www.hmdb.

ca/metabolites/HMDB0011741), which was negatively associated with LTL in the TwinsUK cohort[47]. It is involved in the gamma-glutamyl cycle (as shown in **Figure 3**) and indicates involvement of increased oxidative stress[47], a factor related to LTL shortening[77,78]. Elevated blood levels of the amino acid tyrosine are seen in obese individuals[79,80], and were found to be a novel risk factor for the development of diabetes[38,81]. Type 2 diabetes has been associated with shorter LTL[10,16,17].

Both methionine and tyrosine are correlated to PC aa C32:1, which is the fourth metabolite significantly associated with LTL. PCs are the major phospholipids in cell membranes and lipoproteins[82,83]. They consist of a glycerol backbone with different fatty acid combinations that are linked by ester (a) or ether (e) bonds, resulting in either diacyl (aa) or acyl-alkyl (ae) PCs[84]. We observed a cluster of metabolites (cluster 2) negatively associated with LTL, including methionine, tyrosine, PC aa C32:1 and multiple other diacyl PCs. Various metabolites of this cluster including PC aa C32:1, PC aa C36:1, PC aa C38:3, and PC aa C40:5, have been associated with increased risk of type 2 diabetes[40]. The other PC that surpassed the significance threshold in model 1 was PC ae C38:4. However, PC ae C38:4 was nominally significant after including BMI in the model (model 2) and after excluding the younger QIMR study from the meta-analysis. PC ae C38:4 showed a positive association with LTL and clustered with lysoPC a C17:0 and a series of PC ae metabolites (cluster 1) that also show consistent effect across cohorts such as PC ae C40:3, PC ae C40:4, and PC ae C40:5 (FDR < 0.05). In line with this finding, PC ae's have been shown to have antioxidant properties, protecting lipids from oxidation[85,86], and the metabolites in this cluster showed a reduced risk of type 2 diabetes[40].

Although also C3-OH was found to be associated with LTL when adjusting for multiple testing, the association with LTL was only observed in ERF and LLS. In the other five studies, this metabolite did not pass the quality control. Therefore, this finding and other findings based on data of two studies only, such as PC aa C38:1, should be interpreted with care and more research, including alternative assessments of these metabolites, is needed. C3-OH is a metabolite of interest for further investigation as it is an acylcarnitine and involved in lipid transport as well as lipid and fatty acid metabolism (http://www.hmdb.ca/metabolites/HMDB0013125). Carnitine is mainly absorbed from the diet but can also be synthesized from the amino acids lysine and methionine[87] and is essential for energy metabolism as it transports fatty acids from the cytosol into the mitochondrion for β-oxidation[88,89]. **Figure 2** shows indeed that C3-OH is correlated to methionine, as predicted[87]. Carnitine insufficiency has been implicated as a common trait of insulin-resistant states, including advanced age, genetic diabetes, and diet-induced obesity[90]. However, when clustering the correlations of the metabolites to LTL, we find that C3-OH does not cluster with other metabolites (**Figure 1**).

Pathway analysis using the Taverna workflow revealed that both methionine and PCs are part of homocysteine metabolism. Our results give novel metabolic insights into the findings of previous studies that describe an increase in plasma homocysteine with age and

shortening of LTL with increasing homocysteine levels[91,92]. Our study suggests that lysoPC a C17:0, PC aa C32:1, PC ae C38:4 as well as methionine and tyrosine are key metabolites in the link between the homocysteine pathway and telomere length. Homocysteine metabolism has been implicated in a wide range of age-related diseases, such as cardiovascular diseases[93,94], dementia[95,96], Alzheimer's disease[96,97], diabetes and its associated vascular complications[98-100], and in mortality[101-104]. Taking together the findings of our study with that of previous studies, the endothelium may be the tissue of interest. There is substantial evidence that homocysteine and lysoPC are involved in endothelial dysfunction[77,78] caused by inflammation and oxidative stress[105-109]. In cultured endothelial cells, homocysteine was also shown to accelerate telomere shortening and endothelial senescence[92,110].

A major strength of this study is that both LTL and metabolites were measured centrally, using a standard protocol and blood samples taken at the same time point. Metabolite levels were quantified using the Absolute*IDQ*™ p150 kit (Biocrates Life Sciences, Innsbruck, Austria) that detects biologically relevant metabolites from four compound classes: acylcarnitines, amino acids, hexoses, and phosho- and sphingolipids. This method has been proven to be in conformance with FDA Guideline 'Guidance for Industry—Bioanalytical Method Validation (May 2001)'[111], which implies proof of reproducibility within a given error range. At the same time, measuring metabolites with this specific platform may be considered also a limitation of our study as other metabolites might also be related to LTL.

In conclusion, using data from 7,853 individuals from seven independent cohorts, we found longer LTL associated with higher levels of lysoPC a C17:0 and PC ae C38:4, and with lower levels of methionine, tyrosine, PC aa C32:1, and C3-OH. These metabolites form two clusters, one including lysoPC a C17:0, PC ae C38:4, and a series of PC ae metabolites positively associated with LTL, while the second cluster includes methionine, tyrosine, PC aa C32:1, and a series of PC aa metabolites. These metabolites have been implicated in cardiovascular disease and diabetes, two major drivers of morbidity and mortality. The functional role of these metabolites involves inflammation and oxidative stress. Our pathway analysis links the metabolites to homocysteine metabolism, a pathway linked to cardiovascular disease, diabetes and many other age-related diseases.

References

1 Blackburn, E. H. & Gall, J. G. A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. *J Mol Biol* **120**, 33-53, doi:10.1016/0022-2836(78)90294-2 (1978).

2 Blackburn, E. H., Greider, C. W. & Szostak, J. W. Telomeres and telomerase: the path from maize, Tetrahymena and yeast to human cancer and aging. *Nat Med* **12**, 1133-1138, doi:10.1038/nm1006-1133 (2006).

3 Vaziri, H. *et al.* Evidence for a mitotic clock in human hematopoietic stem cells: loss of telomeric DNA with age. *Proc Natl Acad Sci U S A* **91**, 9857-9860, doi:10.1073/pnas.91.21.9857 (1994).

4 Abdallah, P. *et al.* A two-step model for senescence triggered by a single critically short telomere. *Nat Cell Biol* **11**, 988-993, doi:10.1038/ncb1911 (2009).

5 Blackburn, E. H. Switching and signaling at the telomere. *Cell* **106**, 661-673, doi:10.1016/S0092-8674(01)00492-5 (2001).

6 Campisi, J. & d'Adda di Fagagna, F. Cellular senescence: when bad things happen to good cells. *Nat Rev Mol Cell Biol* **8**, 729-740, doi:10.1038/nrm2233 (2007).

7 Lindsey, J., McGill, N. I., Lindsey, L. A., Green, D. K. & Cooke, H. J. In vivo loss of telomeric repeats with age in humans. *Mutat Res* **256**, 45-48, doi:10.1016/0921-8734(91)90032-7 (1991).

8 Slagboom, P. E., Droog, S. & Boomsma, D. I. Genetic determination of telomere size in humans: a twin study of three age groups. *Am J Hum Genet* **55**, 876-882 (1994).

9 Broer, L. *et al.* Meta-analysis of telomere length in 19,713 subjects reveals high heritability, stronger maternal inheritance and a paternal age effect. *Eur J Hum Genet* **21**, 1163-1168, doi:10.1038/ejhg.2012.303 (2013).

10 Fitzpatrick, A. L. *et al.* Leukocyte telomere length and cardiovascular disease in the cardiovascular health study. *Am J Epidemiol* **165**, 14-21, doi:10.1093/aje/kwj346 (2007).

11 Maubaret, C. G. *et al.* Telomeres are shorter in myocardial infarction patients compared to healthy subjects: correlation with environmental risk factors. *J Mol Med (Berl)* **88**, 785-794, doi:10.1007/s00109-010-0624-3 (2010).

12 Rehkopf, D. H. *et al.* Leukocyte Telomere Length in Relation to 17 Biomarkers of Cardiovascular Disease Risk: A Cross-Sectional Study of US Adults. *PLoS Med* **13**, e1002188, doi:10.1371/journal.pmed.1002188 (2016).

13 Brouilette, S., Singh, R. K., Thompson, J. R., Goodall, A. H. & Samani, N. J. White cell telomere length and risk of premature myocardial infarction. *Arterioscler Thromb Vasc Biol* **23**, 842-846, doi:10.1161/01.ATV.0000067426.96344.32 (2003).

14 Brouilette, S. W. *et al.* Telomere length, risk of coronary heart disease, and statin treatment in the West of Scotland Primary Prevention Study: a nested case-control study. *Lancet* **369**, 107-114, doi:10.1016/S0140-6736(07)60071-3 (2007).

15 Haycock, P. C. *et al.* Leucocyte telomere length and risk of cardiovascular disease: systematic review and meta-analysis. *BMJ* **349**, g4227 (2014).

16 Jeanclos, E. *et al.* Shortened telomere length in white blood cells of patients with IDDM. *Diabetes* **47**, 482-486 (1998).

17 Zhao, J., Miao, K., Wang, H., Ding, H. & Wang, D. W. Association between telomere length and type 2 diabetes mellitus: a meta-analysis. *PLoS One* **8**, e79993, doi:10.1371/journal.pone.0079993 (2013).

18 Honig, L. S., Kang, M. S., Schupf, N., Lee, J. H. & Mayeux, R. Association of shorter leukocyte telomere repeat length with dementia and mortality. *Arch Neurol* **69**, 1332-1339, doi:10.1001/archneurol.2012.1541 (2012).

19 Martin-Ruiz, C. *et al.* Telomere length predicts poststroke mortality, dementia, and cognitive decline. *Ann Neurol* **60**, 174-180, doi:10.1002/ana.20869 (2006).

20 Fitzpatrick, A. L. *et al.* Leukocyte telomere length and mortality in the Cardiovascular Health Study. *J Gerontol A Biol Sci Med Sci* **66**, 421-429, doi:10.1093/gerona/glq224 (2011).

21 Marioni, R. E. *et al.* The epigenetic clock and telomere length are independently associated with chronological age and mortality. *Int J Epidemiol*, doi:10.1093/ije/dyw041 (2016).

22 Bakaysa, S. L. *et al.* Telomere length predicts survival independent of genetic influences. *Aging Cell* **6**, 769-774, doi:10.1111/j.1474-9726.2007.00340.x (2007).

23 Cawthon, R. M., Smith, K. R., O'Brien, E., Sivatchenko, A. & Kerber, R. A. Association between telomere length in blood and mortality in people aged 60 years or older. *Lancet* **361**, 393-395, doi:10.1016/S0140-6736(03)12384-7 (2003).

24 Ehrlenbach, S. *et al.* Influences on the reduction of relative telomere length over 10 years in the population-based Bruneck Study: introduction of a well-controlled high-throughput assay. *Int J Epidemiol* **38**, 1725-1734, doi:10.1093/ije/dyp273 (2009).

25 Kimura, M. *et al.* Telomere length and mortality: a study of leukocytes in elderly Danish twins. *Am J Epidemiol* **167**, 799-806, doi:10.1093/aje/kwm380 (2008).

26 Deelen, J. *et al.* Leukocyte telomere length associates with prospective mortality independent of immune-related parameters and known genetic markers. *Int J Epidemiol* **43**, 878-886, doi:10.1093/ije/dyt267 (2014).

27 Rode, L., Nordestgaard, B. G. & Bojesen, S. E. Peripheral blood leukocyte telomere length and mortality among 64,637 individuals from the general population. *J Natl Cancer Inst* **107**, djv074, doi:10.1093/jnci/djv074 (2015).

28 Njajou, O. T. *et al.* Association between telomere length, specific causes of death, and years of healthy life in health, aging, and body composition, a population-based cohort study. *J Gerontol A Biol Sci Med Sci* **64**, 860-864, doi:10.1093/gerona/glp061 (2009).

29 Bischoff, C. *et al.* No association between telomere length and survival among the elderly and oldest old. *Epidemiology* **17**, 190-194, doi:10.1097/01.ede.0000199436.55248.10 (2006).

30 Harris, S. E. *et al.* The association between telomere length, physical health, cognitive ageing, and mortality in non-demented older people. *Neurosci Lett* **406**, 260-264, doi:10.1016/j.neulet.2006.07.055 (2006).

31 Martin-Ruiz, C. M., Gussekloo, J., van Heemst, D., von Zglinicki, T. & Westendorp, R. G. Telomere length in white blood cells is not associated with morbidity or mortality in the oldest old: a population-based study. *Aging Cell* **4**, 287-290, doi:10.1111/j.1474-9726.2005.00171.x (2005).

32 Strandberg, T. E. *et al.* Association of telomere length in older men with mortality and midlife body mass index and smoking. *J Gerontol A Biol Sci Med Sci* **66**, 815-820, doi:10.1093/gerona/glr064 (2011).

33 Houben, J. M., Giltay, E. J., Rius-Ottenheim, N., Hageman, G. J. & Kromhout, D. Telomere length and mortality in elderly men: the Zutphen Elderly Study. *J Gerontol A Biol Sci Med Sci* **66**, 38-44, doi:10.1093/gerona/glq164 (2011).

34 Cheng, S. *et al.* Distinct metabolomic signatures are associated with longevity in humans. *Nat Commun* **6**, 6791, doi:10.1038/ncomms7791 (2015).

35 Montoliu, I. *et al.* Serum profiling of healthy aging identifies phospho- and sphingolipid species as markers of human longevity. *Aging (Albany NY)* **6**, 9-25, doi:10.18632/aging.100630 (2014).

36 Gonzalez-Covarrubias, V. *et al.* Lipidomics of familial longevity. *Aging Cell* **12**, 426-434, doi:10.1111/acel.12064 (2013).

37 Wang-Sattler, R. *et al.* Novel biomarkers for pre-diabetes identified by metabolomics. *Mol Syst Biol* **8**, 615, doi:10.1038/msb.2012.43 (2012).

38 Liu, J. *et al.* Metabolomics based markers predict type 2 diabetes in a 14-year follow-up study. *Metabolomics* **13**, 104, doi:10.1007/s11306-017-1239-2 (2017).

39 Suhre, K. *et al.* Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. *PLoS One* **5**, e13953, doi:10.1371/journal.pone.0013953 (2010).

40 Floegel, A. *et al.* Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. *Diabetes* **62**, 639-648, doi:10.2337/db12-0495 (2013).

41 Matsumoto, T., Kobayashi, T. & Kamata, K. Role of lysophosphatidylcholine (LPC) in atherosclerosis. *Curr Med Chem* **14**, 3209-3220 (2007).

42 Cao, M., Zhao, L., Chen, H., Xue, W. & Lin, D. NMR-based metabolomic analysis of human bladder cancer. *Anal Sci* **28**, 451-456 (2012).

43 Oresic, M. *et al.* Metabolome in progression to Alzheimer's disease. *Transl Psychiatry* **1**, e57, doi:10.1038/tp.2011.55 (2011).

44 Yu, Z. *et al.* Human serum metabolic profiles are age dependent. *Aging Cell* **11**, 960-967, doi:10.1111/j.1474-9726.2012.00865.x (2012).

45 Menni, C. *et al.* Metabolomic markers reveal novel pathways of ageing and early development in human populations. *Int J Epidemiol* **42**, 1111-1119, doi:10.1093/ije/dyt094 (2013).

46 Zhao, J. *et al.* Metabolic profiles of biological aging in American Indians: the Strong Heart Family Study. *Aging (Albany NY)* **6**, 176-186, doi:10.18632/aging.100644 (2014).

47 Zierer, J. *et al.* Metabolomics profiling reveals novel markers for leukocyte telomere length. *Aging (Albany NY)* **8**, 77-94, doi:10.18632/aging.100874 (2016).

48 Li, J. & Ji, L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb)* **95**, 221-227, doi:10.1038/sj.hdy.6800717 (2005).

49 Hull, D. *et al.* Taverna: a tool for building and running workflows of services. *Nucleic Acids Res* **34**, W729-732, doi:10.1093/nar/gkl320 (2006).

50 Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30 (2000).

51 Wei, T. & Simko, V. R package "corrplot": Visualization of a Correlation Matrix (Version 0.84). Available from https://github.com/taiyun/corrplot. (2017).

52 Blanco, A. & Blanco, G. *Medical biochemistry*. (2017).

53 van Meer, G., Voelker, D. R. & Feigenson, G. W. Membrane lipids: where they are and how they behave. *Nat Rev Mol Cell Biol* **9**, 112-124, doi:10.1038/nrm2330 (2008).

54 Chen, L. *et al.* Oxidative modification of low density lipoprotein in normal and hyperlipidemic patients: effect of lysophosphatidylcholine composition on vascular relaxation. *J Lipid Res* **38**, 546-553 (1997).

55 Weber, C., Erl, W. & Weber, P. C. Enhancement of monocyte adhesion to endothelial cells by oxidatively modified low-density lipoprotein is mediated by activation of CD11b. *Biochem Biophys Res Commun* **206**, 621-628, doi:10.1006/bbrc.1995.1088 (1995).

56 Olofsson, K. E., Andersson, L., Nilsson, J. & Bjorkbacka, H. Nanomolar concentrations of lysophosphatidylcholine recruit monocytes and induce pro-inflammatory cytokine production in macrophages. *Biochem Biophys Res Commun* **370**, 348-352, doi:10.1016/j.bbrc.2008.03.087 (2008).

57 Kume, N., Cybulsky, M. I. & Gimbrone, M. A., Jr. Lysophosphatidylcholine, a component of atherogenic lipoproteins, induces mononuclear leukocyte adhesion molecules in cultured human and rabbit arterial endothelial cells. *J Clin Invest* **90**, 1138-1144, doi:10.1172/JCI115932 (1992).

58 Paapstel, K. *et al.* Inverse relations of serum phosphatidylcholines and lysophosphatidylcholines with vascular damage and heart rate in patients with atherosclerosis. *Nutr Metab Cardiovasc Dis* **28**, 44-52, doi:10.1016/j.numecd.2017.07.011 (2018).

59 Ward-Caviness, C. K. *et al.* Improvement of myocardial infarction risk prediction via inflammation-associated metabolite biomarkers. *Heart* **103**, 1278-1285, doi:10.1136/heartjnl-2016-310789 (2017).

60 O'Donovan, A. *et al.* Cumulative inflammatory load is associated with short leukocyte telomere length in the Health, Aging and Body Composition Study. *PLoS One* **6**, e19687, doi:10.1371/journal.pone.0019687 (2011).

61 Bekaert, S. *et al.* Telomere length and cardiovascular risk factors in a middle-aged population free of overt cardiovascular disease. *Aging Cell* **6**, 639-647, doi:10.1111/j.1474-9726.2007.00321.x (2007).

62 Finkelstein, J. D. Methionine metabolism in mammals. *J Nutr Biochem* **1**, 228-237 (1990).

63 Lee, B. C., Kaya, A. & Gladyshev, V. N. Methionine restriction and life-span control. *Ann N Y Acad Sci* **1363**, 116-124, doi:10.1111/nyas.12973 (2016).

64 Sun, L., Sadighi Akha, A. A., Miller, R. A. & Harper, J. M. Life-span extension in mice by preweaning food restriction and by methionine restriction in middle age. *J Gerontol A Biol Sci Med Sci* **64**, 711-722, doi:10.1093/gerona/glp051 (2009).

65 Orentreich, N., Matias, J. R., DeFelice, A. & Zimmerman, J. A. Low methionine ingestion by rats extends life span. *J Nutr* **123**, 269-274, doi:10.1093/jn/123.2.269 (1993).

66 Zimmerman, J. A., Malloy, V., Krajcik, R. & Orentreich, N. Nutritional control of aging. *Exp Gerontol* **38**, 47-52 (2003).

67 Miller, R. A. *et al.* Methionine-deficient diet extends mouse lifespan, slows immune and lens aging, alters glucose, T4, IGF-I and insulin levels, and increases hepatocyte MIF levels and stress resistance. *Aging Cell* **4**, 119-125, doi:10.1111/j.1474-9726.2005.00152.x (2005).

68 During, M. J., Acworth, I. N. & Wurtman, R. J. Effects of systemic L-tyrosine on dopamine release from rat corpus striatum and nucleus accumbens. *Brain Res* **452**, 378-380 (1988).

69 Evans, R. M. The steroid and thyroid hormone receptor superfamily. *Science* **240**, 889-895 (1988).

70 Brown-Borg, H. M., Borg, K. E., Meliska, C. J. & Bartke, A. Dwarf mice and the ageing process. *Nature* **384**, 33, doi:10.1038/384033a0 (1996).

71 Ooka, H., Fujita, S. & Yoshimoto, E. Pituitary-thyroid activity and longevity in neonatally thyroxine-treated rats. *Mech Ageing Dev* **22**, 113-120 (1983).

72 Tatar, M., Bartke, A. & Antebi, A. The endocrine regulation of aging by insulin-like signals. *Science* **299**, 1346-1351, doi:10.1126/science.1081447 (2003).

73 Atzmon, G., Barzilai, N., Hollowell, J. G., Surks, M. I. & Gabriely, I. Extreme longevity is associated with increased serum thyrotropin. *J Clin Endocrinol Metab* **94**, 1251-1254, doi:10.1210/jc.2008-2325 (2009).

74 Rozing, M. P. *et al.* Familial longevity is associated with decreased thyroid function. *J Clin Endocrinol Metab* **95**, 4979-4984, doi:10.1210/jc.2010-0875 (2010).
75 Rozing, M. P. *et al.* Low serum free triiodothyronine levels mark familial longevity: the Leiden Longevity Study. *J Gerontol A Biol Sci Med Sci* **65**, 365-368, doi:10.1093/gerona/glp200 (2010).
76 Ferguson, A. A. *et al.* TATN-1 mutations reveal a novel role for tyrosine as a metabolic signal that influences developmental decisions and longevity in Caenorhabditis elegans. *PLoS Genet* **9**, e1004020, doi:10.1371/journal.pgen.1004020 (2013).
77 von Zglinicki, T., Saretzki, G., Docke, W. & Lotze, C. Mild hyperoxia shortens telomeres and inhibits proliferation of fibroblasts: a model for senescence? *Exp Cell Res* **220**, 186-193, doi:10.1006/excr.1995.1305 (1995).
78 von Zglinicki, T. Oxidative stress shortens telomeres. *Trends Biochem Sci* **27**, 339-344 (2002).
79 Newgard, C. B. *et al.* A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab* **9**, 311-326, doi:10.1016/j.cmet.2009.02.002 (2009).
80 Moore, S. C. *et al.* Human metabolic correlates of body mass index. *Metabolomics* **10**, 259-269, doi:10.1007/s11306-013-0574-1 (2014).
81 Wang, T. J. *et al.* Metabolite profiles and the risk of developing diabetes. *Nat Med* **17**, 448-453, doi:10.1038/nm.2307 (2011).
82 Skipski, V. P. *et al.* Lipid composition of human serum lipoproteins. *Biochem J* **104**, 340-352 (1967).
83 Cole, L. K., Vance, J. E. & Vance, D. E. Phosphatidylcholine biosynthesis and lipoprotein metabolism. *Biochim Biophys Acta* **1821**, 754-761, doi:10.1016/j.bbalip.2011.09.009 (2012).
84 Magnusson, C. D. & Haraldsson, G. G. Ether lipids. *Chem Phys Lipids* **164**, 315-340, doi:10.1016/j.chemphyslip.2011.04.010 (2011).
85 Engelmann, B. Plasmalogens: targets for oxidants and major lipophilic antioxidants. *Biochem Soc Trans* **32**, 147-150, doi:10.1042/ (2004).
86 Zoeller, R. A. *et al.* Plasmalogens as endogenous antioxidants: somatic cell mutants reveal the importance of the vinyl ether. *Biochem J* **338 (Pt 3)**, 769-776 (1999).
87 Vaz, F. M. & Wanders, R. J. Carnitine biosynthesis in mammals. *Biochem J* **361**, 417-429 (2002).
88 Fritz, I. B. & Yue, K. T. Long-Chain Carnitine Acyltransferase and the Role of Acylcarnitine Derivatives in the Catalytic Increase of Fatty Acid Oxidation Induced by Carnitine. *J Lipid Res* **4**, 279-288 (1963).
89 McGarry, J. D. & Brown, N. F. The mitochondrial carnitine palmitoyltransferase system. From concept to molecular analysis. *Eur J Biochem* **244**, 1-14 (1997).
90 Noland, R. C. *et al.* Carnitine insufficiency caused by aging and overnutrition compromises mitochondrial performance and metabolic control. *J Biol Chem* **284**, 22840-22852, doi:10.1074/jbc.M109.032888 (2009).
91 Richards, J. B. *et al.* Homocysteine levels and leukocyte telomere length. *Atherosclerosis* **200**, 271-277, doi:10.1016/j.atherosclerosis.2007.12.035 (2008).
92 Xu, D., Neville, R. & Finkel, T. Homocysteine accelerates endothelial cell senescence. *FEBS Lett* **470**, 20-24 (2000).
93 Homocysteine Studies, C. Homocysteine and risk of ischemic heart disease and stroke: a meta-analysis. *JAMA* **288**, 2015-2022 (2002).
94 Wald, D. S., Law, M. & Morris, J. K. Homocysteine and cardiovascular disease: evidence on causality from a meta-analysis. *BMJ* **325**, 1202 (2002).
95 Smith, A. D. *et al.* Homocysteine-lowering by B vitamins slows the rate of accelerated brain atrophy in mild cognitive impairment: a randomized controlled trial. *PLoS One* **5**, e12244, doi:10.1371/journal.pone.0012244 (2010).
96 Smith, A. D. *et al.* Homocysteine and Dementia: An International Consensus Statement. *J Alzheimers Dis* **62**, 561-570, doi:10.3233/JAD-171042 (2018).
97 Morris, M. S. Homocysteine and Alzheimer's disease. *Lancet Neurol* **2**, 425-428 (2003).
98 Smulders, Y. M. *et al.* Fasting and post-methionine homocysteine levels in NIDDM. Determinants and correlations with retinopathy, albuminuria, and cardiovascular disease. *Diabetes Care* **22**, 125-132 (1999).
99 Buysschaert, M., Dramais, A. S., Wallemacq, P. E. & Hermans, M. P. Hyperhomocysteinemia in type 2 diabetes: relationship to macroangiopathy, nephropathy, and insulin resistance. *Diabetes Care* **23**, 1816-1822 (2000).
100 Joshi, M. B. *et al.* Elevated homocysteine levels in type 2 diabetes induce constitutive neutrophil extracellular traps. *Sci Rep* **6**, 36362, doi:10.1038/srep36362 (2016).

101 Bostom, A. G. *et al.* Nonfasting plasma total homocysteine levels and stroke incidence in elderly persons: the Framingham Study. *Ann Intern Med* **131**, 352-355 (1999).
102 Dangour, A. D. *et al.* Plasma homocysteine, but not folate or vitamin B-12, predicts mortality in older people in the United Kingdom. *J Nutr* **138**, 1121-1128 (2008).
103 Nygard, O. *et al.* Plasma homocysteine levels and mortality in patients with coronary artery disease. *N Engl J Med* **337**, 230-236, doi:10.1056/NEJM199707243370403 (1997).
104 Vollset, S. E. *et al.* Plasma total homocysteine and cardiovascular and noncardiovascular mortality: the Hordaland Homocysteine Study. *Am J Clin Nutr* **74**, 130-136 (2001).
105 Li, J. J. *et al.* Homocysteine Triggers Inflammatory Responses in Macrophages through Inhibiting CSE-H2S Signaling via DNA Hypermethylation of CSE Promoter. *Int J Mol Sci* **16**, 12560-12577, doi:10.3390/ijms160612560 (2015).
106 Au-Yeung, K. K. *et al.* Hyperhomocysteinemia activates nuclear factor-kappaB in endothelial cells via oxidative stress. *Circ Res* **94**, 28-36, doi:10.1161/01.RES.0000108264.67601.2C (2004).
107 Topal, G. *et al.* Homocysteine induces oxidative stress by uncoupling of NO synthase activity through reduction of tetrahydrobiopterin. *Free Radic Biol Med* **36**, 1532-1541, doi:10.1016/j.freeradbiomed.2004.03.019 (2004).
108 Hoffman, M. Hypothesis: hyperhomocysteinemia is an indicator of oxidant stress. *Med Hypotheses* **77**, 1088-1093, doi:10.1016/j.mehy.2011.09.009 (2011).
109 Pushpakumar, S., Kundu, S. & Sen, U. Endothelial dysfunction: the link between homocysteine and hydrogen sulfide. *Curr Med Chem* **21**, 3662-3672 (2014).
110 Kurz, D. J. *et al.* Chronic oxidative stress compromises telomere integrity and accelerates the onset of senescence in human endothelial cells. *J Cell Sci* **117**, 2417-2426, doi:10.1242/jcs.01097 (2004).
111 US Department of Health and Human Services, F. a. D. A. Center for Drug Evaluation and Research (CDER) Center for Veterinary Medicine (CVM) Guidance for Industry Bioanalytical Method Validation. 2001.
112 Dash, P. K. *et al.* Traumatic Brain Injury Alters Methionine Metabolism: Implications for Pathophysiology. *Front Syst Neurosci* **10**, 36, doi:10.3389/fnsys.2016.00036 (2016).

CHAPTER 2.3

Fat metabolism is associated with telomere length and mortality in six population-based studies

Ashley van der Spek, Hata Karamujić-Čomić, René Pool, Mariska Bot, Marian Beekman, Sanzhima Garmaeva, Pascal P. Arp, Sandra Henkelman, Jun Liu, Alexessander Couto Alves, Gonneke Willemsen, Gerard van Grootheest, Geraldine Aubert, BBMRI Metabolomics Consortium, M. Arfan Ikram, Marjo-Riitta Jarvelin, Peter Lansdorp, André G. Uitterlinden, Alexandra Zhernakova, P. Eline Slagboom, Brenda W.J.H. Penninx, Dorret I. Boomsma, Najaf Amin, Cornelia M. van Duijn

This chapter is accepted for publication in *Human Molecular Genetics*.

The supplemental information for this paper is available at https://drive.google.com/drive/folders/1v9J9CWE4H3npx0kBbdKOUVRvgv1a9-rF?usp=sharing

Abstract

Telomeres are repetitive DNA sequences located at the end of chromosomes, which are associated to biological aging, cardiovascular disease, cancer, and mortality. Lipid and fatty acid metabolism have been associated with telomere shortening. We have conducted an in-depth study investigating the association of metabolic biomarkers with telomere length (LTL).

We performed an association analysis of 226 metabolic biomarkers with LTL using data from 11,775 individuals from six independent population-based cohorts (BBMRI-NL consortium). Metabolic biomarkers include lipoprotein lipids and subclasses, fatty acids, amino acids, glycolysis measures and ketone bodies. LTL was measured by quantitative polymerase chain reaction or FlowFISH. Linear regression analysis was performed adjusting for age, sex, lipid-lowering medication and cohort-specific covariates (model 1) and additionally for body mass index (BMI) and smoking (model 2), followed by inverse variance-weighted meta-analyses (significance threshold $p_{meta} = 6.5x10^{-4}$).

We identified four metabolic biomarkers positively associated with LTL, including two cholesterol to lipid ratios in small VLDL (S-VLDL-C % and S-VLDL-CE %) and two omega-6 fatty acid ratios (FAw6/FA and LA/FA). After additionally adjusting for BMI and smoking, these metabolic biomarkers remained associated with LTL with similar effect estimates. In addition, cholesterol esters in very small VLDL (XS-VLDL-CE) became significantly associated with LTL ($p = 3.6x10^{-4}$). We replicated the association of FAw6/FA with LTL in an independent dataset of 7,845 individuals ($p = 1.9x10^{-4}$).

To conclude, we identified multiple metabolic biomarkers involved in lipid and fatty acid metabolism that may be involved in LTL biology. Longitudinal studies are needed to exclude reversed causation.

Introduction

Telomeres are repetitive DNA sequences located at the end of chromosomes that have an important role in the maintenance of genomic stability[1]. Telomeres gradually shorten as a consequence of cell replication and damage accumulation with increasing age[2,3]. Beyond a minimal critical telomere length, cells enter replicative senescence and this process of cellular senescence gradually affects multiple tissues during aging and the viability of stem cells[4-6]. Telomere length is therefore considered as a marker of biological aging. Although telomere shortening with age has a tissue-specific pace, telomere length in blood is considered a dynamic marker of physiological health and well-being in epidemiological and clinical studies[7]. Both short and long leukocyte telomere length (LTL) have been associated with cancer[8-12], but only short LTL has been associated with several age-related diseases, including cardiovascular diseases[13-19], diabetes[20-22], and dementia[23,24]. Multiple studies have also shown associations of shorter LTL with mortality[25-35], although findings have been inconsistent[13,36-39]. Telomere length is highly heritable (h^2 between 44-86%)[40-42] and various genetic determinants have been identified[43-45]. One of the most intriguing findings in studies investigating genetic determinants of LTL is that the identified genetic variants underscore the association between LTL and cardiovascular and metabolic diseases[18,43], Alzheimer's disease[46], and several cancers[47-49].

Lipid metabolism appears to play a key role in telomere length regulation. A relatively small study in 423 American Indians tested the association of 1,364 distinct mass-to-charge ratio (m/z) features detected by untargeted liquid chromatography–mass spectrometry (LC/MS) with LTL[50]. This study found nineteen metabolites significantly associated with LTL, independent of chronological age and other aging-related factors. These metabolites belong to the classes of glycerophosphoethanolamines, glycerophosphocholines, glycerolipids, bile acids, isoprenoids, fatty amides, and carnitine esters[50]. A second metabolomics study using an untargeted gas chromatography–mass spectrometry (GC/MS) and LC/MS platform showed associations of lysolipids and gamma-glutamyl amino acids with LTL in 3,511 females from the TwinsUK cohort, suggesting the involvement of lipid metabolism, fatty acid metabolism, and oxidative stress in telomere shortening[51]. A third study identified phosphatidylcholines, amino acids, and a carnitine associated with LTL using a targeted electrospray ionization tandem mass spectrometry (MS/MS) metabolomics platform in 7,853 individuals, providing further support for the association of lipid metabolism, fatty acid metabolism and oxidative stress with LTL[52].

In this study, we investigated the association of LTL measured using quantitative polymerase chain reaction (qPCR)[44] or FlowFISH[53], a technique combining flow cytometry with fluorescent in-situ hybridization, with metabolic biomarkers measured on a high-throughput proton nuclear magnetic resonance (NMR) platform[54] that targets lipoprotein subclasses and fatty acids specifically along with other low-weight molecules such as amino acids. We conducted the study in 11,775 participants from six Dutch cohorts, as part of the Biobanking for Medical Research Infrastructure of the Netherlands (BBMRI-NL) consortium.

Results

In this study, we included 11,775 participants of predominantly European descent with data available on LTL as well as on metabolic biomarkers. The following six Dutch cohort studies were included: the Leiden Longevity Study (LLS), the Netherlands Study of Depression and Anxiety (NESDA), the Netherlands Twin Register (NTR), the Erasmus Rucphen Family (ERF) study, the Rotterdam Study (RS), and the LifeLines-DEEP (LLDeep) study. Descriptive statistics of the study participants are shown in **Table 1**. There was a higher proportion of female participants than male participants in all included cohort studies, with the highest percentage of females in NESDA and NTR (67%). The mean age of participants was between 39 and 75 years, covering a wide age range of 18 until 95 years across cohorts. The participants in the LLS and the RS were older ($\bar{x}_{age,\ LLS}$ = 59 years and $\bar{x}_{age,\ RS}$ = 75 years, respectively) than the participants of the other studies ($\bar{x}_{age,\ NESDA}$ = 42 years, $\bar{x}_{age,\ NTR}$ = 39 years, and $\bar{x}_{age,\ ERF}$ = 48 years). Consequently, the mean LTL measured using qPCR was lower in the LLS ($\bar{x}_{LTL}$ = 1.5) and RS ($\bar{x}_{LTL}$ = 0.9) than in the younger cohorts ($\bar{x}_{LTL}$ = 1.8 in ERF study and $\bar{x}_{LTL}$ = 2.8 in NTR), with exception of NESDA ($\bar{x}_{LTL}$ =1.1). The mean LTL of the LLDeep study cannot be directly compared with the other studies as a different measurement technique for LTL was used (FlowFISH vs. qPCR). The proportion of participants that used lipid-lowering medication was on average between 4 – 11%, with the exception of the RS, where 20% of participants used lipid-lowering medication. This may be explained by the older age of the RS participants. BMI was comparable between studies with means of 24.5 – 27.4 kg/m^2, but the proportion of current smokers differed between studies, ranging from 15.1% in the RS to 39.5% in the ERF study.

Top findings of the meta-analyses ($p_{model\ 1} < 6.5 \times 10^{-3}$) are depicted in **Table 2** and **Figure 1**. The complete results of the meta-analyses for both models (model 1 and 2) are available in **Supplementary Table 2**, and individual results per cohort are available in **Supplementary Table 3** (model 1) and **Supplementary Table 4** (model 2). After adjustment for age, sex, lipid-lowering medication, and cohort-specific covariates (model 1), two ratios of very-low-density lipoprotein (VLDL) and two fatty acid ratios showed significant evidence of a positive association with LTL: total cholesterol to total lipids ratio in small VLDL (S-VLDL-C %, $p = 1.5 \times 10^{-4}$), cholesterol esters to total lipids ratio in small VLDL (S-VLDL-CE %, $p = 2.3 \times 10^{-4}$), ratio of omega-6 fatty acids to total fatty acids (FAw6/FA, $p = 4.2 \times 10^{-4}$), and ratio of 18:2 linoleic acid to total fatty acids (LA/FA, $p = 4.4 \times 10^{-4}$). However, these findings are not independent as S-VLDL-C % and S-VLDL-CE % were significantly correlated with each other (r = 0.99), as well as the omega-6 fatty acid measurements FAw6/FA and LA/FA (r = 0.93), as shown in **Figure 2** (data of the ERF study). Additional adjustment for BMI and current smoking (model 2) had minimal effect on all four metabolic biomarkers as effect sizes remained similar (**Table 2**). Although the metabolic biomarker cholesterol esters in very small VLDL (XS-VLDL-CE) was not significantly associated with LTL after adjusting for multiple testing in model 1, this metabolic biomarker was significantly associated with LTL in model 2 ($p = 3.6 \times 10^{-4}$).

Table 1. Descriptive statistics of the study populations.

		LLS	NESDA	NTR	ERF	RS	LLDeep
Model 1	N	1858	2885	4170	1243	683	936
	N females (%)	1022 (55.0)	1920 (66.6)	2804 (67.2)	694 (55.8)	395 (58.0)	541 (57.8)
	Age range (years)	30-80	18-65	18-79	17-87	65-95	18-81
	Age mean (SD)	59.3 (6.6)	41.9 (13.0)	39.2 (13.0)	47.7 (14.0)	75.4 (6.2)	45.1 (13.6)
	LTL range (T/S ratio)	0.74-2.75	0.33-2.85	0.81-4.70	0.79-2.82	0.53-1.56	4.1-9.9*
	LTL mean (SD)	1.46 (0.27)	1.11 (0.31)	2.75 (0.48)	1.79 (0.36)	0.94 (0.16)	7.07 (0.96)*
	N lipid-lowering medication (%)	188 (10.1)	204 (7.1)	158 (3.8)	137 (11.0)	138 (20.2)	41 (4.4)
Model 2	N	1600	2883	4113	1235	650	927
	N females (%)	880 (55.0)	1918 (66.5)	2772 (67.4)	690 (55.9)	376 (57.8)	536 (57.8)
	Age range (years)	30-80	18-65	18-79	17-87	65-95	18-81
	Age mean (SD)	59.3 (6.5)	41.9 (13.0)	39.2 (13.0)	47.7 (14.0)	75.3 (6.2)	45.2 (13.5)
	LTL range (T/S ratio)	0.78-2.75	0.33-2.85	0.81-4.70	0.79-2.82	0.53-1.56	4.1-9.9*
	LTL mean (SD)	1.46 (0.27)	1.11 (0.31)	2.75 (0.48)	1.79 (0.36)	0.94 (0.16)	7.07 (0.95)*
	N lipid-lowering medication (%)	165 (10.3)	204 (7.1)	155 (3.8)	135 (10.9)	132 (20.3)	41 (4.4)
	BMI range (kg/m^2)	17.2-44.6	14.7-55.8	14.6-50.7	15.5-51.1	15.3-44.9	16.7-44.9
	BMI mean (SD)	25.5 (3.6)	25.6 (5.0)	24.5 (4.0)	26.9 (4.8)	27.4 (3.9)	25.1 (4.0)
	N current smokers (%)	210 (13.1)	1119 (38.8)	856 (20.8)	486 (39.5)	98 (15.1)	189 (20.4)

Abbreviations: LLS = Leiden Longevity Study; NESDA = Netherlands Study of Depression and Anxiety; NTR = Netherlands Twin Register; ERF = Erasmus Rucphen Family study; RS = Rotterdam Study; LLDeep = LifeLines Deep study; N = Number of participants; T/S ratio = ratio of telomere repeat length (T) to copy number of a single copy gene 36B4 (S) ; BMI = Body Mass Index. * LTL was measured in kilobases (Kb) using FlowFISH in the LLDeep cohort.

Table 2. Overview results meta-analysis (*p* in model 1 < $6.5x10^{-3}$).

Metabolic biomarkers	Model 1					Model 2					Description
	Beta	SE	P-value	Direction*	N	Beta	SE	P-value	Direction*	N	
S-VLDL-C %	0.015	0.004	**1.51x10-4**	+++-+-	11766	0.014	0.004	8.04x10-4	+++-+-	11399	Total cholesterol to total lipids ratio in small VLDL
S-VLDL-CE %	0.014	0.004	**2.30x10-4**	+++-+-	11766	0.013	0.004	8.27x10-4	+++-+-	11399	Cholesterol esters to total lipids ratio in small VLDL
FAw6/FA	0.013	0.004	**4.15x10-4**	++++++	11514	0.012	0.004	1.32x10-3	+++++-	11157	Ratio of omega-6 fatty acids to total fatty acids
LA/FA	0.013	0.004	**4.45x10-4**	++++++	11515	0.013	0.004	8.48x10-4	+++++-	11158	Ratio of 18:2 linoleic acid to total fatty acids
M-VLDL-CE %	0.013	0.004	8.52x10-4	+++-+-	11772	0.011	0.004	4.78x10-3	+++-+-	11405	Cholesterol esters to total lipids ratio in medium VLDL
S-VLDL-TG %	-0.013	0.004	8.70x10-4	---+--	11766	-0.011	0.004	6.78x10-3	---+-+	11399	Triglycerides to total lipids ratio in small VLDL
XS-VLDL-CE	0.013	0.004	1.06x10-3	+++++-	11774	0.014	0.004	**3.65x10-4**	+++++-	11407	Cholesterol esters in very small VLDL
PUFA/FA	0.012	0.004	1.19x10-3	++++++	11515	0.011	0.004	4.07x10-3	+++++-	11158	Ratio of polyunsaturated fatty acids to total fatty acids
XS-VLDL-C %	0.011	0.004	1.92x10-3	+++-+-	11773	0.010	0.004	7.70x10-3	+++-+-	11406	Total cholesterol to total lipids ratio in very small VLDL
Phe	-0.011	0.004	2.01x10-3	---+-+	11687	-0.010	0.004	5.55x10-3	---+-+	11322	Phenylalanine
XS-VLDL-TG %	-0.011	0.004	2.06x10-3	---+--	11773	-0.010	0.004	1.11x10-2	---+-+	11406	Triglycerides to total lipids ratio in very small VLDL
XS-VLDL-C	0.012	0.004	2.16x10-3	+++++-	11774	0.014	0.004	7.17x10-4	+++++-	11407	Total cholesterol in very small VLDL
S-HDL-PL	-0.011	0.004	2.39x10-3	----++	11775	-0.011	0.004	2.92x10-3	----++	11408	Phospholipids in small HDL
M-VLDL-C %	0.011	0.004	3.06x10-3	+++-+-	11772	0.010	0.004	1.09x10-2	+++-+-	11405	Total cholesterol to total lipids ratio in medium VLDL
XS-VLDL-CE %	0.010	0.004	3.26x10-3	+++-+-	11773	0.009	0.004	1.54x10-2	+++-+-	11406	Cholesterol esters to total lipids ratio in very small VLDL

Table 2. Continued

Metabolic biomarkers	Model 1					Model 2					Description
	Beta	SE	P-value	Direction*	N	Beta	SE	P-value	Direction*	N	
VLDL-D	-0.011	0.004	3.56x10-3	--+---	11775	-0.009	0.004	2.49x10-2	--+---	11408	Mean diameter for VLDL particles
Alb	0.010	0.004	3.60x10-3	++++++	11775	0.008	0.004	2.66x10-2	-+++++	11408	Albumin
S-VLDL-FC %	0.012	0.004	4.30x10-3	++++++	11766	0.009	0.004	4.25x10-2	++++++	11399	Free cholesterol to total lipids ratio in small VLDL
S-LDL-TG %	-0.010	0.004	4.76x10-3	--+---	11735	-0.009	0.004	1.93x10-2	--+---	11369	Triglycerides to total lipids ratio in small LDL
XL-HDL-C	0.011	0.004	5.45x10-3	+++-++	11775	0.009	0.004	2.20x10-2	+-+-+-	11408	Total cholesterol in very large HDL
IDL-FC	0.011	0.004	5.53x10-3	+++++-	11773	0.011	0.004	5.85x10-3	++++++	11406	Free cholesterol in IDL
HDL-D	0.011	0.004	6.31x10-3	++++++	11775	0.008	0.005	7.21x10-2	+++++-	11408	Mean diameter for HDL particles

Abbreviations: SE = standard error, N = sample size.
Model 1: linear regression analysis with LTL as dependent variable adjusted for metabolic biomarker, age, sex, lipid-lowering medication (yes/no), and if necessary for batch effects, case-control status or familial relationships. Model 2: model 1 + additional adjustment for BMI and smoking.
* Order of cohorts in direction column: LLS, NESDA, NTR, ERF, RS, LLDeep; Direction of effect represented by - (negative association) + (positive association).
P-values in bold surpassed the significance threshold.

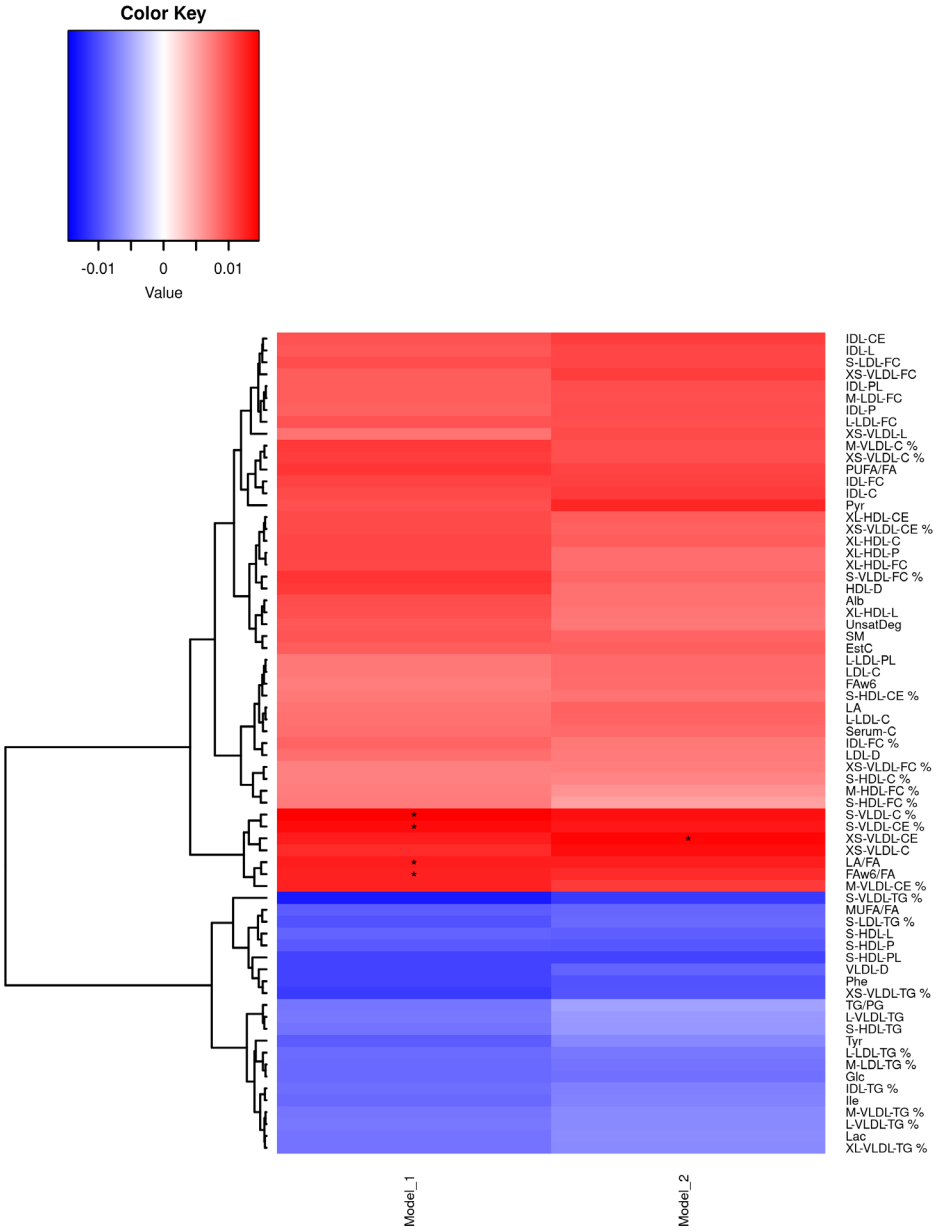


Figure 1. Heat map showing cluster analysis of the metabolic biomarker correlations with LTL (p in model 1 < $1.0x10^{-2}$). Metabolic biomarkers with p < $1.0x10^{-2}$ were included in this Figure. The metabolic biomarkers are displayed vertically (y-axis) and the two models used in this study on the horizontal axis (x-axis). The association analysis was adjusted for age, sex, and lipid-lowering medication in model 1 and additionally for BMI and smoking in model 2. A blue color represents a negative correlation of the metabolite with LTL and a red color represents a positive correlation. Metabolic biomarkers that were significantly associated with LTL (p < $6.5x10^{-4}$) are labelled with a star.

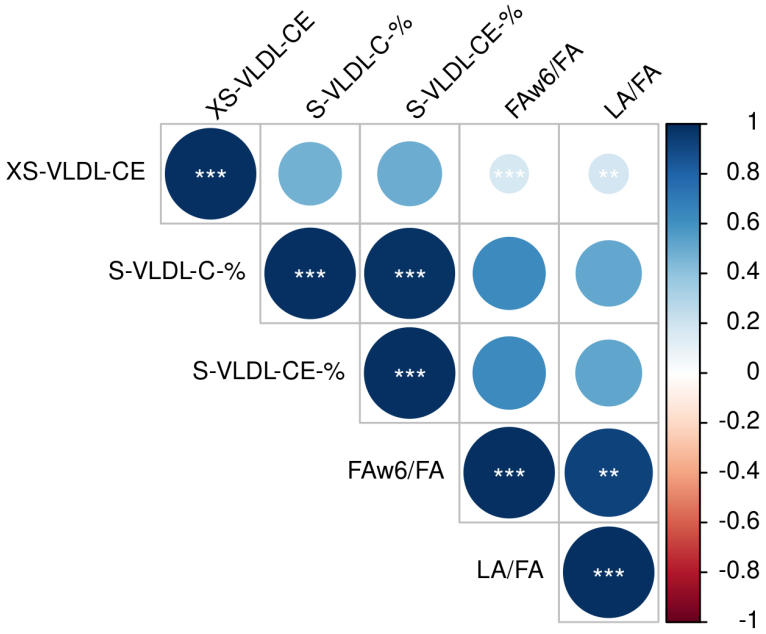


Figure 2. Correlation plot of the metabolic biomarkers associated with LTL after correction for multiple testing using ERF data. Positive correlations are displayed in blue and negative correlations in red. Color intensity and the size of the circle are proportional to the correlation coefficients, with larger and darker circles indicating higher correlation point estimates. * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

We next performed a sensitivity analysis to determine whether the analyses were influenced by the data of the LLDeep cohort, which used a different method to measure LTL (FlowFISH vs. qPCR). After excluding the LLDeep cohort from the meta-analysis, all metabolites remained significantly associated with LTL in model 1 and XS-VLDL-CE was also significantly associated with LTL (**Supplementary Table 5**). In model 2, the results also remained similar and S-VLDL-C % and S-VLDL-CE % became significantly associated with LTL (**Supplementary Table 5**).

In the replication analyses, where a lookup of the significant metabolic biomarkers was performed in the results of the study performed by Couto Alves et al. (see **Materials and Methods** and **Supplementary Material**), we were able to replicate the association FAw6/FA (N = 7,845, $p_{model1} = 1.9 \times 10^{-4}$, $p_{model2} = 1.4 \times 10^{-2}$) with LTL only (**Table 3**). Unfortunately, the other four metabolic biomarkers were not available in the replication data and their association with LTL could not be confirmed. In the sex-stratified analysis, the top findings were consistent in males and females and showed a similar effect size as in the total cohort (**Supplementary Figure 1**), but were not statistically significant, most likely due to a reduction in sample size. Results of the sex-stratified meta-analyses are available in **Supplementary Table 6** and **7** for model 1 and 2, respectively.

Table 3. Replication of the association between LTL and FAw6/FA.

Metabolic biomarkers		Discovery					Replication			
		Beta	SE	P-value	Direction*	N	Beta	P-value	Direction**	N
Model 1	FAw6/FA	0.013	0.004	4.15x10-4	++++++	11514	0.042	1.87x10-4	++++-	7845
Model 2	FAw6/FA	0.012	0.004	1.32x10-3	+++++-	11157	0.029	1.43x10-2	+-+--	7670

* Order of cohorts in direction column: LLS, NESDA, NTR, ERF, RS, LLDeep; Direction of effect represented by - (negative association) or + (positive association).
** Order of cohorts in direction column: NFBC1966, KORA, HBCS, TWINFAT, TWINACTIVE; Direction of effect represented by - (negative association) or + (positive association).

Discussion

We performed an association analysis of NMR based metabolic biomarkers with LTL using data of 11,775 participants from six Dutch cohorts. We found higher levels of five metabolic biomarkers, three lipid subtypes (S-VLDL-C %, S-VLDL-CE %, and XS-VLDL-CE) and two fatty acid ratios (FAw6/FA and LA/FA), associated with higher LTL values. We were able to replicate the association of FAw6/FA in a large sample of 7,845 individuals.

Both fatty acid ratios that were significantly associated with LTL are omega-6 fatty acids, where linoleic acid (LA) is the major dietary omega-6 fatty acid in most Western diets[55]. The association of omega-6 fatty acids with health remains unclear because of the pro-inflammatory as well as anti-inflammatory properties of omega-6 fatty acids[56-58]. LA intake has been shown to be inversely associated with the risk of cardiovascular heart disease[59-61], death of cardiovascular disease, and mortality[62]. This is in agreement with the positive association of the ratio of omega-6 fatty acids to total fatty acids (FAw6/FA) and the ratio of LA to total fatty acids (LA/FA) with LTL. Our findings are in contrast with the findings in the Nurses' Health Study (N = 2,284 females)[63], where LA intake was negatively associated with LTL (*p* for trend 0.05). Of note, the findings of these studies are difficult to compare as we have used quantitative measures of metabolic biomarkers instead of dietary data derived from food frequency questionnaires. More importantly, our data agree with the recent large meta-analysis of the Fatty Acids and Outcomes Research Consortium (FORCE) consortium[64]. This study used data of 68,659 participants from 30 prospective studies and observed that higher circulating LA levels were associated with a lower risk of cardiovascular disease, cardiovascular mortality, and ischemic stroke[64]. The observations with regard to the other type of polyunsaturated fatty acids (PUFA), the omega-3 fatty acids[55], are also of interest. Higher levels of omega-3 fatty acids have been associated with lower levels of pro-inflammatory cytokines, higher levels of anti-inflammatory cytokines and reduced oxidative stress[56]. Higher omega-3 fatty acids levels have also been associated with a reduced rate of

telomere length shortening[65]. Although the PUFA/FA ratio (i.e., the ratio of polyunsaturated fatty acids to total fatty acids) showed a nominally significant association with longer LTL ($p = 1.2x10^{-3}$), this is most likely driven by the omega-6 fatty acid ratios as we observed no association of omega-3 with LTL.

The observed significant association of PUFA ratios (FAw6/FA and LA/FA) with LTL, a predictor of mortality[25-35], is in line with one of our previous studies in which we found that the PUFA/FA was one of the 14 independent circulating biomarkers that were significantly associated with all-cause mortality (HR = 0.78, 95% confidence interval (CI) = 0.75-0.80, $p = 1.1x10^{-47}$)[66]. PUFAs are hydrocarbon chains containing two or more double bonds, which are further classified as either an omega-3 PUFA (FAw3) or omega-6 PUFA (FAw6), which is based on the position of the first double bond relative to the methyl end of the fatty acid[67]. Although we found the strongest association of FAw6/FA with LTL, the PUFA/FA ratio also showed a nominally significant positive association with LTL (β = 0.012, SE = 0.004, $p = 1.2x10^{-3}$). PUFA/FA and FAw6/FA appear to denote the same entity when studying LTL, as their correlation is very high (r = 0.96), and, as a result, when including both PUFA/FA and FAw6/FA in the regression analysis in the Rotterdam Study we found a high level of multicollinearity (variance inflation factor $(VIF)_{PUFA/FA} = 13.3$, $VIF_{FAw6/FA} = 13.4$). When re-examining the association of the FAw6/FA ratio with all-cause mortality[66], we found that FAw6/FA was more strongly associated with mortality (HR = 0.85, 95% CI = 0.83-0.87, $p = 3.0x10^{-39}$) than FAw3/FA (HR = 0.91, 95% CI = 0.89-0.94, $p = 2.8x10^{-11}$)[66].

It is of interest that we identified three VLDL metabolic biomarkers associated with LTL, which were all lipid measures to lipid ratios in VLDL (i.e., S-VLDL-C %, S-VLDL-CE %, and XS-VLDL-CE). We did not find associations of absolute lipid measures of the VLDL sub-fractions, which may imply that type of lipids (i.e., small VLDL) and their composition drive the association with LTL and not the total lipids. Although we were unable to replicate our findings because of the unavailability of these metabolites in other studies, lipid metabolism has been associated before with LTL[51,52]. The mechanisms through which lipids relate to LTL have been discussed before but are far from understood. One mechanism through which lipid as well as fatty acid metabolism may influence telomere length is oxidative stress, which is proposed as a cause of aging[68-70] and is known to attenuate telomere length attrition[71-73]. Fat accumulation has been associated with oxidative stress[74] and previous studies have shown that oxidative stress is involved in the development of age-related diseases, including the metabolic syndrome[74,75]. The metabolic syndrome is defined by clinical and biochemical alterations characterized by multiple components such as obesity, dyslipidaemia, arterial hypertension, hyperglycaemia, and insulin resistance[75,76]. These components have been associated with oxidative damage at DNA and lipid level and also with shorter telomere length[75]. Interestingly, statins have been shown to prevent telomere length shortening through decreasing oxidative stress[77]. Interventions promoting healthy lifestyle may delay telomere shortening and development of the metabolic syndrome. Because of the multifactorial process, development of interventions remains a challenge

and these associations should receive attention in future studies. However, alternative mechanisms are plausible including the effect that lipid metabolism has on inflammation, a major driver of telomere length in blood[78-80]. Also, inflammation and oxidative stress are intertwined[70,78]. As most of the mechanisms are based on epidemiological cross-sectional studies and functional studies are lacking, further functional studies are therefore needed to explore these mechanisms.

The strength of our study is that metabolic biomarkers were measured in a standardized way with the same platform in all cohorts. Using a targeted metabolomics platform is both a strength as well as a limitation for this study. This platform contains a detailed catalogue of lipid sub fractions, cholesterol and triglyceride measures, fatty acids, and various low-molecular metabolites, which enabled us to find supportive evidence for the association of fatty acid metabolism with LTL and more in-depth information on the association of lipid metabolism with LTL. However, using this platform limits us to study other metabolic biomarkers and pathways that might also be related to LTL such as phosphatidylcholines and the methionine-homocysteine pathway which were previously found to be associated with LTL[52]. Another limitation of this study is that participants included in this study were predominantly from European descent, which makes it more difficult to generalize the findings to other populations. Lastly, we did not have longitudinal data to investigate the changes in telomere length over time. A previous study has found an inverse relationship between omega-3 fatty acids intake at baseline and telomere length shortening rate over 5 years, while there was no association between omega-3 fatty acid levels and telomere length at baseline[65]. Longitudinal data might therefore provide further insights into telomere length biology and the opportunity to investigate reverse causality.

To conclude, we found subclasses of VLDL and ratios of omega-6 fatty acids to total fatty acids significantly associated with LTL. We were able to replicate the association of higher FAw6/FA levels with longer LTL in an independent dataset. These findings further support the association of lipid and fatty acid metabolism with LTL and provide more detailed information on the association of specific lipoprotein subclasses and fatty acids with telomere length. In the future, these findings might help to create prevention and therapeutic strategies to increase healthy aging.

Materials and Methods

Discovery populations

Leiden Longevity Study (LLS): long-lived siblings of Dutch descent were recruited together with their offspring and the partners of the offspring. Families were included if at least two long-lived siblings were alive and fulfilled the age criterion of 89 years or older for males and 91 years or older for females. In total, 944 long-lived proband siblings from 421 families with a mean age of 94 years (range, 89-104), 1,671 offspring (61 years, 39-81), and 744 offspring's partners (60 years, 36-79) were included in the study[81]. DNA was extracted from samples

(non-fasted) at baseline using conventional methods[82]. For the current analysis only the offspring and their partners were used of whom LTL and Nightingale metabolomics data were available (N = 1,858).

Netherlands Study of Depression and Anxiety (NESDA): a multi-centre study consisting of 2,981 participants aged 18 to 65 years with depression or anxiety disorders (current and in the past) and healthy controls[83]. Baseline data, collected between 2004 and 2007, of 2,885 participants were included in this analysis. DNA and plasma levels were collected after overnight fast in a standardized manner.

Netherlands Twin Register (NTR): twins and their siblings and parents were recruited to study the causes of individual differences in health, behavior, and lifestyle. Participants are followed longitudinally and details about the cohort have been published previously[84,85]. A subsample of unselected twins and their family members has taken part in the NTR-Biobank[86] in which biological samples, including DNA and RNA, were collected in a standardized manner after overnight fasting. In total, 4,170 NTR participants with LTL measures were included in this study.

Erasmus Rucphen Family (ERF) study: a family-based study consisting of ~3,000 inhabitants from an isolated population in the southwest of the Netherlands[87,88]. ERF participants are descendants of 22 founder couples that had at least six children baptized in the community church. At baseline, participants were screened for many quantitative traits related to common diseases of interest. Samples were collected after overnight fasting in a standardized manner. Baseline data was collected between 2002 and 2005, and data of 1,243 participants were included in this study.

Rotterdam Study (RS): a prospective population-based cohort consisting of 14,926 participants from the Ommoord district in Rotterdam, the Netherlands, aiming at investigating the occurrence and causes of diseases that are frequent in the elderly[89]. The RS cohort was initialised in 1990 including participants of 55 years and older (RS-I, N = 7,983). At baseline, all participants were interviewed at home and had an extensive set of examinations, which were repeated every 3-4 years. The cohort was further extended in 2000 (RS-II) and 2005 (RS-III). For this study, the fourth follow-up visit from the first RS cohort was used (RSI-4, N = 683). Samples were collected after overnight fasting in a standardized manner.

LifeLines-DEEP (LLDeep) study: a population-based cohort study in the northern Netherlands including 1,539 participants of 18 years and older[90]. We included 936 participants in this analysis, for whom both telomere length and NMR data were available. Samples were collected after overnight fasting.

All participants provided written informed consent, and all studies were approved by the relevant institutional boards.

Leukocyte telomere length measurements

In all cohorts, except in the LLDeep study, mean LTL was measured with a quantitative PCR-based technique (qPCR) as previously described[44,52,91,92]. Telomere length was expressed as the ratio (T/S) of telomere repeat length (T) to copy number of the single copy gene *36B4* (S) in each sample. In the LLDeep cohort mean LTL was measured by FlowFISH using the previously published protocols[53]. Telomere length measurements in lymphocytes were used for this analysis.

Metabolic measurements

The metabolic biomarkers were quantified from EDTA plasma and serum samples using a high-throughput proton NMR metabolomics platform (Nightingale Health, Helsinki, Finland), as described earlier[54,66,93]. The NMR platform enables quantification of 14 lipoprotein subclasses, their lipid concentrations and composition, apolipoproteins, various cholesterol and triglyceride measures, albumin, fatty acids and other small metabolites including amino acids, glycolysis related measures and ketone bodies. In this study, we included all 226 available metabolic biomarkers, of which a full list is shown in **Supplementary Table 1**. Quality control of metabolic biomarkers was done in a standardized manner. First, metabolic biomarkers that failed quality control as indicated by Nightingale Health were excluded from the analysis. Second, metabolic biomarkers with more than 10% missing values were removed. Third, a value of one was added to all metabolic biomarkers included in the analysis to take into account metabolic biomarkers with values below the detection limit, followed by a natural logarithm (LN) transformation to adjust for deviation from a normal distribution. Finally, all metabolic biomarkers were scaled to standard deviation units in order to standardize measurements across cohorts.

Statistical analysis

Linear regression analysis was performed over all metabolic traits per cohort using LTL as outcome variable and each metabolite as independent variable, adjusting for age, sex, lipid-lowering medication (yes/no), and if necessary, for batch effects, case-control status or familial relationships (model 1). In the second model (model 2), we additionally adjusted for body mass index (BMI) and smoking (current smoking: yes/no) as these factors might have an effect on both LTL as well as the human metabolome[94-99]. BMI (kg/m^2) was calculated using the standard formula: weight (kg) divided by height in meters squared (m^2). Both models were analysed in the total sample and in males and females separately. Inverse variance-weighted fixed effects meta-analyses were performed using METAL software[100]. Heterogeneity was assessed using Cochran's Q-test as implemented by METAL software. We additionally performed a sensitivity analysis to test the robustness of the results and repeated the meta-analysis excluding the LLDeep cohort as telomere length was measured using a different method (FlowFISH vs. qPCR). To correct for multiple testing, we calculated the number of independent metabolic biomarkers in the Rotterdam Study[101]. There were 77 independent metabolic biomarkers, resulting in a Bonferroni corrected *p-value* of 6.5×10^{-4} (= 0.05/77).

Replication analysis

We performed a lookup of the metabolites that were significantly associated with LTL in the results of Couto Alves et al (Supplementary Material). In this study, data on up to 20,155 individuals of European ancestry from 11 cohorts were used to replicate our findings[102]. Metabolite data was generated on one or two NMR platforms[54,103]. Included cohorts were: Northern Finland Birth Cohort 1966, Northern Finland Birth Cohort 1986, LLS, NTR, ERF study, Cooperative Health Research in the Augsburg Region study, Estonian Genome Center of University of Tartu Cohort, Helsinki Birth Cohort Study, TWINFAT, TWINACTIVE, and HRT Twins. Although three of these cohorts (LLS, NTR, and ERF study) overlap with the cohorts included in the discovery population of the current study, there is no overlap in individuals between the discovery and replication populations for the metabolites examined for replication. Couto Alves et al. used two statistical models in their analyses, where in the first model LTL was regressed on metabolite levels adjusting for age and sex and for family structure if needed, and the second model was additionally adjusted for BMI.

References

1 Blackburn, E. H. Structure and function of telomeres. *Nature* **350**, 569-573, doi:10.1038/350569a0 (1991).
2 Lindsey, J., McGill, N. I., Lindsey, L. A., Green, D. K. & Cooke, H. J. In vivo loss of telomeric repeats with age in humans. *Mutat Res* **256**, 45-48, doi:10.1016/0921-8734(91)90032-7 (1991).
3 Slagboom, P. E., Droog, S. & Boomsma, D. I. Genetic determination of telomere size in humans: a twin study of three age groups. *Am J Hum Genet* **55**, 876-882 (1994).
4 Abdallah, P. *et al.* A two-step model for senescence triggered by a single critically short telomere. *Nat Cell Biol* **11**, 988-993, doi:10.1038/ncb1911 (2009).
5 Blackburn, E. H. Switching and signaling at the telomere. *Cell* **106**, 661-673, doi:10.1016/S0092-8674(01)00492-5 (2001).
6 Campisi, J. & d'Adda di Fagagna, F. Cellular senescence: when bad things happen to good cells. *Nat Rev Mol Cell Biol* **8**, 729-740, doi:10.1038/nrm2233 (2007).
7 Fasching, C. L. Telomere length measurement as a clinical biomarker of aging and disease. *Crit Rev Clin Lab Sci* **55**, 443-465, doi:10.1080/10408363.2018.1504274 (2018).
8 Telomeres Mendelian Randomization, C. *et al.* Association Between Telomere Length and Risk of Cancer and Non-Neoplastic Diseases: A Mendelian Randomization Study. *JAMA Oncol* **3**, 636-651, doi:10.1001/jamaoncol.2016.5945 (2017).
9 Willeit, P. *et al.* Telomere length and risk of incident cancer and cancer mortality. *JAMA* **304**, 69-75, doi:10.1001/jama.2010.897 (2010).
10 McGrath, M., Wong, J. Y., Michaud, D., Hunter, D. J. & De Vivo, I. Telomere length, cigarette smoking, and bladder cancer risk in men and women. *Cancer Epidemiol Biomarkers Prev* **16**, 815-819, doi:10.1158/1055-9965.EPI-06-0961 (2007).
11 Risques, R. A. *et al.* Leukocyte telomere length predicts cancer risk in Barrett's esophagus. *Cancer Epidemiol Biomarkers Prev* **16**, 2649-2655, doi:10.1158/1055-9965.EPI-07-0624 (2007).
12 Wu, X. *et al.* Telomere dysfunction: a potential cancer predisposition factor. *J Natl Cancer Inst* **95**, 1211-1218 (2003).
13 Fitzpatrick, A. L. *et al.* Leukocyte telomere length and cardiovascular disease in the cardiovascular health study. *Am J Epidemiol* **165**, 14-21, doi:10.1093/aje/kwj346 (2007).
14 Rehkopf, D. H. *et al.* Leukocyte Telomere Length in Relation to 17 Biomarkers of Cardiovascular Disease Risk: A Cross-Sectional Study of US Adults. *PLoS Med* **13**, e1002188, doi:10.1371/journal.pmed.1002188 (2016).
15 Maubaret, C. G. *et al.* Telomeres are shorter in myocardial infarction patients compared to healthy subjects: correlation with environmental risk factors. *J Mol Med (Berl)* **88**, 785-794, doi:10.1007/s00109-010-0624-3 (2010).
16 Brouilette, S., Singh, R. K., Thompson, J. R., Goodall, A. H. & Samani, N. J. White cell telomere length and risk of premature myocardial infarction. *Arterioscler Thromb Vasc Biol* **23**, 842-846, doi:10.1161/01.ATV.0000067426.96344.32 (2003).
17 Brouilette, S. W. *et al.* Telomere length, risk of coronary heart disease, and statin treatment in the West of Scotland Primary Prevention Study: a nested case-control study. *Lancet* **369**, 107-114, doi:10.1016/S0140-6736(07)60071-3 (2007).
18 Haycock, P. C. *et al.* Leucocyte telomere length and risk of cardiovascular disease: systematic review and meta-analysis. *BMJ* **349**, g4227 (2014).
19 Samani, N. J., Boultby, R., Butler, R., Thompson, J. R. & Goodall, A. H. Telomere shortening in atherosclerosis. *Lancet* **358**, 472-473, doi:10.1016/S0140-6736(01)05633-1 (2001).
20 Jeanclos, E. *et al.* Shortened telomere length in white blood cells of patients with IDDM. *Diabetes* **47**, 482-486 (1998).
21 Zhao, J., Miao, K., Wang, H., Ding, H. & Wang, D. W. Association between telomere length and type 2 diabetes mellitus: a meta-analysis. *PLoS One* **8**, e79993, doi:10.1371/journal.pone.0079993 (2013).
22 Demissie, S. *et al.* Insulin resistance, oxidative stress, hypertension, and leukocyte telomere length in men from the Framingham Heart Study. *Aging Cell* **5**, 325-330, doi:10.1111/j.1474-9726.2006.00224.x (2006).
23 Honig, L. S., Kang, M. S., Schupf, N., Lee, J. H. & Mayeux, R. Association of shorter leukocyte telomere repeat length with dementia and mortality. *Arch Neurol* **69**, 1332-1339, doi:10.1001/archneurol.2012.1541 (2012).

24 Martin-Ruiz, C. *et al.* Telomere length predicts poststroke mortality, dementia, and cognitive decline. *Ann Neurol* **60**, 174-180, doi:10.1002/ana.20869 (2006).
25 Deelen, J. *et al.* Leukocyte telomere length associates with prospective mortality independent of immune-related parameters and known genetic markers. *Int J Epidemiol* **43**, 878-886, doi:10.1093/ije/dyt267 (2014).
26 Cawthon, R. M., Smith, K. R., O'Brien, E., Sivatchenko, A. & Kerber, R. A. Association between telomere length in blood and mortality in people aged 60 years or older. *Lancet* **361**, 393-395, doi:10.1016/S0140-6736(03)12384-7 (2003).
27 Bakaysa, S. L. *et al.* Telomere length predicts survival independent of genetic influences. *Aging Cell* **6**, 769-774, doi:10.1111/j.1474-9726.2007.00340.x (2007).
28 Kimura, M. *et al.* Telomere length and mortality: a study of leukocytes in elderly Danish twins. *Am J Epidemiol* **167**, 799-806, doi:10.1093/aje/kwm380 (2008).
29 Ehrlenbach, S. *et al.* Influences on the reduction of relative telomere length over 10 years in the population-based Bruneck Study: introduction of a well-controlled high-throughput assay. *Int J Epidemiol* **38**, 1725-1734, doi:10.1093/ije/dyp273 (2009).
30 Epel, E. S. *et al.* The rate of leukocyte telomere shortening predicts mortality from cardiovascular disease in elderly men. *Aging (Albany NY)* **1**, 81-88, doi:10.18632/aging.100007 (2008).
31 Fitzpatrick, A. L. *et al.* Leukocyte telomere length and mortality in the Cardiovascular Health Study. *J Gerontol A Biol Sci Med Sci* **66**, 421-429, doi:10.1093/gerona/glq224 (2011).
32 Marioni, R. E. *et al.* The epigenetic clock and telomere length are independently associated with chronological age and mortality. *Int J Epidemiol*, doi:10.1093/ije/dyw041 (2016).
33 Rode, L., Nordestgaard, B. G. & Bojesen, S. E. Peripheral blood leukocyte telomere length and mortality among 64,637 individuals from the general population. *J Natl Cancer Inst* **107**, djv074, doi:10.1093/jnci/djv074 (2015).
34 Wang, Q., Zhan, Y., Pedersen, N. L., Fang, F. & Hagg, S. Telomere Length and All-Cause Mortality: A Meta-analysis. *Ageing Res Rev* **48**, 11-20, doi:10.1016/j.arr.2018.09.002 (2018).
35 Arbeev, K. G. *et al.* Association of Leukocyte Telomere Length With Mortality Among Adult Participants in 3 Longitudinal Studies. *JAMA Netw Open* **3**, e200023, doi:10.1001/jamanetworkopen.2020.0023 (2020).
36 Martin-Ruiz, C. M., Gussekloo, J., van Heemst, D., von Zglinicki, T. & Westendorp, R. G. Telomere length in white blood cells is not associated with morbidity or mortality in the oldest old: a population-based study. *Aging Cell* **4**, 287-290, doi:10.1111/j.1474-9726.2005.00171.x (2005).
37 Bischoff, C. *et al.* No association between telomere length and survival among the elderly and oldest old. *Epidemiology* **17**, 190-194, doi:10.1097/01.ede.0000199436.55248.10 (2006).
38 Harris, S. E. *et al.* The association between telomere length, physical health, cognitive ageing, and mortality in non-demented older people. *Neurosci Lett* **406**, 260-264, doi:10.1016/j.neulet.2006.07.055 (2006).
39 Njajou, O. T. *et al.* Association between telomere length, specific causes of death, and years of healthy life in health, aging, and body composition, a population-based cohort study. *J Gerontol A Biol Sci Med Sci* **64**, 860-864, doi:10.1093/gerona/glp061 (2009).
40 Broer, L. *et al.* Meta-analysis of telomere length in 19,713 subjects reveals high heritability, stronger maternal inheritance and a paternal age effect. *Eur J Hum Genet* **21**, 1163-1168, doi:10.1038/ejhg.2012.303 (2013).
41 Nersisyan, L., Nikoghosyan, M., Arakelyan, A. & Genome of the Netherlands, c. WGS-based telomere length analysis in Dutch family trios implicates stronger maternal inheritance and a role for RRM1 gene. *Sci Rep* **9**, 18758, doi:10.1038/s41598-019-55109-7 (2019).
42 Njajou, O. T. *et al.* Telomere length is paternally inherited and is associated with parental lifespan. *Proc Natl Acad Sci U S A* **104**, 12135-12139, doi:10.1073/pnas.0702703104 (2007).
43 Codd, V. *et al.* Identification of seven loci affecting mean telomere length and their association with disease. *Nat Genet* **45**, 422-427, 427e421-422, doi:10.1038/ng.2528 (2013).
44 Codd, V. *et al.* Common variants near TERC are associated with mean telomere length. *Nat Genet* **42**, 197-199, doi:10.1038/ng.532 (2010).
45 Li, C. *et al.* Genome-wide Association Analysis in Humans Links Nucleotide Metabolism to Leukocyte Telomere Length. *Am J Hum Genet* **106**, 389-404, doi:10.1016/j.ajhg.2020.02.006 (2020).
46 Zhan, Y. *et al.* Telomere Length Shortening and Alzheimer Disease--A Mendelian Randomization Study. *JAMA Neurol* **72**, 1202-1203, doi:10.1001/jamaneurol.2015.1513 (2015).

47 Zhang, C. *et al.* Genetic determinants of telomere length and risk of common cancers: a Mendelian randomization study. *Hum Mol Genet* **24**, 5356-5366, doi:10.1093/hmg/ddv252 (2015).
48 Iles, M. M. *et al.* The effect on melanoma risk of genes previously associated with telomere length. *J Natl Cancer Inst* **106**, doi:10.1093/jnci/dju267 (2014).
49 Zhan, Y. *et al.* Exploring the Causal Pathway From Telomere Length to Coronary Heart Disease: A Network Mendelian Randomization Study. *Circ Res* **121**, 214-219, doi:10.1161/CIRCRESAHA.116.310517 (2017).
50 Zhao, J. *et al.* Metabolic profiles of biological aging in American Indians: the Strong Heart Family Study. *Aging (Albany NY)* **6**, 176-186, doi:10.18632/aging.100644 (2014).
51 Zierer, J. *et al.* Metabolomics profiling reveals novel markers for leukocyte telomere length. *Aging (Albany NY)* **8**, 77-94, doi:10.18632/aging.100874 (2016).
52 van der Spek, A. *et al.* Metabolomics reveals a link between homocysteine and lipid metabolism and leukocyte telomere length: the ENGAGE consortium. *Sci Rep* **9**, 11623, doi:10.1038/s41598-019-47282-6 (2019).
53 Baerlocher, G. M., Vulto, I., de Jong, G. & Lansdorp, P. M. Flow cytometry and FISH to measure the average length of telomeres (flow FISH). *Nat Protoc* **1**, 2365-2376, doi:10.1038/nprot.2006.263 (2006).
54 Soininen, P., Kangas, A. J., Wurtz, P., Suna, T. & Ala-Korpela, M. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circ Cardiovasc Genet* **8**, 192-206, doi:10.1161/CIRCGENETICS.114.000216 (2015).
55 Wang, D. D. Dietary n-6 polyunsaturated fatty acids and cardiovascular disease: Epidemiologic evidence. *Prostaglandins Leukot Essent Fatty Acids* **135**, 5-9, doi:10.1016/j.plefa.2018.05.003 (2018).
56 Ferrucci, L. *et al.* Relationship of plasma polyunsaturated fatty acids to circulating inflammatory markers. *J Clin Endocrinol Metab* **91**, 439-446, doi:10.1210/jc.2005-1303 (2006).
57 Simopoulos, A. P. The importance of the omega-6/omega-3 fatty acid ratio in cardiovascular disease and other chronic diseases. *Exp Biol Med (Maywood)* **233**, 674-688, doi:10.3181/0711-MR-311 (2008).
58 Weylandt, K. H. & Kang, J. X. Rethinking lipid mediators. *Lancet* **366**, 618-620, doi:10.1016/S0140-6736(05)67119-X (2005).
59 Farvid, M. S. *et al.* Dietary linoleic acid and risk of coronary heart disease: a systematic review and meta-analysis of prospective cohort studies. *Circulation* **130**, 1568-1578, doi:10.1161/CIRCULATIONAHA.114.010236 (2014).
60 Mozaffarian, D., Micha, R. & Wallace, S. Effects on coronary heart disease of increasing polyunsaturated fat in place of saturated fat: a systematic review and meta-analysis of randomized controlled trials. *PLoS Med* **7**, e1000252, doi:10.1371/journal.pmed.1000252 (2010).
61 Sacks, F. M. *et al.* Dietary Fats and Cardiovascular Disease: A Presidential Advisory From the American Heart Association. *Circulation* **136**, e1-e23, doi:10.1161/CIR.0000000000000510 (2017).
62 Wu, J. H. *et al.* Circulating omega-6 polyunsaturated fatty acids and total and cause-specific mortality: the Cardiovascular Health Study. *Circulation* **130**, 1245-1253, doi:10.1161/CIRCULATIONAHA.114.011590 (2014).
63 Cassidy, A. *et al.* Associations between diet, lifestyle factors, and telomere length in women. *Am J Clin Nutr* **91**, 1273-1280, doi:10.3945/ajcn.2009.28947 (2010).
64 Marklund, M. *et al.* Biomarkers of Dietary Omega-6 Fatty Acids and Incident Cardiovascular Disease and Mortality. *Circulation* **139**, 2422-2436, doi:10.1161/CIRCULATIONAHA.118.038908 (2019).
65 Farzaneh-Far, R. *et al.* Association of marine omega-3 fatty acid levels with telomeric aging in patients with coronary heart disease. *JAMA* **303**, 250-257, doi:10.1001/jama.2009.2008 (2010).
66 Deelen, J. *et al.* A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals. *Nature Communications* **10**, 3346, doi:10.1038/s41467-019-11311-9 (2019).
67 Saini, R. K. & Keum, Y. S. Omega-3 and omega-6 polyunsaturated fatty acids: Dietary sources, metabolism, and significance - A review. *Life Sci.* **203**, 255-267, doi:10.1016/j.lfs.2018.04.049 (2018).
68 Osterod, M. *et al.* Age-related and tissue-specific accumulation of oxidative DNA base damage in 7,8-dihydro-8-oxoguanine-DNA glycosylase (Ogg1) deficient mice. *Carcinogenesis* **22**, 1459-1463, doi:10.1093/carcin/22.9.1459 (2001).
69 Saretzki, G. & Von Zglinicki, T. Replicative aging, telomeres, and oxidative stress. *Ann. N. Y. Acad. Sci.* **959**, 24-29, doi:10.1111/j.1749-6632.2002.tb02079.x (2002).
70 Liguori, I. *et al.* Oxidative stress, aging, and diseases. *Clin. Interv. Aging* **13**, 757-772, doi:10.2147/CIA.S158513 cia-13-757 [pii] (2018).

71 von Zglinicki, T., Saretzki, G., Docke, W. & Lotze, C. Mild hyperoxia shortens telomeres and inhibits proliferation of fibroblasts: a model for senescence? *Exp Cell Res* **220**, 186-193, doi:10.1006/excr.1995.1305 (1995).

72 von Zglinicki, T. Oxidative stress shortens telomeres. *Trends Biochem Sci* **27**, 339-344 (2002).

73 Ahmed, W. & Lingner, J. Impact of oxidative stress on telomere biology. *Differentiation* **99**, 21-27, doi:S0301-4681(17)30121-4 [pii] 10.1016/j.diff.2017.12.002 (2018).

74 Furukawa, S. *et al.* Increased oxidative stress in obesity and its impact on metabolic syndrome. *J. Clin. Invest.* **114**, 1752-1761, doi:10.1172/JCI21625 (2004).

75 Gavia-García, G. *et al.* Telomere Length and Oxidative Stress and Its Relation with Metabolic Syndrome Components in the Aging. *Biology (Basel)* **10** (2021).

76 Grundy, S. M. *et al.* Definition of metabolic syndrome: Report of the National Heart, Lung, and Blood Institute/American Heart Association conference on scientific issues related to definition. *Circulation* **109**, 433-438 (2004).

77 Boccardi, V. & Paolisso, G. The association between statins and telomere shortening. *Clin. Lipidol.* **9**, 311-315, doi:10.2217/clp.14.21 (2014).

78 O'Donovan, A. *et al.* Cumulative inflammatory load is associated with short leukocyte telomere length in the Health, Aging and Body Composition Study. *PLoS One* **6**, e19687, doi:10.1371/journal.pone.0019687 (2011).

79 Bekaert, S. *et al.* Telomere length and cardiovascular risk factors in a middle-aged population free of overt cardiovascular disease. *Aging Cell* **6**, 639-647, doi:10.1111/j.1474-9726.2007.00321.x (2007).

80 Wong, J. Y., De Vivo, I., Lin, X., Fang, S. C. & Christiani, D. C. The relationship between inflammatory biomarkers and telomere length in an occupational prospective cohort study. *PLoS One* **9**, e87348, doi:10.1371/journal.pone.0087348 PONE-D-13-42930 [pii] (2014).

81 Schoenmaker, M. *et al.* Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *Eur J Hum Genet* **14**, 79-84, doi:10.1038/sj.ejhg.5201508 (2006).

82 Beekman, M. *et al.* Chromosome 4q25, microsomal transfer protein gene, and human longevity: novel data and a meta-analysis of association studies. *J Gerontol A Biol Sci Med Sci* **61**, 355-362 (2006).

83 Penninx, B. W. *et al.* The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. *Int J Methods Psychiatr Res* **17**, 121-140 (2008).

84 Boomsma, D. I. *et al.* Netherlands Twin Register: from twins to twin families. *Twin Res Hum Genet* **9**, 849-857 (2006).

85 Ligthart, L. *et al.* The Netherlands Twin Register: Longitudinal Research Based on Twin and Twin-Family Designs. *Twin Res Hum Genet*, 1-14, doi:10.1017/thg.2019.93 (2019).

86 Willemsen, G. *et al.* The Netherlands Twin Register biobank: a resource for genetic epidemiological studies. *Twin Res Hum Genet* **13**, 231-245, doi:10.1375/twin.13.3.231 (2010).

87 Aulchenko, Y. S. *et al.* Linkage disequilibrium in young genetically isolated Dutch population. *Eur J Hum Genet* **12**, 527-534, doi:10.1038/sj.ejhg.5201188 (2004).

88 Pardo, L. M., MacKay, I., Oostra, B., van Duijn, C. M. & Aulchenko, Y. S. The effect of genetic drift in a young genetically isolated population. *Annals of Human Genetics* **69**, 288-295, doi:DOI 10.1046/j.1529-8817.2005.00162.x (2005).

89 Ikram, M. A. *et al.* Objectives, design and main findings until 2020 from the Rotterdam Study. *Eur J Epidemiol* **35**, 483-517, doi:10.1007/s10654-020-00640-5 (2020).

90 Tigchelaar, E. F. *et al.* Cohort profile: LifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: study design and baseline characteristics. *BMJ Open* **5**, e006772, doi:10.1136/bmjopen-2014-006772 (2015).

91 Cawthon, R. M. Telomere measurement by quantitative PCR. *Nucleic Acids Res* **30**, e47 (2002).

92 van der Spek, A. *et al.* Exome Sequencing Analysis Identifies Rare Variants in ATM and RPL8 That Are Associated With Shorter Telomere Length. *Front Genet* **11**, 337, doi:10.3389/fgene.2020.00337 (2020).

93 Hagenbeek, F. A. *et al.* Heritability estimates for 361 blood metabolites across 40 genome-wide association studies. *Nat Commun* **11**, 39, doi:10.1038/s41467-019-13770-6 (2020).

94 Thevenot, E. A., Roux, A., Xu, Y., Ezan, E. & Junot, C. Analysis of the Human Adult Urinary Metabolome Variations with Age, Body Mass Index, and Gender by Implementing a Comprehensive Workflow for Univariate and OPLS Statistical Analyses. *J Proteome Res* **14**, 3322-3335, doi:10.1021/acs.jproteome.5b00354 (2015).

95 Yu, Z. *et al.* Human serum metabolic profiles are age dependent. *Aging Cell* **11**, 960-967, doi:10.1111/j.1474-9726.2012.00865.x (2012).
96 Valdes, A. M. *et al.* Obesity, cigarette smoking, and telomere length in women. *Lancet* **366**, 662-664, doi:10.1016/S0140-6736(05)66630-5 (2005).
97 Lee, M., Martin, H., Firpo, M. A. & Demerath, E. W. Inverse association between adiposity and telomere length: The Fels Longitudinal Study. *Am J Hum Biol* **23**, 100-106, doi:10.1002/ajhb.21109 (2011).
98 Nordfjall, K. *et al.* Telomere length is associated with obesity parameters but with a gender difference. *Obesity (Silver Spring)* **16**, 2682-2689, doi:10.1038/oby.2008.413 (2008).
99 Huzen, J. *et al.* Telomere length loss due to smoking and metabolic traits. *J Intern Med* **275**, 155-163, doi:10.1111/joim.12149 (2014).
100 Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191, doi:10.1093/bioinformatics/btq340 (2010).
101 Li, J. & Ji, L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb)* **95**, 221-227, doi:10.1038/sj.hdy.6800717 (2005).
102 Couto Alves, A. *et al.* Circulating metabolite associations with leukocyte telomere length are mostly sex-specific, driven by BMI and interact with age and geographic region. *Manuscript in preparation* (2019).
103 Vaarhorst, A. A. *et al.* A metabolomic profile is associated with the risk of incident coronary heart disease. *Am Heart J* **168**, 45-52 e47, doi:10.1016/j.ahj.2014.01.019 (2014).

Supplementary Information

Supplementary Figure 1. Scatterplot of the results (betas) in the sex-stratified analysis. The betas of the female-specific analysis are shown on the x-axis and betas of the male-specific analysis on the y-axis.

Supplementary Table 1. List of metabolic biomarkers included in this study.

Supplementary Table 2. Complete results of association analysis of metabolic biomarkers with LTL for model 1 and model 2.

Supplementary Table 3. Results of the association analysis in the individual cohorts, model 1.

Supplementary Table 4. Results of the association analysis in the individual cohorts, model 2.

Supplementary Table 5. Results of the sensitivity analysis, excluding the LLDeep cohort from the meta-analysis, model 1 and 2.

Supplementary Table 6. Complete association results of metabolic biomarkers with LTL for model 1, separately for the total sample, males, and females.

Supplementary Table 7. Complete association results of metabolic biomarkers with LTL for model 2, separately for the total sample, males, and females

CHAPTER 2.4

Revisiting the role of insulin-like growth factor-I receptor stimulating activity and the apolipoprotein E in Alzheimer's disease

Sara A. Galle, Ashley van der Spek, Madeleine L. Drent, Michael P. Brugts, Erik J.A. Scherder, Joseph A.M.J.L. Janssen, M. Arfan Ikram, Cornelia M. van Duijn

This chapter was published in *Frontiers in Aging Neuroscience*. 2019 Feb 12;11:20.

Abstract

Background

Alterations in insulin-like growth factor I (IGF-I) signaling have been associated with dementia and Alzheimer's disease (AD). Studies on the association between IGF-I levels and dementia risk have been inconclusive. We reported earlier that higher levels of IGF-I receptor stimulating activity are associated with a higher prevalence and incidence of dementia.

Objective

In the present study, we test the robustness of the association between IGF-I receptor stimulating activity and dementia by extending the follow-up period to 16 years and investigate possible effect modification by apolipoprotein E (ApoE).

Methods

At baseline, circulating IGF-I receptor stimulating activity was determined by the IGF-I kinase receptor activation (KIRA) assay in 1,014 elderly from the Rotterdam Study. Dementia was assessed from baseline (1997-1999) to follow-up in January 2015. Associations of IGF-I receptor stimulating activity and incident dementia were assessed with Cox proportional hazards models.

Results

During 10,752 person-years of follow-up, 174 people developed dementia. In the extended follow-up we no longer observed a dose-response relationship between IGF-I receptor stimulating activity and risk of dementia (adjusted odds ratio 1.11; 95% confidence interval (CI) 0.97-1.28). Interestingly, we found evidence of an interaction between *ApoE-ε4* and tertiles of IGF-I receptor stimulating activity. IGF-I receptor stimulating activity in the median and top tertiles was related to increased dementia incidence in hetero- and homozygotes of the *ApoE-ε4* allele, but did not show any association with dementia risk in people without the *ApoE-ε4* allele (adjusted odds ratio medium vs. low IGF-I receptor stimulating activity in *ApoE-ε4* carriers: 1.45; 95% CI 1.00-2.12). These findings suggest a threshold effect in *ApoE-ε4* carriers. In line with the hypothesis that downregulation of IGF-I signaling is associated with increased dementia risk, *ApoE-ε4* homozygotes without prevalent dementia displayed lower levels of IGF-I receptor stimulating activity than heterozygotes and non-carriers.

Conclusion

The findings shed new light on the association between IGF-I signaling and the neuropathology of dementia and ask for replication in other cohorts, using measures of IGF-I receptor stimulating activity rather than total serum levels as putative markers of dementia risk.

Introduction

Insulin-like growth factor I (IGF-I) is a multifunctional peptide hormone known to modulate multiple cellular processes including proliferation, differentiation, energy metabolism, glucose homeostasis, stress resistances and apoptosis. Downregulation of IGF-I signaling is found in the elderly and in patients with type 2 diabetes. In contrast, elevated concentrations of circulating IGF-I have been associated with an increased risk of prostate, breast[1-3], colorectal[2,4,5], and lung[6] cancer. IGF-I signaling is also markedly disturbed in the brain of patients with Alzheimer's Disease (AD)[7-9] with alterations in both the levels and phosphorylation state of IGF-I receptor (IGF-IR) as well as the levels and distribution of IGF-I and IGF-IR mRNA in the brain[10,11]. Dysregulation progresses as the disease advances[12]. It remains unclear whether alterations in IGF-I signaling are a causal factor in the pathogenesis of AD or rather a consequence. Findings of experimental and observational studies have been controversial.

In experimental studies, reduced IGF-I signaling has been linked to increased amyloid β (Aβ) deposition[13,14], development of phosphorylated tau[15-17], increased oxidative stress, neuro-inflammation and apoptosis[18]. IGF-I can increase the transport of Aβ carrier proteins albumin and transthyretin into the brain. Upon systemic administration, brain levels of albumin and transthyretin increased and the fraction of Aβ bound to carrier proteins in the CSF and blood was elevated. Suggesting that IGF-I reduced brain Aβ load, in part by enhancing its clearance through carrier proteins such as albumin and transthyretin[14]. Systemic administration of IGF-I has also been shown to lower the toxicity of Aβ in wild type mice[19] and restore cognitive function in mouse models of AD[20], supporting the potential of IGF-I as a therapeutic target in human patients. Peripheral administration has, however, failed to alter Aβ levels in trials with transgenic rats, mice and dogs[21-24]. In contrast to the neuroprotective role of IGF-I, it has also been suggested that the downregulation of IGF-I signaling attenuates the effects of aging and neurodegeneration. Suppression of IGF-I signaling has been associated with longevity in humans[25] and has shown to delay the process of aging and increase lifespan in model organisms[26-28]. In AD mouse models long term suppression of IGF-IR signaling has been linked to reduced neuronal loss, greater resistance to oxidative stress, neuro-inflammation and Aβ aggregation, and has been associated with prolonged preservation of spatial memory and a reduction of behavioral deficits, even when plasma Aβ levels increased[29-32]. Last but not least, lowering serum IGF-I via a protein restriction diet ameliorated Alzheimer pathology in transgenic mouse models[22].

In human observational studies the role of IGF-I signaling in the risk of AD and dementia remains open to question. Longitudinal analyses of the cumulative dementia incidence in 3,582 participants of the Framingham Heart Study, spanning middle and old age, indicated that for those with the lowest levels of serum IGF-I at baseline dementia risk was increased by 51%[33]. No such relation was found by Green et al.[34] examining the prospective association between total IGF-I, IGF-II, and IGF-I Binding Protein 3 (IGFBP-3) and cognitive function in

724 males participating in the Caerphilly Prospective Study. In this study, both total serum IGF-II and IGFBP-3 were associated with age-related cognitive decline and cognitive impairment, but previous associations of total serum IGF-I with cognitive decline and dementia were not replicated[34]. Correspondingly, a meta-analysis of epidemiological studies on the association between total serum IGF-I and dementia nullified the results of previous studies. Five studies suggested that increased levels of circulating total IGF-I predict a higher risk of AD, while three studies suggested an inverse association and two studies reported no significant differences between groups[12]. Differences in findings across studies are speculatively attributed to differences in age of onset and stage of disease progression, comorbid diabetes, or the differential influence of IGF-I gene polymorphisms. Although, the majority of studies report a contribution of alterations in IGF-I signaling to the prediction of dementia risk independent of apolipoprotein E (*ApoE*) genotype[35-38], Deelen et al. reported an association between the *ApoE-ε4* allele and lowered total serum IGF-I levels in middle-aged women[39]. However, a recent Mendelian randomization study by Williams et al. did not provide any evidence for an association between genetically predicted variation in total IGF-I or its binding protein IGFBP-3 and risk of AD[40]. These findings decrease the probability that total serum IGF-I is the relevant determinant of AD and dementia.

As most of the circulating IGF-I measured in serum is bound to IGF-I binding proteins and therefore biologically inactive, levels of total IGF-I poorly reflect the actual IGF-I bioactivity. We therefore applied an IGF-I specific kinase receptor activation assay (KIRA) to assess IGF-I bioactivity, by measuring IGF-I receptor stimulating activity[41,42]. IGF-I receptor stimulating activity takes into account the modifying effect of IGF-I binding proteins on the interaction between IGF-I and the IGF-I receptor and measures the net effects on IGF-I receptor activation. In a previous study we have shown that IGF-I bioactivity is positively related to total and free IGF-I levels obtained by IGF-I immunoassays. Interestingly, correlations were relatively weak (0.52 for total IGF-I and 0.20 for free IGF-I respectively), suggesting that the IGF-I KIRA assay produces new information about IGF-I signaling[43].

We reported earlier that higher levels of IGF-I receptor stimulating activity were associated with a higher prevalence and a higher incidence of dementia[44]. In light of the conflicting results of the experimental and human studies, we aimed to test the long-term robustness of the association between IGF-I receptor stimulating activity and dementia risk by extending the follow up period with another 4 years and investigate possible effect modification by *ApoE*, the major genetic driver of AD and dementia risk.

Materials and Methods

Setting

This study was embedded within the prospective, population-based Rotterdam Study, designed to study risk factors and determinants of chronic diseases in the elderly population.

The Rotterdam Study began in 1990, with an invitation to inhabitants of 55 years and older residing in Ommoord, a district of Rotterdam in the Netherlands. Of the 10,215 people invited, 7,983 agreed to participate in the examinations at baseline. Up until 2015, there have been five follow-up examinations. Details of the study are described elsewhere[45]. Because IGF-I receptor stimulating activity was measured in blood samples collected at the second follow-up examination, between 1997 and 1999, this visit was used as baseline for the current study. Of the 5,990 participants that were alive in 1997–1999, 4,797 persons participated in the second follow-up assessment. IGF-I receptor stimulating activity levels were measured in blood samples of 1,050 randomly selected participants due to financial constraints. Five participants were excluded because their blood samples could not be correctly matched and 14 participants were excluded because measurements did not pass prior defined assay acceptance criteria (inter-assay coefficient of variation <10%). Another 17 participants were excluded because dementia screening was incomplete. Eventually, 1,014 participants were included in the analyses. The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC (registration number MEC 02.1015) and by the Dutch Ministry of Health, Welfare and Sport (Population Screening Act WBO, license number 1071272-159521-PG). The Rotterdam Study has been entered into the Netherlands National Trial Register (NTR; www.trialregister.nl) and into the WHO International Clinical Trials Registry Platform (ICTRP; www.who.int/ictrp/network/primary/en/) under shared catalogue number NTR6831. All participants provided written informed consent to participate in the study and to have their information obtained from treating physicians.

Assessment of IGF-I receptor stimulating activity

IGF-I receptor stimulating activity levels were measured using an IGF-I kinase receptor activation assay (KIRA) (intra- and inter-assay coefficients of variation of 5.2 and 12.2%, respectively; cross-reactivity of 15% for IGF-II)[42,43]. Details of the assessment are described previously[44].

Assessment of dementia

Participants were screened for dementia at baseline and follow-up examinations using a 3-step protocol[46]. First, screening was performed using the Mini-Mental State Examination (MMSE)[47] and the Geriatric Mental Schedule (GMS) organic level[48]. People with a MMSE score lower than 26 or GMS organic level higher than 0 were subsequently subjected to further examination and informant interview including the Cambridge Examination for Mental Disorders in the Elderly (CAMDEX)[49]. When necessary, participants underwent further neuropsychological assessment. When information on neuro-imaging was available, it was used as an aid for decision-making. For all suspected cases of dementia, the diagnosis was made by a consensus panel, led by a neurologist. During follow-up the cohort was under continuous surveillance for dementia incidence through electronic linkage of the database of the Rotterdam Study with medical records from general practitioners and the regional institute for outpatient mental health care[50]. The applied criteria for the diagnosis of

dementia and probable AD are in accordance with the standard criteria for dementia (Diagnostic and Statistical Manual of Mental Disorders III-revised)[51] and AD (National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association)[52]. The total cohort was continuously monitored for incidence of dementia through linkage to the digitized medical records from general practitioners and the Regional Institute for Outpatient Mental Health Care. Follow-up for incident dementia is complete until January 2015.

Other measurements

Information on *ApoE* genotype was obtained using polymerase chain reaction on coded DNA samples. *ApoE-ε4* carrier status was defined as carrier of one or two ε4 alleles. Blood pressure was calculated as the average of two measurements at the right brachial artery using a random-zero sphygmomanometer. Hypertension was defined as a blood pressure ≥140/90 mmHg or use of blood pressure lowering medication, prescribed for the indication of hypertension. Waist circumference was measured in centimeters. Serum glucose, total cholesterol, and high-density lipoprotein (HDL)-cholesterol levels were acquired by an automated enzymatic procedure (Boehringer Mannheim System). Missing values in covariates (for *ApoE-ε4* carrier status 4.8%, for all other covariates less than 3.5%) were imputed based on age and sex.

Statistical analyses

We examined the association between IGF-I receptor stimulating activity and incident dementia using Cox proportional hazards models. IGF-I receptor stimulating activity was entered per standard deviation (SD) into the models. We also studied IGF-I receptor stimulating activity in tertiles, using the lowest tertile as reference. All models were adjusted for age and sex (Model I) and additionally for hypertension, glucose, waist circumference, *ApoE-ε4* carrier status, total cholesterol, and HDL-cholesterol (Model II) for being potential confounders. To investigate possible effect modification by *ApoE*, the (multiplicative) interaction between *ApoE-ε4* carrier status and IGF-I receptor stimulating activity on dementia risk was tested using interaction terms and separate analyses on data stratified on *ApoE-ε4* carrier status were performed. The underlying time-scale in the Cox proportional hazards models was the follow-up time, which was defined from time at blood sample collection (1997–1999) until the end of December 2015. Participants were censored within this time period when they were diagnosed with dementia, died, or decided to terminate their participation in the study. We separately investigated the association between IGF-I receptor stimulating activity and AD. Analyses were performed using IBM SPSS statistics version 24.0 (IBM Corp, Armonk, NY, USA).

Results

Baseline characteristics of the study population are provided in **Table 1**. At baseline, 31 participants suffered from prevalent dementia, of which 23 had AD. During a follow-up of 10,752 person-years (mean follow-up of 11 years, SD 5.2 years), 174 participants developed dementia, of whom 140 were diagnosed with AD.

In the overall proportional hazard analyses, there was no statistically significant evidence for a relation between the level of IGF-I receptor stimulating activity at baseline and risk of dementia. However, the hazard ratio (HR) per SD increase in IGF-I receptor stimulating activity (1.11; 95% confidence interval (CI) 0.97-1.28; see **Table 2**) was very similar to the HR reported in our previous analyses (1.15; 95% CI 1.00-1.33) with shorter follow-up[44]. A congruent HR was found for the incidence of AD (HR 1.10 (95% CI 0.95-1.28)).

Table 1. Baseline characteristics.

	Prevalent dementia N = 30	At risk for incident dementia N = 984
Age, years	81.54 (8.33)	72.04 (7.1)
Females	70%	55.8%
IGF-IRSA, pmol/L	208.13 (77.59)	179.06 (55.48)
Apolipoprotein E-ε4 carrier status	71.4%	27.4%
Hypertension	82.1%	75.5%
Waist circumference, cm	94.17 (8.33)	93.84 (11.1)
Glucose, mmol/L	6.03 (1.07)	6.01 (1.51)
Total cholesterol, mmol/L	5.52 (1.17)	5.83 (1)
HDL-cholesterol, mmol/L	1.34 (0.46)	1.38 (0.37)

Data are presented as means (standard deviations) or percentages. N = number of people, IGF-IRSA = insulin-like growth factor I stimulating receptor activity, HDL = high-density lipoprotein.

Table 2. IGF-I receptor stimulating activity and risk of incident dementia.

	Dementia	Alzheimer's disease
	HR (95% CI) n/N 174/973	HR (95% CI) n/N 140/973
Model I	1.09 (0.95-1.25)	1.07 (0.92-1.25)
Model II	1.11 (0.97-1.28)	1.10 (0.95-1.28)

Data are presented as hazard ratios (HR) and 95% confidence intervals (CI). N = number of people at risk for incident dementia, n = number of cases of incident dementia.
Model I: adjusted for age and sex
Model II: adjusted for age, sex, hypertension, glucose, waist circumference, Apolipoprotein E-ε4 status, total cholesterol, and high-density lipoprotein cholesterol.

Figure 1 shows the cumulative incidence curves of dementia per tertile of IGF-I receptor stimulating activity. Proportional hazard analyses of dementia incidence revealed that those in the lowest tertile of IGF-I receptor stimulating activity at baseline had the lowest risk of dementia (HR moderate vs. low: 1.45 (95% CI 1.00-2.12); HR high vs. low 1.40 (95% CI 0.96-2.04)), while there was no difference in risk of dementia between the medium and highest tertiles (**Figure 1**).

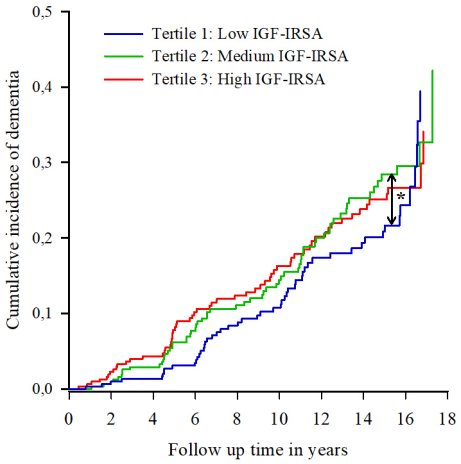


Figure 1. Cumulative incidence curves of dementia per tertile of IGF-I receptor stimulating activity (IGF-IRSA). Tertile 1 represents the lowest levels of IGF-I receptor stimulating activity, tertile 2 medium levels and tertile 3 the highest levels. * indicates significant at $p < 0.05$.

IGF-I receptor activity and ApoE-ε4

At baseline there was a statistically significant difference in IGF-I receptor stimulating activity between *ApoE-ε4* genotype groups without dementia (non-carrier, heterozygote and homozygote), after adjustment for age and sex ($p = 0.04$; see **Table 3**). The levels of IGF-I receptor stimulating activity were significantly lower in homozygotes for *ApoE-ε4* than in people with no copies of the *ApoE-ε4* allele ($p = 0.04$). There were no statistically significant differences in level of IGF-I receptor stimulating activity between non-carriers and *ApoE-ε4* heterozygotes or *ApoE-ε4* heterozygotes and homozygotes.

When testing for effect modification, significant evidence for a multiplicative interaction between IGF-I receptor stimulating activity and *ApoE-ε4* carrier status was observed ($\chi 2$ (2) = 10.85, $p = 0.004$). In those without the *ApoE-ε4* variant, the level of IGF-I receptor stimulating activity was not associated with the risk of dementia (medium vs. low: HR 0.97 (95% CI 0.59-1.60); high vs. low: HR 1.09 (95% CI 0.67-1.77); $\chi 2$ (2) = 0.24, $p = 0.89$). For those with one or more copies of the *ApoE-ε4* allele, level of IGF-I receptor stimulating activity was positively associated with dementia risk. Dementia risk was significantly increased in people with one

or more copies of the *ApoE-ε4* allele and IGF-1 receptor stimulating activity in the median and top tertiles compared to those with IGF-I receptor stimulating activity in the bottom tertile (medium vs. low: HR 3.80 (95% CI 1.90-7.60); high vs. low: HR 2.71 (95% CI 1.37-5.38)). Similar results were found for the incidence of AD (**Table 4**). **Figure 2a** and **2b** show the cumulative incidence of dementia per tertile group of IGF-I receptor stimulating activity, stratified by *ApoE-ε4* genotype.

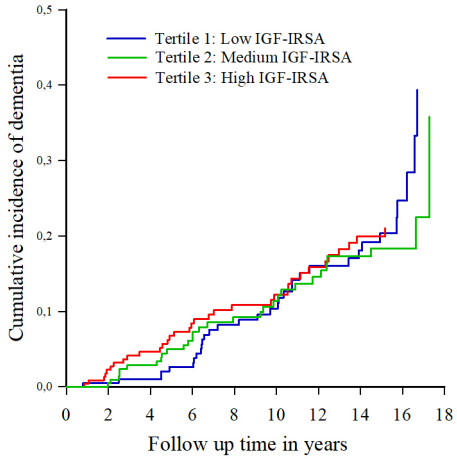


Figure 2(A). Cumulative incidence curves of dementia per tertile of IGF-I receptor stimulating activity (IGF-IRSA), for persons without *ApoE-ε4*. Tertile 1 represents the lowest levels of IGF-I receptor stimulating activity, tertile 2 medium levels and tertile 3 the highest levels.

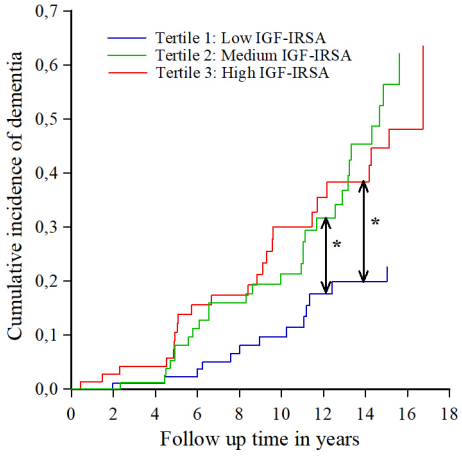


Figure 2(B). Cumulative incidence curves of dementia per tertile of IGF-I receptor stimulating activity (IGF-IRSA), for hetero- and homozygotes of the *ApoE-ε4* allele. Tertile 1 represents the lowest levels of IGF-I receptor stimulating activity, tertile 2 medium levels and tertile 3 the highest levels. * indicates significant at $p < 0.05$.

Table 3. IGF-I receptor stimulating activity stratified by apolipoprotein E (*ApoE*) group.

	Non-carriers N = 680	ApoE-ε4 heterozygotes N = 240	ApoE-ε4 homozygotes N = 17	P for trend N = 937
IGF-IRSA (pmol/L)	181.92 (58.58)	174.31 (48.24)	151.82 (34.98)	F(2) 3.20, p = 0.04

Data are presented as means (standard deviations). N = number of persons, IGF-IRSA = insulin-like growth factor I stimulating receptor activity. Adjusted for age and sex.

Table 4. IGF-I receptor stimulating activity tertile groups and risk of incident dementia.

	Dementia		Alzheimer's disease	
IGF-IRSA groups	ApoE-ε4 + n/N 65/255 HR (95% CI)	ApoE-ε4 - n/N 97/669 HR (95% CI)	ApoE-ε4 + n/N 51/255 HR (95% CI)	ApoE-ε4 - n/N 78/669 HR (95% CI)
Medium	3.80 (1.90-7.60)	0.97 (0.59-1.60)	3.44 (1.61-7.34)	1.00 (0.57-1.74)
High	2.71 (1.37-5.38)	1.09 (0.67-1.77)	2.38 (1.12-5.08)	1.01 (0.86-1.75)
Medium vs. High	1.40 (0.79-2.48)	0.89 (0.54-1.46)	1.44 (0.75-2.76)	0.99 (0.57-1.73)

Data are presented as hazard ratios (HR) and 95% confidence intervals (CI). N = number of people at risk for incident dementia, n = number of cases of incident dementia, IGF-IRSA = insulin-like growth factor I receptor stimulating activity divided in tertile groups. The lowest tertile group is used as reference. Adjusted for age, sex, hypertension, glucose, waist circumference, total cholesterol, and high-density lipoprotein cholesterol.

Discussion

In the extended follow-up period of 16 years, our study did not find evidence for a long-term dose response association between circulating IGF-I receptor stimulating activity at baseline and the future risk of dementia. Interestingly, we found evidence of an interaction between *ApoE-ε4* and IGF-I receptor stimulating activity. In those hetero- and homozygous for the *ApoE-ε4* allele, dementia risk was increased in persons with medium and high levels of IGF-I receptor stimulating activity at baseline, compared to those with low IGF-I receptor stimulating activity at baseline. However, no relation between IGF-I receptor stimulating activity and dementia risk was observed in non-carriers of the *ApoE-ε4* allele. This suggests that, in *ApoE-ε4* carriers, there is a certain threshold above which IGF-I receptor stimulating activity becomes associated with dementia at long-term follow up. In addition, in individuals without dementia, IGF-I receptor stimulating activity was lower in homozygote carriers of *ApoE-ε4* than in people with other *ApoE* genotypes.

To our knowledge the Rotterdam Study is still the only study that has investigated the role of circulating IGF-I receptor stimulating activity in relation to dementia. As circulating IGF-I receptor stimulating activity is only modestly correlated to total serum IGF-I and IGF-I/IGFBP-3 ratio[43] comparison to other studies on serum total IGF-I, measured by immunoassays, and dementia risk, described by Ostrowski et al. is difficult[12].

We found a modifying effect of *ApoE-ε4* on circulating IGF-I receptor stimulating activity at baseline and an interaction on the relation between IGF-I receptor stimulating activity and the risk of dementia. The observed interaction between circulating IGF-I receptor stimulating activity and *ApoE* isoforms in our study could reflect opposing influences on shared pathways involved in Alzheimer pathology. Both ApoE and IGF-I are involved in the regulation of AD biomarkers: IGF-I is an important mediator in the clearance and regulation of Aβ in the brain, enhances survival of neurons exposed to Aβ and inhibits tau phosphorylation[9,14,17,20,36,53,54]. The *ApoE-ε4* allele, on the other hand, is associated with decreased $A\beta_{1-42}$ and higher tau and p-tau in the CSF and increased cerebral amyloid deposition across the AD spectrum[55-60]. In mice expressing human *ApoE-ε4*, increased tau phosphorylation has been demonstrated[61]. *ApoE-ε4* and IGF-I also have an opposing influence on NMDA receptor signaling. The NR2B subunit of the NMDA receptor, in particular, is suggested to be of specific importance for spatial learning and long-term potentiation, impaired in AD[62-64]. The *ApoE-ε4* genotype is associated with decreased NR2B NMDA receptor subunit levels and enhances age-related decline in cognitive function by down-regulating signaling in mice[65]. In contrast, IGF-I positively affects the NMDA receptor pathway by increasing the NR2B subunit mRNA transcript of the hippocampal NMDA receptor in rats[62,64]. The observed association between elevated levels of IGF-I receptor stimulating activity and increased risk of dementia in *ApoE-ε4* carriers might thus be a reflection of a compensatory response to neuropathological changes associated with the *ApoE-ε4* genotype and a preclinical loss of sensitivity of the IGF-I receptor.

The strengths of our study are its prospective, population-based design, the long follow-up period, and the use of a direct measure of circulating IGF-I receptor stimulating activity. However, there are also some limitations. First, IGF-I receptor stimulating activity was only measured in peripheral blood sample. Even though circulating IGF-I crosses the blood-brain barrier, we could not assess the extent to which our measurements of circulating IGF-I receptor stimulating activity are related to actual IGF-I receptor stimulating activity in the brain[66]. In addition, IGF-I has important autocrine and paracrine actions at the tissue level. However, IGF-I measured by the KIRA assay may not necessarily reflect IGF-I bioactivity at the local tissue level[42]. Second, no total serum IGF-I levels were measured in our study, therefore we were unable to compare the relationship of IGF-I receptor stimulating activity and total IGF-I with dementia and to show that measuring IGF-I bioactivity with the IGF-I KIRA assay provides other insights about the role of IGF-I in dementia than the measurement of total IGF-I. Third, IGF-I receptor stimulating activity was assessed at the second follow-up visit of the Rotterdam Study, which might have led to survival effects in the study population which was included at baseline.

In conclusion, our current study sheds new light on the association between IGF-I signaling and the neuropathology of dementia, suggesting a threshold effect of IGF-I receptor stimulating activity moderated by *ApoE* genotype, since only for those with one or more copies of the *ApoE-ε4* allele and in the lowest tertile of IGF-I receptor stimulating activity the risk of future dementia is decreased. Our study suggests that the *ApoE-ε4* genotype modifies

the relationship between IGF-I receptor stimulating activity and dementia and elevated IGF-I receptor stimulating activity levels mark a compensatory response to neuropathological changes associated with the *ApoE-ε4* genotype. In line with the hypothesis that low IGF-I activity increases the risk of dementia, we found the *ApoE-ε4* homozygotes, with a lifetime risk of AD of 80%[67], have the lowest IGF-I levels. This provides a genetic benchmark for the hypothesis that low IGF-I receptor stimulating activity is associated with an increased risk of AD. However, our findings require replication in other cohorts, reusing measures of IGF-I receptor stimulating activity rather than total IGF-I serum levels as putative predictors of dementia risk.

References

1 Hankinson, S. E. *et al.* Circulating concentrations of insulin-like growth factor-I and risk of breast cancer. *Lancet* **351**, 1393-1396, doi:10.1016/S0140-6736(97)10384-1 (1998).

2 Kaaks, R. Nutrition, Insulin, IGF-1 Metabolism and Cancer Risk: A Summary of Epidemiological Evidence Novartis Foundation Symposium 262. 247-264 (2008).

3 Renehan, A. G. *et al.* Insulin-like growth factor (IGF)-I, IGF binding protein-3, and cancer risk: systematic review and meta-regression analysis. *Lancet* **363**, 1346-1353, doi:10.1016/S0140-6736(04)16044-3 (2004).

4 Ma, J. *et al.* Prospective study of colorectal cancer risk in men and plasma levels of insulin-like growth factor (IGF)-I and IGF-binding protein-3. *J Natl Cancer Inst* **91**, 620-625, doi:10.1093/jnci/91.7.620 (1999).

5 Wu, Y., Yakar, S., Zhao, L., Hennighausen, L. & LeRoith, D. Circulating insulin-like growth factor-I levels regulate colon cancer growth and metastasis. *Cancer Res* **62**, 1030-1035 (2002).

6 Yu, H. *et al.* Plasma levels of insulin-like growth factor-I and lung cancer risk: a case-control analysis. *J Natl Cancer Inst* **91**, 151-156, doi:10.1093/jnci/91.2.151 (1999).

7 Frolich, L. *et al.* Brain insulin and insulin receptors in aging and sporadic Alzheimer's disease. *J Neural Transm (Vienna)* **105**, 423-438, doi:10.1007/s007020050068 (1998).

8 Frolich, L., Blum-Degen, D., Riederer, P. & Hoyer, S. A disturbance in the neuronal insulin receptor signal transduction in sporadic Alzheimer's disease. *Ann N Y Acad Sci* **893**, 290-293, doi:10.1111/j.1749-6632.1999.tb07839.x (1999).

9 Moloney, A. M. *et al.* Defects in IGF-1 receptor, insulin receptor and IRS-1/2 in Alzheimer's disease indicate possible resistance to IGF-1 and insulin signalling. *Neurobiol Aging* **31**, 224-243, doi:10.1016/j.neurobiolaging.2008.04.002 (2010).

10 Steen, E. *et al.* Impaired insulin and insulin-like growth factor expression and signaling mechanisms in Alzheimer's disease--is this type 3 diabetes? *J Alzheimers Dis* **7**, 63-80, doi:10.3233/jad-2005-7107 (2005).

11 Rivera, E. J. *et al.* Insulin and insulin-like growth factor expression and function deteriorate with progression of Alzheimer's disease: link to brain reductions in acetylcholine. *J Alzheimers Dis* **8**, 247-268, doi:10.3233/jad-2005-8304 (2005).

12 Ostrowski, P. P., Barszczyk, A., Forstenpointner, J., Zheng, W. & Feng, Z. P. Meta-Analysis of Serum Insulin-Like Growth Factor 1 in Alzheimer's Disease. *PLoS One* **11**, e0155733, doi:10.1371/journal.pone.0155733 (2016).

13 Ashpole, N. M., Sanders, J. E., Hodges, E. L., Yan, H. & Sonntag, W. E. Growth hormone, insulin-like growth factor-1 and the aging brain. *Exp Gerontol* **68**, 76-81, doi:10.1016/j.exger.2014.10.002 (2015).

14 Carro, E., Trejo, J. L., Gomez-Isla, T., LeRoith, D. & Torres-Aleman, I. Serum insulin-like growth factor I regulates brain amyloid-beta levels. *Nat Med* **8**, 1390-1397, doi:10.1038/nm1202-793 (2002).

15 Gasparini, L. *et al.* Stimulation of beta-amyloid precursor protein trafficking by insulin reduces intraneuronal beta-amyloid and requires mitogen-activated protein kinase signaling. *J Neurosci* **21**, 2561-2570, doi:21/8/2561 [pii] (2001).

16 Gasparini, L., Netzer, W. J., Greengard, P. & Xu, H. Does insulin dysfunction play a role in Alzheimer's disease? *Trends Pharmacol Sci* **23**, 288-293, doi:10.1016/s0165-6147(02)02037-0 (2002).

17 Cheng, C. M. *et al.* Tau is hyperphosphorylated in the insulin-like growth factor-I null brain. *Endocrinology* **146**, 5086-5091, doi:10.1210/en.2005-0063 (2005).

18 Bedse, G., Di Domenico, F., Serviddio, G. & Cassano, T. Aberrant insulin signaling in Alzheimer's disease: current knowledge. *Front Neurosci* **9**, 204, doi:10.3389/fnins.2015.00204 (2015).

19 Aguado-Llera, D., Arilla-Ferreiro, E., Campos-Barros, A., Puebla-Jimenez, L. & Barrios, V. Protective effects of insulin-like growth factor-I on the somatostatinergic system in the temporal cortex of beta-amyloid-treated rats. *J Neurochem* **92**, 607-615, doi:10.1111/j.1471-4159.2004.02889.x (2005).

20 Carro, E. *et al.* Therapeutic actions of insulin-like growth factor I on APP/PS2 mice with severe brain amyloidosis. *Neurobiol Aging* **27**, 1250-1257, doi:10.1016/j.neurobiolaging.2005.06.015 (2006).

21 Lanz, T. A. *et al.* Peripheral elevation of IGF-1 fails to alter Abeta clearance in multiple in vivo models. *Biochem Pharmacol* **75**, 1093-1103, doi:10.1016/j.bcp.2007.11.001 (2008).

22 Parrella, E. *et al.* Protein restriction cycles reduce IGF-1 and phosphorylated Tau, and improve behavioral performance in an Alzheimer's disease mouse model. *Aging Cell* **12**, 257-268, doi:10.1111/acel.12049 (2013).

23 Sevigny, J. J. *et al.* Growth hormone secretagogue MK-677: no clinical effect on AD progression in a randomized trial. *Neurology* **71**, 1702-1708, doi:10.1212/01.wnl.0000335163.88054.e7 (2008).

24 Trueba-Saiz, A. *et al.* Loss of serum IGF-I input to the brain as an early biomarker of disease onset in Alzheimer mice. *Transl Psychiatry* **3**, e330, doi:10.1038/tp.2013.102 (2013).
25 Suh, Y. *et al.* Functionally significant insulin-like growth factor I receptor mutations in centenarians. *Proc Natl Acad Sci U S A* **105**, 3438-3442, doi:10.1073/pnas.0705467105 (2008).
26 Milman, S. *et al.* Low insulin-like growth factor-1 level predicts survival in humans with exceptional longevity. *Aging Cell* **13**, 769-771, doi:10.1111/acel.12213 (2014).
27 Tatar, M. *et al.* A mutant Drosophila insulin receptor homolog that extends life-span and impairs neuroendocrine function. *Science* **292**, 107-110, doi:10.1126/science.1057987 (2001).
28 Tazearslan, C., Huang, J., Barzilai, N. & Suh, Y. Impaired IGF1R signaling in cells expressing longevity-associated human IGF1R alleles. *Aging Cell* **10**, 551-554, doi:10.1111/j.1474-9726.2011.00697.x (2011).
29 Cohen, E. *et al.* Reduced IGF-1 signaling delays age-associated proteotoxicity in mice. *Cell* **139**, 1157-1169, doi:10.1016/j.cell.2009.11.014 (2009).
30 Freude, S., Schilbach, K. & Schubert, M. The role of IGF-1 receptor and insulin receptor signaling for the pathogenesis of Alzheimer's disease: from model organisms to human disease. *Curr Alzheimer Res* **6**, 213-223, doi:10.2174/156720509788486527 (2009).
31 George, C. *et al.* The Alzheimer's disease transcriptome mimics the neuroprotective signature of IGF-1 receptor-deficient neurons. *Brain* **140**, 2012-2027, doi:10.1093/brain/awx132 (2017).
32 Gontier, G., George, C., Chaker, Z., Holzenberger, M. & Aid, S. Blocking IGF Signaling in Adult Neurons Alleviates Alzheimer's Disease Pathology through Amyloid-beta Clearance. *J Neurosci* **35**, 11500-11513, doi:10.1523/JNEUROSCI.0343-15.2015 (2015).
33 Westwood, A. J. *et al.* Insulin-like growth factor-1 and risk of Alzheimer dementia and brain atrophy. *Neurology* **82**, 1613-1619, doi:10.1212/WNL.0000000000000382 (2014).
34 Green, C. J. *et al.* The role of IGF-I, IGF-II, and IGFBP-3 in male cognitive aging and dementia risk: the Caerphilly Prospective Study. *J Alzheimers Dis* **41**, 867-875, doi:10.3233/JAD-132183 (2014).
35 Lane, E. M., Hohman, T. J., Jefferson, A. L. & Alzheimer's Disease Neuroimaging, I. Insulin-like growth factor binding protein-2 interactions with Alzheimer's disease biomarkers. *Brain Imaging Behav* **11**, 1779-1786, doi:10.1007/s11682-016-9636-0 (2017).
36 Talbot, K. *et al.* Demonstrated brain insulin resistance in Alzheimer's disease patients is associated with IGF-1 resistance, IRS-1 dysregulation, and cognitive decline. *J Clin Invest* **122**, 1316-1338, doi:10.1172/JCI59903 (2012).
37 van Exel, E. *et al.* Insulin-like growth factor-1 and risk of late-onset Alzheimer's disease: findings from a family study. *Neurobiol Aging* **35**, 725 e727-710, doi:10.1016/j.neurobiolaging.2013.08.014 (2014).
38 Vargas, T. *et al.* IGF-I gene variability is associated with an increased risk for AD. *Neurobiol Aging* **32**, 556 e553-511, doi:10.1016/j.neurobiolaging.2010.10.017 (2011).
39 Deelen, J. *et al.* Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited. *Aging Cell* **10**, 686-698, doi:10.1111/j.1474-9726.2011.00705.x (2011).
40 Williams, D. M., Karlsson, I. K., Pedersen, N. L. & Hagg, S. Circulating insulin-like growth factors and Alzheimer disease: A mendelian randomization study. *Neurology* **90**, e291-e297, doi:10.1212/WNL.0000000000004854 (2018).
41 Brugts, M. P. *et al.* Igf-I bioactivity in an elderly population: relation to insulin sensitivity, insulin levels, and the metabolic syndrome. *Diabetes* **59**, 505-508, doi:10.2337/db09-0583 (2010).
42 Chen, J. W. *et al.* A highly sensitive and specific assay for determination of IGF-I bioactivity in human serum. *Am J Physiol Endocrinol Metab* **284**, E1149-1155, doi:10.1152/ajpendo.00410.2002 (2003).
43 Brugts, M. P. *et al.* Normal values of circulating insulin-like growth factor-I bioactivity in the healthy population: comparison with five widely used IGF-I immunoassays. *J Clin Endocrinol Metab* **93**, 2539-2545, doi:10.1210/jc.2007-2454 (2008).
44 de Bruijn, R. F. *et al.* Insulin-like growth factor-I receptor stimulating activity is associated with dementia. *J Alzheimers Dis* **42**, 137-142, doi:10.3233/JAD-140186 (2014).
45 Ikram, M. A. *et al.* The Rotterdam Study: 2018 update on objectives, design and main results. *Eur J Epidemiol* **32**, 807-850, doi:10.1007/s10654-017-0321-4 (2017).
46 Ott, A., Breteler, M. M., van Harskamp, F., Stijnen, T. & Hofman, A. Incidence and risk of dementia. The Rotterdam Study. *Am J Epidemiol* **147**, 574-580, doi:10.1093/oxfordjournals.aje.a009489 (1998).
47 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**, 189-198, doi:10.1016/0022-3956(75)90026-6 (1975).
48 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**, 439-449, doi:10.1017/s0033291700015889 (1976).
49 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, doi:10.1192/bjp.149.6.698 (1986).

50 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, doi:10.1186/s12916-015-0377-5 (2015).
51 Association, A. P. *Diagnostic and statistical manual of mental disorders : DSM-III-R*. (APA Press, 1987).
52 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. *Alzheimers Dement* **7**, 263-269, doi:10.1016/j.jalz.2011.03.005 (2011).
53 Dore, S., Kar, S. & Quirion, R. Insulin-like growth factor I protects and rescues hippocampal neurons against beta-amyloid- and human amylin-induced toxicity. *Proc Natl Acad Sci U S A* **94**, 4772-4777, doi:10.1073/pnas.94.9.4772 (1997).
54 Engel, T., Hernandez, F., Avila, J. & Lucas, J. J. Full reversal of Alzheimer's disease-like phenotype in a mouse model with conditional overexpression of glycogen synthase kinase-3. *J Neurosci* **26**, 5083-5090, doi:10.1523/JNEUROSCI.0604-06.2006 (2006).
55 Kanekiyo, T., Xu, H. & Bu, G. ApoE and Abeta in Alzheimer's disease: accidental encounters or partners? *Neuron* **81**, 740-754, doi:10.1016/j.neuron.2014.01.045 (2014).
56 Leoni, V. The effect of apolipoprotein E (ApoE) genotype on biomarkers of amyloidogenesis, tau pathology and neurodegeneration in Alzheimer's disease. *Clin Chem Lab Med* **49**, 375-383, doi:10.1515/CCLM.2011.088 (2011).
57 Morris, J. C. *et al.* APOE predicts amyloid-beta but not tau Alzheimer pathology in cognitively normal aging. *Ann Neurol* **67**, 122-131, doi:10.1002/ana.21843 (2010).
58 Risacher, S. L. & Saykin, A. J. Neuroimaging and other biomarkers for Alzheimer's disease: the changing landscape of early detection. *Annu Rev Clin Psychol* **9**, 621-648, doi:10.1146/annurev-clinpsy-050212-185535 (2013).
59 Tapiola, T. *et al.* Relationship between apoE genotype and CSF beta-amyloid (1-42) and tau in patients with probable and definite Alzheimer's disease. *Neurobiol Aging* **21**, 735-740, doi:10.1016/s0197-4580(00)00164-0 (2000).
60 Wildsmith, K. R., Holley, M., Savage, J. C., Skerrett, R. & Landreth, G. E. Evidence for impaired amyloid beta clearance in Alzheimer's disease. *Alzheimers Res Ther* **5**, 33, doi:10.1186/alzrt187 (2013).
61 Tesseur, I. *et al.* Expression of human apolipoprotein E4 in neurons causes hyperphosphorylation of protein tau in the brains of transgenic mice. *Am J Pathol* **156**, 951-964, doi:10.1016/S0002-9440(10)64963-2 (2000).
62 Le Greves, M., Le Greves, P. & Nyberg, F. Age-related effects of IGF-1 on the NMDA-, GH- and IGF-1-receptor mRNA transcripts in the rat hippocampus. *Brain Res Bull* **65**, 369-374, doi:10.1016/j.brainresbull.2005.01.012 (2005).
63 Reiman, E. M. *et al.* Fibrillar amyloid-beta burden in cognitively normal people at 3 levels of genetic risk for Alzheimer's disease. *Proc Natl Acad Sci U S A* **106**, 6820-6825, doi:10.1073/pnas.0900345106 (2009).
64 Sonntag, W. E. *et al.* Age and insulin-like growth factor-1 modulate N-methyl-D-aspartate receptor subtype expression in rats. *Brain Res Bull* **51**, 331-338, doi:10.1016/s0361-9230(99)00259-2 (2000).
65 Liu, D. S. *et al.* APOE4 enhances age-dependent decline in cognitive function by down-regulating an NMDA receptor pathway in EFAD-Tg mice. *Mol Neurodegener* **10**, 7, doi:10.1186/s13024-015-0002-2 (2015).
66 Reinhardt, R. R. & Bondy, C. A. Insulin-like growth factors cross the blood-brain barrier. *Endocrinology* **135**, 1753-1761, doi:10.1210/endo.135.5.7525251 (1994).
67 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, doi:10.1016/S1474-4422(18)30053-X (2018).

CHAPTER 3

Late onset depression and sleep in the general population

CHAPTER 3.1

Circulating metabolites related to diet, medication and the gut microbiome are associated with depression

Ashley van der Spek*, Isobel D. Stewart*, Brigitte Kühnel*, Maik Pietzner*, Tahani Alshehri*, Friederike Gauß, Pirro G. Hysi, Siamak Mahmoudian Dehkordi, Almut Heinken, Annemarie I. Luik, Karl-Heinz Ladwig, Gabi Kastenmüller, Cristina Menni, Johannes Hertel, M. Arfan Ikram, Renée de Mutsert, Karsten Suhre, Christian Gieger, Konstantin Strauch, Henry Völzke, Thomas Meitinger, Massimo Mangino, Antonia Flaquer, Melanie Waldenberger, Annette Peters, Ines Thiele, Rima Kaddurah-Daouk, Boadie W Dunlop, Frits R. Rosendaal, Nicholas J. Wareham, Tim D. Spector, Sonja Kunze**, Hans Jörgen Grabe**, Dennis O. Mook-Kanamori**, Claudia Langenberg**, Cornelia M. van Duijn**, Najaf Amin**

* These authors contributed equally to this manuscript (shared first)

** These authors contributed equally to this manuscript (shared last)

This chapter is submitted.

The supplemental information for this paper is available at https://drive.google.com/drive/folders/1DXPeusi-c30DdDtnx1Y2cA6v8i4c__8l?usp=sharing

Abstract

Background

Depression is a complex psychiatric disorder with an elusive pathogenesis that spans beyond the brain, involving metabolic changes in the blood circulation. A comprehensive analysis of circulating metabolites could provide insights into the metabolic causes and consequences of depression.

Methods

The metabolome-wide association study was performed in 13,596 participants (age range 20-96 years) of European descent from five independent population-based cohort studies. Circulating metabolites were assayed using a mass spectrometry-based semi-untargeted metabolomics platform (Metabolon).

Outcomes

Major depression/depressive symptoms were assessed using self-reports/questionnaire.

Interpretation

We identified 53 metabolites including 41 novel ones not earlier associated with depression. To disentangle the effect of medication use and lifestyle factors, we integrated data on medication use, lifestyle factors, genomics and gut microbiome. For 41 of the 53 metabolites associated to depression, the association was explained by antidepressant use and another four by medication use for cardiovascular health outcomes and lifestyle factors. However, increased levels of retinol (vitamin A), 1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1) (lecithin) and mannitol/sorbitol, and lower levels of hippurate, 4-hydroxycoumarin, 2-aminooctanoate, 10-undecenoate (11:1n1), and 1-linoleoyl-GPA (18:2) were significantly associated with depression independent of antidepressant use, cardiovascular medication and lifestyle factors. Retinol, hippurate, mannitol/sorbitol, 4-hydroxycoumarin and 1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1) (lecithin) are either directly food derived or are products of host and gut microbial metabolism of food-derived products. Integrating the data with genetic determinants suggests that hippurate may be causally related with depression. Our findings highlight putative actionable targets for the prevention of depression that may be modulated by simple dietary interventions.

Introduction

Depression is the most common psychiatric disorder with an average lifetime prevalence of 11-15%[1]. A sharp increase in the prevalence of depression worldwide (33.7%; confidence interval 27.5–40.6) has been observed during the recent COVID-19 pandemic[2] and is predicted to increase as the effects of the pandemic unfold further[3]. The molecular mechanisms underlying depression remain elusive. The heritability is estimated to be 40%[4] and 87 genetic variants have been identified to be associated with depression[5]. There is also a range of environmental risk factors for morbidity including low education, diet and smoking[6]. Depression often co-occurs with other neuro-psychiatric pathologies[7,8] and dementia[9], but also clusters strongly with systemic disorders such as cardiometabolic disease, diabetes and arthritis[10,11-13]. Treatment success for depression is poor and mortality is high[12,14,15].

While depression is primarily considered as a disorder of the brain[16], it is associated with metabolic changes in the blood circulation that may be explained by weight loss/gain, changes in diet and altered gut metabolism[17]. There is increasing interest in metabolomic studies of depression that capture the downstream effects of genes, lifestyle factors, pathology and medication[18-20]. A novel hypothesis why circulating metabolites may be involved in depression is that these metabolites are involved in the gut-brain axis, i.e., the bi-directional signalling between the gut, its microbiome and the brain[21,22].

Metabolomic studies on depression have been small and findings have not always been consistent[23]. Yet, consensus is building that depression is associated with increased levels of glutamate, lactate, alanine, isobutyrate and sorbitol and with decreased levels of kynurenine, gamma aminobutyric acid (GABA), phenylalanine, tyrosine, creatinine, hypoxanthine, leucine, tryptophan, N-methylnicotinamide, β-aminoisobutyric acid, hippurate, amino-ethanol and malonate[23]. Our study of 5,283 patients with depression and 10,145 controls from nine Dutch cohorts[24] using a proton Nuclear Magnetic Resonance (NMR) metabolomics platform (Nightingale Health Ltd., Helsinki, Finland) identified 21 cardiometabolic metabolites that are significantly related to depression. These include an unfavorable spectrum of metabolites associated to cardiometabolic morbidity and mortality[25-27] including apolipoprotein A1 and B, very-low-density and high-density lipoprotein cholesterol, di- and triglycerides, (mono-) unsaturated fatty acids, fatty acid chain length, acetate, glycoprotein acetyls, tyrosine, and isoleucine[24].

A problem hampering the translation of findings of metabolomics studies into preventive and therapeutic interventions is that metabolites in the blood circulation are strongly influenced by medication and comorbidity[22]. Although their effects are well recognized, the potential bias is not controlled for in most studies conducted to date. Another problem to be tackled is to disentangle metabolic changes that occur as a cause from those that occur as a consequence of depression progression. To control for confounding, we conducted a

comprehensive analysis of the relation between the blood metabolome and depression in five large scale epidemiologic cohorts including a total of 13,596 participants. This setting allows us to control for confounding effects of medication and co-morbidity. The metabolome in the circulation was characterized by mass spectrometry (MS) using Metabolon. To identify the origin of metabolites (gut and/or human) we integrate our findings with those of the Virtual Metabolic Human (VMH) and Assembly of Gut Organisms through Reconstruction and Analysis (AGORA2) databases. To separate potential causal effects from the consequences of the disease, we integrate genomic and metabolomic data using the NIHR BioResource (NBR). We then examine the impact of anti-depressive therapy on the metabolites in the Predictors of Remission in Depression to Individual and Combined Treatments (PReDICT) study. Finally, we study the association of the diet-based sources of these metabolites with depression and brain pathology in the UK Biobank.

Methods

Study populations

The association analysis of metabolite levels with depression was performed in 13,596 participants separately recruited in five different cohort studies. The following cohort studies were included: the Rotterdam Study (RS), the Study of Health in Pomerania (SHIP-TREND), the Cooperative Health Research in the Region of Augsburg (KORA) study, the European Prospective Investigation into Cancer (EPIC)-Norfolk Study, and the Netherlands Epidemiology of Obesity (NEO) study. Detailed information on these cohorts is provided in the **Supplementary Materials**. All participants provided written informed consent, studies were approved by their local ethics committees and conformed to the principles of the declaration of Helsinki.

Association of depression with the dietary sources of the depression-associated metabolites was performed in the UK Biobank study. UK Biobank is a prospective cohort study including ~500,000 participants aged 40-69 years at baseline recruited between 2006 and 2010. The aim of the study is to investigate the effects of genetic and environmental factors on the risk of common multifactorial diseases. Participants have provided detailed information on lifestyle, medical history and nutritional habits; basic variables such weight, height, blood pressure etc. were measured; and blood and urine samples were taken. Detailed information about the cohort is provided in the **Supplementary Materials**.

To ascertain the effects of various depression treatments including cognitive behavioural therapy (CBT) and antidepressants SSRI (escitalopram) and SNRI (duloxetine) on the depression-associated metabolites, we performed a lookup in the PReDICT study. The design of PReDICT study has been published previously[28]. Details on the study and the metabolomics assessments are provided in the **Supplementary Materials**.

To select instruments/proxies for metabolites for Mendelian Randomization we used the results of the genome-wide association study (GWAS) performed using the NIHR BioResource (NBR). NIHR BioResource (NBR) – Rare Disease Study is a multi-center whole-exome and whole-genome sequencing study including up to 13,600 patients (http://bioresource.nihr.ac.uk/rare-diseases/rare-diseases/). The NBR–Rare Diseases study was approved by the East of England Cambridge South national research ethics committee (REC) under reference number: 13/EE/0325. The inclusion and exclusion criteria, as well as other steps of quality control, adjustment and transformations followed exactly the same analytical steps as described before[29].

Depression assessment

In the RS, depressive symptoms were assessed with the 20-item version of the Centre for Epidemiologic Studies Depression (CES-D) scale, a self-report measure of depressive symptoms experienced during the prior week[30]. The total score ranges from 0 – 60, where a higher score indicates more depressive symptoms. In the SHIP-trend and KORA cohorts, depressive symptoms were assessed with the Patient Health Questionnaire 9 (PHQ-9)[31], where each of the nine DSM-IV criteria for depression are scored from 0 – 3. The total score ranges from 0 – 27 where higher score indicates a greater depression severity. In KORA a brief interview version of PHQ-9 called Patient Health Questionnaire Depression (PHQ-D) module was used to measure depression[31,32]. In the EPIC-Norfolk study depression was assessed using the following question: "Has the doctor ever told you that you have any of the following: depression requiring treatment?" with answers "yes" or "no". In the NEO cohort, depressive symptoms were assessed using the Inventory Depressive Symptomatology Self Report questionnaire (IDS-SR30)[33], which assesses specific depressive symptoms (via a 4-level response system) during the last week and their severity. The total score ranges from 0 to 84, with higher scores indicating higher severity. Thus, in all cohorts, except EPIC-Norfolk, depression in participants was measured on a quantitative scale and used as such in the analysis.

In the UKB study, we used the derived lifetime probable major depressive disorder measure as described in Smith et al. 2013[34]. We further defined current depressive symptoms by summing the responses to four questions related to mood in the past two weeks. These include, 1) Over the past two weeks, how often have you felt down, depressed or hopeless?, 2) Over the past two weeks, how often have you had little interest or pleasure in doing things?, 3) Over the past two weeks, how often have you felt tense, fidgety or restless? and 4) Over the past two weeks, how often have you felt tired or had little energy? Answers could be given on a four-point scale ranging from 0-3 (0 = not at all, 1 = several days, 2 = more than half of the days and 3 = nearly every day). The total score ranged from 0-12 where higher scores indicate more severe depression.

In the PReDICT study, participants were treatment-naive adults defined as having never previously received a minimally adequate course of treatment with an antidepressant

medication or evidence-based psychotherapy for a mood disorder, aged 18 to 65 years with moderate-to-severe, non-psychotic MDD depression as assessed by the Structured Clinical Interview for DSM-IV[35] and a psychiatrist's evaluation, and if they scored ≥18 on the HRSD17. Eligible patients were randomized equally to one of three 12-week treatment arms: (1) cognitive behavior therapy (CBT, 16 sessions); (2) duloxetine (30–60 mg/d); or (3) escitalopram (10–20 mg/d).

Metabolomics measurements

In all studies, the metabolome was quantified using the untargeted Metabolon platform (Metabolon Inc., Durham, USA). Different versions of the platform have been used and details on the platforms are included in the **Supplementary Materials**. In all studies, metabolites with ≥ 40% missing values were removed and for the remaining metabolites missing metabolite values were replaced with half of the detection limit for that particular metabolite[36]. Subsequently, a natural logarithm transformation was applied to all metabolites and metabolites were scaled to standard deviation units.

In the PReDICT study, metabolites were quantified using targeted metabolomics platforms including ultra-performance liquid chromatography triple quadrupole mass spectrometry (UPLC-TQMS) (Waters XEVO TQ-S, Milford, USA) and gas chromatography time-of-flight mass spectrometry (GC-TOFMS) (Leco Corporation, St Joseph, USA). Metabolites with >20% missing values were excluded. Then, metabolites were log-transformed, imputed and scaled to mean zero and variance 1. Details are provided in the **Supplementary Materials**.

Non-targeted metabolite detection and quantification was conducted by the metabolomics provider Metabolon, Inc. (Durham, USA) on fasting plasma samples of 10,654 participants from the UK Bioresource. The metabolomic dataset measured by Metabolon included 1069 compounds of known structural identity belonging to the following broad categories: amino-acids, peptides, carbohydrates, energy intermediates, lipids, nucleotides, cofactors and vitamins, and xenobiotics. Metabolite data were day-median normalized, and inverse normalized, as the metabolite concentrations were not normally distributed. Metabolic traits with more than 20% missing values were excluded leaving 722 metabolites of known chemical identity for analysis.

Genotyping

For the GWAS of metabolites, genotyping in the UK bioresource was carried out with a high-density array data (Affymetrix UK Biobank Axiom® Array). Genotypes were subsequently imputed using information from the Human Reference Consortium imputation panel (version r1.1, 2016)[37]. Only individuals of full European ancestry (N = 8,809) were included in the analyses in the discovery cohort.

Statistical analyses

Metabolites association analysis

All cohorts used linear regression analysis to test the association between the metabolite (dependent variable) and depression. Three different models were tested, where the first model (model 1) was adjusted for age and sex only, the second model (model 2) was additionally adjusted for antidepressant medication usage, and the third model was an extension of the second model (model 3) with additional adjustment for lipid-lowering medication (yes/no), antihypertensive medication (yes/no), antidiabetic medication (yes/no), body mass index (BMI, kg/m^2), and current smoking (yes/no). The summary statistics from all cohorts were combined in a sample size-weighted meta-analysis using METAL software[38]. Sample size weighted meta-analysis was used since the depression measurement scales were different among cohorts. Only metabolites that were present in two or more studies were included. To investigate the robustness of our findings, a sensitivity analysis was performed by including only cohorts that assessed metabolites with the most recent version of the Metabolon platform (HD4).

Association analysis of major depressive disorder with dietary sources of the metabolites in the UK Biobank

We used logistic regression analysis to test the association between major depressive disorder and dietary sources of metabolites (vitamin A supplements, retinol intake estimated from food, fresh fruits intake and vitamin K antagonists). Age, sex and principal components were used as covariates in the analysis. For the association of current depressive symptoms, we used linear regression analysis. We further tested the association of volume of white matter hyperintensities (WMH) with vitamin supplements to ascertain the impact of these supplements on brain pathology. Linear regression analysis was used with the volume of WMH as the dependent variable, vitamin supplements as the independent variable, and age, sex, BMI, head size and principal components as covariates. All analyses were performed in R.

Metabolite GWAS for Mendelian Randomization (MR) analysis

To test for association between metabolite levels and genotypes, we built linear regression models where the outcome was defined as the transformed level of each metabolite, predicted by the allele dosage at each polymorphic (MAF > 0.01) genotyped or imputed genetic variant. In addition, analyses were adjusted for age, sex and BMI. All analyses were conducted using the PLINK software (https://www.cog-genomics.org/plink/2.0/).

Mendelian Randomization (MR) analysis

To understand the relationship between the identified metabolites and major depression we performed bidirectional two-sample MR analysis. For major depression we used the independent genome-wide significant single nucleotide polymorphisms (SNPs) reported by Howard et al. 2019[39] as instrumental variables (IVs). Summary statistics for these IVs were extracted from Howard et al. The summary statistics for the metabolites were extracted

from the GWAS performed in UK Bioresource. Of the identified metabolites in this study (model 3), GWAS results were available for six metabolites including 2-aminooctanoate, 10-undecenoate (11:1n1), 1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1), hippurate, mannitol/sorbitol and retinol (**Supplementary Table 1**). The IVs for these six metabolites and their summary statistics were extracted from the same GWAS. Because of scarcity in GWAS-grade significance for SNPs associated with these metabolites, we used independent SNPs that showed the strongest association with a *p*-value < 10^{-06} as instruments (**Supplementary Table 2**). The summary statistics for depression for these IVs were extracted from the publicly available dataset (2019 PGC UKB Depression Genome-wide; https://www.med.unc.edu/pgc/download-results/mdd/). For the analysis we used the 'mr_allmethods' option of the R (https://cran.r-project.org/) library "MendelianRandomization"[40] that reports the results from the median method (simple, weighted and penalized), Inverse variance weighted and Egger methods (penalized, robust and penalized & robust).

Effect of antidepressant therapy on metabolites in PReDICT study
To examine the strength and significance of metabolite concentration changes within each of the three treatment arms, i.e., (1) CBT (16 sessions); (2) duloxetine (30–60 mg/d); or (3) escitalopram (10–20 mg/d), linear mixed effect models (with random intercept) with metabolite levels (in log scale) as the dependent variable, were fitted while correcting for age, sex, BMI, and baseline HRSD17. Then, the R package "emmeans" was used to compute the least squared means of the contrasts of interest (week 12 vs. baseline) and their corresponding *p*-values.

To detect whether metabolites levels were associated with clinical outcomes, linear regression analyses were conducted corrected for age, sex and treatment arm. Dependent variables (Baseline HRSD17, Week 12 HRSD17, and 12 weeks change in HRSD17) were regressed on either of following independent variables: 1) baseline metabolite, 2) week 12 metabolite, 3) 2 weeks change in metabolites and 4) 12 weeks change in metabolites.

Linking metabolites to human and/or gut metabolism
To assess whether the identified metabolites are products of human metabolism, gut microbial metabolism, or both, we matched the metabolites associated with depression to the namespace of the Virtual Metabolic Human (VMH) database[41] (www.vmh.life). The VMH database is a resource combining human and gut microbiome metabolism, nutrition, and disease. It enables browsing and querying the content of genome-scale reconstructions of human and gut microbial metabolism. A genome-scale reconstruction is a metabolic network containing the reactions, metabolites, and genes present in a given organism that was manually curated based on genomic, biochemical, and physiological data. It serves as a knowledge base for the target organism and can be converted into a mathematical model that can predict biological properties.

To identify specific gut microbial taxa potentially implicated in depression, we queried a resource of 7,206 genome-scale reconstructions of human gut microbes, AGORA2[42]. AGORA2 is an expansion of the previously published resource of 818 genome-scale reconstructions, AGORA[43], in simulations. The capacity of all 7,206 strains to consume and/or secrete 16 microbial metabolites implicated in depression (see above) was computed as follows. Each AGORA model was allowed to take up each compound it could potentially transport. The range of flux through all exchange reactions corresponding to the 16 microbial metabolites was then computed for each model with the flux variability analysis (FVA)[44] implementation in the COBRA Toolbox[45]. By convention, negative flux through the exchange reactions corresponds to the capability to consume the metabolite, while positive flux corresponds to the capability to secrete the metabolite. The uptake and secretion capabilities for each individual metabolite were consequently extracted from all 7,206 models. The simulations were performed in MATLAB (Mathworks, Inc.) version R2018b with IBM CPLEX (IBM) as the linear programming solver.

Results

This study includes 13,596 participants from five independent cohorts including the Rotterdam Study (RS), the Study of Health in Pomerania (SHIP-TREND), the Cooperative Health Research in the Region of Augsburg (KORA) study, the European Prospective Investigation into Cancer (EPIC)-Norfolk Study, and the Netherlands Epidemiology of Obesity (NEO) study. A detailed description of the study participants is provided in **Table 1**. Depression was measured on a quantitative scale in all cohorts except the EPIC-Norfolk study, where the participants reported depression on a yes/no scale. The mean age ranged from 50.1 years in SHIP-Trend to 73.1 years in the Rotterdam Study. The percentage of female participants (51-56%) and mean BMI (between 26-28 kg/m^2) were comparable between studies. There were differences in the percentage of smokers between the cohorts, ranging from 11% in EPIC-Norfolk and to 22% in SHIP-Trend.

When testing for an association with depression adjusting for age and sex, 53 (41 novel) metabolites were significantly associated with depression after adjusting for multiple testing (false discovery rate (FDR) < 0.05; **Table 2** & **Figure 1**). These include nine metabolites in the amino acid metabolism pathway including five previously associated with depression (leucine, kynurenate, citrulline, glutamate and serotonin)[23,46,47] and four novel metabolites (N-acetylputrescine, 5-methylthioadenosine (MTA), 2-aminobutyrate and indolepropionate). In addition, significant association was found for six carbohydrates (one novel), six cofactors and vitamins, all of which were novel, 26 lipids (25 novel), and six xenobiotics (five novel) (**Table 2**).

Table 1. Descriptive statistics of the study populations.

	RS	SHIP-trend	KORA	EPIC-Norfolk B2	EPIC-Norfolk B3	NEO
N	484	965	1688	4639	5163	599
Ncases/Ncontrols	-	-	-	638/4001	685/4478	-
Mean age (years) (SD)	73.1 (6.3)	50.1 (13.6)	61 (8.8)	59.9 (8.8)	59.6 (8.9)	55.8 (6.0)
Age range (years)	62-96	20-81	32-77	40-78	40-78	45-66
Females (%)	52.5	56.0	51.4	52.4	52.8	52.6
Mean BMI (kg/m^2) (SD)	26.8 (3.7)	27.4 (4.6)	28.2 (4.8)	26.2 (3.7)	26.2 (3.8)	25.9 (4.0)
Smoking (%)	12.6	22.0	14.5	11.4	10.9	11.9
Medication						
Antidepressants (%)	3.7	4.0	5.6	4.5	3.8	5.3
Lipid-lowering medication (%)	10.5	7.8	16.7	1.4	1.5	7.7
Antihypertensives (%)	0.6	28.2	37.9	19.5	17.0	19.7
Antidiabetics (%)	5.4	0	7.5	1.9	2.0	2.7

Abbreviations: N = sample size, SD = standard deviation.

When adjusting for antidepressant use (model 2), 12 metabolites remained significantly associated (FDR < 0.05) with depression (**Table 2**, **Figure 1**), suggesting that most associations observed with depression were confounded by antidepressant medication use. Of the amino acids, only citrulline maintained significance after adjustment for antidepressant medication (**Table 2**, **Figure 1**). Other metabolites that remained significantly associated with depression in the extended model included four xenobiotics (4-hydroxycoumarin, hippurate, 3-phenylpropionate (hydrocinnamate) and cinnamoylglycine), four lipids (2-aminooctanoate, 10-undecenoate (11:1n1), 1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1) and 1-linoleoyl-GPA (18:2)), and three cofactors and vitamins (retinol (vitamin A), bilirubin (Z,Z), bilirubin (E,Z or Z,E)). Among these, higher levels of 1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1) and retinol (vitamin A) were associated with an increased risk of depression, while the others were associated with a decreased risk (**Figure 1**).

We subsequently build a more conservative model, further adjusting for other medication use, including lipid-lowering medication, antihypertensive medication, antidiabetic medication, BMI and current smoking (model 3). Seven out of the 12 metabolites remained significantly associated with depression (**Table 2**). These included retinol (vitamin A), hippurate, 4-hydroxycoumarin, 2-aminooctanoate, 10-undecenoate (11:1n1), 1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1), and 1-linoleoyl-GPA (18:2). Additionally, mannitol/sorbitol appeared statistically significant in model 3. Complete results of the meta-analysis are available in the **Supplementary Table 3**.

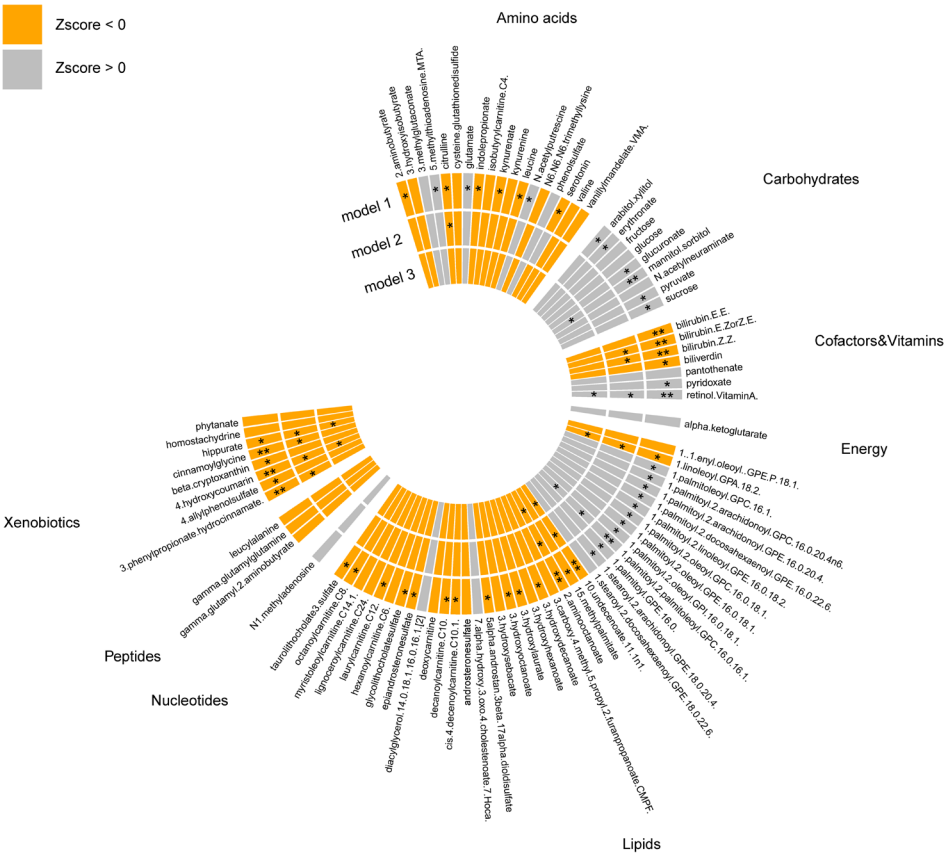


Figure 1: Association plot of metabolites with depression. This plot shows the top findings of the association analysis of metabolites with depressive symptoms, for all three models tested. Only metabolites with FDR *p*-value < 0.1 are shown in this Figure. The associations with a negative Z-score are depicted in grey, while the positive associations are depicted in orange. The plot is divided per metabolite subgroup. Significance levels: **: FDR < 0.001, *: FDR < 0.05. Script for Figure modified from Nath et al.[101].

There was no significant evidence for effect modification by sex (**Supplementary Table 4**) and the directionality of effects tended to be consistent in men and women. Effect sizes appeared to be stronger in women. Results were consistent across various versions of the Metabolon platform and depression assessing instruments and a sensitivity meta-analysis, which only included results from cohorts that had assessed metabolites on the most recent (HD4) platform, showed that they remained essentially unchanged (**Supplementary Table 5**).

Table 2. Top findings of the association analysis of metabolites with depression (FDR corrected *p*-value < 0.05 in model 1).

			Model 1				Model 2				Model 3			
Chemical ID	**Name**	**Super pathway**	**N**	**Zscore**	**Direction***	**FDR**	**N**	**Zscore**	**Direction***	**FDR**	**N**	**Zscore**	**Direction***	**FDR**
100001197	10-undecenoate (11:1n1)	Lipid	13596	-5.12	+-----	8.02E-05	13556	-3.94	+-----	1.7E-02	13549	-3.79	+-----	2.71E-02
100001740	mannitol/sorbitol	Carbohydrate	12631	5.14	+?-+++	8.02E-05	12592	3.39	+?-+++	5.02E-02	12586	3.58	+?-+++	4.47E-02
1090	bilirubin (Z,Z)	Cofactors and Vitamins	13596	-5.33	------	8.02E-05	13556	-3.60	------	3.5E-02	13549	-2.84	------	1.83E-01
100001950	bilirubin (E,E)*	Cofactors and Vitamins	13596	-5.25	------	8.02E-05	13556	-3.33	------	5.13E-02	13549	-2.73	------	2.08E-01
100002049	4-hydroxycoumarin	Xenobiotics	10885	-4.99	-??---	1.30E-04	10847	-4.48	-??---	4.0E-03	10847	-4.17	-??---	1.12E-02
100008984	1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1)*	Lipid	10885	4.84	+??+++	1.94E-04	10847	3.58	+??+++	3.5E-02	10847	3.51	+??+++	4.47E-02
100001951	bilirubin (E,Z or Z,E)*	Cofactors and Vitamins	12631	-4.74	-?----	2.83E-04	12592	-3.50	-?----	3.5E-02	12586	-2.98	-?----	1.37E-01
498	retinol (Vitamin A)	Cofactors and Vitamins	10885	4.65	+??+++	3.58E-04	10847	3.89	+??+++	1.7E-02	10847	4.14	+??+++	1.12E-02
100004227	2-aminooctanoate	Lipid	11850	-4.65	--?---	3.58E-04	11811	-4.00	--?---	1.7E-02	11810	-3.92	--?---	1.87E-02
100002253	cinnamoylglycine	Xenobiotics	10885	-4.40	-??---	9.38E-04	10847	-3.50	-??---	3.5E-02	10847	-3.32	-??---	6.71E-02
100000010	3-phenylpropionate (hydrocinnamate)	Xenobiotics	13596	-4.42	------	9.38E-04	13556	-3.73	------	2.9E-02	13549	-3.24	--+---	7.89E-02
100001251	decanoylcarnitine (C10)	Lipid	13596	-4.35	+-----	1.02E-03	13556	-3.35	+-----	5.13E-02	13549	-3.15	+-----	8.53E-02
1526	1-palmitoyl-2-oleoyl-GPE (16:0/18:1)	Lipid	10885	4.36	+??+++	1.02E-03	10847	2.64	-??+++	2.12E-01	10847	2.44	-??+++	2.68E-01
250	biliverdin	Cofactors and Vitamins	13596	-4.29	------	1.24E-03	13556	-2.88	------	1.40E-01	13549	-2.41	------	2.75E-01
504	serotonin	Amino acid	12631	-4.24	-?----	1.49E-03	12592	0.18	-?-+++	9.55E-01	12586	-0.02	-?-+++	9.92E-01
192	N-acetylputrescine	Amino acid	10885	4.06	-??+++	3.01E-03	10847	2.09	-??-++	3.52E-01	10847	2.16	-??+++	3.44E-01
100002259	cis-4-decenoylcarnitine (C10:1)	Lipid	13596	-4.02	------	3.27E-03	13556	-2.80	------	1.58E-01	13549	-2.47	------	2.68E-01
212	5-methylthioadenosine (MTA)	Amino acid	10885	4.03	+??+++	3.27E-03	10847	2.09	+??+++	3.52E-01	10847	1.78	+??+++	4.91E-01
100000997	3-hydroxydecanoate	Lipid	11850	-3.97	+-?---	3.89E-03	11811	-2.60	+-?---	2.23E-01	11810	-2.39	+-?---	2.75E-01
1539	1-palmitoyl-2-oleoyl-GPC (16:0/18:1)	Lipid	10885	3.85	+??++-	6.01E-03	10847	2.93	+??++-	1.36E-01	10847	2.76	+??++-	2.03E-01

Table 2. Continued

			Model 1				Model 2				Model 3			
Chemical ID	Name	Super pathway	N	Zscore	Direction*	FDR	N	Zscore	Direction*	FDR	N	Zscore	Direction*	FDR
100000014	hippurate	Xenobiotics	13596	-3.80	------	7.10E-03	13556	-4.17	------	1.1E-02	13549	-3.72	------	2.99E-02
100001392	laurylcarnitine (C12)	Lipid	13596	-3.73	+-----	8.97E-03	13556	-2.68	+-----	1.96E-01	13549	-2.61	+-----	2.48E-01
1128	2-aminobutyrate	Amino acid	13596	-3.68	------	9.61E-03	13556	-2.08	---+--	3.54E-01	13549	-1.84	---+--	4.74E-01
561	glutamate	Amino acid	13596	3.70	++++++	9.61E-03	13556	1.92	++++-+	4.13E-01	13549	1.15	++++-+	7.26E-01
98	kynurenate	Amino acid	10885	-3.61	-??---	1.22E-02	10847	-2.79	-??---	1.58E-01	10847	-2.50	-??---	2.68E-01
100001658	taurolithocholate 3-sulfate	Lipid	13596	-3.52	------	1.45E-02	13556	-3.09	------	9.31E-02	13549	-3.09	---+--	1.01E-01
100001511	1-palmitoleoyl-GPC (16:1)*	Lipid	13596	3.54	+++++-	1.45E-02	13556	2.43	+++++-	2.74E-01	13549	2.57	++++++	2.48E-01
100001112	3-hydroxylaurate	Lipid	10885	-3.53	-??---	1.45E-02	10847	-2.54	-??---	2.42E-01	10847	-2.43	-??---	2.69E-01
100008990	1-palmitoyl-2-arachidonoyl-GPE (16:0/20:4)*	Lipid	10885	3.52	-??+++	1.45E-02	10847	2.22	-??+++	3.00E-01	10847	2.23	-??+++	3.27E-01
100000773	3-hydroxyoctanoate	Lipid	11850	-3.54	+-?---	1.45E-02	11811	-2.25	+-?---	3.00E-01	11810	-2.13	+-?---	3.63E-01
100001868	4-allylphenol sulfate	Xenobiotics	10885	-3.48	-??---	1.58E-02	10847	-2.86	-??-+-	1.46E-01	10847	-2.46	-??-+-	2.68E-01
100001247	octanoylcarnitine (C8)	Lipid	13596	-3.47	+-----	1.58E-02	13556	-2.62	+-----	2.14E-01	13549	-2.48	+-----	2.68E-01
100001083	indolepropionate	Amino acid	13596	-3.48	------	1.58E-02	13556	-2.95	------	1.35E-01	13549	-2.12	------	3.66E-01
100001977	beta-cryptoxanthin	Xenobiotics	6246	-3.46	-???--	1.60E-02	6208	-3.25	-???--	5.74E-02	6208	-2.59	-???--	2.48E-01
100006430	arabitol/xylitol	Carbohydrate	12631	3.44	+?++++	1.65E-02	12592	2.48	+?++++	2.51E-01	12586	2.69	+?++++	2.15E-01
100009082	1-linoleoyl-GPA (18:2)*	Lipid	10885	-3.39	-??---	1.91E-02	10847	-3.57	-??---	3.5E-02	10847	-3.53	-??---	4.47E-02
100001870	1-palmitoyl-2-linoleoyl-GPE (16:0/18:2)	Lipid	10885	3.39	+??+++	1.91E-02	10847	2.02	+??+++	3.78E-01	10847	2.18	+??+++	3.41E-01
100000257	glucuronate	Carbohydrate	10885	3.35	+??+++	2.18E-02	10847	2.09	+??+++	3.52E-01	10847	1.93	+??+++	4.51E-01
391	citrulline	Amino acid	13596	-3.31	------	2.45E-02	13556	-3.53	------	3.5E-02	13549	-3.29	------	7.22E-02
100001121	pyridoxate	Cofactors and Vitamins	13596	3.29	++-++-	2.50E-02	13556	2.71	++-++-	1.83E-01	13549	3.17	++-++-	8.53E-02
100002021	5alpha-androstan-3beta,17alpha-diol disulfate	Lipid	10885	-3.25	-??---	2.79E-02	10847	-2.05	-??---	3.67E-01	10847	-1.63	-??---	5.35E-01
100001287	epiandrosterone sulfate	Lipid	13596	-3.16	-+----	3.59E-02	13556	-1.71	-+----	4.70E-01	13549	-1.73	-+----	5.03E-01

Table 2. Continued

			Model 1				Model 2				Model 3			
Chemical ID	**Name**	**Super pathway**	**N**	**Zscore**	**Direction***	**FDR**	**N**	**Zscore**	**Direction***	**FDR**	**N**	**Zscore**	**Direction***	**FDR**
100008914	1-palmitoyl-2-arachidonoyl-GPC (16:0/20:4n6)	Lipid	10885	3.13	-??+++	3.89E-02	10847	2.92	-??+++	1.36E-01	10847	2.94	-??++-	1.52E-01
100001320	erythronate*	Carbohydrate	12631	3.12	+?++++	3.89E-02	12592	1.76	+?++++	4.55E-01	12586	1.77	+?++++	4.91E-01
935	sucrose	Carbohydrate	10885	3.13	+??+++	3.89E-02	10847	1.61	+??+-+	4.99E-01	10847	1.16	+??+-+	7.21E-01
100001567	1-palmitoyl-GPE (16:0)	Lipid	13596	3.10	++-++-	4.01E-02	13556	2.15	++-++-	3.28E-01	13549	2.57	++++++	2.48E-01
100008991	1-palmitoyl-2-docosahexaenoyl-GPE (16.0/22.6)	Lipid	10401	3.06	???+++	4.26E-02	10401	2.43	???+++	2.74E-01	10401	2.71	???+++	2.15E-01
100001657	glycolithocholate sulfate*	Lipid	11850	-3.06	--?--+	4.26E-02	11811	-2.46	--?---	2.60E-01	11810	-2.29	--?---	3.09E-01
100008977	1-stearoyl-2-arachidonoyl-GPE (18:0/20:4)	Lipid	10885	3.07	-??+++	4.26E-02	10847	1.94	-??+++	4.07E-01	10847	1.93	-??+++	4.51E-01
823	pyruvate	Carbohydrate	13596	3.06	++++-+	4.26E-02	13556	2.28	++++-+	2.91E-01	13549	1.84	+-++-+	4.74E-01
397	leucine	Amino acid	13596	-3.04	--+--+	4.44E-02	13556	-2.73	--+--+	1.76E-01	13549	-3.33	------	6.71E-02
100002945	15-methylpalmitate	Lipid	13596	-3.04	+-----	4.44E-02	13556	-2.77	+-----	1.63E-01	13549	-2.40	+-----	2.75E-01
100009066	1-palmitoyl-2-oleoyl-GPI (16:0/18:1)*	Lipid	6246	3.03	+???++	4.50E-02	6208	1.96	+???+-	4.01E-01	6208	1.86	+???+-	4.66E-01

Abbreviations: N = sample size, FDR = false discovery rate.
* The order of the direction column: RS, SHIP-trend, KORA, EPIC-Norfolk B2, EPIC-Norfolk B3, NEO.

Table 3. Results of association of depression outcomes with dietary sources of metabolites.

	Current depressive symptoms				Major Depressive Disorder (MDD)						
	N	Beta	SE	p-value	N	Beta	SE	p-value	OR	Lower 95% CI	Upper 95% CI
Vitamin A supplements	304399	0.23	0.02	1.25E-25	189800	0.34	0.04	9.72E-18	1.40	1.30	1.52
Vitamin D supplement	313100	0.19	0.02	5.07E-32	194681	0.34	0.03	1.22E-34	1.41	1.33	1.48
Retinol intake from food	61363	2.29E-04	4.02E-05	1.26E-08	38758	1.76E-04	5.50E-05	1.37E-03	1.00	1.00	1.00
Fresh fruits intake	434770	-0.06	0.002	1.61E-205	264796	-0.04	0.004	3.27E-22	0.96	0.96	0.97
Vitamin K antagonists use	435867	0.43	0.03	1.04E-46	265648	0.14	0.06	0.016	1.15	1.03	1.28

Abbreviations: N = sample size, SE = standard error, OR = odds ratio, 95% CI = 95% confidence interval.

Association of depression with dietary sources of metabolites in the UK Biobank

To evaluate the association of food sources of the identified metabolites with major depression we conducted a series of analyses in the UK Biobank (UKB). In the UKB information on vitamin supplements including vitamin A, retinol intake from food, consumption of fresh fruits – a major source of hippurate, and medication use including vitamin k antagonist (a proxy for 4-hydroxycoumarin) was available. In a cross-sectional analysis, we found a significant positive association of vitamin A intake from supplements with both measures of depression including current depressive symptoms (beta = 0.23, *p*-value = $1.25x10^{-25}$) and lifetime major depressive disorder (MDD, OR = 1.40, *p*-value = $9.72x10^{-18}$). However, vitamin D supplement intake was also significantly positively associated with both measures of depression (**Table 3**), suggesting that depressed individuals take more vitamin supplements than non-depressed individuals. Since both vitamin A and vitamin D are fat-soluble and can cross the blood-brain barrier, we performed additional association with the measure of brain pathology, i.e., white matter hyperintensity (WMH) volume. Only vitamin A supplement intake was found to be associated with higher volume of WMH (beta = 479.09, *p*-value = 0.04, **Supplementary Table 6**), suggesting a possible role of vitamin A in brain diseases. To address the issue of reverse causality, we additionally tested the association of depression with retinol intake estimated from the food consumed in the previous 24 hours. Significant positive association of estimated retinol intake was observed with both measures of depression (current depressive symptoms, *p*-value = $1.26x10^{-08}$; lifetime MDD, *p*-value = $1.4x10^{-03}$). However, the effect estimates were small (**Table 3**), which may in part be explained by the imprecision of food consumption questionnaires. Fresh fruit intake, a major source of hippurate, was negatively associated with both measures of depression (current depressive symptoms, beta = -0.06, *p*-value = $1.61x10^{-205}$; lifetime MDD, OR = 0.96, *p*-value = $3.27x10^{-22}$) and vitamin K antagonists, a proxy for 4-hydroxycoumarin, was positively associated with both measures of depression (current depressive symptoms, beta = 0.43, *p*-value = $1.04x10^{-46}$; lifetime MDD, OR = 1.15, *p*-value = 0.016) (**Table 3**).

Mendelian randomization analysis

Testing the hypothesis that major depression results in changes of circulating metabolites in the Mendelian randomization analysis (MR), nominally significant results were obtained for 2-aminooctanoate and 10-undecenoate (11:1n1), under the MR-Egger method and weighted median method, respectively. However, these findings did not remain significant after adjustment for multiple testing (**Supplementary Table 7**). MR models in which we tested the hypothesis that levels of circulating metabolites increase the risk of depression provided significant evidence for a causal relation between hippurate and the risk of depression, both in the MR-Egger robust and penalized-robust methods (**Supplementary Table 8**). The effect estimate was consistent with the inverse relationship observed between hippurate and major depression in this study. However, a significant intercept was also observed suggesting pleiotropy. To exclude a pleiotropic effect, we studied the effect of intervention on the metabolite in the PReDICT trial.

Effect of antidepressant therapy on hippurate

To further evaluate the impact of antidepressant therapy including cognitive behavioral therapy (CBT), duloxetine – a serotonin-norepinephrine reuptake inhibitor (SNRI) and escitalopram – a selective serotonin reuptake inhibitor (SSRI) on hippurate we consulted the PReDICT study. The PReDICT study allowed us to test the effect of antidepressant therapy on the metabolite levels in circulation by measuring the metabolite levels before and after the antidepressant therapy. In PReDICT, we found that levels of hippurate in the circulation increased significantly from baseline to week 12 only after treatment with escitalopram (estimated week 12 vs. baseline difference = 0.45, 95% confidence interval (CI; 0.16,0.74), *p*-value = 0.002; **Supplementary Figure 1**), but not in the cognitive behavior therapy (CBT) and duloxetine treatment arms (CBT: estimated difference = -0.02, 95% CI (-0.39,0.33) and *p*-value = 0.87; duloxetine: estimated difference = 0.13, 95% CI (-0.17,0.44) and *p*-value = 0.38). In this study, we could not show a relation between hippurate and depression as the study recruited patients only and lacked healthy controls. In patients receiving pharmacotherapy (escitalopram and duloxetine arms), the association of baseline depression as measured by the 17-item Hamilton Rating Scale for Depression (HRSD17) and baseline hippurate was not statistically significant (beta = 0.04, 95% CI (-0.03,0.11), *p*-value = 0.27). Further, no significant association was observed between depression in week 12 as measured by the HRSD17 and week 12 hippurate (beta = 0.09, *p*-value = 0.45) and 12 weeks change in HRSD17 and 12 weeks change in hippurate (beta = 0.02, 95% CI (-0.65,1.57), *p*-value = 0.85).

Linking the human circulating metabolome to gut microbiome metabolism

Of the 53 metabolites identified in this study in model 1, 28 metabolites could be matched to a unique VMH metabolite ID. For each metabolite, the presence or absence in the global human reconstruction, Recon3D[48], and a resource of 7,206 reconstructions of human gut microbes, AGORA2 (https://www.biorxiv.org/content/10.1101/2020.11.09.375451v1) was retrieved. In total, 12 metabolites were present in both the human and gut microbial metabolic networks, three were only present in gut microbes, and 13 were only present in human (**Supplementary Table 9**). To further investigate potential links between the microbiome and metabolites associated with depression, the potential of the 7,206 AGORA2 strains to consume or secrete the 15 microbial metabolites identified in this study was computed. Since hippurate is synthesized in the liver and renal cortex from the microbial metabolite benzoate[49], the uptake and secretion potential for benzoate was also predicted for the 7,206 strains.

Of the 16 depression-related analyzed metabolites, 12 are consumed or secreted by at least one AGORA2 strain (**Supplementary Tables 10** and **11**). A wide range of genera and species were involved in the uptake of mannitol (**Supplementary Table 10**). Mannitol is largely secreted by several species of the genus *Bacteroides* followed by *Lactobacillus,* among others (**Supplementary Table 11**). Both genera have previously been found to be associated with depression[17]. In total, 3,616 AGORA2 strains mainly of the Gammaproteobacteria and Bacilli

classes (**Supplementary Table 11**) synthesized benzoate as a product of benzamide (VMH reaction ID: BZAMAH). Interestingly, benzamides are a class of antipsychotic medication.

Similarly, a wide variety of species distributed across all major gut microbial phyla secrete L-glutamate, L-leucine, and serotonin (**Supplementary Figure 2**). The nucleoside 5-methylthioadenosine is secreted and consumed by a limited number of species including *Escherichia coli* and *Clostridioides difficile* (**Supplementary Figure 2**). Glucuronate can be consumed and secreted by a variety of taxa belonging mainly to the Proteobacteria and Bacteroidetes phyla (**Supplementary Figure 2**). While sucrose is widely utilized, the sugar alcohols L-arabitol and xylitol are consumed by a limited number of species including *Klebsiella pneumoniae* and *Streptococcus mutants*, respectively (**Supplementary Figure 2**). Finally, the tryptophan metabolite indole-3-propionate, which is also neuroactive, can be produced only by four species of the Clostridiales order (**Supplementary Figure 2**). Taken together, these simulation results highlight that these measured metabolites could indeed by of microbial origin and thus, that human microbial strains may influence the bioavailability of metabolites implicated in depression.

Discussion

In this study, we identified 53 metabolites significantly associated with depression, most of which, including those in the monoamine and neurotransmitter pathways (serotonin, kynurenate and glutamate), were explained by antidepressant use. We identified novel associations with depression for six metabolites, including retinol (vitamin A), 4-hydroxycoumarin, 2-aminooctanoate, 10-undecenoate (11:1n1), 1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1), 1-linoleoyl-GPA (18:2) and confirmed the association of hippurate and mannitol/sorbitol. We found that the relation of hippurate and depression is not confounded by medication and may be causal. We show that hippurate levels can be modified by a specific antidepressant, escitalopram. Analyzing the major dietary sources of these metabolites in the UKB study, we found that retinol (vitamin A) intake was significantly higher and fresh fruits intake significantly lower in depressed individuals compared to those who were not depressed.

One of the most interesting findings of this study is the identification of the association of higher levels of retinol (active form of vitamin A) with depression. There have been several case reports of individuals with vitamin A intoxication with no previous history of depression, who developed symptoms of depression and even psychosis when overdosed with vitamin A[50,51]. Depressive symptoms resolved upon discontinuation of vitamin A, implying that depression may be a side effect of vitamin A intake[50]. Animal models have suggested elevated monoamine oxidase enzyme activity and depression-related behavior upon vitamin A supplementation[52,53]. Our study is the first to link higher levels of retinol in blood with depression in the general population. Retinol and its derivatives known as retinoids are lipid

soluble and are able to cross the blood-brain barrier. Vitamin A is required for brain development and functioning[54,55]. However, excess of vitamin A is neurotoxic and may result in brain shrinkage[55]. Brain areas high in retinoic acid signaling and receptors overlap with areas of relevance to stress and depression[56]. Further, vitamin A is known to increase the synthesis of triglyceride-rich very low-density lipoproteins (VLDLs) and apolipoproteins in the liver[57,58], which we found associated with depression in our previous study[24]. An important question to answer whether the increased levels of vitamin A are related to changes in food intake related to depression. In the UK Biobank we found significant increase in dietary retinol intake in individuals with depression. Thus, our findings ask for intervention studies that evaluate prospectively the effect of vitamin A reduction in depressed patients.

Two of the most strongly associated metabolites with depression were xenobiotics, hippurate and 4-hydroxycoumarin. In line with the findings of our study, decreased levels of hippurate have been previously reported in urine and plasma of individuals with unipolar and bipolar depression consistently in several studies and it has been suggested as a biomarker for depressive disorders[23]. Our MR analysis suggests that low hippurate levels in circulation are a part of the causal pathway leading to depression. However, as the MR could not exclude a pleiotropic effect, our findings yield a hypothesis that requires further evaluation in a larger trial. While we could not show an association between hippurate and depression in the PReDICT study, as the study lacked controls, hippurate levels were higher 12 weeks after initiation of selective serotonin reuptake inhibitor (SSRI) therapy (escitalopram) but not for SNRI or CBT, raising the question whether blood levels of hippurate can be used in clinical trials for compliance and efficiency of SSRIs specifically. It should be noted that hippurate is derived from benzoate which is part of the human diet. In line with our finding hippurate was predicted to be synthesized by a wide range of microbes, a previous study shows that hippurate was strongly associated to increased gut microbiome diversity[59]. Polyphenol-rich foods including coffee, whole grains and fruit have been reported to increase levels of hippurate. In the UKB we found significantly decreased fresh fruit intake among individuals with depression.

The metabolite 4-hydroxycoumarin is a fungal derivative of coumarin. Coumarins are found naturally in plants and spices[60] and coumarin is converted into 4-hydroxycoumarin by fungi[61]. 4-hydroxycoumarin is then converted into dicoumarol in the presence of formaldehyde[61]. Dicoumarol is an anticoagulant (warfarin) that inhibits the synthesis of vitamin K, also called vitamin K antagonist, and is commonly used to treat thromboembolic diseases[62]. In the UKB, we found significant positive association of anticoagulant use (vitamin K antagonists) with major depression. A history of depression is a risk factor for thromboembolism[63-65]. Antidepressants are also known to interact with warfarin[66] and are also associated with increased risk of thromboembolism[67]. Taking all findings together, we hypothesize that depression/antidepressant use depletes 4-hydroxycoumarin in the circulation leading to thromboembolism. Vitamin K has been shown to act in the nervous system as it is involved in sphingolipid synthesis[68]. Sphingolipids are present in high concentrations in cell

membranes of neuronal and glial cells[69]. Sphingolipids are essential for important cellular events, including proliferation, differentiation, senescence, cell-cell interactions, and transformation[70] and they have been linked to aging, Alzheimer's disease, and Parkinson's disease[71-73]. Further, sphingolipids were found to play a crucial role in the development of depression- and anxiety-related behaviors in mice[74,75] and depression is seen often in patients with sphingolipid storage diseases[76-80]. Treatment with escitalopram/citalopram is also associated with changes in sphingolipids[81]. In our study, we did not find an association of depression with circulating sphingolipids present on the Metabolon platform. However, we cannot exclude that 4-hydroxycoumarin in the blood affects sphingolipid metabolism in the brain specifically.

Other metabolites that were found to be significant in our study include mannitol/sorbitol, of which increased levels were associated with depression. Higher levels of sorbitol in plasma and urine have previously been consistently reported in patients with unipolar and bipolar depression and, like hippurate, it has been suggested as a diagnostic biomarker of depression[23]. Mannitol/sorbitol are sugar alcohols found in food such as fruits and berries and often used in diet/sugar free foods as sweeteners[82]. Fructose reduced diets have been shown to improve gastrointestinal disorders, depression and mood disorders[83]. Our AGORA2 analysis suggests that mannitol is mainly secreted by several species of *Bacteroides*, *Lactobacillus*, *Fructobacillus*, *Alistipes* and *Bifidobacterium*. Interestingly, all genera, except for *Fructobacillus* have previously been associated with depression[17], asking for further studies on the role of the microbiome, circulating levels of mannitol and depression.

Finally, there were four lipids identified in our study (2-aminooctanoate, 10-undecenoate (11:1n1), 1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1) and 1-linoleoyl-GPA (18:2)) significantly associated with depression. 1-Palmitoyl-2-palmitoleoyl-GPC (16:0/16:1) also known as phosphatidylcholine (16:0/16:1) or lecithin (HMDB0007969) is commonly found in foods like eggs, soyabean, liver, nuts and seeds and is a precursor of choline. Lecithin is believed to cause depression by increasing the production of acetylcholine in the brain[84]. When fed to animals and humans, lecithin significantly increases the levels of choline in blood and brain and of acetylcholine in brain[85-87]. Our study is the first to show higher circulating levels of lecithin in depressed individuals from the general population. The other three lipids 2-aminooctanoate, 10-undecenoate (11:1n1) and 1-linoleoyl-GPA (18:2) were negatively associated with depression. 2-Aminooctanoate (alpha-aminocaprylic acid) and 10-undecenoate (11:1n1) (undecylenic acid) are neutral hydrophobic molecules for which there is not much known in the literature. Lower levels of 10-undecenoate (11:1n1) have been found in individuals with non-alcoholic fatty liver disease[88]. 1-linoleoyl-GPA (18:2) is a lysophosphatidic acid (LPA 18:2). LPA is a bioactive membrane lipid that acts on at least six distinct G protein□coupled receptors (LPA_{1-6}) and plays a role in pain sensitivity and emotional regulation[89]. LPA knock out mice exhibit anxiety-related behaviour[89,90].

We found that decreased plasma levels of serotonin, kynurenate, leucine and citrulline and higher levels of glutamate were associated with depression. Lower plasma/serum levels of serotonin, kynurenate, citrulline and leucine and higher levels of glutamate have been reported in relationship to depression in earlier studies[23,46,47,91], which also appears consistent with our findings of model 1. However, we and others have shown that antidepressants affect plasma/serum levels of serotonin, glutamate, leucine and kynurenine[92-95]. An important finding of our study is that only citrulline remained significantly associated with depression after adjusting for antidepressant medication use. Lower levels of citrulline and its precursor arginine were previously associated to depression in unmedicated individuals[47,96]. Interestingly, treatment with SSRIs significantly increase the levels of plasma citrulline[81]. Further, levels of plasma citrulline were found to be significantly increased two hours post treatment with ketamine, suggesting a possible mechanism of action of the rapid acting drug[97]. Citrulline is an intermediate in the urea cycle and linked to nitric-oxide synthesis[98]. It is absorbed by the gut and has useful therapeutic effects against cardiovascular diseases[99]. In our study the association of citrulline with depression lost its significance, albeit not completely, after adjusting for cardiovascular medication use and BMI.

Our study is the first large-scale effort combining metabolites measured on assorted, untargeted metabolomics platforms (Metabolon) studied in relationship to depression. In addition to confirming several previously identified metabolites in smaller studies, we successfully identified novel metabolites that are associated with depression. Our findings are robust across different versions of the Metabolon platform or the criteria assessing presence of clinical or subclinical depression. A possible limitation of our study is that differences in metabolomics platforms and technologies that were used by different cohorts to assess depression may have resulted in a reduction of statistical power. Older versions of the Metabolon platform reported significantly fewer known metabolites compared to the more recent implementations. Another possible limitation of our study is the presence of residual confounding. After adjusting for medication use and the lifestyle factors smoking and BMI, confounding may still be present and may influence the results[100]. Also, our MR analysis was most likely underpowered lacking good instrumental variables for both depression and associated metabolites.

Analyzing circulating levels of 806 metabolites from untargeted metabolomics platforms in 13,596 individuals, we identified six new associations of metabolites with depression including retinol (vitamin A), 4-hydroxycoumarin and four lipids, 2-aminooctanoate, 10-undecenoate (11:1n1), 1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1) and 1-linoleoyl-GPA (18:2), while confirming known associations of hippurate and mannitol/sorbitol. We further show that previously identified associations of depression with metabolites belonging to the amino-acid pathways including serotonin, kynurenate, leucine and glutamate are likely explained by antidepressant medication. Our findings point to effective preventive targets, as most of these metabolites are food derived and thus can be altered in patients by modifying diet.

References

1 Bromet, E. *et al.* Cross-national epidemiology of DSM-IV major depressive episode. *BMC Med* **9**, 90, doi:10.1186/1741-7015-9-90 (2011).

2 Salari, N. *et al.* Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: a systematic review and meta-analysis. *Global Health* **16**, 57, doi:10.1186/s12992-020-00589-w (2020).

3 Sher, L. The impact of the COVID-19 pandemic on suicide rates. *QJM*, doi:10.1093/qjmed/hcaa202 (2020).

4 Sullivan, P. F., Neale, M. C. & Kendler, K. S. Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry* **157**, 1552-1562, doi:10.1176/appi.ajp.157.10.1552 (2000).

5 Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci* **22**, 343-352, doi:10.1038/s41593-018-0326-7 (2019).

6 Batterham, P. J., Christensen, H. & Mackinnon, A. J. Modifiable risk factors predicting major depressive disorder at four year follow-up: a decision tree approach. *Bmc Psychiatry* **9**, doi:10.1186/1471-244x-9-75 (2009).

7 Beekman, A. T. *et al.* Anxiety and depression in later life: Co-occurrence and communality of risk factors. *Am J Psychiatry* **157**, 89-95, doi:10.1176/ajp.157.1.89 (2000).

8 Moffitt, T. E. *et al.* Depression and generalized anxiety disorder: cumulative and sequential comorbidity in a birth cohort followed prospectively to age 32 years. *Arch Gen Psychiatry* **64**, 651-660, doi:10.1001/archpsyc.64.6.651 (2007).

9 Kitching, D. Depression in dementia. *Aust Prescr* **38**, 209-211, doi:DOI 10.18773/austprescr.2015.071 (2015).

10 Kang, H. J. *et al.* Comorbidity of depression with physical disorders: research and clinical implications. *Chonnam Med J* **51**, 8-18, doi:10.4068/cmj.2015.51.1.8 (2015).

11 Judd, L. L. The clinical course of unipolar major depressive disorders. *Arch Gen Psychiatry* **54**, 989-991 (1997).

12 Lopez, A. D., Mathers, C. D., Ezzati, M., Jamison, D. T. & Murray, C. J. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. *Lancet* **367**, 1747-1757, doi:10.1016/S0140-6736(06)68770-9 (2006).

13 Wittchen, H. U. *et al.* The size and burden of mental disorders and other disorders of the brain in Europe 2010. *Eur Neuropsychopharmacol* **21**, 655-679, doi:10.1016/j.euroneuro.2011.07.018 (2011).

14 Angst, F., Stassen, H. H., Clayton, P. J. & Angst, J. Mortality of patients with mood disorders: follow-up over 34-38 years. *J Affect Disord* **68**, 167-181, doi:S0165032701003779 [pii] (2002).

15 Korte, S. M. *et al.* The many different faces of major depression: it is time for personalized medicine. *Eur J Pharmacol* **753**, 88-104, doi:10.1016/j.ejphar.2014.11.045 (2015).

16 Pandya, M., Altinay, M., Malone, D. A., Jr. & Anand, A. Where in the brain is depression? *Curr Psychiatry Rep* **14**, 634-642, doi:10.1007/s11920-012-0322-7 (2012).

17 Cheung, S. G. *et al.* Systematic Review of Gut Microbiota and Major Depression. *Front Psychiatry* **10**, 34, doi:10.3389/fpsyt.2019.00034 (2019).

18 Sethi, S. & Brietzke, E. Omics-Based Biomarkers: Application of Metabolomics in Neuropsychiatric Disorders. *Int J Neuropsychopharmacol* **19**, pyv096, doi:10.1093/ijnp/pyv096 (2015).

19 Quinones, M. P. & Kaddurah-Daouk, R. Metabolomics tools for identifying biomarkers for neuropsychiatric diseases. *Neurobiol Dis* **35**, 165-176, doi:10.1016/j.nbd.2009.02.019 (2009).

20 Patti, G. J., Yanes, O. & Siuzdak, G. Innovation: Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol* **13**, 263-269, doi:10.1038/nrm3314 (2012).

21 Valles-Colomer, M. *et al.* The neuroactive potential of the human gut microbiota in quality of life and depression. *Nat Microbiol* **4**, 623-632, doi:10.1038/s41564-018-0337-x (2019).

22 Liu, J. *et al.* Integration of epidemiologic, pharmacologic, genetic and gut microbiome data in a drug-metabolite atlas. *Nat. Med.* **26**, 110-117, doi:10.1038/s41591-019-0722-x (2020).

23 MacDonald, K. *et al.* Biomarkers for major depressive and bipolar disorders using metabolomics: A systematic review. *Am J Med Genet B* **180**, 122-137, doi:10.1002/ajmg.b.32680 (2019).

24 Bot, M. *et al.* Metabolomics Profile in Depression: A Pooled Analysis of 230 Metabolic Markers in 5283 Cases With Depression and 10,145 Controls. *Biol Psychiatry* **87**, 409-418, doi:10.1016/j.biopsych.2019.08.016 (2020).

25 Deelen, J. *et al.* A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals. *Nature Communications* **10**, 3346, doi:10.1038/s41467-019-11311-9 (2019).

26 Wurtz, P. *et al.* Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation* **131**, 774-785, doi:10.1161/CIRCULATIONAHA.114.013116 (2015).

27 Ahola-Olli, A. V. *et al.* Circulating metabolites and the risk of type 2 diabetes: a prospective study of 11,896 young adults from four Finnish cohorts. *Diabetologia* **62**, 2298-2309, doi:10.1007/s00125-019-05001-w (2019).

28 Dunlop, B. W. *et al.* Predictors of remission in depression to individual and combined treatments (PReDICT): study protocol for a randomized controlled trial. *Trials* **13**, 106, doi:10.1186/1745-6215-13-106 (2012).

29 Shin, S. Y. *et al.* An atlas of genetic influences on human blood metabolites. *Nat Genet* **46**, 543-550, doi:10.1038/ng.2982 (2014).

30 Lewinsohn, P. M., Seeley, J. R., Roberts, R. E. & Allen, N. B. Center for Epidemiologic Studies Depression Scale (CES-D) as a screening instrument for depression among community-residing older adults. *Psychol Aging* **12**, 277-287 (1997).

31 Kroenke, K., Spitzer, R. L. & Williams, J. B. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* **16**, 606-613, doi:10.1046/j.1525-1497.2001.016009606.x (2001).

32 Loewe, B., Spitzer, R. L., Zipfel, S. & Herzog, W. *Gesundheitsfragebogen fuer Patienten (PHQ-D). Manual und Testunterlagen*. Second edn, (Pfizer, 2002).

33 Rush, A. J., Gullion, C. M., Basco, M. R., Jarrett, R. B. & Trivedi, M. H. The Inventory of Depressive Symptomatology (IDS): psychometric properties. *Psychol Med* **26**, 477-486, doi:10.1017/s0033291700035558 (1996).

34 Smith, D. J. *et al.* Prevalence and characteristics of probable major depression and bipolar disorder within UK biobank: cross-sectional study of 172,751 participants. *PLoS One* **8**, e75362, doi:10.1371/journal.pone.0075362 (2013).

35 First, M. B., Spitzer, R. L., Gibbon, M. & Williams, J. B. (ed NY State Psychiatric Institute Biometrics Research Department) (New York, 1996).

36 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, doi:10.1016/j.jalz.2018.07.217 (2019).

37 McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-1283, doi:10.1038/ng.3643 (2016).

38 Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191, doi:10.1093/bioinformatics/btq340 (2010).

39 Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature Neuroscience* **22**, 343-+, doi:10.1038/s41593-018-0326-7 (2019).

40 Yavorska, O. O. & Burgess, S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int J Epidemiol* **46**, 1734-1739, doi:10.1093/ije/dyx034 (2017).

41 Noronha, A. *et al.* The Virtual Metabolic Human database: integrating human and gut microbiome metabolism with nutrition and disease. *Nucleic Acids Res.* **47**, D614-D624, doi:10.1093/nar/gky992 (2019).

42 Heinken, A. *et al.* AGORA2: Large scale reconstruction of the microbiome highlights wide-spread drug-metabolising capacities. *bioRxiv*, 2020.2011.2009.375451, doi:10.1101/2020.11.09.375451 (2020).

43 Magnusdottir, S. *et al.* Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat. Biotechnol.* **35**, 81-89, doi:10.1038/nbt.3703 (2017).

44 Gudmundsson, S. & Thiele, I. Computationally efficient flux variability analysis. *BMC Bioinformatics* **11**, 489, doi:10.1186/1471-2105-11-489 (2010).

45 Heirendt, L. *et al.* Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat. Protoc.* **14**, 639-702, doi:10.1038/s41596-018-0098-2 (2019).

46 Bakkaloglu, B. *et al.* Platelet and Plasma Serotonin Levels and Platelet Monoamine Oxidase Activity in Patients with Major Depression: Effects of Sertraline Treatment. *Turk J Biochem* **33**, 97-103 (2008).

47 Hess, S. *et al.* Decreased serum L-arginine and L-citrulline levels in major depression. *Psychopharmacology (Berl)* **234**, 3241-3247, doi:10.1007/s00213-017-4712-8 (2017).

48 Brunk, E. *et al.* Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat Biotechnol* **36**, 272-281, doi:10.1038/nbt.4072 (2018).

49 Lees, H. J., Swann, J. R., Wilson, I. D., Nicholson, J. K. & Holmes, E. Hippurate: the natural history of a mammalian-microbial cometabolite. *J Proteome Res* **12**, 1527-1546, doi:10.1021/pr300900b (2013).

50 Bremner, J. D., Shearer, K. D. & McCaffery, P. J. Retinoic acid and affective disorders: the evidence for an association. *J Clin Psychiatry* **73**, 37-50, doi:10.4088/JCP.10r05993 (2012).

51 Landy, D. Pibloktoq (hysteria) and Inuit nutrition: possible implication of hypervitaminosis A. *Soc Sci Med* **21**, 173-185, doi:10.1016/0277-9536(85)90087-5 (1985).

52 O'Reilly, K., Bailey, S. J. & Lane, M. A. Retinoid-mediated regulation of mood: possible cellular mechanisms. *Exp Biol Med (Maywood)* **233**, 251-258, doi:10.3181/0706-MR-158 (2008).

53 de Oliveira, M. R., da Rocha, R. F., Pasquali, M. A. & Moreira, J. C. The effects of vitamin A supplementation for 3 months on adult rat nigrostriatal axis: increased monoamine oxidase enzyme activity, mitochondrial redox dysfunction, increased beta-amyloid(1-40) peptide and TNF-alpha contents, and susceptibility of mitochondria to an in vitro H2O2 challenge. *Brain Res Bull* **87**, 432-444, doi:10.1016/j.brainresbull.2012.01.005 (2012).

54 Lane, M. A. & Bailey, S. J. Role of retinoid signalling in the adult brain. *Prog Neurobiol* **75**, 275-293, doi:10.1016/j.pneurobio.2005.03.002 (2005).

55 Mey, J. & McCaffery, P. Retinoic acid signaling in the nervous system of adult vertebrates. *Neuroscientist* **10**, 409-421, doi:10.1177/1073858404263520 (2004).

56 Bremner, J. D. & McCaffery, P. The neurobiology of retinoic acid in affective disorders. *Prog Neuropsychopharmacol Biol Psychiatry* **32**, 315-331, doi:10.1016/j.pnpbp.2007.07.001 (2008).

57 Gerber, L. E. & Erdman, J. W., Jr. Changes in lipid metabolism during retinoid administration. *J Am Acad Dermatol* **6**, 664-674, doi:10.1016/s0190-9622(82)80047-9 (1982).

58 Klor, H. U. *et al.* The impact of oral vitamin A derivatives on lipid metabolism - What recommendations can be derived for dealing with this issue in the daily dermatological practice? *J Dtsch Dermatol Ges* **9**, 600-606, doi:10.1111/j.1610-0387.2011.07637.x (2011).

59 Pallister, T. *et al.* Hippurate as a metabolomic marker of gut microbiome diversity: Modulation by diet and relationship to metabolic syndrome. *Sci Rep* **7**, 13670, doi:10.1038/s41598-017-13722-4 (2017).

60 Lončar, M. *et al.* Coumarins in Food and Methods of Their Determination. *Foods* **9**, doi:10.3390/foods9050645 (2020).

61 Liu, B., Raeth, T., Beuerle, T. & Beerhues, L. A novel 4-hydroxycoumarin biosynthetic pathway. *Plant Mol Biol* **72**, 17-25, doi:10.1007/s11103-009-9548-0 (2010).

62 Manolov, I., Maichle-Moessmer, C. & Danchev, N. Synthesis, structure, toxicological and pharmacological investigations of 4-hydroxycoumarin derivatives. *Eur J Med Chem* **41**, 882-890, doi:10.1016/j.ejmech.2006.03.007 (2006).

63 von Kanel, R. *et al.* Depressive symptoms as a novel risk factor for recurrent venous thromboembolism: a longitudinal observational study in patients referred for thrombophilia investigation. *PLoS One* **10**, e0125858, doi:10.1371/journal.pone.0125858 (2015).

64 Kunutsor, S. K., Seidu, S. & Khunti, K. Depression, antidepressant use, and risk of venous thromboembolism: systematic review and meta-analysis of published observational evidence. *Ann Med* **50**, 529-537, doi:10.1080/07853890.2018.1500703 (2018).

65 Lee, C. W. *et al.* Depression and risk of venous thromboembolism: a population-based retrospective cohort study. *Psychosom Med* **77**, 591-598, doi:10.1097/PSY.0000000000000193 (2015).

66 Sansone, R. A. & Sansone, L. A. Warfarin and Antidepressants: Happiness without Hemorrhaging. *Psychiatry (Edgmont)* **6**, 24-29 (2009).

67 Parkin, L. *et al.* Antidepressants, Depression, and Venous Thromboembolism Risk: Large Prospective Study of UK Women. *J Am Heart Assoc* **6**, doi:10.1161/JAHA.116.005316 (2017).

68 Ferland, G. Vitamin K and the nervous system: an overview of its actions. *Adv Nutr* **3**, 204-212, doi:10.3945/an.111.001784 (2012).

69 Bartke, N. & Hannun, Y. A. Bioactive sphingolipids: metabolism and function. *J Lipid Res* **50 Suppl**, S91-96, doi:10.1194/jlr.R800080-JLR200 (2009).

70 Zeidan, Y. H. & Hannun, Y. A. Translational aspects of sphingolipid metabolism. *Trends Mol Med* **13**, 327-336, doi:10.1016/j.molmed.2007.06.002 (2007).

71 Cutler, R. G. *et al.* Involvement of oxidative stress-induced abnormalities in ceramide and cholesterol metabolism in brain aging and Alzheimer's disease. *Proc Natl Acad Sci U S A* **101**, 2070-2075, doi:10.1073/pnas.0305799101 (2004).

72 Jana, A., Hogan, E. L. & Pahan, K. Ceramide and neurodegeneration: susceptibility of neurons and oligodendrocytes to cell damage and death. *J Neurol Sci* **278**, 5-15, doi:10.1016/j.jns.2008.12.010 (2009).

73 Posse de Chaves, E. & Sipione, S. Sphingolipids and gangliosides of the nervous system in membrane function and dysfunction. *FEBS Lett* **584**, 1748-1759, doi:10.1016/j.febslet.2009.12.010 (2010).
74 Gulbins, E. *et al.* Acid sphingomyelinase-ceramide system mediates effects of antidepressant drugs. *Nat Med* **19**, 934-938, doi:10.1038/nm.3214 (2013).
75 Muller, C. P. *et al.* Brain membrane lipids in major depression and anxiety disorders. *Biochim Biophys Acta* **1851**, 1052-1065, doi:10.1016/j.bbalip.2014.12.014 (2015).
76 MacQueen, G. M., Rosebush, P. I. & Mazurek, M. F. Neuropsychiatric aspects of the adult variant of Tay-Sachs disease. *J Neuropsychiatry Clin Neurosci* **10**, 10-19, doi:10.1176/jnp.10.1.10 (1998).
77 Cole, A. L. *et al.* Depression in adults with Fabry disease: a common and under-diagnosed problem. *J Inherit Metab Dis* **30**, 943-951, doi:10.1007/s10545-007-0708-6 (2007).
78 Laney, D. A. *et al.* Social-adaptive and psychological functioning of patients affected by Fabry disease. *J Inherit Metab Dis* **33 Suppl 3**, S73-81, doi:10.1007/s10545-009-9025-6 (2010).
79 Sadek, J., Shellhaas, R., Camfield, C. S., Camfield, P. R. & Burley, J. Psychiatric findings in four female carriers of Fabry disease. *Psychiatr Genet* **14**, 199-201, doi:00041444-200412000-00006 [pii] (2004).
80 Packman, W., Wilson Crosbie, T., Riesner, A., Fairley, C. & Packman, S. Psychological complications of patients with Gaucher disease. *J Inherit Metab Dis* **29**, 99-105, doi:10.1007/s10545-006-0154-x (2006).
81 MahmoudianDehkordi, S. *et al.* Alterations in Acylcarnitines, Amines, and Lipids Inform about Mechanism of Action of Citalopram/Escitalopram in Major Depression. *bioRxiv*, 2020.2002.2010.927012, doi:10.1101/2020.02.10.927012 (2020).
82 Awuchi, C. Sugar Alcohols: Chemistry, Production, Health Concerns and Nutritional Importance of Mannitol, Sorbitol, Xylitol, and Erythritol. *International Journal of Advanced Academic Research* **3**, 2488-9849 (2017).
83 M. Ledochowski, B. W. H. B. T. P. D. F. Fructose- and Sorbitol-reduced Diet Improves Mood and Gastrointestinal Disturbances in Fructose Malabsorbers. *Scandinavian Journal of Gastroenterology* **35**, 1048-1052, doi:10.1080/003655200451162 (2000).
84 Janowsky, D. S., el-Yousef, M. K. & Davis, J. M. Acetylcholine and depression. *Psychosom Med* **36**, 248-257, doi:10.1097/00006842-197405000-00008 (1974).
85 Hirsch, M. J. & Wurtman, R. J. Lecithin consumption increases acetylcholine concentrations in rat brain and adrenal gland. *Science* **202**, 223-225, doi:10.1126/science.694529 (1978).
86 Magil, S. G., Zeisel, S. H. & Wurtman, R. J. Effects of ingesting soy or egg lecithins on serum choline, brain choline and brain acetylcholine. *J Nutr* **111**, 166-170, doi:10.1093/jn/111.1.166 (1981).
87 Loffelholz, K., Klein, J. & Koppen, A. Choline, a precursor of acetylcholine and phospholipids in the brain. *Prog Brain Res* **98**, 197-200, doi:10.1016/s0079-6123(08)62399-7 (1993).
88 Puri, P. *et al.* A lipidomic analysis of nonalcoholic fatty liver disease. *Hepatology* **46**, 1081-1090, doi:10.1002/hep.21763 (2007).
89 Callaerts-Vegh, Z., Leo, S., Vermaercke, B., Meert, T. & D'Hooge, R. LPA5 receptor plays a role in pain sensitivity, emotional exploration and reversal learning. *Genes Brain Behav* **11**, 1009-1019, doi:10.1111/j.1601-183X.2012.00840.x (2012).
90 Moreno-Fernandez, R. D. *et al.* maLPA1-null mice as an endophenotype of anxious depression. *Transl Psychiatry* **7**, e1077, doi:10.1038/tp.2017.24 (2017).
91 Saldanha, D., Kumar, N., Ryali, V., Srivastava, K. & Pawar, A. A. Serum Serotonin Abnormality in Depression. *Med J Armed Forces India* **65**, 108-112, doi:10.1016/S0377-1237(09)80120-2 (2009).
92 Wood, K., Harwood, J. & Coppen, A. The effect of antidepressant drugs on plasma kynurenine in depressed patients. *Psychopharmacology (Berl)* **59**, 263-266, doi:10.1007/bf00426632 (1978).
93 Alvarez, J. C. *et al.* Plasma serotonin level after 1 day of fluoxetine treatment: a biological predictor for antidepressant response? *Psychopharmacology (Berl)* **143**, 97-101, doi:10.1007/s002130050924 (1999).
94 Inoshita, M. *et al.* Elevated peripheral blood glutamate levels in major depressive disorder. *Neuropsychiatr Dis Treat* **14**, 945-953, doi:10.2147/NDT.S159855 (2018).
95 Ogawa, S. *et al.* Plasma amino acid profile in major depressive disorder: Analyses in two independent case-control sample sets. *J Psychiatr Res* **96**, 23-32, doi:10.1016/j.jpsychires.2017.09.014 (2018).
96 Ali-Sisto, T. *et al.* Global arginine bioavailability ratio is decreased in patients with major depressive disorder. *J Affect Disord* **229**, 145-151, doi:10.1016/j.jad.2017.12.030 (2018).
97 Rotroff, D. M. *et al.* Metabolomic signatures of drug response phenotypes for ketamine and esketamine in subjects with refractory major depressive disorder: new mechanistic insights for rapid acting antidepressants. *Transl Psychiatry* **6**, e894, doi:10.1038/tp.2016.145 (2016).

98 Guerreiro, J. R. *et al.* Argininosuccinate synthetase is a functional target for a snake venom anti-hypertensive peptide: role in arginine and nitric oxide production. *J Biol Chem* **284**, 20022-20033, doi:10.1074/jbc.M109.021089 (2009).
99 Bahri, S. *et al.* Citrulline: from metabolism to therapeutic use. *Nutrition* **29**, 479-484, doi:10.1016/j.nut.2012.07.002 (2013).
100 Fewell, Z., Davey Smith, G. & Sterne, J. A. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. *Am. J. Epidemiol.* **166**, 646-655, doi:10.1093/aje/kwm165 (2007).
101 Nath, A. P. *et al.* An interaction map of circulating metabolites, immune gene networks, and their genetic regulation. *Genome Biol.* **18**, 146, doi:10.1186/s13059-017-1279-y (2017).

Supplementary Information

Supplementary Figure 1. Change in hippuric acid from baseline to week 12 after treatment with CBT, duloxetine, and escitalopram in the PReDICT study.

Supplementary Figure 2. Uptake and secretion of the nine depression-related metabolites that are consumed or secreted by at least one AGORA strain. Each figure shows the metabolite-secreting strains or the metabolite-consuming strains by genus.

Supplementary Table 1. GWAS results that were used to perform the MR analysis.

Supplementary Table 2. List of SNPs that were used as instruments in the MR analysis (p-value < 10^{-6}).

Supplementary Table 3. Results of the association analyses of all metabolites with depression. This Table includes results for model 1, 2, and model 3. The order of the direction column: RS, SHIP-trend, KORA, EPIC-Norfolk B2, EPIC-Norfolk B3, NEO. Significant results are highlighted (FDR corrected p-value < 0.05).

Supplementary Table 4. Results of the sex-stratified association analysis of all metabolites with depression for model 2. The order of the direction column: RS, SHIP-trend, KORA, EPIC-Norfolk B2, EPIC-Norfolk B3, NEO. Significant results are highlighted (FDR corrected p-value < 0.05).

Supplementary Table 5. Results of the sensitivity analysis, excluding the cohorts which measured metabolites on older Metabolon platforms, including the SHIP-trend and KORA studies. Results are shown for model 2. The order of the direction column: RS, EPIC-Norfolk B2, EPIC-Norfolk B3, NEO. Significant results are highlighted (FDR corrected p-value < 0.05).

Supplementary Table 6. Results of the association analysis of white matter hyperintensity volume with vitamin supplements.

Supplementary Table 7. Results of the MR analysis with major depression as exposure and metabolites as outcome.

Supplementary Table 8. Results of the MR analysis with metabolites as exposure and major depression as outcome.

Supplementary Table 9. Results of the human and gut microbiome metabolic network analysis.

Supplementary Table 10. Depression-associated metabolites included in AGORA2 that could be consumed by at least one AGORA2 strain.

Supplementary Table 11. Depression-associated metabolites included in AGORA2 that could be secreted by at least one AGORA2 strain.

CHAPTER 3.2

Genetic variants in *RBFOX3* are associated with sleep latency

Najaf Amin, Karla V. Allebrandt, Ashley van der Spek, Bertram Müller-Myhsok, Karin Hek, Maris Teder-Laving, Caroline Hayward, Tõnu Esko, Josine G. van Mill, Hamdi Mbarek, Nathaniel F. Watson, Scott A. Melville, Fabiola M. Del Greco, Enda M. Byrne, Edwin Oole, Ivana Kolcic, Ting-hsu Chen, Daniel S. Evans, Josef Coresh, Nicole Vogelzangs, Juha Karjalainen, Gonneke Willemsen, Sina A. Gharib, Lina Zgaga, Evelin Mihailov, Katie L. Stone, Harry Campbell, Rutger W.W. Brouwer, Ayse Demirkan, Aaron Isaacs, Zoran Dogas, Kristin D. Marciante, Susan Campbell, Fran Borovecki, Annemarie I. Luik, Man Li, Jouke Jan Hottenga, Jennifer E. Huffman, Mirjam C.G.N. van den Hout, Steven R. Cummings, Yurii S. Aulchenko, Philip R. Gehrman, Andre G. Uitterlinden, Heinz-Erich Wichmann, Martina Müller-Nurasyid, Rudolf S.N. Fehrmann, Grant W. Montgomery, Albert Hofman, Wen Hong Linda Kao, Ben A. Oostra, Alan F. Wright, Jacqueline M. Vink, James F. Wilson, Peter P. Pramstaller, Andrew A. Hicks, Ozren Polasek, Naresh M. Punjabi, Susan Redline, Bruce M. Psaty, Andrew C. Heath, Martha Merrow, Gregory J. Tranah, Daniel J. Gottlieb, Dorret I. Boomsma, Nicholas G. Martin, Igor Rudan, Henning Tiemeier, Wilfred F.J. van IJcken, Brenda W. Penninx, Andres Metspalu, Thomas Meitinger, Lude Franke, Till Roenneberg*, Cornelia M. van Duijn*

* These authors contributed equally to this work (shared last)

This chapter was published in *European Journal of Human Genetics.* 2016 Oct;24(10):1488-95.

The supplemental information for this paper is available online at https://doi.org/10.1038/ejhg.2016.31.

Abstract

Time to fall asleep (sleep latency) is a major determinant of sleep quality. Chronic, long sleep latency is a major characteristic of sleep-onset insomnia and/or delayed sleep phase syndrome. In this study we aimed to discover common polymorphisms that contribute to the genetics of sleep latency. We performed a meta-analysis of genome-wide association studies (GWAS) including 2,572,737 single nucleotide polymorphisms (SNPs) established in seven European cohorts including 4242 individuals. We found a cluster of three highly correlated variants (rs9900428, rs9907432 and rs7211029) in the RNA binding protein fox-1 homolog 3 gene (*RBFOX3*) associated with sleep latency (*p*-values = $5.77x10^{-08}$, $6.59x10^{-08}$ and $9.17x10^{-08}$). These SNPs were replicated in up to 12 independent populations including 30,377 individuals (*p*-values = $1.5x10^{-02}$, $7.0x10^{-03}$ and $2.5x10^{-03}$; combined meta-analysis *p*-values = $5.5x10^{-07}$, $5.4x10^{-07}$ and $1.0x10^{-07}$). A functional prediction of *RBFOX3* based on co-expression with other genes shows that this gene is predominantly expressed in brain (*p*-value = $1.4x10^{-316}$) and the central nervous system (*p*-value = $7.5x10^{-321}$). The predicted function of *RBFOX3* based on co-expression analysis with other genes shows that this gene is significantly involved in the release cycle of neurotransmitters including gamma-aminobutyric acid and various monoamines (*p*-values < $2.9x10^{-11}$) that are crucial in triggering the onset of sleep. To conclude, in this first large-scale GWAS of sleep latency we report a novel association of variants in *RBFOX3* gene. Further, a functional prediction of *RBFOX3* supports the involvement of *RBFOX3* with sleep latency.

Introduction

Sleep occurs during defined periods within the 24 h cycle[1]. Its timing and duration is considered to be regulated by two processes; (i) a circadian process (i.e., being awake during the day and asleep during night) and (ii) a homeostatic process that represents the sleep pressure accumulating during wakefulness, i.e., the longer one is awake, the greater sleep pressure and the greater will be the duration of sleep when permitted[2]. Sleep latency, i.e., the duration it takes to fall asleep, is a measure of sleep quality computed as the time interval between 'lights out' until the onset of sleep. Photic information from the retina is projected to the suprachiasmatic nucleus (SCN) via the retino-hypothalamic tract. In humans, the SCN is considered the pacemaker for the timing of daily sleep-wake behavior and consists of ~10,000 neurons located on both sides of the midline above the optic chiasma, ~3 cm behind the eyes[3,4]. At night, the SCN signals the release of melatonin – a hormone produced by the pineal gland that promotes sleep. Sleep latency may be assessed using self-reports, actigraphy or with polysomnography. Normal sleepers are known to provide estimates of sleep latency that correlate well with polysomnographic measures[5,6]. Previous studies show that sleep latency is largely independent of an individual's chronotype[7].

Inter-individual differences in sleep latency are caused by both genetic and non-genetic factors. The latter include gender[8,9], age[8], consumption of stimulants[10], dietary intake[11], sedentary life[12] and illnesses, such as depression[13]. Persistent increased sleep latency is a major characteristic of delayed sleep phase syndrome[14] and/or sleep-onset insomnia[15-18]. Prolonged sleep latency may shorten sleep duration and may lead to a wide range of problems including irritability, cognitive impairment, depression and loss of productivity as well as accident rates due to sleepiness. Increased sleep latency has also been associated with poor academic performance in children and adolescents[19]. Although very long sleep latency can lead to sleep deprivation, very short sleep latency can reflect sleep deprivation (i.e., 'sleep debt' due to insufficient sleep normally shortens sleep latency). Unusually short sleep latency also may indicate disorders of excessive sleepiness (e.g., narcolepsy). When giving individuals the opportunity to sleep during the day, sleep latency is used as an objective measure of daytime sleepiness in the diagnosis of sleep disorders.

Heritability of sleep latency is estimated to be ~17-44%[20-22]. Although much is known about the environmental factors that can prolong sleep latency, we know little about the genetic influences on sleep latency. In this study, we performed a meta-analysis of unpublished genome-wide association studies (GWAS) on sleep latency in order to elucidate genetic associations with this trait.

Materials and Methods

Study populations (Stage 1/ Discovery cohorts)

We meta-analyzed data from seven GWAS comprising of 4242 subjects with European ancestry (**Supplementary Tables S1** and **S2**). The participating cohorts in the gene discovery phase included the Erasmus Rucphen Family (ERF), Estonian Genome Center (EGP/EGCUT), CROATIA-Korcula, the Micro-isolates in South Tyrol Study (MICROS), Cooperative health research in the Region of Augsburg (KORA), the Netherlands Study of Depression and Anxiety (NESDA) and the Orkney Complex Disease Study (ORCADES) (**Supplementary Tables S1** and **S2**). A detailed description of the included studies is provided in the **Supplementary text**. All studies in the discovery cohort used the Munich Chronotype Questionnaire (MCTQ)[23] to assess sleep latency. Subjects were asked to report how long they take to fall asleep on free and workdays. Free days sleep latency was used in the analyses of those cohorts, where a person's sleep pattern is not influenced by professional duties[24] (**Supplementary Figure S1**). The question from the MCTQ used to assess sleep latency (in minutes): 'I need … minutes to fall asleep'. Sleep duration was calculated by subtracting sleep onset from sleep end. Mid-sleep was calculated as the midpoint between sleep onset and waking on free days[25]. The quality control was centralized and the inclusion criteria were: (i) no use of an alarm clock on free days; (ii) no shift-work during the last three months; and (iii) no use of sleep medication (benzodiazepines and other pharmacological agents that influence sleep; see **Supplementary Table S3**). Informed consent was obtained from all study participants and an appropriate local committee approved the study protocols.

Study populations (Stage 2/ Replication cohorts)

The replication stage included 12 independent cohorts (see **Supplementary text**; **Supplementary Figure S1**). The descriptive statistics are provided in **Supplementary Table S1**. Inclusion criteria were: (1) sleep assessment with any available tool and (2) No use of sleep medication (**Supplementary Figure S1**). Five replication cohorts (MrOS, RS-I, RS-II, RS-III and SOF) used the Pittsburg Sleep Quality Index to assess sleep (**Supplementary Table S2**), which uses only a single measure for sleep assessment and does not distinguish between sleep on free and working days.

Genotyping and imputation

Both discovery and replication cohorts were genotyped on a variety of platforms (Affymetrix 250 K, Illumina 318 K, Illumina 370 K, Illumina 610 k; Perlegen 600 K; Affymetrix 1000 K). Quality control was done in each group separately. The overall criteria were to exclude individuals with low call rate, excess heterozygosity and gender mismatch, and exclude variants that were out of Hardy-Weinberg equilibrium, had low minor allele frequency (MAF) or low call rate (**Supplementary Table S2**). In EGCUT1 study where the genome-wide data was not available, the two most significant single nucleotide polymorphisms (SNPs) for which a TaqMan assay was available were genotyped (**Supplementary Table S4**). Imputations of non-genotyped SNPs in the discovery cohorts were carried out within each study using

either MACH[26,27] or IMPUTE[28,29], and HapMap CEU v21a or v22 as reference (**Supplementary Table S2**). Genetic imputations in the replication cohorts were performed using MACH, IMPUTE, minimac or BimBam (**Supplementary Table S2**). Of the three SNPs rs9907432 was genotyped in most replication cohorts, whereas rs9900428 was imputed in all replication cohorts (**Supplementary Table S4**). The data are available in the GWAS Central database, under the accession number HGVST1836 (http://www.gwascentral.org/study/HGVST1836).

Methylation and expression

The Infinium Human Methylation 450 array (Illumina, San Diego, CA, USA) was used to quantify genome-wide DNA methylation for 748 samples from the Rotterdam Study (RS-III) covering 485,577 CpG sites in the genome. Bisulfite conversion followed by amplification, hybridization and imaging were performed according to standard protocols. Illumina GenomeStudio software was used to estimate β score from intensity. Quality-control (QC) steps included removal of signal probes with a detection of *p*-value > 0.01; and exclusion of probes with low intensity. Samples with a call rate < 99% were removed from the dataset. SWAN package for R software was used to normalize remaining CpG sites and correct for batch effects.

RNA from the same samples was obtained (PAXgene) and hybridized to Illumina HumanHT-12 arrays. Raw probe intensity was obtained using BeadStudio. Gene expression data were quantile-normalized to the median distribution, and subsequently log2 transformed. The probe and sample means were centered to zero. Probes that had a detection *p*-value < 0.05 in > 10% of the samples were removed from the analysis. The final analysis included 21,328 probes, which were significantly expressed in blood.

Statistical analysis

Heritability analysis

Heritability analysis of sleep latency was performed in the family-based ERF cohort using Sequential Oligogenic Linkage Analysis Routines (SOLAR)[30]. SOLAR uses likelihood ratio tests to evaluate heritability by comparing a purely polygenic model with a sporadic model in the case of testing heritability. The ERF cohort forms one large family, which consists of > 23,000 members spanning over 23 generations. Since this uniquely large family is too large to be analyzed with SOLAR, we cut it into smaller pedigrees (three to five generations) using the Pedcut software[31] for the heritability analysis. A natural log transformation was applied to the trait before estimating the heritability. The heritability was estimated with the 'polygenic screen' option and using age and sex as covariates in the model. The estimate was then compared with the heritability estimate derived from the polygenic analysis of GenABEL that uses genome-wide genotype data instead of the pedigree to estimate heritability[32].

Genome-wide association analysis

Individual GWAS was performed using linear regression (under additive model), natural log of sleep latency as the dependent variable, SNP allele dosage as predictor and age and sex

as covariates. CROATIA-Korcula and ORCADES additionally used first three principal components as covariates in the association model. The association analyses were conducted in ProbABEL[32-34] or SNPTEST[35]. For cohorts with related individuals (ERF, MICROS, CROATIA-Korcula, ORCADES), a linear mixed model in ProbABEL using the 'mmscore' option was used to account for familial relationships. The mmscore option performs the score test that uses the inverse variance-covariance object estimated from the genetic data and returned from the 'polygenic' function of GenABEL[32] to correct for familial relationships. This is a slightly modified FASTA method developed by Abecasis et al.[36].

As all three SNPs show no heterogeneity across the cohorts (*p*-values > 0.7) (**Table 1**), a fixed effects meta-analysis was conducted using the inverse variance weighted method as implemented in METAL (http://www.sph.umich.edu/csg/abecasis/metal/). All SNPs that had a MAF < 0.01 and low imputation quality (Rsq/proper_info < 0.3) were dropped from the meta-analysis. Genomic control correction was also applied to all cohorts prior to the meta-analysis.

For the SNPs that approached genome-wide significance (*p*-value < $5x10^{-08}$) we performed replication in up to 12 independent cohorts using the same model as in the initial GWAS. In the replication cohorts, among the family-based studies, ERF_ext used SOLAR, FHS used LMEKIN package of R software (http://www.r-project.org/), CROATIA-Split used the 'mmscore' option in ProbABEL while QIMR used MERLIN to account for family structure. Other population-based cohorts used SPSS, PLINK (http://pngu.mgh.harvard.edu/purcell/plink/)[37] or R software to perform association analysis (**Supplementary Table S2**). Meta-analysis was performed using square-root of the sample size as weights[38].

Methylome-wide association analysis
Top SNPs from the meta-analysis of GWAS were tested for association with genome-wide methylation (CpG) sites by performing linear regression analyses of methylation sites on each of the three SNPs while adjusting for age, sex, technical covariates including batch effects and blood cell counts (granulocytes, lymphocytes, monocytes, erythrocytes and platelets) in R software.

Gene functional prediction and network analysis
Prediction of gene function can be conducted using a guilt-by-association approach: e.g., if there are 100 genes that are known to be involved in apoptosis, identification of a gene that is strongly co-expressed with these 100 genes suggests that this gene is likely to be involved in apoptosis as well. As such co-expression data can be used to predict likely functions for genes. However, important to realize is that some phenomena exert very strong transcriptomic effects and therefore will overshadow more subtle effects. In order to be able to identify such subtle relationships as well, we conducted a principal component analysis (PCA) on an unprecedented scale[39]. We collected gene expression data for three different species (homo sapiens, mus musculus and rattus norvegicus) from the Gene

Expression Omnibus. We confined analyses to four different Affymetrix expression platforms (Affymetrix Human Genome U133A Array, Affymetrix Human Genome U133 Plus 2.0 Array, Affymetrix Mouse Genome 430 2.0 Array and Affymetrix Rat Genome 230 2.0 Array). For each of these platforms we downloaded the raw CEL files (20,108, 43,278, 18,639 and 6124 arrays, respectively), and used RMA for normalization. We could run RMA on all samples at once for the 20,108 Human Genome U133A Array, 18,639 Mouse Genome 430 2.0 Array and 6123 Rat Genome 230 2.0 Array. For the 43,278 Human Genome U133 Plus 2.0 Array samples we ran RMA in eight batches due to its size, by randomly assigning the samples to one of these batches. We subsequently conducted QC on the data. We first removed duplicate samples, and then conducted a PCA on the sample correlation matrix. The first principal component (PCqc) on such a matrix describes nearly always a constant pattern (dominating the data), which explains ~80-90% of the total variance[40,41]. This pattern can be regarded as probe-specific variance, independent of the biological sample hybridized to the array. The correlation of each individual microarray with this PCqc can be used to detect outliers, as arrays of lesser quality will have a lower correlation with the PCqc. We removed samples that had a correlation $R < 0.75$. After QC in total 77,840 different samples remained for downstream analysis (54,736 human samples, 17,081 mouse samples, 6023 rat samples). Although this QCed data set can be well used for the aforementioned guilt-by-association co-expression analysis, we reasoned that the presence of profound effects on many genes will make it difficult to identify the more subtle relationships that exist between genes. Therefore, we conducted a PCA on the probe correlation matrix, resulting in the identification of in total 2206 robustly estimated principal components (377 for Human Genome U133A, 777 for Human Genome U133 Plus 2.0, 677 for Mouse Genome 430 2.0 and 375 for Rat Genome 230 2.0) by requiring a Cronbach's alpha > 0.70 for each individual principal component. Jointly these components explain between 79% and 90% of the variance in the data per Affymetrix expression platform, and many of these are well conserved across the three species.

Subsequent Gene Set Enrichment Analysis revealed that each of these 2206 components are significantly enriched (false discovery rate < 0.05) for at least one GO term, KEGG, BioCarta or Reactome pathway, indicating that these components are describing biologically relevant but often diverse phenomena. While per species the very first components describe profound effects on expression (i.e., many enriched pathways and GO terms), the other components are potentially equally biologically relevant, as each of the components describe certain biological phenomena. We therefore used the individual components and integrated the different platforms and species by collapsing the probe identifiers to human Ensembl genes and used orthology information from Ensembl for the mouse and rat platform, resulting in a harmonized matrix of 19,997 unique Ensembl genes x 2206 principal components.

We subsequently predicted the most likely Gene Ontology (GO) biological process using the following strategy: (i) we first ascertained each individual GO term and assessed per PC

whether the genes that were explicitly annotated with this GO term showed a significant difference from the genes that were not annotated with this GO term using a *t*-test; (ii) we converted the resulting *p*-value into an 'enrichment' Z-Score (to ensure normality); and (iii) we subsequently investigated RNA-binding protein fox-1 homolog 3 gene (*RBFOX3*) and correlated the 2206 PC eigenvector coefficients of *RBFOX3* with each GO term by taking the 2206 'enrichment' Z-Scores as the expression profile for that GO term. A significant positive correlation means *RBFOX3* has an expression profile that is comparable to the GO term. We have visualized this method at www.genenetwork.nl/genenetwork (click on 'Method'). In order to correct for multiple testing, we permuted Ensembl gene identifiers: using permuted data we redid the 'enrichment' Z-score calculation and investigated how strong *RBFOX3* correlated with permuted pathway. We repeated this analysis 100 times, allowing us to determine which of these predictions were significant (controlling to the false discovery rate of 0.05). We used the same procedure to predict in which BioCarta and Reactome pathways *RBFOX3* is involved.

Results

A basic description of the study populations is given in **Supplementary Table S1**. The heritability of sleep latency using the pedigree data of the ERF study was estimated to be 0.18 (SE = 0.112, *p*-value = 0.05) (see Materials and methods section), which is consistent with earlier findings of heritability for this trait[21] and also comparable to the heritability estimate (h^2 = 0.20) derived from the genome-wide genotype data (see Materials and methods section for details) of the same cohort. The quantile-quantile plot of the meta-analysis shows no inflation of the chi-square statistic as the genomic control inflation factor (λ) is 1.01 (**Supplementary Figure S2**). The meta-analysis of GWAS produced a cluster of three borderline genome-wide significant SNPs (**Table 1**; **Figure 1** and **Supplementary Figure S3**) on chromosome 17q25. The highest-ranking SNP rs9900428 hg18. chr17:g.74651323G>A) (*p*-value = 5.7×10^{-08}) had a MAF of 0.20-0.33, with the imputation quality ranging from 0.85 to 0.99 across all cohorts. The effect size (β = -0.094) and the direction of the risk were consistent across all cohorts (**Table 1**). The other two SNPs rs9907432 (hg18.chr17:g.74651967:G>A) (*p*-value = 6.5×10^{-08}) and rs7211029 (hg18. chr17:g.74652903C>T) (*p*-value = 9.1×10^{-08}) were in linkage disequilibrium (LD; r^2 = 0.95) with rs9900428 (**Figure 1**, **Supplementary Figure S4**). **Figure 2** shows the mean sleep latency per genotype category for all the discovery cohorts for the three SNPs unadjusted for age and gender. Compared with the homozygote carriers of the reference allele (pooled average sleep latency = 16.4 min), the homozygote carriers (pooled average sleep latency = 13.5 min) of the minor allele are estimated to have a sleep latency of on average approximately 3 min less (a reduction of 18%) and the heterozygote carriers (pooled average sleep latency = 15.5 min) ~1 min less (a reduction of ~6%) (**Figure 2**).

Table 1. Top SNPs from the genome-wide association analysis for sleep latency.

SNP	Position (B36)	Allele	Gene	Maf	Chrom	Cohort	EGP	ERF	KORA	KORCULA	MICROS	NESDA	ORKNEY	Meta-analysis
						N	933	740	548	610	693	540	206	4270
rs9900428	74651323	G>A	*RBFOX3*	0.20-0.33	17	*β*	-0.106	-0.085	-0.158	-0.133	-0.082	-0.03	-0.122	-0.094
						SE	0.051	0.024	0.057	0.067	0.055	0.064	0.104	0.017
						p-value	0.037	0.00036	0.0057	0.045	0.137	0.632	0.239	$5.77x10^{-08}$
						Rsq	0.97	0.90	0.90	0.97	0.96	0.85	0.99	
						P_{HET}								0.823
rs9907432	74651967	G>A	*RBFOX3*	0.20-0.34	17	β	-0.106	-0.083	-0.153	-0.127	-0.084	-0.0248	-0.1216	-0.092
						SE	0.0501	0.0234	0.0551	0.066	0.0543	0.0616	0.1031	0.017
						p-value	0.0354	0.00038	0.00568	0.053	0.1206	0.6874	0.2382	$6.59x10^{-08}$
						Rsq	1.0	0.94	0.95	1.0	0.99	0.90	1.0	
						P_{HET}								0.814
rs7211029	74652903	C>T	*RBFOX3*	0.21-0.34	17	β	-0.107	-0.083	-0.1499	-0.127	-0.083	-0.0119	-0.1215	-0.091
						SE	0.050	0.0234	0.0538	0.066	0.0544	0.0596	0.1034	0.017
						p-value	0.035	0.00038	0.0054	0.054	0.1249	0.8416	0.2397	$9.17x10^{-08}$
						Rsq	0.97	0.93	1.0	0.98	0.98	0.96	0.99	
						P_{HET}								0.739

Abbreviations: β = effect of the genetic variant, SE = standard error of the effect estimate, Rsq = imputation quality, P_{HET} = *p*-value for heterogeneity.

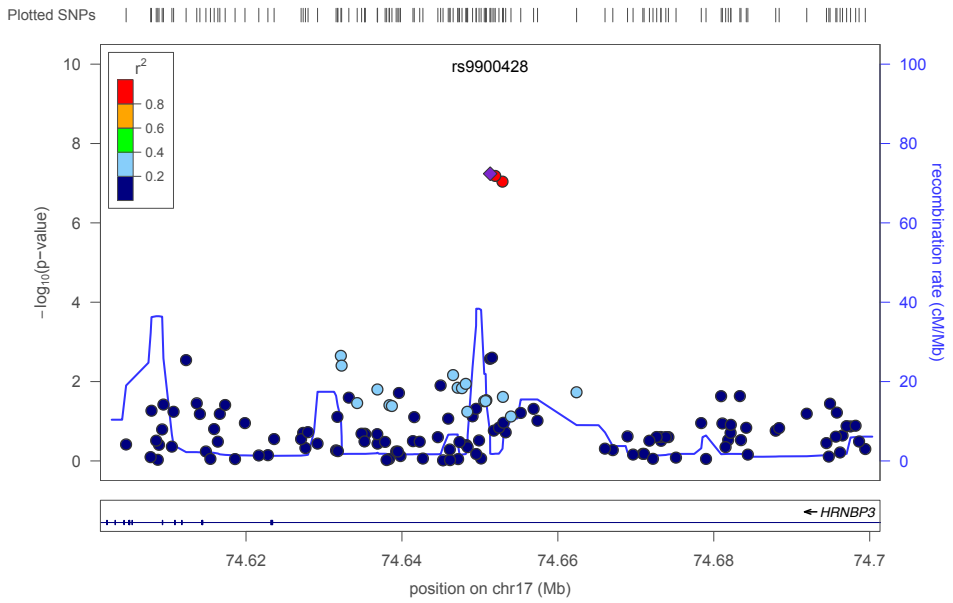


Figure 1. Regional association plot of sleep latency for the region 17q25. The plot was constructed using Locus Zoom (http://csg.sph.umich.edu/locuszoom/). The most significant SNP is depicted as a diamond and other SNPs in the region are depicted by circles. Various colors represent the extent of linkage disequilibrium with most significant SNP. The x-axis gives the position in mega bases and the y-axis shows the negative logarithm of the *p*-values from the meta-analysis.

We attempted to replicate these three SNPs in up to 12 independent cohorts (**Supplementary Tables S2** and **S4**). SNP rs9900428 was available *in-silico* in 11 cohorts, whereas rs9907432 and rs7211029 were available in all 12 cohorts. The replication analyses showed consistency in the direction of the effect across most replication cohorts (**Figure 3**). The meta-analysis of the replication cohorts yielded significant evidence of association of rs9900428 (*p*-value = 1.5×10^{-02}), rs9907432 (*p*-value = 7.1×10^{-03}) and rs7211029 (*p*-value = 2.5×10^{-03}) with sleep latency (**Table 2**). The three SNPs are intronic to the *RBFOX3* (also known as *HRNBP3*; **Figure 1**) and lie in the region with active regulatory elements (ENCODE) containing the H3K27ac mark and methylation marks. H3K27ac is an important enhancer mark that can distinguish between active and poised enhancer elements. Such enhancer elements are known to affect the expression of proximal genes and cluster near the genes they regulate[42]. We investigated the three SNPs further using the methylome data of RS. The three SNPs were methylome-wide significantly associated (*p*-value = 8.1×10^{-9}, FDR = 0.004) with the CpG site: cg16185152 in the *RBFOX3* gene. We attempted to investigate the effect of methylation on gene expression, there was, however, only one probe for the *RBFOX3* gene in the HumanHT-12_V4_Illumina 450K RNA expression array, which was removed in the quality control.

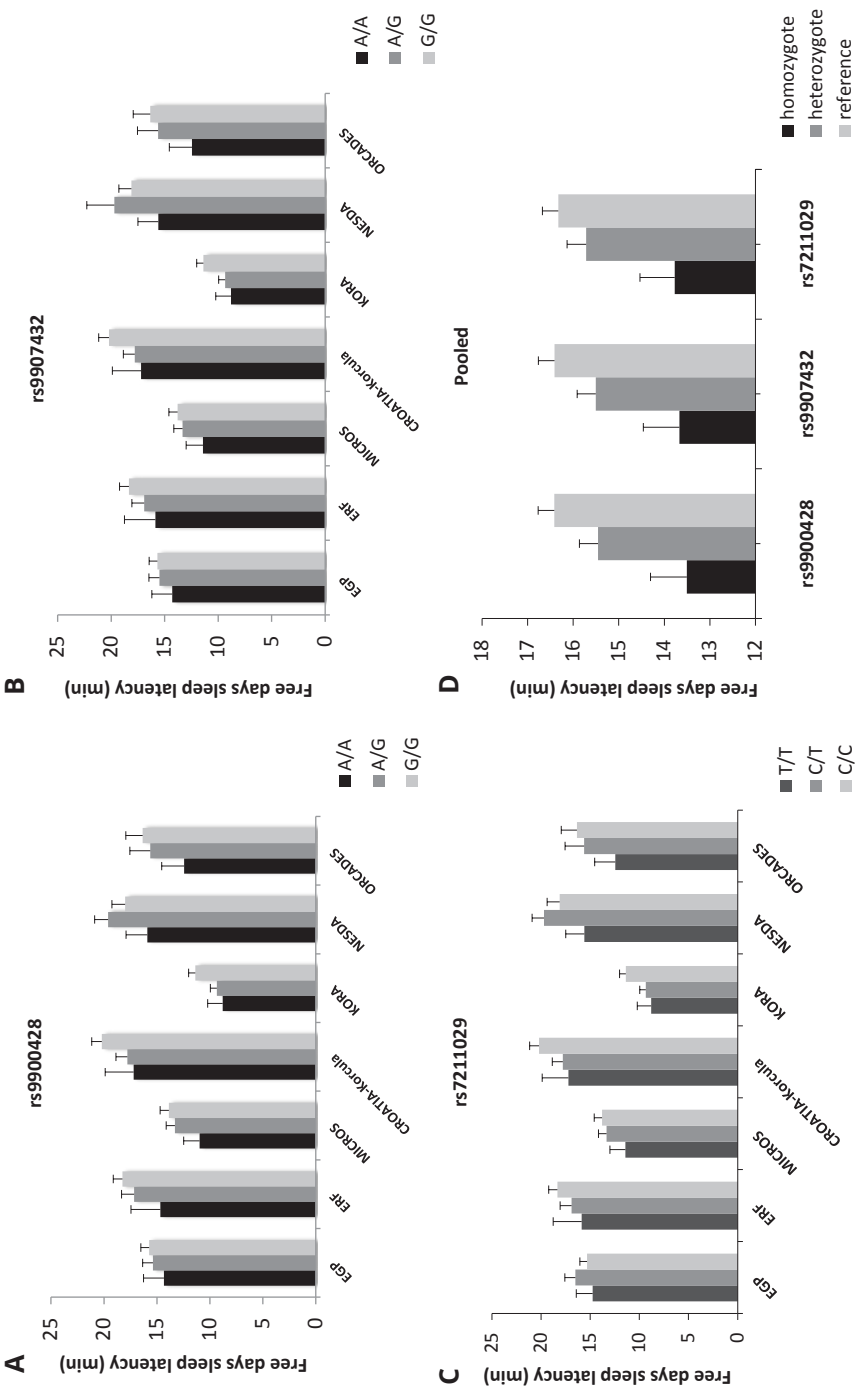


Figure 2. Means per genotype category for all the discovery cohorts for the three most significant SNPs. Black bars represent the mean sleep latency for the homozygous carriers of the effect allele. Y-axis depicts the sleep latency time in minutes. Panels A, B and C represent the effects of the SNPs in individual cohorts. Panel D represents the pooled average sleep latency per genotype category across all cohorts for the three SNPs (x-axis). Data used to generate this figure were not adjusted for age and sex.

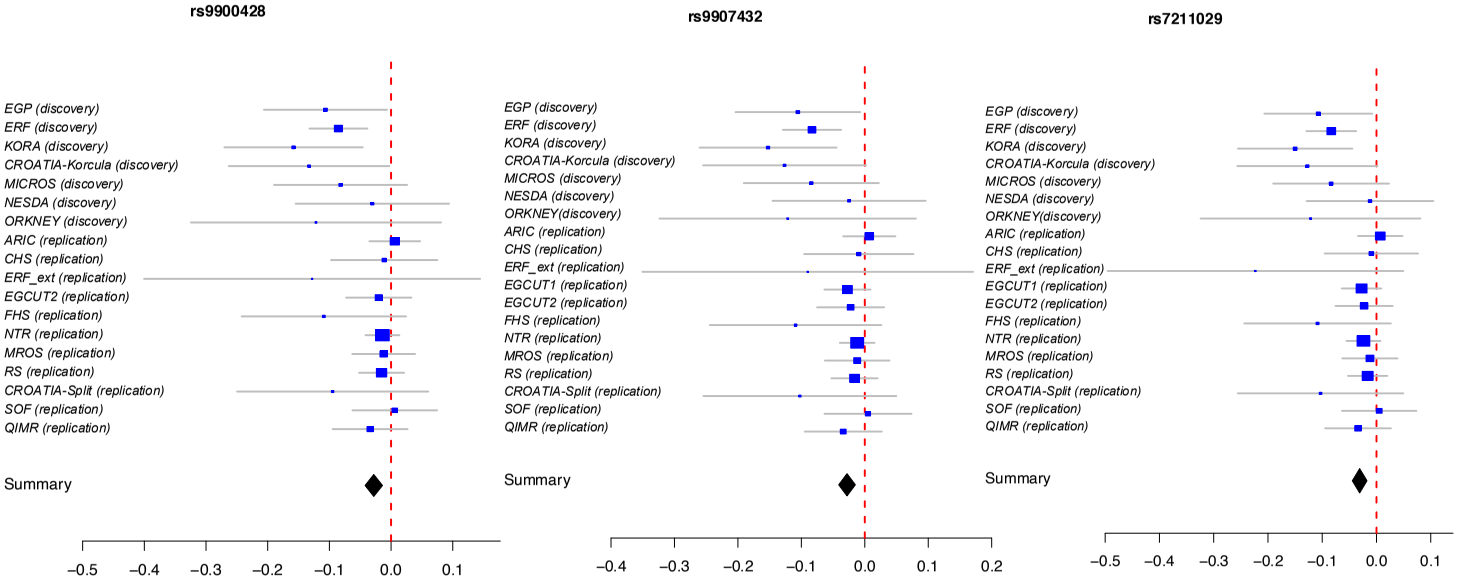


Figure 3. Forest plot for the three SNPs at chromosome 17q25. On the left, the populations including the populations in the GWAS and the replication phase. The boxes represent the precision and horizontal lines representing the confidence intervals. Pooled estimate is the effect estimate from the meta-analysis of all cohorts. The horizontal axis represents the scale of effects.

Table 2. Results of the replication of the top SNPs.

SNP	cohort	ARIC	CHS	EGCUT1	EGCUT2	ERF_ext	FHS	MROS	NTR	QIMR	RS	SOF	SPLIT	Meta-analysis Replication
	N	**3583**	**1533**	**5925**	**3540**	**143**	**2192**	**1849**	**1795**	**2280**	**5641**	**1480**	**416**	**30377**
rs9900428	*β*	0.006	-0.011	NA	-0.020	-0.128	-0.109	-0.012	-0.014	-0.034	-0.015	0.006	-0.095	-0.026
	SE	0.021	0.044	NA	0.027	0.139	0.068	0.026	0.014	0.031	0.019	0.035	0.079	0.010
	p-value	0.775	0.802	NA	0.461	0.357	0.109	0.644	0.317	0.273	0.408	0.864	0.231	0.015
	P_{HET}													0.919
rs9907432	*β*	0.007	-0.009	-0.027	-0.022	-0.090	-0.109	-0.012	-0.012	-0.034	-0.016	0.005	-0.102	-0.026
	SE	0.021	0.044	0.019	0.027	0.133	0.069	0.026	0.014	0.031	0.0186	0.035	0.078	0.0097
	p-value	0.739	0.831	0.140	0.408	0.499	0.114	0.644	0.391	0.273	0.384	0.886	0.187	0.007
	P_{HET}													0.949
rs7211029	*β*	0.007	-0.009	-0.028	-0.023	-0.223	-0.109	-0.012	-0.024	-0.034	-0.016	0.005	-0.103	-0.030
	SE	0.021	0.044	0.019	0.027	0.139	0.069	0.026	0.016	0.031	0.019	0.035	0.078	0.0098
	p-value	0.739	0.831	0.142	0.398	0.109	0.116	0.644	0.133	0.272	0.384	0.886	0.186	0.0025
	P_{HET}													0.792

Abbreviations: β = effect of the genetic variant, SE = standard error of the effect estimate, P_{HET} = *p*-value for heterogeneity.

A gene network analysis of *RBFOX3* using gene network tool (http://genenetwork.nl:8080/GeneNetwork/) shows strongest co-expression with the hippocalcin gene *HPCA* followed by *SNCB, CABP1, JPH3, CPLX2, GABRA6, GABRD, NRXN3, RBFOX1, RTN4R, CNTN2* and *WSCD2* among others (**Supplementary Figure S5**). A functional prediction of the gene showed involvement in the biological processes of synaptic functioning, membrane depolarization, gamma-aminobutyric acid (GABA) signaling and nervous system development (**Supplementary Tables S5** and **S6**) and in dopamine, serotonin and glutamate neurotransmitters release cycle (**Supplementary Table S7**). *RBFOX3* is expressed most strongly in various parts of the brain (brain, p-value = 1.4×10^{-316}; central nervous system, p-value = 7.6×10^{-321}; cerebral cortex, p-value = 1.0×10^{-174}) including the hypothalamus (p-value = 9.7×10^{-10}, AUC = 0.96) (**Supplementary Table S8**), the locale of the SCN and a central part of the circadian clock. None of the three SNPs showed strong association with sleep duration or mid-sleep, which is an indicator of the chronotype of an individual, **Supplementary Table S9**.

Discussion

In this first large-scale GWAS of sleep latency, we report the association of novel variants located in the gene *RBFOX3* with sleep latency. Our gene discovery phase consisted of a sample of 4242 individuals from seven European populations, where sleep latency was assessed according to a common protocol. With this sample we discovered a cluster of three borderline genome-wide significant SNPs that were intronic to the gene *RBFOX3*. The replication of the three SNPs in up to 30,377 individuals from 12 independent cohorts showed significant association of the three variants and consistency in the direction of the effect estimates across most cohorts. A functional prediction of *RBFOX3* based on a gene network analysis suggests significant involvement in the release cycle of neurotransmitters including GABA and various monoamines that are core to the human circadian clock, thus supporting the involvement of *RBFOX3* with sleep latency.

The *RBFOX3* gene (also called *Fox-3, Hrnbp3, Neun*) is located on the long arm of chromosome 17 (17q25). It belongs to the Fox-1 family of genes and shows high homology to *RBFOX1* (also called *Fox-1, A2bp1, Hrnbp1*) and *RBFOX2* (also called *Fox-2, Rbm9, Hrnbp2*). *RBFOX3* is a relatively new member of this family and was recently identified to code for neuronal nuclei (NeuN) protein[43]. The Fox proteins are a highly conserved family of tissue-specific splicing regulators[44]. Although *RBFOX1* is expressed in neurons, muscles and heart, and *RBFOX2* in ovary, whole embryo, neurons and muscles[45,46], the expression of *RBFOX3* was detected exclusively in the post-mitotic regions of embryonic mouse brain[44,47]. *RBFOX3* is believed to have a role in neuron-specific alternative splicing[43]. Alternative splicing occurs most frequently in the brain[48,49], presumably to generate large numbers of neuronal cell types and to support their diverse functions[43]. A recent analysis of rodent SCN (the biological clock) anatomy using antibody against NeuN protein shows that *RBFOX3* has a unique distribution which is limited to a particular sector of the SCN[50]. Mutations in the *FOX-1* genes lead to

severe neuro-developmental phenotypes exhibiting mental retardation, epilepsy and autism spectrum disorder[51-54]. Further, a very recent study on patients with developmental delay detected a translocation disrupting the intron 2 of the *RBFOX3* gene[55]. Interestingly, this patient additionally had sleeping difficulties.

Gene-network analysis of *RBFOX3* showed strong co-expression with genes involved in calcium channel activity and GABA signaling. GABA-ergic sleep neurons of the ventro-lateral pre-optic nucleus are activated by the circadian clock and adenosine, which progressively accumulates in the brain during waking. In turn these sleep-active neurons begin to inhibit the wake-active neurons via the neurotransmitter GABA[56]. *GABA-A* receptors are the site of action of a number of important pharmacologic agents including barbiturates, benzodiazepines (sleep inducing drugs) and ethanol[57,58]. Polymorphisms in *GABA-A* receptors have been associated with insomnia[59]. Functional prediction based on gene network analysis shows significant involvement of *RBFOX3* in the release cycle of various neurotransmitters including dopamine, serotonin and glutamate, *GABA-A* receptor activation and Ras-activation upon Ca^{2+} influx through the NMDA receptor. Photic information is communicated to the molecular clockworks by release of glutamate from retino-hypothalamic nerve terminals and stimulation of glutamate receptors on SCN neurons[60]. Glutamate stimulation is followed by intra-cellular increases of cyclic adenosine monophosphate and Ca^{2+} and activation of a Ras-dependent signal cascade in the circadian clockwork. Non-photic signaling to the core subdivision of the SCN is conveyed through two major pathways including GABA-containing neurons derived from the thalamus and serotonin or 5-hydroxytryptamine-containing neurons derived from the midbrain. Serotonergic input to the SCN shifts the timing of the clock[60].

Our study shows strong evidence of association of sleep latency with *RBFOX3*. The fact that the association signal was stronger in the discovery sample compared with the replication sample, even though the replication sample was sixfold larger, may be explained by the differences in phenotyping as well as by different cohort characteristics (e.g. age). Notably, all discovery cohorts were European, whereas the replication cohorts were drawn additionally from North America and Asia; it is possible that socio-cultural factors may have influenced sleep latency or how sleep latency was reported across cohorts. For instance, the phenotyping and quality control for the discovery cohorts was synchronized and centralized; all cohorts were assessed with MCTQ, which, unlike other such instruments, assesses information separately for free days and working days. Our GWAS was based on sleep latency on free days, as sleep latency on workdays is heavily influenced by daily professional activities. Unfortunately, this distinction was not available for most of the replication cohorts. Moreover, exclusions in the GWAS discovery cohorts were based not only on sleep medication use but also on other drugs that are known to influence sleep. Sleep medication or for that matter any medication that has a sleep-inducing effect reduces sleep latency[24], thereby introducing a potential bias on genetic studies of sleep latency. We also removed shift-workers and those using alarm clocks from the analysis in the discovery

phase, which was not done in most replication cohorts. All of these factors likely affected the results in the replication phase. Nevertheless, despite a small effect size, the direction of association signal was consistent in most replication cohorts indicating the robustness of our finding. Furthermore, gene functional prediction and network analysis support the association of *RBFOX3* variants with sleep latency. The predicted functioning of *RBFOX3*, including neurotransmitters' release cycle and GABA-receptor activation strongly implicates a chronobiological explanation. However, further tests of association with various sleep disorders and functional analyses will provide a better insight into the relationship between *RBFOX3* and sleep.

References

1 Krauchi, K. & Wirz-Justice, A. Circadian clues to sleep onset mechanisms. *Neuropsychopharmacology* **25**, S92-96, doi:10.1016/S0893-133X(01)00315-3 (2001).
2 Cirelli, C. The genetic and molecular regulation of sleep: from fruit flies to humans. *Nat Rev Neurosci* **10**, 549-560, doi:10.1038/nrn2683 (2009).
3 Hastings, M. The brain, circadian rhythms, and clock genes. *BMJ* **317**, 1704-1707, doi:10.1136/bmj.317.7174.1704 (1998).
4 Hastings, M. H. Central clocking. *Trends Neurosci* **20**, 459-464 (1997).
5 Baekeland, F. & Hoy, P. Reported vs recorded sleep characteristics. *Arch Gen Psychiatry* **24**, 548-551 (1971).
6 Hoch, C. C. *et al.* Empirical note: self-report versus recorded sleep in healthy seniors. *Psychophysiology* **24**, 293-299 (1987).
7 Lehnkering, H. & Siegmund, R. Influence of chronotype, season, and sex of subject on sleep behavior of young adults. *Chronobiol Int* **24**, 875-888, doi:10.1080/07420520701648259 (2007).
8 Geisler, P. *et al.* The influence of age and sex on sleep latency in the MSLT-30--a normative study. *Sleep* **29**, 687-692 (2006).
9 Krishnan, V. & Collop, N. A. Gender differences in sleep disorders. *Curr Opin Pulm Med* **12**, 383-389 (2006).
10 Bonnet, M. H. & Webb, W. B. The return to sleep. *Biol Psychol* **8**, 225-233 (1979).
11 Afaghi, A., O'Connor, H. & Chow, C. M. High-glycemic-index carbohydrate meals shorten sleep onset. *Am J Clin Nutr* **85**, 426-430 (2007).
12 Nixon, G. M. *et al.* Falling asleep: the determinants of sleep latency. *Arch Dis Child* **94**, 686-689, doi:10.1136/adc.2009.157453 (2009).
13 Gupta, R., Dahiya, S. & Bhatia, M. S. Effect of depression on sleep: Qualitative or quantitative? *Indian J Psychiatry* **51**, 117-121 (2009).
14 Blader, J. C., Koplewicz, H. S., Abikoff, H. & Foley, C. Sleep problems of elementary school children. A community survey. *Arch Pediatr Adolesc Med* **151**, 473-480 (1997).
15 Eddy, M. & Walbroehl, G. S. Insomnia. *Am Fam Physician* **59**, 1911-1916, 1918 (1999).
16 Weitzman, E. D. *et al.* Delayed sleep phase syndrome. A chronobiological disorder with sleep-onset insomnia. *Arch Gen Psychiatry* **38**, 737-746 (1981).
17 Freedman, R. R. EEG power spectra in sleep-onset insomnia. *Electroencephalogr Clin Neurophysiol* **63**, 408-413 (1986).
18 van den Heuvel, C. J. & Lushington, K. Chronobiology and insomnia: pathophysiology and treatment of circadian rhythm sleep disorders. *Expert Rev Neurother* **2**, 249-260, doi:10.1586/14737175.2.2.249 (2002).
19 Touchette, E. *et al.* Associations between sleep duration patterns and behavioral/cognitive functioning at school entry. *Sleep* **30**, 1213-1219 (2007).
20 Heath, A. C., Kendler, K. S., Eaves, L. J. & Martin, N. G. Evidence for genetic influences on sleep disturbance and sleep pattern in twins. *Sleep* **13**, 318-335 (1990).
21 Klei, L. *et al.* Heritability of morningness-eveningness and self-report sleep measures in a family-based sample of 521 hutterites. *Chronobiol Int* **22**, 1041-1054, doi:10.1080/07420520500397959 (2005).
22 Partinen, M., Kaprio, J., Koskenvuo, M., Putkonen, P. & Langinvainio, H. Genetic and environmental determination of human sleep. *Sleep* **6**, 179-185, doi:10.1093/sleep/6.3.179 (1983).
23 Kantermann, T., Juda, M., Merrow, M. & Roenneberg, T. The human circadian clock's seasonal adjustment is disrupted by daylight saving time. *Curr Biol* **17**, 1996-2000, doi:10.1016/j.cub.2007.10.025 (2007).
24 DeMartinis, N. A. & Winokur, A. Effects of psychiatric medications on sleep and sleep disorders. *CNS Neurol Disord Drug Targets* **6**, 17-29 (2007).
25 Wittmann, M., Dinich, J., Merrow, M. & Roenneberg, T. Social jetlag: misalignment of biological and social time. *Chronobiol Int* **23**, 497-509, doi:10.1080/07420520500545979 (2006).
26 Li, Y. & Abecasis, G. R. Mach 1.0: Rapid Haplotype Reconstruction and Missing Genotype Inference. *Am J Hum Genet* **S79**, 2290 (2006).
27 Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype imputation. *Annu Rev Genomics Hum Genet* **10**, 387-406 (2009).
28 Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529, doi:10.1371/journal.pgen.1000529 (2009).

29 Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906-913, doi:10.1038/ng2088 (2007).

30 Almasy, L. & Blangero, J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* **62**, 1198-1211, doi:10.1086/301844 (1998).

31 Liu, F., Kirichenko, A., Axenovich, T. I., van Duijn, C. M. & Aulchenko, Y. S. An approach for cutting large and complex pedigrees for linkage analysis. *Eur J Hum Genet* **16**, 854-860, doi:10.1038/ejhg.2008.24 (2008).

32 Aulchenko, Y. S., Ripke, S., Isaacs, A. & van Duijn, C. M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294-1296 (2007).

33 Amin, N., van Duijn, C. M. & Aulchenko, Y. S. A genomic background based method for association analysis in related individuals. *PLoS One* **2**, e1274, doi:10.1371/journal.pone.0001274 (2007).

34 Aulchenko, Y. S., Struchalin, M. V. & van Duijn, C. M. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics* **11**, 134, doi:10.1186/1471-2105-11-134 (2010).

35 Wellcome Trust Case Control, C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-678, doi:10.1038/nature05911 (2007).

36 Chen, W. M. & Abecasis, G. R. Family-based association tests for genomewide association scans. *Am J Hum Genet* **81**, 913-926, doi:10.1086/521580 (2007).

37 Orwoll, E. *et al.* Design and baseline characteristics of the osteoporotic fractures in men (MrOS) study--a large observational study of the determinants of fracture in older men. *Contemp Clin Trials* **26**, 569-585, doi:10.1016/j.cct.2005.05.006 (2005).

38 Zaykin, D. V. Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis. *J Evol Biol* **24**, 1836-1841, doi:10.1111/j.1420-9101.2011.02297.x (2011).

39 Fehrmann, R. S. *et al.* Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat. Genet.* **47**, 115-125, doi:10.1038/ng.3173 (2015).

40 Alter, O., Brown, P. O. & Botstein, D. Singular value decomposition for genome-wide expression data processing and modeling. *Proc Natl Acad Sci U S A* **97**, 10101-10106 (2000).

41 Sherlock, G. Analysis of large-scale gene expression data. *Brief Bioinform* **2**, 350-362 (2001).

42 Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-21936, doi:10.1073/pnas.1016071107 (2010).

43 Kim, K. K., Adelstein, R. S. & Kawamoto, S. Identification of neuronal nuclei (NeuN) as Fox-3, a new member of the Fox-1 gene family of splicing factors. *J Biol Chem* **284**, 31052-31061, doi:10.1074/jbc.M109.052969 (2009).

44 Dredge, B. K. & Jensen, K. B. NeuN/Rbfox3 nuclear and cytoplasmic isoforms differentially regulate alternative splicing and nonsense-mediated decay of Rbfox2. *PLoS One* **6**, e21585, doi:10.1371/journal.pone.0021585 (2011).

45 Kiehl, T. R., Shibata, H., Vo, T., Huynh, D. P. & Pulst, S. M. Identification and expression of a mouse ortholog of A2BP1. *Mamm Genome* **12**, 595-601 (2001).

46 Underwood, J. G., Boutz, P. L., Dougherty, J. D., Stoilov, P. & Black, D. L. Homologues of the Caenorhabditis elegans Fox-1 protein are neuronal splicing regulators in mammals. *Mol Cell Biol* **25**, 10005-10016, doi:10.1128/MCB.25.22.10005-10016.2005 (2005).

47 McKee, A. E. *et al.* A genome-wide in situ hybridization map of RNA-binding proteins reveals anatomically restricted expression in the developing mouse brain. *BMC Dev Biol* **5**, 14, doi:10.1186/1471-213X-5-14 (2005).

48 Modrek, B., Resch, A., Grasso, C. & Lee, C. Genome-wide detection of alternative splicing in expressed sequences of human genes. *Nucleic Acids Res* **29**, 2850-2859 (2001).

49 Yeo, G., Holste, D., Kreiman, G. & Burge, C. B. Variation in alternative splicing across human tissues. *Genome Biol* **5**, R74, doi:10.1186/gb-2004-5-10-r74 (2004).

50 Morin, L. P., Hefton, S. & Studholme, K. M. Neurons identified by NeuN/Fox-3 immunoreactivity have a novel distribution in the hamster and mouse suprachiasmatic nucleus. *Brain Res* **1421**, 44-51, doi:10.1016/j.brainres.2011.09.015 (2011).

51 Bhalla, K. *et al.* The de novo chromosome 16 translocations of two patients with abnormal phenotypes (mental retardation and epilepsy) disrupt the A2BP1 gene. *J Hum Genet* **49**, 308-311, doi:10.1007/s10038-004-0145-4 (2004).

52 Barnby, G. *et al.* Candidate-gene screening and association analysis at the autism-susceptibility locus on chromosome 16p: evidence of association at GRIN2A and ABAT. *Am J Hum Genet* **76**, 950-966, doi:10.1086/430454 (2005).

53 Martin, C. L. *et al.* Cytogenetic and molecular characterization of A2BP1/FOX1 as a candidate gene for autism. *Am J Med Genet B Neuropsychiatr Genet* **144B**, 869-876, doi:10.1002/ajmg.b.30530 (2007).
54 Sebat, J. *et al.* Strong association of de novo copy number mutations with autism. *Science* **316**, 445-449, doi:10.1126/science.1138659 (2007).
55 Utami, K. H. *et al.* Detection of chromosomal breakpoints in patients with developmental delay and speech disorders. *PLoS One* **9**, e90852, doi:10.1371/journal.pone.0090852 (2014).
56 Luppi, P. H. Neurochemical aspects of sleep regulation with specific focus on slow-wave sleep. *World J Biol Psychiatry* **11 Suppl 1**, 4-8, doi:10.3109/15622971003637611 (2010).
57 Whiting, P. J. The GABA-A receptor gene family: new targets for therapeutic intervention. *Neurochem Int* **34**, 387-390 (1999).
58 Whiting, P. J. *et al.* Molecular and functional diversity of the expanding GABA-A receptor gene family. *Ann N Y Acad Sci* **868**, 645-653 (1999).
59 Buhr, A. *et al.* Functional characterization of the new human GABA(A) receptor mutation beta3(R192H). *Hum Genet* **111**, 154-160, doi:10.1007/s00439-002-0766-7 (2002).
60 Charney, D. S. & Nestler, E. J. *Neurobiology of mental illness*. (Oxford University Press, 2004).

CHAPTER 3.3

Exome-wide Meta-analysis Identifies Rare 3'-UTR Variant in ERCC1/CD3EAP Associated with Symptoms of Sleep Apnea

Ashley van der Spek, Annemarie I. Luik, Desana Kocevska, Chunyu Liu, Rutger W.W. Brouwer, Jeroen G.J. van Rooij, Mirjam C.G.N. van den Hout, Robert Kraaij, Albert Hofman, André G. Uitterlinden, Wilfred F.J. van IJcken, Daniel J. Gottlieb, Henning Tiemeier, Cornelia M. van Duijn, Najaf Amin

This chapter was published in *Frontiers in Genetics*. 2017 Oct 18;8:151.

Abstract

Obstructive sleep apnea (OSA) is a common sleep breathing disorder associated with an increased risk of cardiovascular and cerebrovascular diseases and mortality. Although OSA is fairly heritable (~40%), there have been only few studies looking into the genetics of OSA. In the present study, we aimed to identify genetic variants associated with symptoms of sleep apnea by performing a whole-exome sequence meta-analysis of symptoms of sleep apnea in 1,475 individuals of European descent. We identified 17 rare genetic variants with at least suggestive evidence of significance. Replication in an independent dataset confirmed the association of a rare genetic variant (rs2229918; minor allele frequency = 0.3%) with symptoms of sleep apnea (p-value$_{meta}$ = 6.98×10^{-9}, β_{meta} = 0.99). Rs2229918 overlaps with the 3' untranslated regions of *ERCC1* and *CD3EAP* genes on chromosome 19q13. Both genes are expressed in tissues in the neck area, such as the tongue, muscles, cartilage and the trachea. Further, *CD3EAP* is localized in the nucleus and mitochondria and involved in the tumor necrosis factor-alpha/nuclear factor kappa B signaling pathway. Our results and biological functions of *CD3EAP*/*ERCC1* genes suggest that the 19q13 locus is interesting for further OSA research.

Introduction

Sleep is a complex and essential biological process that has been conserved across diverse animal species throughout evolution[1]. Although normal healthy sleep highly varies within and between adults[2-4], it has to consist of adequate duration, good quality, proper timing and regularity, and the absence of sleep disturbances or disorders[5]. Several large epidemiological studies have shown that short or disturbed sleep is associated with various cognitive[6,7], psychiatric[8-10] and health consequences e.g., diabetes mellitus[11,12], activation of pro-inflammatory pathways[13], and cardiovascular diseases[14]. One of the most common causes of short and disturbed sleep is sleep apnea.

Sleep apnea is a highly prevalent[15] sleep breathing disorder, with obstructive sleep apnea (OSA) as the most common type[16]. OSA affects up to 38% of the general adult population[17] and untreated OSA has been associated with severe health problems[18] such as hypertension[19,20], cardiovascular disease[21-23], stroke[24], type 2 diabetes[25-27], impaired cognitive function[28,29], depression[30], and increased mortality[31-33]. The main characteristic of OSA is the partial or complete obstruction of the upper airways during sleep, causing oxyhemoglobin desaturations and arousals from sleep. This leads to sleep fragmentation and decreased periods of slow wave and REM sleep[34,35]. Consequently, the two most common signs and symptoms of OSA are snoring and excessive daytime sleepiness[36] where the latter can result in personal and occupational problems, and an increased risk of traffic and work-related accidents[18,34,35].

OSA is a complex trait influenced by both environment and genetics[37,38] with obesity, age, and sex as most important risk factors[15,18,39-42]. About 40% of the variance in apneic activity can be explained by genetic factors[38]. At least half of the genetic contribution to sleep apnea acts through mechanisms independent of obesity[43]. Previous genetic studies have focused on several candidate genes for breathing disorders, where the most studied genes are the angiotensin-converting enzyme gene (*ACE*)[44-46]; apolipoprotein, allele E4 (*APOE-ε4*)[47,48]; serotonin receptors and transporters genes (*5-HT2A*, *5-HT2C*, *5-HTT*)[49-53]; adrenergic receptors (*ADRB2/3*)[54,55]; and tumor necrosis factor (*TNF*)[56-58]. However, the results of these studies have been inconsistent or have yet to be confirmed[59]. Using linkage analysis, a method to identify the chromosomal location of the disease influencing genes, two regions on chromosome 2p16 and 19q13 were found to be suggestively linked with OSA independent of obesity[60]. Genome wide association studies (GWASs) could provide more information on common variants involved in the pathogenesis of OSA. Until now only a few GWASs have been reported for OSA. Loci in *GPR83* and *C6ORF183/CCD162P* were found to be significantly associated with OSA[61], and a locus in the neuregulin-1 (*NRG1*) gene was suggestively implicated[62]. Two other studies used customized or targeted genotyping arrays and identified loci in *PPARGC1B*[63], *PTGER3*[64], *PLEK*[64], and *LPAR1*[64] to be associated with OSA. However, most of these findings were not replicated. Consequently, the genetic architecture of OSA remains largely unexplored.

In the present study we aimed to identify genetic variants associated with symptoms of sleep apnea, assessed using the Pittsburgh Sleep Quality Index (PSQI). We performed a GWAS using whole-exome sequence (WES) data of 1,475 individuals from two Dutch studies. Subsequently, we replicated our findings in an independent sample.

Materials and Methods

Study populations

Discovery cohorts

The discovery sample consists of participants from two cohorts including the Erasmus Rucphen Family (ERF) study and the Rotterdam Study (RS) from The Netherlands.

ERF is a family-based study that includes inhabitants of a genetically isolated community in the Southwest of the Netherlands, ascertained as part of the Genetic Research in Isolated Population program. The ERF cohort includes ~3,000 living descendants of 22 founder couples, who had at least six children baptized in the community church. Individuals who were 18 years or older were invited to participate in the study. Data was collected between 2002 and 2005[65]. The study was approved by the Medical Ethics Committee of the Erasmus Medical Center (EMC), Rotterdam, The Netherlands. All participants provided written informed consents and all investigations were carried out in accordance with the Declaration of Helsinki.

RS is a prospective cohort study ongoing since 1990, which aims to investigate determinants of disease occurrence and progression in the elderly[66]. Initially, the RS included 7,983 individuals of 55 years of age or over, living in the well-defined Ommoord district in Rotterdam, The Netherlands. All participants were examined at baseline by an at home interview and an extensive set of examinations in the research facility in Ommoord. The RS was approved by the Medical Ethics Committee of the EMC and by the Ministry of Health, Welfare and Sport of the Netherlands. All participants provided written informed consent to participate in the study. All investigations were carried out in accordance with the Declaration of Helsinki.

Study participants from ERF and RS were assessed for sleep phenotypes using a self-administered questionnaire including questions from the PSQI[67]. The PSQI has been specifically designed to measure sleep quality and sleep disturbances over a 1-month time interval. Symptoms of sleep apnea were assessed by asking the participants "How often did you or your partner notice long pauses between breaths while asleep?" Answers were provided on a categorical scale ranging from 1 to 4 (1. not during the past month; 2. less than once per week; 3. once or twice per week; 4. more than twice per week). Symptoms of sleep apnea were assessed in 1,366 ERF participants and 2,660 RS participants, where for the latter data of the fourth visit was used as it had the largest participation.

Replication cohort
The replication sample included participants from the offspring cohort of the population-based prospective Framingham Heart Study (FHS)[68]. The offspring cohort was recruited between 1971 and 1975, including 5,124 offspring of the original FHS cohort and their spouses[69]. The study was approved by the Institutional Review Board for Human Research of the Boston University Medical Center, Boston, MA, USA. Each participant provided written informed consent.

FHS has collected sleep data using the Sleep Heart Health Study sleep habits questionnaire[70]. Symptoms of sleep apnea scores were constructed as a combination of the following questions: "A. Are there times when you stop breathing during your sleep?" with answers "yes", "no", "I don't know" and "B. If yes to question A: How often do you have times when you stop breathing during your sleep?". Answers to question B were provided on a categorical scale ranging from 1 to 5 (1. Rarely, less than one night per week; 2. Sometimes, one or two nights per week; 3. Frequently, three to five nights per week; 4. Always or almost always, six or seven nights per week; 5. I don't know). Individuals with answers "I don't know" were excluded, since this option is not available in the PSQI. The constructed symptoms of sleep apnea score had answers ranging from 1 to 4, matching the PSQI: 1. not during the past month (A2); 2. less than once per week (A1 and B1); 3. once or twice per week (A1 and B2); 4. more than twice per week (A1 and B3 or A1 and B4).

Sequencing and quality control
Discovery cohorts
In ERF Genomic DNA was extracted from peripheral venous blood utilizing the salting out method[71]. Exomes of 1,336 ERF participants were sequenced at the Erasmus Center for Biomics of the Cell Biology department of the EMC, The Netherlands, using the Agilent V4 capture kit on an Illumina HiSeq2000 sequencing machine with the TruSeq Version 3 protocol[72]. The sequence reads were aligned to the human genome build 19 (hg19) using Burrows Wheeler Aligner (BWA)[73] and the NARWHAL pipeline[74]. Aligned reads were further processed using IndelRealigner, MarkDuplicates and TableRecalibration tools from the Genome Analysis Toolkit (GATK)[75], and Picard (http://broadinstitute.github.io/picard/). Genetic variants were called using the GATK UnifiedGenotyper tool. Individuals with low concordance to genotyping array or with a low call rate and low-quality variants (Phred quality score <30, call rate <90%) and out of Hardy-Weinberg equilibrium (HWE) (p-value < 10^{-6}), were removed. The final dataset for ERF included 528,617 single nucleotide variants (SNVs) in 1,308 individuals[72] of whom 654 individuals also had phenotype data on symptoms of sleep apnea available.

Exomes of 2,628 individuals from the RS population were sequenced at the Human Genotyping facility of the Internal Medicine department at the EMC, the Netherlands, to an average depth of 54x using the Nimblegen SeqCap EZ V2 capture kit on an Illumina Hiseq2000 sequencer using the TruSeq Version 3 protocol[72]. The sequenced reads were aligned to hg19 using BWA[73]. Subsequently, the aligned reads were processed further using

Picard's MarkDuplicates, SAMtools[76], and GATK[75]. Genetic variants were called using the Haplotypecaller from GATK. Samples with low concordance to genotyping array (<95%), low transition to transversion ratio (<2.3) and high heterozygote to homozygote ratio (>2.0) were removed and additionally SNVs with a low call rate (<90%) and out of HWE (p-value < 10^{-8}) were also removed from the data. The final dataset included 600,806 SNVs in 2,356 individuals[77] of whom 821 individuals also had phenotype data on symptoms of sleep apnea available.

For both ERF and RS, file handling and formatting was done using vcftools[78] and PLINK[79] (http://pngu.mgh.harvard.edu/purcell/plink/). Annotation of the variants was performed using SeattleSeq annotation 138 (http://snp.gs.washington.edu/SeattleSeq Annotation138/).

Replication cohort

In FHS exomes of 1,271 participants were sequenced using Illumina HiSeq2000 and 2500 platforms. DNA samples were constructed into Illumina paired-end pre-capture libraries according to the manufacturer's protocol. For exome capture, two, four or six pre-capture libraries were pooled together and hybridized to the HGSC VCRome 2.1 design[80] (42 Mb, NimbleGen). After sequencing the HGSC Mercury analysis pipeline (https://www.hgsc.bcm.edu/content/mercury) and CASAVA software were used to perform sequencing analysis and to de-multiplex the pooled samples. Sequenced reads were aligned to Genome Reference Consortium Human Build 37 (GRCh37) using BWA[73] producing BAM[76] files. The aligned reads were recalibrated using GATK[81] together with BAM sorting, duplicate read marking, and realignment near insertions or deletions. SNVs, insertions and deletions were called using Atlas2[82]. SNVs were excluded with low SNV posterior probability (<0.95), low variant read count (<3), variant read ratio <0.25 or >0.75, strand-bias of more than 99% variant reads in a single strand direction, or total coverage <10. Reference calls with <10x coverage were also set to missing. Variants were excluded outside exon capture regions (VCRome 2.1), multi-allelic sites, monomorphic sites, missing rate >20%, mappability score <0.8, mean depth of coverage >500, or not fulfilling HWE (p-value < $5x10^{-6}$). Samples were excluded with missingness >20%, less than 6 SD from mean depth, more than 6 SD for singleton count, or outside of 6 SD for heterozygous to homozygous ratio or transition to transversion ratio. Variants were annotated using ANNOVAR[83] and dbNSFP v2.0 (https://sites.google.com/site/jpopgen/dbNSFP) according to the GRCh37 reference genome and National Center for Biotechnology Information RefSeq. The final dataset included 1,749,755 SNVs in 1,271 individuals of whom 472 individuals also had phenotype data on symptoms of sleep apnea available.

Statistical analyses

Descriptive analysis was performed using IBM SPSS Statistics version 21 (IBM Corp. Released 2012. IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp.). Study specific exome analyses and meta-analysis of the individual study data were performed using the seqMeta v1.5 library of the R software (http://cran.r-project.org/web/packages/seqMeta/). Single variant association analysis was performed by assuming an additive effect. In ERF

and FHS a linear mixed effects model was used adjusting for familial relationships by including the kinship matrix. To account for population stratification in the RS, we tested the association of ten principal components with the phenotype. None of them was significantly associated with symptoms of sleep apnea and we did not include them in the analysis. The regression analysis was performed using the four categories of symptoms of sleep apnea score as a continuous trait, adjusting for the three main risk factors for OSA; age, sex and body mass index (BMI) (kg/m^2). Meta-analysis was performed using a fixed effects model. Variants that were present in both discovery cohorts (ERF and RS, 115,526 variants) were tested for association, giving a Bonferroni corrected *p*-value threshold of 4.3×10^{-7}. All variants that showed significant or suggestive (*p*-value < 1.0×10^{-6}) association signals in the discovery samples, were tested for replication in FHS. Bonferroni correction was also applied to correct for multiple testing in the replication stage.

Results

Descriptive statistics of the study populations are presented in **Table 1**. The mean age in RS was 75 years ($\bar{x}_{BMI}$ = 27.4 kg/m^2), where the mean age in ERF was 46 years ($\bar{x}_{BMI}$ = 26.7 kg/m^2) and 59 years in FHS ($\bar{x}_{BMI}$ = 27.5 kg/m^2). The prevalence of symptoms of sleep apnea was higher in the ERF population, where 16.8% of the participants reported to have experienced apneas during the last month, compared to 11.6% and 6.6% of the RS and FHS participants, respectively (**Table 2**).

Table 1. Descriptive statistics of the study populations.

	ERF	RS	FHS
N	654	821	472
Age (years), mean ± SD	46.4 ± 13.4	75.0 ± 6.1	59.2 ± 9.4
Male	42.5%	46.8%	48.5%
BMI (kg/m^2), mean ± SD	26.7 ± 4.4	27.4 ± 4.0	27.5 ± 4.7

Abbreviations: ERF = Erasmus Rucphen Family study, RS = Rotterdam Study, FHS = Framingham Heart Study, N = number of participants, BMI = body mass index.

Table 2. Answers to the sleep apnea question for the discovery and replication populations.

		ERF (%)	RS (%)	FHS (%)
How often did you or your partner notice long pauses between breaths while asleep (a so-called sleep apnea)?	Not during the last month	544 (83.2)	726 (88.4)	441 (93.4)
	Less than once a week	48 (7.3)	44 (5.4)	16 (3.4)
	Once or twice a week	32 (4.9)	32 (3.9)	6 (1.3)
	More than twice a week	30 (4.6)	19 (2.3)	9 (1.9)
	Total	654	821	472

Abbreviations: ERF = Erasmus Rucphen Family study, RS = Rotterdam Study, FHS = Framingham Heart Study.

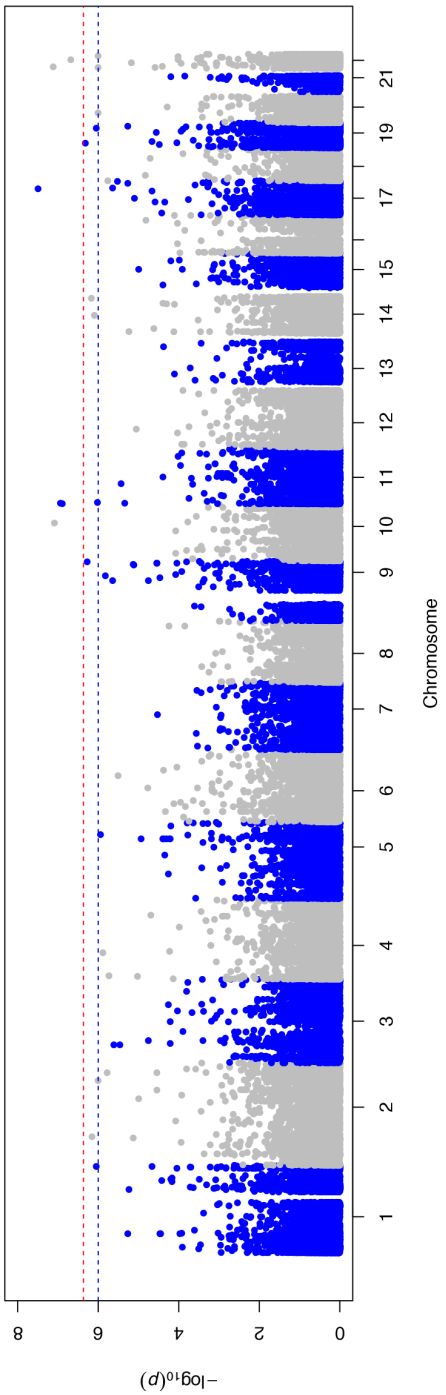


Figure 1. Manhattan plot of the meta-analysis of symptoms of sleep apnea. This plot shows $-\log_{10}$ transformed *p*-values (y-axis) for all SNPs present in the meta-analysis according to their position on each chromosome (x-axis). The red dashed line represents the Bonferroni corrected *p*-value threshold for significance (*p*-value < 4.3×10^{-7}) and the blue dashed line indicates the threshold for suggestive associations (*p*-value < 1.0×10^{-6}).

The exome-wide association results and the distribution of the test statistic (λ = 1.02) are illustrated in **Figures 1** and **2** respectively. Significant associations of symptoms of sleep apnea were observed with six rare variants (minor allele frequency (MAF) < 1%) (located in *ACE*, *AIFM3*, *LIPJ*, *MUC2*, *AP2A2*, *SH3BP1)* (**Table 3**). Suggestive associations of symptoms of sleep apnea were observed with 11 rare variants (located in *KANK2*, *LCN6*, *TRAF3*, *PLEK*, *HIF1A*, *SLC45A3*, *ERCC1/CD3EAP*, *MRGPRE*, *GRAMD4*, *TYW5*, *CST5*) (**Table 3**). Of all 17 variants, only seven were polymorphic in the replication sample and could be tested for association (**Table 4**). Of the six significantly associated variants, two could be tested for association with symptoms of sleep apnea in the FHS (located in *MUC2* and *SH3BP1*).

A significant association of symptoms of sleep apnea with rs2229918, located on chromosome 19q13 in the overlapping 3'-untranslated region (UTR) of the *ERRC1* and *CD3EAP* genes (**Figure 3**), was observed in the replication sample (p-value = 1.84×10^{-3}). Moreover, both the frequency (MAF_{FHS} = 0.3%) and the effect size of the minor allele (G; β_{FHS} = 0.87) were consistent with that of the discovery cohorts (MAF = 0.3%, β = 1.07) suggesting that each copy of the minor allele (G) can result in a shift to a higher category in self-reported apnea symptoms (PSQI). Meta-analysing the discovery and replication cohorts yielded an increased significance of the association of rs2229918 with symptoms of sleep apnea (p-value = 6.98×10^{-9}, β = 0.99).

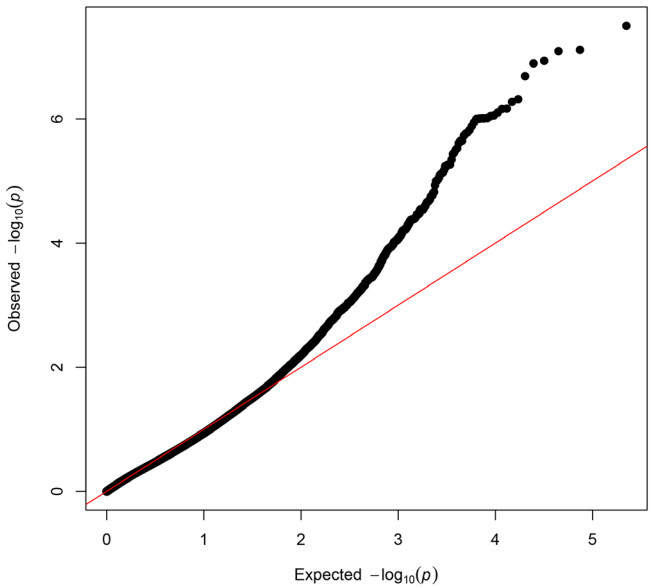


Figure 2. Quantile-Quantile plot of the meta-analysis of symptoms of sleep apnea. The QQ-plot shows the observed p-values plotted on the y-axis against the expected values of the test statistics on the x-axis (X^2-distribution). The red line shows the distribution under the null hypothesis.

Table 3. Meta-analysis association results, filtered on *p-value* < $1.0x10^{-6}$.

Marker name	Gene	Chr	Position	Minor/ major	CADD*	Function GVS*	Poly Phen2*	GERP score*
rs137910205	*ACE*	17	61561775	A/G	0.50	synonymous	--	-9.94
rs178276	*AIFM3*	22	21331950	C/G	6.73	intron	--	-3.74
rs77091298	*LIPJ*	10	90356568	G/T	15.40	missense	0.60	4.12
rs9735156	*MUC2*	11	1093641	C/T	1.95	synonymous	--	-2.97
11:977099	*AP2A2*	11	977099	G/A	15.33	missense	0.30	2.96
rs149928566	*SH3BP1*	22	38039746	T/C	13.28	missense	0.72	-2.34
rs117057052	*KANK2*	19	11277278	C/T	3.31	missense	0.37	3.11
9:139642861	*LCN6*	9	139642861	C/T	2.75	non-coding exon	--	-2.03
rs148461790	*TRAF3*	14	103369593	A/G	15.27	missense near splice	0.43	4.53
rs34515106	*PLEK*	2	68607978	C/A	14.50	missense	--	5.80
rs149348765	*HIF1A*	14	62204819	T/G	14.84	missense	0.59	5.41
rs139592793	*SLC45A3*	1	205632166	T/C	6.52	synonymous	--	1.32
rs2229918	*ERCC1,CD3EAP*	19	45912924	G/C	7.63	3-prime-UTR	--	0.57
rs191846883	*MRGPRE*	11	3249162	A/G	5.89	synonymous	--	3.55
22:47058906	*GRAMD4*	22	47058906	T/C	1.44	intron	--	-5.20
2:200803697	*TYW5*	2	200803697	A/G	38.00	stop-gained	--	4.45
rs142729279	*CST5*	20	23858232	A/G	1.78	synonymous	--	0.46

Abbreviations: CADD = Combined Annotation Dependent Depletion, GERP = Genomic Evolutionary Rate Profiling, ERF = Erasmus Rucphen Family study, RS = Rotterdam Study, Chr = Chromosome, MAF = Minor Allele Frequency, SE = Standard Error, -- = unknown.
***** SeattleSeq Annotation Database 138
† All effects are reported for the minor allele

Discussion

This study aimed at identifying genetic variants associated with symptoms of sleep apnea by performing a meta-analysis of WES data. We identified a rare genetic variant (MAF = 0.3%), rs2229918, located in the shared 3'-UTR region of the *ERCC1* and *CD3EAP* genes with a large effect on symptoms of sleep apnea. We show significant replication of rs2229918 in an independent sample.

The *CD3EAP* gene is located in antisense orientation to *ERCC1* where the 3'-UTRs of both genes overlap. This type of 3'-UTR overlap is conserved in mice and yeast suggesting an important biological function (OMIM #107325). 3'-UTRs can be highly enriched for regulatory elements such as binding sites for regulatory proteins and microRNAs and therefore are most likely involved in post-transcriptional regulation[84]. *ERCC1* encodes Excision Repair

ERF (N=654)				RS (N=821)				Meta-analysis (N=1,475)			
MAF	Beta†	SE	p-value	MAF	Beta†	SE	p-value	MAF	Beta†	SE	p-value
0.001	2.46	0.73	$7.37x10^{-04}$	0.001	2.63	0.60	$1.15x10^{-05}$	0.001	2.56	0.46	$3.15x10^{-08}$
0.002	-0.29	0.52	$5.75x10^{-01}$	0.004	1.53	0.25	$5.03x10^{-10}$	0.003	1.19	0.22	$7.68x10^{-08}$
0.002	1.89	0.52	$2.74x10^{-04}$	0.001	1.68	0.42	$7.61x10^{-05}$	0.001	1.76	0.33	$8.04x10^{-08}$
0.001	2.46	0.73	$7.78x10^{-04}$	0.001	1.77	0.42	$2.93x10^{-05}$	0.001	1.94	0.37	$1.16x10^{-07}$
0.001	2.25	0.73	$2.14x10^{-03}$	0.001	2.59	0.60	$1.61x10^{-05}$	0.001	2.45	0.46	$1.27x10^{-07}$
0.002	0.03	0.53	$9.52x10^{-01}$	0.006	1.11	0.20	$3.14x10^{-08}$	0.004	0.97	0.19	$2.04x10^{-07}$
0.001	2.39	0.73	$1.09x10^{-03}$	0.001	1.67	0.42	$8.60x10^{-05}$	0.001	1.85	0.37	$4.78x10^{-07}$
0.005	1.06	0.30	$4.81x10^{-04}$	0.001	2.54	0.60	$2.47x10^{-05}$	0.002	1.36	0.27	$5.31x10^{-07}$
0.001	1.66	0.73	$2.28x10^{-02}$	0.001	2.73	0.60	$5.23x10^{-06}$	0.001	2.30	0.46	$6.83x10^{-07}$
0.002	0.94	0.52	$7.27x10^{-02}$	0.001	2.10	0.42	$7.80x10^{-07}$	0.001	1.64	0.33	$6.87x10^{-07}$
0.008	0.89	0.24	$1.46x10^{-04}$	0.001	1.41	0.42	$8.62x10^{-04}$	0.004	1.02	0.21	$7.89x10^{-07}$
0.002	1.22	0.43	$4.24x10^{-03}$	0.001	2.67	0.60	$8.42x10^{-06}$	0.001	1.71	0.35	$8.80x10^{-07}$
0.003	0.49	0.37	$1.80x10^{-01}$	0.003	1.37	0.27	$3.39x10^{-07}$	0.003	1.07	0.22	$8.98x10^{-07}$
0.002	1.61	0.52	$1.85x10^{-03}$	0.004	0.85	0.21	$5.87x10^{-05}$	0.003	0.96	0.20	$9.61x10^{-07}$
0.002	1.33	0.52	$1.05x10^{-02}$	0.001	2.71	0.60	$6.16x10^{-06}$	0.001	1.92	0.39	$9.78x10^{-07}$
0.001	1.71	0.73	$1.94x10^{-02}$	0.001	2.64	0.60	$1.02x10^{-05}$	0.001	2.27	0.46	$9.83x10^{-07}$
0.001	1.71	0.73	$1.94x10^{-02}$	0.001	2.64	0.60	$1.02x10^{-05}$	0.001	2.27	0.46	$9.89x10^{-07}$

Cross-Complementation Group 1, a protein functioning in the nucleotide excision repair pathway and needed for the repair of DNA lesions but also involved in recombinational DNA repair and the repair of inter-strand crosslinks[85]. Mutations in *ERCC1* have been linked to cerebro-oculo-facio-skeletal syndrome 4, a severe autosomal recessive disorder characterized by growth retardation, dysmorphic facial features, arthrogryposis, and neurologic abnormalities (OMIM #610758). *CD3EAP* is a component of RNA polymerase I which synthesizes ribosomal RNA precursors and is involved in poly(A) RNA binding and DNA-directed RNA polymerase activity[85]. *CD3EAP* is localized in the nucleus and mitochondria and has two isoforms, isoform 1 is involved in UBTF-activated (Upstream Binding Transcription Factor, RNA Polymerase 1) transcription, while isoform 2 is a component of preformed T-cell receptor complex. *CD3EAP* is involved in multiple pathways including rRNA expression and RNA Polymerase 1 transcription related pathways; RNA polymerase I promotor escape and transcription; gene expression; and the TNF-alpha/NF-kB signaling pathway[85]. Previous genetic studies have associated NF-kB-dependent genes, especially

TNF-α, with OSA[56-58,86]. Moreover, NF-kB is thought to play a key role in mediation of the inflammatory and cardiovascular consequences of OSA[87,88]. GeneNetwork[89] (http://129.125.135.180:8080/Gene Network/) shows that both *ERCC1* and *CD3EAP* are expressed in tissues that may be related to obstruction of the upper airway or diseases of tissues/organs associated with OSA, such as muscle cells, cartilage, trachea, salivary glands, heart and heart ventricles, glucagon secreting cells, the neck and the tongue. This further supports that *ERCC1* and *CD3EAP* are interesting candidate genes for symptoms of sleep apnea.

Rs2229918 is located on chromosome 19q13, a previously identified region with suggestive evidence for linkage to OSA in European-Americans, independently of BMI[60]. Although *APOE*, a known candidate gene for OSA, is also located in this region, it did not show association with OSA in the present study. A previous study fine-mapped the *APOE* region and concluded that *APOE* does not explain the linkage signal, suggesting that *APOE* is not the causative locus[90]. Although the linkage analysis performed by Palmer et al.[60] was redone by adding additional family members and families, the chromosome 19 region was not confirmed. However, this could be due to the genetic or disease heterogeneity[91].

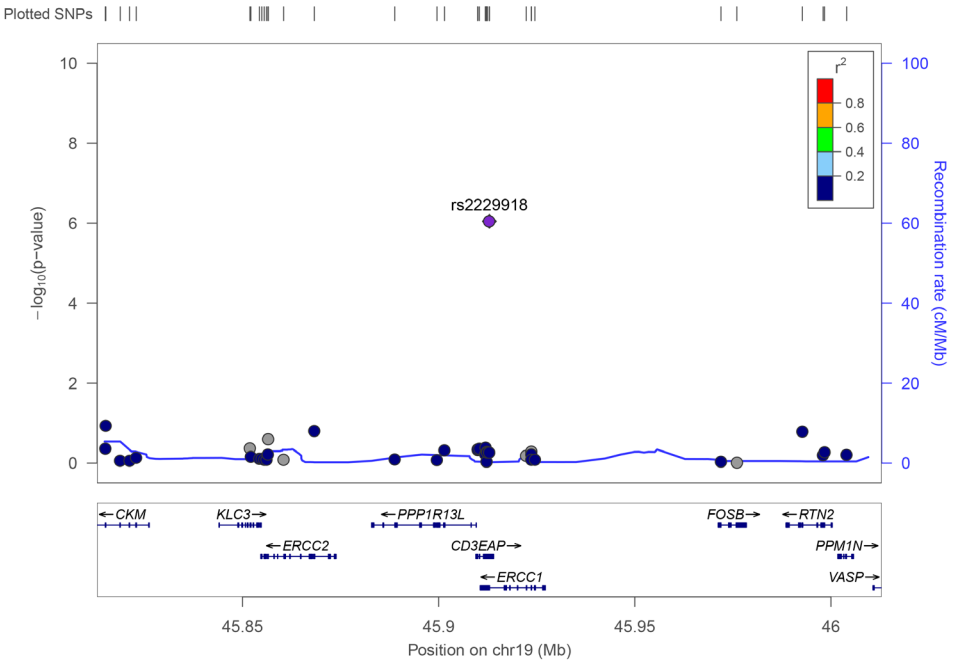


Figure 3. Regional association plot for rs2229918. Rs2229918 is located in purple. The dots show the variants tested in this region on chromosome 19. The –log10 transformed *p*-values are plotted on the y-axis and the genes and positions of the variants (Mb) in this region are depicted on the x-axis.

Table 4. Replication results, filtered on *p-value* < $1.0x10^{-6}$.

Marker name	Gene	Meta-analysis (N = 1,475)				Replication FHS (N = 472)				Meta-analysis (discovery and replication, N = 1,947)			
		MAF	Beta†	SE	*p*-value	MAF	Beta†	SE	p-value	MAF	Beta†	SE	p-value
rs137910205	*ACE*	0.001	2.56	0.46	$3.15x10^{-08}$	--	--	--	--	--	--	--	--
rs77091298	*LIPJ*	0.001	1.76	0.33	$8.04x10^{-08}$	--	--	--	--	--	--	--	--
rs9735156	*MUC2*	0.001	1.94	0.37	$1.16x10^{-07}$	0.002	-0.14	0.34	0.69	0.001	0.83	0.25	$9.09x10^{-04}$
11:977099	*AP2A2*	0.001	2.45	0.46	$1.27x10^{-07}$	--	--	--	--	--	--	--	--
rs149928566	*SH3BP1*	0.004	0.97	0.19	$2.04x10^{-07}$	0.010	-0.16	0.16	0.33	0.005	0.33	0.12	$7.69x10^{-03}$
rs117057052	*KANK2*	0.001	1.85	0.37	$4.78x10^{-07}$	0.005	-0.12	0.22	0.59	0.002	0.39	0.19	$3.66x10^{-02}$
9:139642861	*LCN6*	0.002	1.36	0.27	$5.31x10^{-07}$	--	--	--	--	--	--	--	--
rs148461790	*TRAF3*	0.001	2.30	0.46	$6.83x10^{-07}$	--	--	--	--	--	--	--	--
rs34515106	*PLEK*	0.001	1.64	0.33	$6.87x10^{-07}$	--	--	--	--	--	--	--	--
rs149348765	*HIF1A*	0.004	1.02	0.21	$7.89x10^{-07}$	0.002	-0.15	0.34	0.66	0.004	0.71	0.18	$6.09x10^{-05}$
rs139592793	*SLC45A3*	0.001	1.71	0.35	$8.80x10^{-07}$	0.002	-0.27	0.34	0.43	0.002	0.70	0.24	$3.91x10^{-03}$
rs2229918	*ERCC1,CD3EAP*	0.003	1.07	0.22	$8.98x10^{-07}$	0.003	0.87	0.28	$1.84x10^{-03}$	0.003	0.99	0.17	$6.98x10^{-09}$
rs191846883	*MRGPRE*	0.003	0.96	0.20	$8.98x10^{-07}$	--	--	--	--	--	--	--	--
22:47058906	*GRAMD4*	0.001	1.92	0.39	$9.61x10^{-07}$	--	--	--	--	--	--	--	--
2:200803697	*TYW5*	0.001	2.27	0.46	$9.83x10^{-07}$	--	--	--	--	--	--	--	--
rs142729279	*CST5*	0.001	2.27	0.46	$9.83x10^{-07}$	0.002	-0.02	0.34	0.95	0.001	0.78	0.28	$4.35x10^{-03}$

Abbreviations: MAF = Minor Allele Frequency, SE = Standard Error, -- = Not available. † All effects are reported for the minor allele

Additionally, there were six rare variants (MAF < 0.4%) that surpassed the Bonferroni corrected *p*-value threshold, of which three (located in *ACE*, *LIPJ* and *AP2A2*) were monomorphic in the FHS and could not be tested for replication. Our top finding, rs137910205, a synonymous variant, is located in the *ACE* (angiotensin converting enzyme) gene, one of the most studied genes for OSA. Previous studies found an association between the *ACE* insertion/deletion polymorphism and an increased risk of hypertension in OSA patients[44,45], although results are conflicting[46]. Further, plasma activity of ACE has been found to be increased in untreated OSA patients[92]. Both carriers of rs137910205 (1 in each cohort) reported the highest score for symptoms of sleep apnea, i.e., these individuals have experienced pauses in breathing at least twice per week. The second variant is the missense variant, rs77091298, located in the *LIPJ* (Lipase Family Member J) gene. GeneNetwork showed that *LIPJ* is expressed in the nasopharynx, neck, and muscle cells, all highly relevant tissues in the pathogenesis of OSA[89]. The third variant that could not be tested for replication, 11:977099, has not been identified before. The variant is located in the *AP2A2* gene (Adaptor Related Protein Complex 2 Alpha 2 Subunit), which is related to lipid binding[85]. However, we caution against the interpretation of statistics when the number of carriers of the genetic variants is less than five. Larger sample sizes are needed to further investigate the possible association of these rare genetic variants with OSA.

This study has some limitations regarding the study design. We have used questionnaire data for the assessment of sleep apnea, which could introduce bias[93]. Although reports of breathing pauses more than twice per week are highly predictive of polysomnographic sleep apnea, self- or partner-reported breathing pauses have low sensitivity[42]. Individuals with sleep apnea who experience predominantly hypopneas (shallow breathing) rather than apneas may be less likely to be identified with questionnaire data, as these individuals and their partners may be less likely to recognize these events. Another limitation of using questionnaire data is that the discrimination between OSA, central sleep apnea and mixed sleep apnea is not possible. Although the prevalence of central sleep apnea is generally much lower than OSA in particular in general population samples[94]. Another limitation is that our findings might not be generalizable to other populations as all studies used in this analysis are predominantly European or European American populations. Previous studies have shown a difference in prevalence of sleep apnea between populations, where young African Americans may be at increased risk for sleep apnea[95] and had a higher apnea/hypopnea index relative to European Americans with OSA/hypopnea syndrome[96]. The frequency of the rs2229918 minor allele (G), based on the 1000 Genomes data, also differs across populations (https://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/). Lastly, sleep apnea is a complex and heterogeneous disease influenced by many risk factors such as obesity, age, gender[15,18,39-42], craniofacial and upper airway abnormalities[97,98], race[95,99], alcohol intake[18], smoking[100], and reduced nasal patency due to congestion and respiratory allergies[101]. Despite this phenotypic complexity, we have identified and replicated a rare variant associated with symptoms of sleep apnea. However, we have only used one replication sample and additional studies should further investigate the association of rs2229918 with sleep apnea using objective measurements.

To conclude, this first meta-analysis of symptoms of sleep apnea using WES data identified a rare genetic variant, rs2229918 (MAF 0.3%), located in the 3'-UTR of *ERCC1* and *CD3EAP*, associated with symptoms of sleep apnea. Both genes are interesting candidate genes for (symptoms of) sleep apnea based on their function and expression in tissues relevant for the pathogenesis of the disease. However, the involvement of rs2229918 in OSA pathology should be further examined in larger datasets with more objective measurements.

References

1 Rechtschaffen, A. Current perspectives on the function of sleep. *Perspect. Biol. Med.* **41**, 359-390 (1998).
2 Knutson, K. L., Rathouz, P. J., Yan, L. L., Liu, K. & Lauderdale, D. S. Intra-individual daily and yearly variability in actigraphically recorded sleep measures: the CARDIA study. *Sleep* **30**, 793-796 (2007).
3 Van Dongen, H. P., Vitellaro, K. M. & Dinges, D. F. Individual differences in adult human sleep and wakefulness: Leitmotif for a research agenda. *Sleep* **28**, 479-496 (2005).
4 Mezick, E. J. *et al.* Intra-individual variability in sleep duration and fragmentation: associations with stress. *Psychoneuroendocrinology* **34**, 1346-1354, doi:10.1016/j.psyneuen.2009.04.005 (2009).
5 Consensus Conference, P. *et al.* Recommended Amount of Sleep for a Healthy Adult: A Joint Consensus Statement of the American Academy of Sleep Medicine and Sleep Research Society. *J. Clin. Sleep Med.* **11**, 591-592, doi:10.5664/jcsm.4758 (2015).
6 Pilcher, J. J. & Huffcutt, A. I. Effects of sleep deprivation on performance: a meta-analysis. *Sleep* **19**, 318-326 (1996).
7 Yaffe, K., Falvey, C. M. & Hoang, T. Connections between sleep and cognition in older adults. *Lancet Neurol.* **13**, 1017-1028, doi:10.1016/S1474-4422(14)70172-3 (2014).
8 Peters van Neijenhof, R. J. G. *et al.* Correlates of sleep disturbances in depressed older persons: the Netherlands study of depression in older persons (NESDO). *Aging Ment Health*, 1-6, doi:10.1080/13607863.2016.1247421 (2016).
9 Lovato, N. & Gradisar, M. A meta-analysis and model of the relationship between sleep and depression in adolescents: recommendations for future research and clinical practice. *Sleep Med. Rev.* **18**, 521-529, doi:10.1016/j.smrv.2014.03.006 (2014).
10 Cosgrave, J. *et al.* The interaction between subclinical psychotic experiences, insomnia and objective measures of sleep. *Schizophr. Res.*, doi:10.1016/j.schres.2017.06.058 (2017).
11 Gottlieb, D. J. *et al.* Association of sleep time with diabetes mellitus and impaired glucose tolerance. *Arch. Intern. Med.* **165**, 863-867, doi:10.1001/archinte.165.8.863 (2005).
12 Yaggi, H. K., Araujo, A. B. & McKinlay, J. B. Sleep duration as a risk factor for the development of type 2 diabetes. *Diabetes Care* **29**, 657-661, doi:29/3/657 [pii] (2006).
13 Patel, S. R. *et al.* Sleep duration and biomarkers of inflammation. *Sleep* **32**, 200-204 (2009).
14 Hoevenaar-Blom, M. P., Spijkerman, A. M., Kromhout, D., van den Berg, J. F. & Verschuren, W. M. Sleep duration and sleep quality in relation to 12-year cardiovascular disease incidence: the MORGEN study. *Sleep* **34**, 1487-1492, doi:10.5665/sleep.1382 (2011).
15 Peppard, P. E. *et al.* Increased prevalence of sleep-disordered breathing in adults. *Am. J. Epidemiol.* **177**, 1006-1014, doi:10.1093/aje/kws342 (2013).
16 Mehra, R. *et al.* Prevalence and correlates of sleep-disordered breathing in older men: osteoporotic fractures in men sleep study. *J. Am. Geriatr. Soc.* **55**, 1356-1364, doi:10.1111/j.1532-5415.2007.01290.x (2007).
17 Senaratna, C. V. *et al.* Prevalence of obstructive sleep apnea in the general population: A systematic review. *Sleep Med. Rev.*, doi:10.1016/j.smrv.2016.07.002 (2016).
18 Young, T., Peppard, P. E. & Gottlieb, D. J. Epidemiology of obstructive sleep apnea: a population health perspective. *Am. J. Respir. Crit. Care Med.* **165**, 1217-1239 (2002).
19 Pedrosa, R. P. *et al.* Obstructive sleep apnea: the most common secondary cause of hypertension associated with resistant hypertension. *Hypertension* **58**, 811-817, doi:10.1161/HYPERTENSIONAHA.111.179788 (2011).
20 Peppard, P. E., Young, T., Palta, M. & Skatrud, J. Prospective study of the association between sleep-disordered breathing and hypertension. *N. Engl. J. Med.* **342**, 1378-1384, doi:10.1056/NEJM200005113421901 (2000).
21 Gottlieb, D. J. *et al.* Prospective study of obstructive sleep apnea and incident coronary heart disease and heart failure: the sleep heart health study. *Circulation* **122**, 352-360, doi:10.1161/CIRCULATIONAHA.109.901801 (2010).
22 Shamsuzzaman, A. S., Gersh, B. J. & Somers, V. K. Obstructive sleep apnea: implications for cardiac and vascular disease. *JAMA* **290**, 1906-1914, doi:10.1001/jama.290.14.1906 (2003).
23 Marin, J. M., Carrizo, S. J., Vicente, E. & Agusti, A. G. Long-term cardiovascular outcomes in men with obstructive sleep apnoea-hypopnoea with or without treatment with continuous positive airway pressure: an observational study. *Lancet* **365**, 1046-1053, doi:10.1016/S0140-6736(05)71141-7 (2005).

24 Yaggi, H. K. *et al.* Obstructive sleep apnea as a risk factor for stroke and death. *N. Engl. J. Med.* **353**, 2034-2041, doi:10.1056/NEJMoa043104 (2005).
25 Kendzerska, T., Gershon, A. S., Hawker, G., Tomlinson, G. & Leung, R. S. Obstructive sleep apnea and incident diabetes. A historical cohort study. *Am. J. Respir. Crit. Care Med.* **190**, 218-225, doi:10.1164/rccm.201312-2209OC (2014).
26 Shaw, J. E. *et al.* Sleep-disordered breathing and type 2 diabetes: a report from the International Diabetes Federation Taskforce on Epidemiology and Prevention. *Diabetes Res. Clin. Pract.* **81**, 2-12, doi:10.1016/j.diabres.2008.04.025 (2008).
27 Aurora, R. N. & Punjabi, N. M. Obstructive sleep apnoea and type 2 diabetes mellitus: a bidirectional association. *Lancet Respir Med* **1**, 329-338, doi:10.1016/S2213-2600(13)70039-0 (2013).
28 Kim, H. C. *et al.* Sleep-disordered breathing and neuropsychological deficits. A population-based study. *Am. J. Respir. Crit. Care Med.* **156**, 1813-1819, doi:10.1164/ajrccm.156.6.9610026 (1997).
29 Yaffe, K. *et al.* Sleep-disordered breathing, hypoxia, and risk of mild cognitive impairment and dementia in older women. *JAMA* **306**, 613-619, doi:10.1001/jama.2011.1115 (2011).
30 Peppard, P. E., Szklo-Coxe, M., Hla, K. M. & Young, T. Longitudinal association of sleep-related breathing disorder and depression. *Arch. Intern. Med.* **166**, 1709-1715, doi:10.1001/archinte.166.16.1709 (2006).
31 Punjabi, N. M. *et al.* Sleep-disordered breathing and mortality: a prospective cohort study. *PLoS Med.* **6**, e1000132, doi:10.1371/journal.pmed.1000132 (2009).
32 Young, T. *et al.* Sleep disordered breathing and mortality: eighteen-year follow-up of the Wisconsin sleep cohort. *Sleep* **31**, 1071-1078 (2008).
33 Marshall, N. S. *et al.* Sleep apnea as an independent risk factor for all-cause mortality: the Busselton Health Study. *Sleep* **31**, 1079-1085 (2008).
34 American Academy of Sleep Medicine. *International classification of sleep disorders*. (American Academy of Sleep Medicine, 2014).
35 McNicholas, W. T. Diagnosis of obstructive sleep apnea in adults. *Proc. Am. Thorac. Soc.* **5**, 154-160, doi:10.1513/pats.200708-118MG (2008).
36 Gottlieb, D. J. *et al.* Relation of sleepiness to respiratory disturbance index: the Sleep Heart Health Study. *Am. J. Respir. Crit. Care Med.* **159**, 502-507, doi:10.1164/ajrccm.159.2.9804051 (1999).
37 Redline, S. & Tishler, P. V. The genetics of sleep apnea. *Sleep Med. Rev.* **4**, 583-602, doi:10.1053/smrv.2000.0120 (2000).
38 Redline, S. *et al.* The familial aggregation of obstructive sleep apnea. *Am. J. Respir. Crit. Care Med.* **151**, 682-687, doi:10.1164/ajrccm.151.3.7881656 (1995).
39 Redline, S., Kump, K., Tishler, P. V., Browner, I. & Ferrette, V. Gender differences in sleep disordered breathing in a community-based sample. *Am. J. Respir. Crit. Care Med.* **149**, 722-726, doi:10.1164/ajrccm.149.3.8118642 (1994).
40 Bixler, E. O. *et al.* Prevalence of sleep-disordered breathing in women: effects of gender. *Am. J. Respir. Crit. Care Med.* **163**, 608-613, doi:10.1164/ajrccm.163.3.9911064 (2001).
41 Young, T., Skatrud, J. & Peppard, P. E. Risk factors for obstructive sleep apnea in adults. *Jama-Journal of the American Medical Association* **291**, 2013-2016, doi:DOI 10.1001/jama.291.16.2013 (2004).
42 Young, T. *et al.* Predictors of sleep-disordered breathing in community-dwelling adults: the Sleep Heart Health Study. *Arch. Intern. Med.* **162**, 893-900, doi:ioi10232 [pii] (2002).
43 Patel, S. R., Larkin, E. K. & Redline, S. Shared genetic basis for obstructive sleep apnea and adiposity measures. *Int. J. Obes. (Lond.)* **32**, 795-800, doi:10.1038/sj.ijo.0803803 (2008).
44 Lin, L., Finn, L., Zhang, J., Young, T. & Mignot, E. Angiotensin-converting enzyme, sleep-disordered breathing, and hypertension. *Am. J. Respir. Crit. Care Med.* **170**, 1349-1353, doi:10.1164/rccm.200405-616OC (2004).
45 Bostrom, K. B. *et al.* Interaction between the angiotensin-converting enzyme gene insertion/deletion polymorphism and obstructive sleep apnoea as a mechanism for hypertension. *J. Hypertens.* **25**, 779-783, doi:10.1097/HJH.0b013e328017f6d5 (2007).
46 Patel, S. R., Larkin, E. K., Mignot, E., Lin, L. & Redline, S. The association of angiotensin converting enzyme (ACE) polymorphisms with sleep apnea and hypertension. *Sleep* **30**, 531-533 (2007).
47 Kadotani, H. *et al.* Association between apolipoprotein E epsilon4 and sleep-disordered breathing in adults. *JAMA* **285**, 2888-2890, doi:jbr10003 [pii] (2001).
48 Gottlieb, D. J. *et al.* APOE epsilon4 is associated with obstructive sleep apnea/hypopnea: the Sleep Heart Health Study. *Neurology* **63**, 664-668, doi:63/4/664 [pii] (2004).

49 Bayazit, Y. A. *et al.* Association of the -1438G/A polymorphism of the 5-HT2A receptor gene with obstructive sleep apnea syndrome. *ORL J. Otorhinolaryngol. Relat. Spec.* **68**, 123-128, doi:10.1159/000091216 (2006).

50 Ylmaz, M. *et al.* Association of serotonin transporter gene polymorphism with obstructive sleep apnea syndrome. *Laryngoscope* **115**, 832-836 (2005).

51 Sakai, K. *et al.* Serotonin-2A and 2C receptor gene polymorphisms in Japanese patients with obstructive sleep apnea. *Intern. Med.* **44**, 928-933, doi:JST.JSTAGE/internalmedicine/44.928 [pii] (2005).

52 Qin, B., Sun, Z., Liang, Y., Yang, Z. & Zhong, R. The association of 5-HT2A, 5-HTT, and LEPR polymorphisms with obstructive sleep apnea syndrome: a systematic review and meta-analysis. *PLoS One* **9**, e95856 (2014).

53 Larkin, E. K. *et al.* A candidate gene study of obstructive sleep apnea in European Americans and African Americans. *Am. J. Respir. Crit. Care Med.* **182**, 947-953, doi:10.1164/rccm.201002-0192OC (2010).

54 Grote, L., Kraiczi, H. & Hedner, J. Reduced alpha- and beta(2)-adrenergic vascular response in patients with obstructive sleep apnea. *Am. J. Respir. Crit. Care Med.* **162**, 1480-1487, doi:10.1164/ajrccm.162.4.9912028 (2000).

55 Mills, P. J. *et al.* Beta 2-adrenergic receptor characteristics in sleep apnea patients. *Sleep* **18**, 39-42 (1995).

56 Riha, R. L. *et al.* Tumour necrosis factor-alpha (-308) gene polymorphism in obstructive sleep apnoea-hypopnoea syndrome. *Eur. Respir. J.* **26**, 673-678, doi:10.1183/09031936.05.00130804 (2005).

57 Bhushan, B., Guleria, R., Misra, A., Luthra, K. & Vikram, N. K. TNF-alpha gene polymorphism and TNF-alpha levels in obese Asian Indians with obstructive sleep apnea. *Respir. Med.* **103**, 386-392, doi:10.1016/j.rmed.2008.10.001 (2009).

58 Popko, K. *et al.* Frequency of distribution of inflammatory cytokines IL-1, IL-6 and TNF-alpha gene polymorphism in patients with obstructive sleep apnea. *J. Physiol. Pharmacol.* **59 Suppl 6**, 607-614 (2008).

59 Sleiman, P. & Hakonarson, H. Genetic underpinnings of obstructive sleep apnea: are we making progress? *Sleep* **34**, 1449-1452, doi:10.5665/sleep.1366 (2011).

60 Palmer, L. J. *et al.* A whole-genome scan for obstructive sleep apnea and obesity. *Am. J. Hum. Genet.* **72**, 340-350, doi:10.1086/346064 (2003).

61 Cade, B. E. *et al.* Genetic Associations with Obstructive Sleep Apnea Traits in Hispanic/Latino Americans. *Am. J. Respir. Crit. Care Med.* **194**, 886-897, doi:10.1164/rccm.201512-2431OC (2016).

62 Baik, I., Seo, H. S., Yoon, D., Kim, S. H. & Shin, C. Associations of Sleep Apnea, NRG1 Polymorphisms, Alcohol Consumption, and Cerebral White Matter Hyperintensities: Analysis with Genome-Wide Association Data. *Sleep* **38**, 1137-1143, doi:10.5665/sleep.4830 (2015).

63 Kripke, D. F. *et al.* Genetic variants associated with sleep disorders. *Sleep Med.* **16**, 217-224, doi:10.1016/j.sleep.2014.11.003 (2015).

64 Patel, S. R. *et al.* Association of genetic loci with sleep apnea in European Americans and African-Americans: the Candidate Gene Association Resource (CARe). *PLoS One* **7**, e48836 (2012).

65 Pardo, L. M., MacKay, I., Oostra, B., van Duijn, C. M. & Aulchenko, Y. S. The effect of genetic drift in a young genetically isolated population. *Ann. Hum. Genet.* **69**, 288-295, doi:DOI 10.1046/j.1529-8817.2005.00162.x (2005).

66 Hofman, A. *et al.* The Rotterdam Study: 2016 objectives and design update. *Eur. J. Epidemiol.* **30**, 661-708, doi:10.1007/s10654-015-0082-x (2015).

67 Buysse, D. J., Reynolds, C. F., 3rd, Monk, T. H., Berman, S. R. & Kupfer, D. J. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res.* **28**, 193-213, doi:0165-1781(89)90047-4 [pii] (1989).

68 Dawber, T. R., Meadors, G. F. & Moore, F. E., Jr. Epidemiological approaches to heart disease: the Framingham Study. *Am J Public Health Nations Health* **41**, 279-281 (1951).

69 Kannel, W. B., Feinleib, M., McNamara, P. M., Garrison, R. J. & Castelli, W. P. An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol* **110**, 281-290 (1979).

70 Quan, S. F. *et al.* The Sleep Heart Health Study: design, rationale, and methods. *Sleep* **20**, 1077-1085 (1997).

71 Miller, S. A., Dykes, D. D. & Polesky, H. F. A Simple Salting out Procedure for Extracting DNA from Human Nucleated Cells. *Nucleic Acids Res.* **16**, 1215-1215, doi:DOI 10.1093/nar/16.3.1215 (1988).

72 Amin, N. *et al.* Exome-sequencing in a large population-based study reveals a rare Asn396Ser variant in the LIPG gene associated with depressive symptoms. *Mol. Psychiatry* **22**, 537-543, doi:10.1038/mp.2016.101 (2016).

73 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

74 Brouwer, R. W., van den Hout, M. C., Grosveld, F. G. & van Ijcken, W. F. NARWHAL, a primary analysis pipeline for NGS data. *Bioinformatics* **28**, 284-285, doi:10.1093/bioinformatics/btr613 (2012).

75 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297-1303, doi:10.1101/gr.107524.110 (2010).

76 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

77 Amin, N. *et al.* Nonsynonymous Variation in NKPD1 Increases Depressive Symptoms in the European Populations. *Biol. Psychiatry* **81**, 702-707, doi:10.1016/j.biopsych.2016.08.008 (2016).

78 Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158, doi:10.1093/bioinformatics/btr330 (2011).

79 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559-575, doi:10.1086/519795 (2007).

80 Bainbridge, M. N. *et al.* Targeted enrichment beyond the consensus coding DNA sequence exome reveals exons with higher variant densities. *Genome Biol.* **12**, R68, doi:10.1186/gb-2011-12-7-r68 (2011).

81 DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491-498, doi:10.1038/ng.806 (2011).

82 Challis, D. *et al.* An integrative variant analysis suite for whole exome next-generation sequencing data. *BMC Bioinformatics* **13**, 8, doi:10.1186/1471-2105-13-8 (2012).

83 Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164, doi:10.1093/nar/gkq603 (2010).

84 Xie, X. *et al.* Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* **434**, 338-345, doi:10.1038/nature03441 (2005).

85 Stelzer, G. *et al.* The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr Protoc Bioinformatics* **54**, 1 30 31-31 30 33, doi:10.1002/cpbi.5 (2016).

86 Ryan, S., Taylor, C. T. & McNicholas, W. T. Predictors of elevated nuclear factor-kappaB-dependent genes in obstructive sleep apnea syndrome. *Am. J. Respir. Crit. Care Med.* **174**, 824-830, doi:10.1164/rccm.200601-066OC (2006).

87 Garvey, J. F., Taylor, C. T. & McNicholas, W. T. Cardiovascular disease in obstructive sleep apnoea syndrome: the role of intermittent hypoxia and inflammation. *Eur. Respir. J.* **33**, 1195-1205, doi:10.1183/09031936.00111208 (2009).

88 Ryan, S., Taylor, C. T. & McNicholas, W. T. Selective activation of inflammatory pathways by intermittent hypoxia in obstructive sleep apnea syndrome. *Circulation* **112**, 2660-2667, doi:10.1161/CIRCULATIONAHA.105.556746 (2005).

89 Fehrmann, R. S. *et al.* Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat. Genet.* **47**, 115-125, doi:10.1038/ng.3173 (2015).

90 Larkin, E. K. *et al.* Apolipoprotein E and obstructive sleep apnea: evaluating whether a candidate gene explains a linkage peak. *Genet. Epidemiol.* **30**, 101-110, doi:10.1002/gepi.20127 (2006).

91 Larkin, E. K. *et al.* Using linkage analysis to identify quantitative trait loci for sleep apnea in relationship to body mass index. *Ann. Hum. Genet.* **72**, 762-773, doi:10.1111/j.1469-1809.2008.00472.x (2008).

92 Barcelo, A. *et al.* Angiotensin converting enzyme in patients with sleep apnoea syndrome: plasma activity and gene polymorphisms. *Eur. Respir. J.* **17**, 728-732 (2001).

93 Fedson, A. C., Pack, A. I. & Gislason, T. Frequently used sleep questionnaires in epidemiological and genetic research for obstructive sleep apnea: a review. *Sleep Med. Rev.* **16**, 529-537, doi:10.1016/j.smrv.2011.12.002 (2012).

94 Donovan, L. M. & Kapur, V. K. Prevalence and Characteristics of Central Compared to Obstructive Sleep Apnea: Analyses from the Sleep Heart Health Study Cohort. *Sleep* **39**, 1353-1359, doi:10.5665/sleep.5962 (2016).

95 Redline, S. *et al.* Racial differences in sleep-disordered breathing in African-Americans and Caucasians. *Am. J. Respir. Crit. Care Med.* **155**, 186-192, doi:10.1164/ajrccm.155.1.9001310 (1997).

96 Pranathiageswaran, S., Badr, M. S., Severson, R. & Rowley, J. A. The influence of race on the severity of sleep disordered breathing. *J. Clin. Sleep Med.* **9**, 303-309, doi:10.5664/jcsm.2572 (2013).

97 White, D. P. Pathogenesis of obstructive and central sleep apnea. *Am. J. Respir. Crit. Care Med.* **172**, 1363-1370, doi:10.1164/rccm.200412-1631SO (2005).

98 Mayer, P. *et al.* Relationship between body mass index, age and upper airway measurements in snorers and sleep apnoea patients. *Eur. Respir. J.* **9**, 1801-1809 (1996).
99 Li, K. K., Kushida, C., Powell, N. B., Riley, R. W. & Guilleminault, C. Obstructive sleep apnea syndrome: a comparison between Far-East Asian and white men. *Laryngoscope* **110**, 1689-1693, doi:10.1097/00005537-200010000-00022 (2000).
100 Wetter, D. W., Young, T. B., Bidwell, T. R., Badr, M. S. & Palta, M. Smoking as a risk factor for sleep-disordered breathing. *Arch. Intern. Med.* **154**, 2219-2224 (1994).
101 Young, T., Finn, L. & Kim, H. Nasal obstruction as a risk factor for sleep-disordered breathing. The University of Wisconsin Sleep and Respiratory Research Group. *J. Allergy Clin. Immunol.* **99**, S757-762 (1997).

CHAPTER 4

General discussion

The overall aim of this thesis was to identify genetic and metabolic determinants of age-related outcomes. In the projects included in this thesis, genomics and metabolomics datasets have been studied from multiple cohort studies across the world. This Chapter summarizes and discusses the main findings of this thesis followed by a reflection on the methodology of the studies and possibilities for future research.

Findings of telomere length studies

The first three studies described in **Chapter 2** investigate the association of genetic variants and circulating metabolites with telomere length.

Rare variants

Telomere regulation is a complex process[1] and the underlying molecular mechanisms remain largely unknown. Previous studies have tried to elucidate the genetic architecture of leukocyte telomere length (LTL) in blood by performing genome-wide association studies (GWASs) in Europeans[2-9] and South Asians[10,11]. Even though these studies provided information on telomere length biology, by design, they did not take rare genetic variants into account. In the first project included in this thesis (**Chapter 2.1**), I aimed to identify rare genetic variants associated with LTL by performing the first whole-exome sequencing (WES) analysis of LTL. I identified nine rare variants (minor allele frequency (MAF) < 1%) associated with shorter LTL in the Erasmus Rucphen Family (ERF) study, a family-based study including inhabitants of a genetically isolated community[12]. Eight out of nine variants were located in a region on chromosome 11q22.3. These variants are in high linkage disequilibrium and segregate together with short LTL in a family of the ERF study. This chromosomal region contains the *ATM* (Ataxia Telangiectasia Mutated) locus, which was also found to be associated with LTL in the two most recent LTL GWASs in European[9] and Singaporean Chinese[11] populations. The GWAS performed in the European population additionally used prediction tools to prioritize genes that are most likely influenced by the genetic variants within each locus[9]. This approach highlighted *ATM* as one of the prioritized likely causal genes, together with other genes that are also involved in telomere regulation or DNA damage repair (*PARP1*, *POT1*, *TERF2*, *DCAF4*, *SENP7*, and *RFWD3*)[9]. From a biological perspective, the *ATM* gene is interesting as it encodes ATM kinase that is involved in telomere biology[13-15], where it has conflicting effects. ATM kinase is involved in the localization of telomerase to telomeres and telomere elongation[13] but after activation by critically short telomeres, it can also induce a DNA damage response pathway that leads to cell cycle arrest followed by cellular senescence or apoptosis[16,17]. Our finding of an association of rare variants in the *ATM* locus with shorter LTL highlights the importance of well-balanced DNA repair factors, such as ATM kinase. Future studies should focus on this locus, as a better understanding of telomere length regulation may provide important insights in the association with age-related diseases, and cancers in particular. Activation of telomerase for telomere elongation is not common in somatic cells but expression of telomerase is seen in 85-90% of all

cancers[18,19], and may therefore be a potential therapeutic target to prevent cancer cells from obtaining cellular immortality[20].

The rare variants located on chromosome 11q22.3 were not replicated in two other studies, the Rotterdam Study and the British Heart Foundation Family Heart Study, suggesting this finding may be study-specific. To increase statistical power, I also performed a meta-analysis of the summary statistics of all three studies in **Chapter 2.1**. The meta-analysis resulted in suggestive evidence for the association of rare variants in *RPL8* and *RTEL1* with shorter LTL. Although *RTEL1* is a known telomere length gene[3,9-11], which is involved in telomere length regulation and in the stability and protection of telomeres[16,21], *RPL8* has not been genome-wide significantly associated with any outcome. Interestingly, *RPL8* has been associated with age in a transcriptomic study[22]. A look-up in the phenome-wide association studies (PheWAS) atlas[23] (https://atlas.ctglab.nl/PheWAS) showed the strongest evidence for an association of *RPL8* with type 2 diabetes ($p = 1.7 \times 10^{-6}$), a known age-related disease which is also associated with shorter LTL[24,25]. This finding is, therefore, in need of replication in future studies. Altogether, **Chapter 2.1** showed the strengths but also challenges of rare variant studies. The advantage of using data of family-based studies and more specifically genetic isolates, such as the ERF study, is that rare variants become more frequent over generations due to genetic drift and inbreeding[12]. On the other hand, replication of rare variants in the general population remains a challenge as very large studies are needed. As we provided evidence of the association of rare genetic variants in known and biologically relevant genes, increasing the sample size may enhance the discovery of rare variants that are associated with LTL. This may further help to unravel the genetic architecture of LTL and the relation with age-related diseases.

Metabolomics

Evidence is increasing that metabolomic biomarkers are associated with aging and mortality (see general introduction, page 15). Building up on earlier findings, the Biobanking for Medical Research Infrastructure of the Netherlands (BBMRI-NL) consortium identified 14 metabolomic biomarkers that are associated with mortality. A gap in our knowledge is whether these markers are also associated with LTL. As metabolites, small molecular products of the metabolism, directly reflect the effects of biochemical activity, metabolomics studies may provide more insights in the molecular pathways involved in telomere length regulation and the relation with age-related diseases. In the study described in **Chapter 2.2**, we conducted a metabolomics analysis of LTL using data of 7,853 individuals from seven cohorts, as part of the European Network for Genetic and Genomic Epidemiology (ENGAGE) consortium. Metabolites were measured in a standardized way across cohorts using a mass spectrometry-based platform (Biocrates Life Sciences, Innsbruck, Austria), which includes sugars, amino acids, acyl-carnitines, and phospholipids. We found six metabolites associated with LTL: three phosphatidylcholines (lysophosphatidylcholine acyl C17:0 (lysoPC a C17:0), phosphatidylcholine acyl-alkyl C38:4 (PC ae C38:4), and phosphatidylcholine diacyl C32:1 (PC aa C32:1)), two amino acids (methionine and tyrosine), and one acylcarnitine

(hydroxypropionylcarnitine (C3-OH)). These metabolites are involved in two processes that are known to be involved in LTL shortening[26], namely inflammation and oxidative stress. Interestingly, the two phosphatidylcholines and two of the previous found metabolites[27], 1-stearoylglycerophosphoinositol and 1-palmitoylglycerophospho-inositol, are part of the glycerophospholipids family, which are involved in fatty acid metabolism and membrane composition in biological aging[27,28]. Additionally, pathway analysis pointed to the involvement of homocysteine metabolism, which has been implicated in a wide range of age-related diseases including cardiovascular disease[29,30], dementia[31,32], Alzheimer's disease (AD)[32,33], osteoporotic fracture[34], chronic kidney disease[35], diabetes and vascular complications[36-38]. Increased plasma homocysteine levels have also been associated with age and shortening of LTL[39,40]. An increase of homocysteine levels (with age) can be caused by multiple factors[41] including lifestyle factors (e.g., smoking, alcohol and caffeine intake, nutritional deficiencies of vitamin B and folate[42,43]), the use of drugs that interfere with vitamins involved in homocysteine metabolism[44,45], and also with decreased renal functions[46,47]. These factors were not studied in **Chapter 2.2** and are therefore recommended targets for future ageing research.

As metabolomics panels focus on specific classes of metabolites, in **Chapter 2.3**, I performed a large metabolomics study of LTL using metabolites measured on an NMR platform (Nightingale Health Ltd., Helsinki, Finland), which includes lipoprotein subclasses, cholesterol and triglyceride measures, fatty acids, and various low-molecular metabolites[48]. The meta-analysis performed in this study included data of 11,775 individuals from six Dutch cohorts and was performed as part of the BBMRI-NL consortium. In this study, positive associations of four metabolic ratios with LTL were found, including two cholesterol to lipid ratios in small VLDL (S-VLDL-C % and S-VLDL-CE %) and two omega-6 fatty acid ratios (ratio of omega-6 fatty acids to total fatty acids (FAw6/FA) and ratio of 18:2 linoleic acid to total fatty acids (LA/FA)). We were able to confirm the association of FAw6/FA with LTL in an independent dataset of 7,845 individuals. As data on the other metabolite ratios was not present in the replication dataset, associations could not be confirmed and should be investigated further. Additionally, we performed a lookup of the association of FAw6/FA with mortality in the largest study to date on metabolites and all-cause mortality performed by Deelen et al.[49] (N = 44,168). This study was also performed as part of the BBMRI-NL consortium using metabolite data measured on the same metabolomics platform (Nightingale Health Ltd., Helsinki, Finland)[49]. Indeed, FAw6/FA was significantly associated with mortality in the unadjusted model, however, significance was lost after adjusting for the 14 metabolites that were identified to be associated with all-cause mortality. This is not unexpected as one of these 14 metabolites is the ratio of polyunsaturated fatty acids to total fatty acids (PUFA/FA) and PUFA consists of omega-6 (FAw6) and omega-3 (FAw3) fatty acids. The results of our study point to an important role of FAw6 in ageing, which should be investigated in more detail. Additionally, reviews and meta-analyses of clinical trials and (prospective) cohort studies have shown a positive effect of FAw6, and especially linoleic acid (LA), on coronary artery disease risk[50-52]. After reviewing the scientific evidence, the American Heart Association also concluded that

a reduction of saturated fat intake and replacing it with unsaturated fats (especially polyunsaturated fats) can lower the incidence of cardiovascular disease[53]. Although there is evidence of the potential beneficial effect of FAw6 intake on cardiovascular disease and mortality, recommendations on the FAw6 intake for the prevention of cardiovascular disease remain inconsistent[54] and should receive more investigation since cardiovascular disease is a major health burden.

One of the avenues for finding clues on the potential effects of FAw6 is studying the effects of genes that determine plasma levels of FAw6 fatty acids. Until now, only one genetic variant in the lipase C (*LIPC*) gene has been identified for FAw6/FA[55]. *LIPC* encodes hepatic triglyceride lipase, which is synthesised in the liver after which it enters the bloodstream. It is involved in lipoprotein metabolism, where it is involved in the conversion of fat-transporting molecules and assists in transporting high-density lipoproteins (HDLs), which bring cholesterol and triglycerides to the liver for excretion or catabolism (https://ghr.nlm.nih.gov/gene/LIPC#resources). Interestingly, fine mapping of the metabolite GWAS results revealed that *LIPC* was one of 13 master metabolic regulator genes of metabolite levels[55]. Although all 13 genes show strong pleiotropic effects, *LIPC* showed the strongest effects after correcting for the correlation between the metabolites. Longitudinal data further showed that *LIPC* was positively associated with changes in metabolic activity of FAw6/FA with age[55]. Further fine mapping of the association of *LIPC* with FAw6/FA may provide novel information on their association and can help to provide targets for drug treatment in the future. Taken together, in **Chapter 2.2 and 2.3**, I showed that large samples sizes in metabolomics studies, together with standardized metabolite measurements, provide sufficient power to detect novel associations. Follow-up of these newly identified metabolites using other omics data may reveal more insights in the molecular mechanisms underlying these associations. Another way forward is the use of a wider range of circulating metabolites and the collection of blood samples and metabolite measurements at multiple time points to investigate the changes in metabolite levels over time.

The insulin/insulin-like growth factor I pathway

The insulin/insulin-like growth factor (IGF) signalling (IIS) pathway is involved in cell survival, growth and proliferation in humans and is one of most studied pathways involved in growth- and lifespan regulation[56-58]. Although decreased IIS signalling has been associated with a prolonged lifespan in animal models[59-64], findings in humans are inconsistent[65-67]. Genetic studies of longevity point to the apolipoprotein E gene (*APOE*) as the major driver of longevity[56,68-70]. *APOE* is also the strongest genetic risk factor for AD, the most common form of dementia[71]. There have been multiple speculations on the relationship of lifespan and AD because most patients develop AD late in life, and therefore older adults are at the highest risk[72,73]. The last project of Chapter 2 (**Chapter 2.4**) focused on the association of insulin/insulin-like growth factor I (IGF-I) with dementia and AD. A previous study has found

a significant association of higher IGF-I receptor stimulating activity, a measure of bioactivity of the IGF-I receptor, with increased risk of dementia and AD in the population-based Rotterdam Study[74]. In this thesis, I further investigated the association of IGF-I with dementia and AD using updated information on incident dementia/AD by extending the follow-up time of the Rotterdam Study participants from 12 to 16 years.

Although we no longer found a significant association between IGF-I receptor stimulating activity and risk of dementia, a significant interaction between the apolipoprotein E allele 4 (*APOE*-ε4; rs429358), the major genetic driver of dementia and AD, and tertiles of IGF-I receptor stimulating activity was observed. Increased IGF-I receptor stimulating activity was associated with increased incidence of dementia and AD in carriers of the *APOE*-ε4 allele but not in non-carriers. This finding corresponds to the results of the recent longevity GWAS, which showed that *APOE*-ε4 carriers had lower odds of being long-lived, while carriers of the *APOE*-ε2 allele (rs7412), which is known to be protective for dementia and AD[71], lived longer[70]. As decreased IGF-I plasma levels were previously also found in long-lived individuals[75,76] and research in mice showed that targeting IGF-I receptor signalling using monoclonal antibodies effectively improved female lifespan with 9%[77], this study provides evidence for the favourable effect of reducing IGF-I levels in *APOE*-ε4 carriers. However, an unexpected finding is that at baseline *APOE*-ε4 homozygotes, who have the highest lifetime risk of AD[71], had lower IGF-I receptor stimulating activity than non-carriers of the *APOE*-ε4 allele and *APOE*-ε4 heterozygotes but at follow-up the *APOE*-ε4 carriers had a higher risk of dementia and AD. This may be explained by selection bias in the small group of *APOE*-ε44 carriers in the Rotterdam study: carriers of the *APOE*-ε44 genotype have a high risk of AD by the age of 64.8 years[71] (mean age at baseline visit of the Rotterdam Study). Thus, those with the *APOE*-ε44 genotype and high IGF-I receptor stimulating activity may be selected out from the study at baseline measurement of **Chapter 2.4**, where the IGF-I receptor stimulating activity was assessed at the second follow-up visit of the Rotterdam Study. The mean age of persons at risk for incident dementia at this timepoint was 72 years. This may have resulted in bias in the analysis of incident patients (see discussion section of **Chapter 2.4**). Indeed Table 1 of **Chapter 2.4** shows that the prevalent patients have on average higher levels of IGF-I receptor stimulating activity than persons at risk for dementia. Another consideration to be taken into account is that the results were based on a relatively small number of cases. Although the study in **Chapter 2.4** provided new clues for therapeutic targets that may be of interest for future patient risk stratification, the study also suggests that investigating a cohort consisting of older adults, with a mean age of 72 years, may be biased.

Metabolic determinants of late onset depression in the general population

In **Chapter 3**, I focus on depression, a disease that is frequently present in older individuals[78-80]. Depression is a complex multifactorial disease that has a large effect on an

individual's quality of life and on society[81]. A previous metabolomics-depression (BBMRI-NL) study was performed using the same metabolomic platform as used in **Chapter 2.3** (Nightingale Health Ltd., Helsinki, Finland). This platform focuses specifically on lipoprotein subclasses, cholesterol and triglyceride measures, fatty acids, and various low-molecular metabolites, and is therefore limited in the number of metabolites. We, therefore, explored a more extensive metabolomics platform in relation to depression. In **Chapter 3.1**, I performed a large-scale metabolomics analysis using a mass spectrometry-based untargeted metabolomics platform (Metabolon Inc., Durham, North Carolina, USA). This study included data of 13,596 participants from five independent cohorts. In total I explored the association of 806 circulating metabolites (i.e., 425 lipids, 169 amino acids, 89 xenobiotics, 34 nucleotides, 30 peptides, 23 carbohydrates, 20 cofactors and vitamins, 12 metabolites related to energy metabolism, and 4 unclassified metabolites) with depression. I further explored whether these associations were explained by medication use, lifestyle factors or the disease itself – hypotheses that were not explored thoroughly before. We found 53 metabolites significantly associated with depression in the baseline model, of which 85% (including all amino acids) were explained by the use of antidepressants, medication use for cardiovascular health outcomes, and lifestyle factors including smoking and body mass index (BMI). The remaining eight metabolites, for which the association with depression was not explained by medication use or smoking and BMI, included retinol (vitamin A), 1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1), mannitol/sorbitol, hippurate, 4-hydroxycoumarin, 2-aminooctanoate, 10-undecenoate (11:1n1), and 1-linoleoyl-GPA (18:2). Of these, two metabolites, hippurate and sorbitol, have been consistently associated with depression[82].

Depression and hippurate

Of the eight metabolites associated with depression, we found a weak causal relationship between hippurate and depression in a Mendelian randomization analysis, suggesting that lower levels of hippurate may be causal for depression. We obtained further evidence from the Predictors of Remission in Depression to Individual and Combined Treatments (PReDICT) study, which is a clinical trial aimed at identifying clinical and biological moderators of outcomes to cognitive behaviour therapy (CBT) and antidepressant medication[83]. In the PReDICT study, hippurate levels in blood were not associated with depression at baseline, while hippurate levels increased significantly after 12 weeks of treatment with the antidepressant escitalopram, a selective serotonin reuptake inhibitor (SSRI). This effect on hippurate levels in blood seems to be specific for escitalopram as this effect was not observed after CBT or treatment with duloxetine, a serotonin norepinephrine reuptake inhibitor. This raises the hypothesis whether SSRIs exert their impact on depression by increasing levels of hippurate in the circulation. The formation of hippurate consists of two steps, first metabolism of dietary aromatic compounds by the gut microbiota results in the formation of benzoic acid, followed by conjugation of benzoic acid with glycine in the liver and renal cortex[84,85]. Hippurate has been positively associated with microbiome diversity and with higher fruit and whole grains intake[85,86]. In the UK Biobank study, we found fresh fruit intake to be significantly reduced in depressed individuals, thus complementing the

findings of our study. The causal and therapeutic impact of levels of hippurate in circulation on depression should be investigated further in larger clinical trials.

Depression and retinol

Another highlight of this study was the association of higher levels of retinol (vitamin A) with depression. Retinol and its derivatives (retinoids) are able to cross the blood-brain barrier and although they are necessary for brain development and functioning, an excess of vitamin A is neurotoxic[87,88]. Animal studies also provided evidence for disturbances in pathways involved in depression after vitamin A supplementation, such as monoamine oxidase, resulting in depression-related behaviour[89,90]. Vitamin A is derived from food and regulated by the gut microbiome. Our analysis of food and supplement intake in the UK Biobank showed significantly increased supplementation of vitamin A and a significantly increased retinol intake from food among depressed individuals. Further, we also found significantly higher volume of brain white matter hyperintensities in individuals who were taking vitamin A supplements, which may imply a pathological impact of excess of vitamin A on the human brain. Given the evidence provided by the previous animal studies and case reports[89-92], the association of vitamin A may be causal. We did not confirm causality in the Mendelian randomization analysis, which may be explained by weak instrumental variables and therefore a lack of power.

Depression and 4-hydroxycoumarin, lecithin and mannitol/sorbitol

In addition, we identified 4-hydroxycoumarin, lecithin and mannitol/sorbitol to be associated with depression. These metabolites are also largely derived through food intake and biologically relevant for depression. For instance, 4-hydroxycoumarin is a derivative of coumarin and a precursor of dicoumarol, a vitamin K antagonist, and can be found in vegetables, fruits, plants and spices[93]. Vitamin K is involved in the synthesis of sphingolipids, which are present in the membranes of neuronal and glial cells, and have been implicated in depression and anxiety-related behaviours in mice[94,95]. Further, depression is known to be associated with thromboembolic diseases[96,97], which is consistent with our findings of decreased levels of 4-hydroxycoumarin in depression. Lecithin (1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1)) is derived from eggs, soyabean, liver, nuts and seeds and is a precursor of choline[98]. Previous studies have suggested that lecithin increases the production of acetylcholine in the brain[98-100] leading to depression[101]. Mannitol/sorbitol are sugar alcohols frequently used as sugar substitutes, the sugar-free sweeteners[102], but sorbitol is also commonly present in fruits and mannitol in vegetables[103]. Fructose reduced diets have been shown to improve gastrointestinal disorders, depression and mood disorders[104]. Additionally, sorbitol has been consistently associated with depression[82]. Like other metabolites mannitol and sorbitol are also regulated by the gut microbiome. We found a wide range of genera involved in the update of mannitol, and a limited number in the secretion of mannitol, including *Bacteroides*, *Lactobacillus, Fructobacillus, Alistipes* and *Bifidobacterium*. Interestingly the abundance of all species, except for *Fructobacillus*, have previously been associated with depression[105].

Highlights and conclusions of the study

The most interesting highlight of this study is that most of the metabolites that were significantly associated with depression are derived from the diet and are regulated by the gut-microbiota. Dietary interventions can provide cheap and low risk alternatives to the existing antidepressants. A next step is to design prospective trials to evaluate the effects of these interventions. Another interesting observation of this study is that the associations of serotonin, kynurenate, leucine, and glutamate, which all have been previously reported to be associated with depression[82,106,107], disappeared after adjustment for antidepressants. After adjustment for the medication use for depression, hypertension, hypercholesterolemia, and diabetes, and adjustment for smoking and BMI, only a fraction (15%) of the metabolites remained significantly associated with depression. Therefore, an important conclusion of our study is that (antidepressant) medication has a major effect on the circulating metabolome. Medication as well as potential risk factors such as smoking and BMI should be taken into account in metabolomics studies. There is also the potential effect of residual confounding as confounding may still be present after adjustment for covariates and not all (cardiovascular) risk factors have been investigated. Further, our results also support the findings of a recent paper by Liu et al. where an atlas was constructed of the metabolic effects of medications that are prescribed in type 2 diabetes and other disorders using well-characterized cohorts that are part of the BBMRI-NL consortium[108]. They found that confounders such as BMI and smoking accounted for a fifth of the associations that were seen between medications and metabolites, indicating that these should be considered confounding factors in metabolomics research. These associations may be relevant for adverse untargeted effects of antidepressants and other medication that may be relevant for (cardiometabolic) health to be examined in future studies.

Last but not least, the results of **Chapter 3.1** showed that using an untargeted metabolomics approach can reveal novel biological information and therefore increase the understanding of depression pathophysiology[109]. Our study has provided potential biomarkers to monitor the effectiveness of treatments[110] and for the development of new medication, which would be beneficial as up to 50% of patients do not respond to the current first-line antidepressant therapies[111,112].

Genetic determinants of sleep traits in the general population

Sleep problems are common in the elderly population (~80%), where most older individuals have difficulty falling asleep, increased sleep fragmentation, have the need for naps during the day and do not feel rested[113]. Fragmented sleep can cause a decrease in sleep quality, which has been associated with serious health consequences such as cognitive decline, cardiovascular disease, diabetes but also with decreased quality of life[114]. The last two studies included in **Chapter 3** of this thesis focus on the genetic mechanisms involved in two sleep-related outcomes, sleep latency and sleep apnea.

Sleep latency

The time it takes to fall asleep, i.e., sleep latency, can be caused by both non-genetic and genetic factors, and the heritability of sleep latency is estimated to be between 17-44%[115-117]. Age is an important and complex driver of sleep latency. The association of age with sleep latency follows a quadratic function, with the lowest sleep latency in middle-aged subjects after which sleep latency goes up with increasing age[118]. However, as data on the genetic variants involved in sleep latency were lacking, I conducted a GWAS of sleep latency in 4,242 participants from seven European cohorts in **Chapter 3.2**. In this study, a cluster of three highly correlated intronic single nucleotide polymorphisms (SNPs) mapped to the *RBFOX3* gene was associated with sleep latency. Subsequently, this association was replicated using data from 12 independent cohorts. *RBFOX3* is part of the Fox-1 gene family. This gene is located on chromosome 17q25 and is involved in regulation of alternative pre-mRNA splicing (https://www.ncbi.nlm.nih.gov/gene/?term=146713). Interestingly, using gene network and functional prediction analysis, we found that *RBFOX3* is expressed in brain regions that are involved in sleep and in the release cycle of neurotransmitters, such as gamma-aminobutyric acid and various monoamines that have an important role in sleep onset. Our GWAS was relatively small and the *RBFOX3* gene has not been implicated in sleep in the large GWAS studies at the genome-wide significant level. This underscores the fact that genome-wide significant findings in relatively small studies may not replicate in large studies using data of biobanks. Yet, the gene was associated with alcohol dependence (p = 1.9x10^{-9}), a trait know to be implicated in sleep[23]. Being one of the first studies examining genetic variants involved in sleep latency, makes it difficult to replicate findings. The findings of this study remain to be replicated and should be studied more thoroughly, also in relation to other correlated sleep measures such as sleep quality.

In addition, instead of self-reported sleep latency, it is worthwhile to further investigate the genetic architecture of sleep latency using data of quantitative measures including polysomnography (PSG)[119] and actigraphy[120,121]. A previous study has performed a genetic study of sleep latency using PSG data[122]. The authors found one gene (*CSNK1D*) significantly associated with sleep latency defined by PSG (N = 723) and another gene with subjective sleep latency (*RORA*, N = 733). However, the sample sizes were small and the findings were not replicated. Although PSG is the golden standard, it is not easy to perform and not practical to perform in large cohorts. Actigraphy, on the other hand, is a more cost-effective method to obtain objective sleep data in large studies. One of the recently published studies performed within the large UK Biobank[123] investigated sleep traits (sleep quality, quantity and timing) assessed using accelerometers, which are actigraphy devices (N = 85,670). The authors identified 47 loci at genome-wide significance and provided important novel insights into sleep biology. Unfortunately, sleep diaries, containing information on for example bedtime, were not collected by the UK Biobank participants and without this information it is not possible to define sleep latency[123]. It would therefore be interesting to perform a similar study using data of a large biobank which has collected information on bedtime, either by interviews but preferably by wearable measurement devices.

Sleep apnea

Besides sleep measures such as sleep latency, sleep disorders can also cause a reduction in sleep quality. One of the sleep disorders that increases with age is sleep apnea. Sleep apnea is a common sleep breathing disorder with obstructive sleep apnea (OSA) being the most common form[124]. OSA affects up to 38% of the general adult population and can cause serious health consequences[125,126]. In the last project included in this thesis (**Chapter 3.3**), I investigated the association of genetic variants with symptoms of sleep apnea using WES data. The meta-analysis included data of 1,475 participants of the ERF study and the Rotterdam study. We identified a significant association of a rare genetic variant (rs2229918) with symptoms of sleep apnea, which was assessed using questionnaire data (Pittsburgh Sleep Quality Index[127]). This finding was subsequently replicated in the Framingham Heart Study. The genetic variant is located on chromosome 19q13, a region that has previously been associated with OSA[128]. Further, it is located in the shared 3'-UTR region of the *ERCC1* and *CD3EAP* genes, which are expressed in tissues relevant for the pathogenesis of OSA, such as the tongue, muscles, cartilage and the trachea. The genes are involved in pathways that have been suggested to mediate the inflammatory and cardiovascular consequences of OSA[129-131]. This is the first study that identified and replicated a genetic variant associated with sleep apnea by using a subjective measure. As quantitative measurements (e.g., polysomnography[119]) are the golden standard for the diagnosis of sleep disorders, it would be useful to further look into the genetic architecture of OSA by studying the apnea–hypopnea index (AHI), which is assessed by overnight polysomnography and by studying other intermediate phenotypes such as obesity, craniofacial morphology and smoking[132].

Methodological considerations

Genetic studies

Three studies included in this thesis focused on the identification of genetic variants with age-related outcomes. Two of these studies (**Chapter 2.1 and 3.3**) used WES data to study rare genetic variants, while common variants were studied in **Chapter 3.2** by performing a GWAS. The identification of rare genetic variants can help to elucidate biological mechanisms underlying age-related outcomes. One approach to study rare variants is by using a family-based design, such as the ERF study, as was done in **Chapter 2.1**. A limitation of this design is that replication of these findings is difficult because very large sample sizes are needed to have sufficient statistical power to detect the associations with rare variants in the general population. It would be interesting to repeat this analysis using larger sample sizes by extending the number of studies involved. However, this is currently not feasible because WES data is not always available or only for a subset of the participants. As genetic data measured on the relatively inexpensive SNP arrays is widely available, one solution is the use of imputation panels, like the Haplotype Reference Consortium (HRC), that allows a more accurate imputation of common as well as low-frequency genetic variants[133]. There

are also new panels emerging as part of large-scale sequencing efforts, such as TOPMed[134]. Here, whole-genome sequencing data of 60,039 individuals were used to construct the TOPMed imputation panel, resulting in well-imputed ($r^2 > 0.3$) rare variants with a frequency down to 0.004-0.006%[134]. Interestingly, this imputation panel consists of sequencing data from diverse ethnic backgrounds[134]. This provides possibilities to increase the imputation quality compared to the state-of-the-art reference panels (1000 Genomes Project and HRC) in admixed populations, which was shown by a recent study in African and Hispanic/Latino ancestry samples[135]. This study identified two rare/low-frequency variants associated with blood traits that would otherwise not have been discovered because of low imputation quality[135]. Future studies can use the SNP array data, in combination with these imputation panels, to increase the sample size and statistical power. This can help to identify novel genetic loci and may provide relevant insights in the biological pathways underlying age-related outcomes. Another method to increase the statistical power is the (meta) analysis of gene-based tests, which use the cumulative effect of (rare) variants within genes to test for an association with the outcome[134,136,137].

Metabolomic studies

Three studies included in this thesis provide insights in the metabolites and metabolic pathways involved in LTL biology (**Chapter 2.2** and **Chapter 2.3)** and depression pathophysiology (**Chapter 3.1**). All studies used metabolite data measured on different metabolomics platforms, however, all three are validated and are widely used. **Chapter 2.2** focused on phospholipids and carnitines (Biocrates Life Sciences, Innsbruck, Austria), **Chapter 2.3** focused on circulating lipoprotein particles (Nightingale Health Ltd., Helsinki, Finland), and **Chapter 3.1** used a mixture of a targeted and untargeted approach and therefore assessed a broad range of metabolites (Metabolon Inc., Durham, North Carolina, USA). The advantages of the targeted metabolomics approach are that it is cheaper and more standardized and thus easier to use in the setting of clinical and epidemiological research. In contrast, the advantage of the untargeted approach is that it measures more (classes of) metabolites, but it is more difficult to replicate findings and to pool across cohorts. The lesson learned from the study performed in **Chapter 3.1** is that standardization across cohorts is a challenge for different versions of the Metabolon platform. This problem has hampered the comparison of the effects of metabolites across all metabolites measured. The solution we have opted for is analyzing individual studies and pooling results using meta-analysis.

LTL assessment

LTL in **Chapters 2.1-2.3** was measured using a quantitative polymerase chain reaction (qPCR) method in the majority of cohorts included in these studies. The qPCR method is widely used in population-based cohorts as it is relatively inexpensive, easy to perform, and only requires small amounts of DNA, which makes it feasible to measure larger amounts of samples. However, the T/S ratio (the ratio of telomere amplification product (T) to the amplification product of a single reference gene (S)), which is obtained by the qPCR method, is proportional to the average LTL but does not provide an estimate of the amount of base

pairs[138]. Previous studies have shown that the length of the shortest telomeres trigger DNA damage responses leading to replicative senescence and that these should be studied instead of average LTL[139]. Moreover, the qPCR results reported are based on (small) changes in mean LTL and although large studies are able to detect statistically significant differences, the biological meaning remains unclear[139]. Other methods like STELA (Single TElomere Length Analysis) and TeSLA (Telomere Shortest Length Assay) can be used to measure the shortest telomers but have a low throughput and these methods are labour intensive[139].

Epidemiological studies often use blood, and occasionally saliva, as a DNA source. However, when studying aging one may be interested in an array of tissues beyond blood as changes in telomere length may be tissue specific. It is very encouraging that a recent study compared telomere length in human leukocytes, assessed by qPCR and Southern blot, with telomere length in other tissue types[140]. In this study, more than 20 tissue types were examined in 952 individuals from the Genotype-Tissue Expression (GTEx) project. The results showed that, with a few exceptions such as the ovary, breast, thyroid, esophagus, coronary artery, aorta, cerebellum and skeletal muscle, telomere length in whole blood correlates significantly with that in other tissues (**Figure 1**). Further, genetic variation was found to determine telomere length in multiple tissue types and also that telomere length in tissue may mediate the effect of age on gene expression[140]. This finding is very encouraging for studying telomeres in blood, which happens to be the most convenient tissue to use in epidemiology. Together, these findings support the use of leukocytes and qPCR in large scale epidemiological studies.

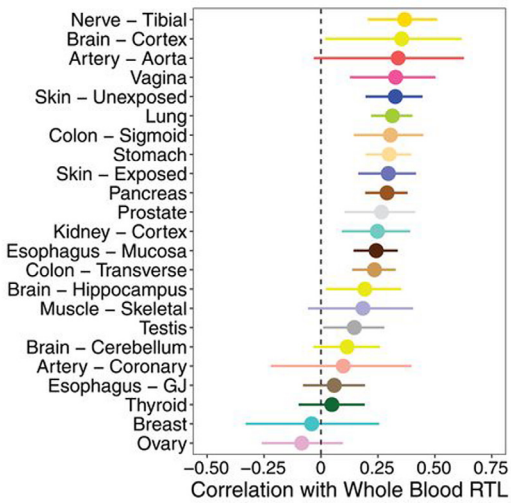


Figure 1. Pearson correlations between whole blood relative telomere length (RTL) and tissue-specific RTL measurements (with 95% confidence intervals). The x-axis describes the correlation of RLT in tissue types (y-axis) with RTL in whole blood. Source: Demanelis et al.[140]

Other methodological problems to consider are the (pre)analytical factors that influence qPCR, such as the type of sample, sample storage, DNA extraction, and assay conditions[141]. In the LTL studies included in this thesis, we observed differences in mean LTL between studies and standardization of the LTL values across cohorts using z-transformation was needed. An example is the TwinsUK registry in **Chapter 2.2**, which had a high mean LTL value compared to the other cohorts. There was no clear explanation for this, except for the fact that the study contained many women who have in general longer LTL[142]. Although the sensitivity analysis where the TwinsUK cohort was excluded from the analysis showed that the high mean LTL value did not have a substantial effect on the results, the distribution of LTL across cohort should be taken into account.

Challenges studying diseases at old age

Studies on aging and age-related mortality face many challenges. The advantage of the Rotterdam Study is that the incidence of diseases at old age can be studied in a powerful setting. However, our study on IGF-I receptor stimulating activity in **Chapter 2.4** provided evidence that these studies may be prone to selection bias against genetic subgroups that are at high risk of early onset of disease. While both prevalent and incident dementia and AD patients showed high levels of IGF-I receptor stimulating activity, *APOE*-ε44 carriers showed lower levels than other *APOE* groups. These low levels may be explained by those with a high risk of disease at early age[71] and not selected out of the analysis of incident patients. This problem should be taken into account when studying late-onset diseases. Another point of concern is that the trend in Table 4 of **Chapter 2.4** (page 116), showing IGF-I receptor stimulating activity tertile groups and risk of incident dementia, is not consistent with a dose effect in that those patients in the medium IGF-I receptor stimulating activity tertile group have higher risk of incident dementia and AD than patients in the high IGF-I receptor stimulating activity tertile group. Again, this may be attributed in part to survival bias.

Heterogeneity among phenotypes

Another challenge in genetic epidemiologic research resulting from the meta-analysis of summary data of multiple studies is the increase in heterogeneity among phenotypes. This is particularly relevant for outcomes where subjective measurements are used such as in psychiatric research. For example, depression in epidemiological studies is mostly assessed using interview/questionnaire data. In the ERF study and the Rotterdam Study, depression symptoms were assessed with the Centre for Epidemiologic Studies Depression (CES-D) scale, a self-report measure of depressive symptoms experienced during the prior week[143]. However, multiple (versions of the) questionnaires are used across cohorts. An alternative approach is to use a stricter clinical diagnosis. Also, previous genetic studies of clinical depression had limited success in the identification of genetic variants. This has been attributed to the high degree of measurement error and heterogeneity in the depression phenotype. At the end of the day, the problem was overcome by a substantial increase of the sample size, as performed by Howard et al.[144], where the meta-analysis included data

of 807,553 individuals and led to the discovery of a series of 102 independent genetic variants[144]. The same problem related to heterogeneity is encountered for sleep phenotypes that were used in the studies included in this thesis. Both sleep latency and symptoms of sleep apnea were assessed using self-reported questionnaire data. The advantages of using questionnaire data are the wide-spread availability and that data gathering can be done at low costs, the disadvantage is the subjectivity of the data, which should be taken into account when interpreting the results.

Future research

The genetic studies included in this thesis were successful in the identification of genetic variants associated with the outcomes including LTL, sleep latency, and symptoms of sleep apnea. However, to obtain further insights into the biological mechanisms underlying these associations, fine mapping of the findings and identification of the causal genetic variant or gene is necessary. As for many outcomes, the disease is the result of many genes acting in concert, and the functional studies in animal and cellular studies are hampered as they cannot capture the small effects of many genes at the same time. Platforms have been developed that use multiple sources for annotation of genetic findings. One example is the online FUMA (Functional Mapping And Annotation of Genome-Wide Association Studies) platform[145], which can facilitate annotation, prioritization, visualization and interpretation of GWAS findings. Further, there have been various other statistical approaches developed to identify causal variants[146] and recent developments in the "omics" field make it possible to combine the results of multiple "omics" studies, such as genetics, epigenetics, metabolomics, and proteomics. For example, a large epigenetics study of depression, performed within the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium, has identified three DNA methylation sites associated with depression, which were located in two genes: *CDC42BPB* and *ARGHEF3*, and an intergenic region at chromosome 15q26.1[147]. The result combined suggested that the axon guidance pathway is involved in depression[147]. Further enlargement of the sample size may bring to light new loci and additional information on the pathogenesis of the disease.

Most genetic studies reported to date, and also the studies included in this thesis, have mainly focused on populations of European ancestry, while only a limited number of studies have included participants of other ancestries such as Hispanic/Latino, Asian and African ancestry. A recent study investigating the differences in telomere length across tissues showed differences in telomere length across ethnicities, where individuals of African ancestry had longer telomeres than individuals of other ancestries[140]. In addition, most results of the telomere length GWAS performed in European individuals[9] were supported by the GWAS performed in Singaporean Chinese individuals, although some variants were population-specific. Also, even if variants are shared between populations, their frequencies may differ and also the effect sizes can be very different[148]. Investigating multiple ancestries

in one study, the so-called multi-ethnic studies, can therefore provide insights in shared causal variants and population-specific variants, and improve generalizability of findings. Multi-ethnic genetic studies are also supported by the new developments in imputation panels, for example TOPMed. These panels further improve imputation quality for Hispanic/Latino and African populations, which have a more complex linkage disequilibrium structure than European populations[134]. These developments will increase the statistical power for the identification of genetic variants and genes associated with (complex) traits in admixed populations.

In addition to the multi-ethnic analysis, future studies of LTL will benefit from longitudinal data, which can be obtained from prospective cohort studies such as the Rotterdam Study[149]. Although we find that LTL predicts mortality, the change in LTL may be a much better predictor. The same argument can be made for metabolomics research, where changes in the metabolome may be of interest and better predictors of aging and age-related pathology. However, this requires precision of measurements over time. Furthermore, prospective cohort studies collect information on exposures before the outcome occurs and the temporality of the associations can be assessed (i.e., reverse causality). However, longitudinal studies are not easy to interpret[150]. For example, particular patients or persons at high risk of cardiovascular diseases are often treated, which may affect the changes in metabolites over time. The challenges encountered should be tackled in future methodological research.

Future studies investigating metabolites and metabolic pathways in relation to LTL may provide additional insights in the biological mechanisms underlying biological aging, such as the homocysteine pathway, and into their relation with age-related diseases such as cardiovascular disease. These studies should also take lifestyle factors and drug usage into account as for example elevated homocysteine levels can be caused by lifestyle factors and the use of drugs. A key issue to resolve is the causal direction of associations. An important tool for future studies is Mendelian randomization[146,151], which has been used to bench mark the causal relationship between cardiovascular pathology and telomere length[3,9,152]. However, for many outcomes this successful application awaits the discovery of the underlying genetic architecture.

Mental health is one of the most important health problems worldwide (https://www.who.int/health-topics/mental-health) and the costs related to depression are high and increasing, e.g., between 2005 and 2010 the costs related to depression in the United States increased by 21.5% to $210,5 billion[153]. Any avenue for prevention is worth to follow-up. One of the most urgent follow-up studies of this thesis concerns the relation of metabolites such as hippurate and retinol (vitamin A) to depression (**Chapter 3.1**), as findings from these association studies can be investigated in dietary intervention studies. Subsequently, when proven, findings from intervention studies can be translated into public health messages. In addition, from a geriatric perspective, an important finding of my thesis that deserves further attention is the polytherapy that is often seen in older individuals. The strong effect

of medication on the metabolome (**Chapter 3.1**) asks for more detailed pharmacometabolomic studies addressing the question whether these metabolites are implicated in side effects of drugs.

Concluding remarks

In this thesis, I have studied genetic and metabolic determinants of age-related outcomes. I have used genomics as well as metabolomics approaches in cohort studies across the world and identified promising genetic variants, genes, metabolites, and metabolic pathways. For me, one of the most important highlights of this thesis is that I was on one hand able to elucidate long suspected relationships, e.g., the role of *ATM* in telomere length, and on the other hand I discovered new findings such as the relation of FAw6 with telomere length. Another highlight of my work as a PhD student is that my research brought to surface new avenues for dietary and lifestyle interventions to prevent (telomere) aging and depression, e.g., homocysteine, FAw6, hippurate, mannitol/sorbitol and vitamin A. My ultimate goal of research is to help patients or those at high risk of disease. This can be achieved by earlier and more accurate diagnosis, the development of new treatments, finding the most optimal treatment for specific patient groups and ultimately the prevention of disease development.

References

1 Blackburn, E. H. Telomere states and cell fates. *Nature* **408**, 53-56, doi:10.1038/35040500 (2000).
2 Codd, V. *et al.* Common variants near TERC are associated with mean telomere length. *Nat Genet* **42**, 197-199, doi:10.1038/ng.532 (2010).
3 Codd, V. *et al.* Identification of seven loci affecting mean telomere length and their association with disease. *Nat Genet* **45**, 422-427, 427e421-422, doi:10.1038/ng.2528 (2013).
4 Levy, D. *et al.* Genome-wide association identifies OBFC1 as a locus involved in human leukocyte telomere biology. *Proc Natl Acad Sci U S A* **107**, 9293-9298, doi:10.1073/pnas.0911494107 (2010).
5 Gu, J. *et al.* A genome-wide association study identifies a locus on chromosome 14q21 as a predictor of leukocyte telomere length and as a marker of susceptibility for bladder cancer. *Cancer Prev Res (Phila)* **4**, 514-521, doi:10.1158/1940-6207.CAPR-11-0063 (2011).
6 Pooley, K. A. *et al.* A genome-wide association scan (GWAS) for mean telomere length within the COGS project: identified loci show little association with hormone-related cancer risk. *Hum Mol Genet* **22**, 5056-5064, doi:10.1093/hmg/ddt355 (2013).
7 Mangino, M. *et al.* Genome-wide meta-analysis points to CTC1 and ZNF676 as genes regulating telomere homeostasis in humans. *Hum Mol Genet* **21**, 5385-5394, doi:10.1093/hmg/dds382 (2012).
8 Mangino, M. *et al.* DCAF4, a novel gene associated with leucocyte telomere length. *J Med Genet* **52**, 157-162, doi:10.1136/jmedgenet-2014-102681 (2015).
9 Li, C. *et al.* Genome-wide Association Analysis in Humans Links Nucleotide Metabolism to Leukocyte Telomere Length. *Am J Hum Genet* **106**, 389-404, doi:10.1016/j.ajhg.2020.02.006 (2020).
10 Delgado, D. A. *et al.* Genome-wide association study of telomere length among South Asians identifies a second RTEL1 association signal. *J Med Genet* **55**, 64-71, doi:10.1136/jmedgenet-2017-104922 (2018).
11 Dorajoo, R. *et al.* Loci for human leukocyte telomere length in the Singaporean Chinese population and trans-ethnic genetic studies. *Nat Commun* **10**, 2491, doi:10.1038/s41467-019-10443-2 (2019).
12 Pardo, L. M., MacKay, I., Oostra, B., van Duijn, C. M. & Aulchenko, Y. S. The effect of genetic drift in a young genetically isolated population. *Annals of Human Genetics* **69**, 288-295, doi:DOI 10.1046/j.1529-8817.2005.00162.x (2005).
13 Tong, A. S. *et al.* ATM and ATR Signaling Regulate the Recruitment of Human Telomerase to Telomeres. *Cell Rep* **13**, 1633-1646, doi:10.1016/j.celrep.2015.10.041 (2015).
14 Pandita, T. K. ATM function and telomere stability. *Oncogene* **21**, 611-618, doi:10.1038/sj.onc.1205060 (2002).
15 Lee, S. S., Bohrson, C., Pike, A. M., Wheelan, S. J. & Greider, C. W. ATM Kinase Is Required for Telomere Elongation in Mouse and Human Cells. *Cell Rep* **13**, 1623-1632, doi:10.1016/j.celrep.2015.10.035 (2015).
16 de Lange, T. Shelterin-Mediated Telomere Protection. *Annu Rev Genet* **52**, 223-247, doi:10.1146/annurev-genet-032918-021921 (2018).
17 d'Adda di Fagagna, F. *et al.* Effects of DNA nonhomologous end-joining factors on telomere length and chromosomal stability in mammalian cells. *Curr Biol* **11**, 1192-1196, doi:10.1016/s0960-9822(01)00328-1 (2001).
18 Kim, N. W. *et al.* Specific association of human telomerase activity with immortal cells and cancer. *Science* **266**, 2011-2015, doi:10.1126/science.7605428 (1994).
19 Shay, J. W. & Bacchetti, S. A survey of telomerase activity in human cancer. *Eur J Cancer* **33**, 787-791, doi:10.1016/S0959-8049(97)00062-2 (1997).
20 Shay, J. W. & Wright, W. E. Role of telomeres and telomerase in cancer. *Semin Cancer Biol* **21**, 349-353, doi:10.1016/j.semcancer.2011.10.001 (2011).
21 Vannier, J. B., Sarek, G. & Boulton, S. J. RTEL1: functions of a disease-associated helicase. *Trends Cell Biol* **24**, 416-425, doi:10.1016/j.tcb.2014.01.004 (2014).
22 Peters, M. J. *et al.* The transcriptional landscape of age in human peripheral blood. *Nat Commun* **6**, 8570, doi:10.1038/ncomms9570 (2015).
23 Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet* **51**, 1339-1348, doi:10.1038/s41588-019-0481-0 (2019).
24 Fitzpatrick, A. L. *et al.* Leukocyte telomere length and cardiovascular disease in the cardiovascular health study. *Am J Epidemiol* **165**, 14-21, doi:10.1093/aje/kwj346 (2007).
25 Zhao, J., Miao, K., Wang, H., Ding, H. & Wang, D. W. Association between telomere length and type 2 diabetes mellitus: a meta-analysis. *PLoS One* **8**, e79993, doi:10.1371/journal.pone.0079993 (2013).

26 Houben, J. M., Moonen, H. J., van Schooten, F. J. & Hageman, G. J. Telomere length assessment: biomarker of chronic oxidative stress? *Free Radic Biol Med* **44**, 235-246, doi:10.1016/j.freeradbiomed.2007.10.001 (2008).

27 Zierer, J. *et al.* Metabolomics profiling reveals novel markers for leukocyte telomere length. *Aging (Albany NY)* **8**, 77-94, doi:10.18632/aging.100874 (2016).

28 Hishikawa, D., Hashidate, T., Shimizu, T. & Shindou, H. Diversity and function of membrane glycerophospholipids generated by the remodeling pathway in mammalian cells. *J Lipid Res* **55**, 799-807, doi:10.1194/jlr.R046094 (2014).

29 Homocysteine Studies, C. Homocysteine and risk of ischemic heart disease and stroke: a meta-analysis. *JAMA* **288**, 2015-2022 (2002).

30 Wald, D. S., Law, M. & Morris, J. K. Homocysteine and cardiovascular disease: evidence on causality from a meta-analysis. *BMJ* **325**, 1202 (2002).

31 Smith, A. D. *et al.* Homocysteine-lowering by B vitamins slows the rate of accelerated brain atrophy in mild cognitive impairment: a randomized controlled trial. *PLoS One* **5**, e12244, doi:10.1371/journal.pone.0012244 (2010).

32 Smith, A. D. *et al.* Homocysteine and Dementia: An International Consensus Statement. *J Alzheimers Dis* **62**, 561-570, doi:10.3233/JAD-171042 (2018).

33 Morris, M. S. Homocysteine and Alzheimer's disease. *Lancet Neurol* **2**, 425-428 (2003).

34 van Meurs, J. B. *et al.* Homocysteine levels and the risk of osteoporotic fracture. *N Engl J Med* **350**, 2033-2041, doi:10.1056/NEJMoa032546 (2004).

35 Perna, A. F. & Ingrosso, D. Homocysteine and chronic kidney disease: an ongoing narrative. *J Nephrol* **32**, 673-675, doi:10.1007/s40620-019-00622-1 (2019).

36 Smulders, Y. M. *et al.* Fasting and post-methionine homocysteine levels in NIDDM. Determinants and correlations with retinopathy, albuminuria, and cardiovascular disease. *Diabetes Care* **22**, 125-132 (1999).

37 Buysschaert, M., Dramais, A. S., Wallemacq, P. E. & Hermans, M. P. Hyperhomocysteinemia in type 2 diabetes: relationship to macroangiopathy, nephropathy, and insulin resistance. *Diabetes Care* **23**, 1816-1822 (2000).

38 Joshi, M. B. *et al.* Elevated homocysteine levels in type 2 diabetes induce constitutive neutrophil extracellular traps. *Sci Rep* **6**, 36362, doi:10.1038/srep36362 (2016).

39 Richards, J. B. *et al.* Homocysteine levels and leukocyte telomere length. *Atherosclerosis* **200**, 271-277, doi:10.1016/j.atherosclerosis.2007.12.035 (2008).

40 Xu, D., Neville, R. & Finkel, T. Homocysteine accelerates endothelial cell senescence. *FEBS Lett* **470**, 20-24 (2000).

41 Ostrakhovitch, E. A. & Tabibzadeh, S. Homocysteine and age-associated disorders. *Ageing Res Rev* **49**, 144-164, doi:10.1016/j.arr.2018.10.010 (2019).

42 Jacques, P. F. *et al.* Determinants of plasma total homocysteine concentration in the Framingham Offspring cohort. *Am J Clin Nutr* **73**, 613-621, doi:10.1093/ajcn/73.3.613 (2001).

43 Malaguarnera, M. *et al.* Elevated plasma total homocysteine in centenarians. *Clin Chem Lab Med* **42**, 307-310, doi:10.1515/CCLM.2004.056 (2004).

44 Schwaninger, M. *et al.* Elevated plasma concentrations of homocysteine in antiepileptic drug treatment. *Epilepsia* **40**, 345-350, doi:10.1111/j.1528-1157.1999.tb00716.x (1999).

45 Kulkarni, K. & Richard, B. C. Lifestyle, homocysteine, and the metabolic syndrome. *Metab Syndr Relat Disord* **1**, 141-147, doi:10.1089/154041903322294461 (2003).

46 Levi, A. *et al.* Elevated serum homocysteine is a predictor of accelerated decline in renal function and chronic kidney disease: A historical prospective study. *Eur J Intern Med* **25**, 951-955, doi:10.1016/j.ejim.2014.10.014 (2014).

47 O'Sullivan, E. D., Hughes, J. & Ferenbach, D. A. Renal Aging: Causes and Consequences. *J Am Soc Nephrol* **28**, 407-420, doi:10.1681/ASN.2015121308 (2017).

48 Soininen, P., Kangas, A. J., Wurtz, P., Suna, T. & Ala-Korpela, M. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circ Cardiovasc Genet* **8**, 192-206, doi:10.1161/CIRCGENETICS.114.000216 (2015).

49 Deelen, J. *et al.* A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals. *Nature Communications* **10**, 3346, doi:10.1038/s41467-019-11311-9 (2019).

50 Mozaffarian, D., Micha, R. & Wallace, S. Effects on coronary heart disease of increasing polyunsaturated fat in place of saturated fat: a systematic review and meta-analysis of randomized controlled trials. *PLoS Med* **7**, e1000252, doi:10.1371/journal.pmed.1000252 (2010).

51 Farvid, M. S. *et al.* Dietary linoleic acid and risk of coronary heart disease: a systematic review and meta-analysis of prospective cohort studies. *Circulation* **130**, 1568-1578, doi:10.1161/CIRCULATIONAHA.114.010236 (2014).

52 Marklund, M. *et al.* Biomarkers of Dietary Omega-6 Fatty Acids and Incident Cardiovascular Disease and Mortality. *Circulation* **139**, 2422-2436, doi:10.1161/CIRCULATIONAHA.118.038908 (2019).

53 Sacks, F. M. *et al.* Dietary Fats and Cardiovascular Disease: A Presidential Advisory From the American Heart Association. *Circulation* **136**, e1-e23, doi:10.1161/CIR.0000000000000510 (2017).

54 Harris, W. S. *et al.* Omega-6 fatty acids and risk for cardiovascular disease: a science advisory from the American Heart Association Nutrition Subcommittee of the Council on Nutrition, Physical Activity, and Metabolism; Council on Cardiovascular Nursing; and Council on Epidemiology and Prevention. *Circulation* **119**, 902-907, doi:10.1161/CIRCULATIONAHA.108.191627 (2009).

55 Gallois, A. *et al.* A comprehensive study of metabolite genetics reveals strong pleiotropy and heterogeneity across time and context. *Nat Commun* **10**, 4788, doi:10.1038/s41467-019-12703-7 (2019).

56 Deelen, J. *et al.* Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited. *Aging Cell* **10**, 686-698, doi:10.1111/j.1474-9726.2011.00705.x (2011).

57 Broer, L. (Genetic) Epidemiology of Aging. Erasmus University Rotterdam. Retrieved from http://hdl.handle.net/1765/50038. (2013).

58 Laviola, L., Natalicchio, A. & Giorgino, F. The IGF-I signaling pathway. *Curr Pharm Des* **13**, 663-669, doi:10.2174/138161207780249146 (2007).

59 Kenyon, C., Chang, J., Gensch, E., Rudner, A. & Tabtiang, R. A C. elegans mutant that lives twice as long as wild type. *Nature* **366**, 461-464, doi:10.1038/366461a0 (1993).

60 Bohni, R. *et al.* Autonomous control of cell and organ size by CHICO, a Drosophila homolog of vertebrate IRS1-4. *Cell* **97**, 865-875, doi:10.1016/s0092-8674(00)80799-0 (1999).

61 Kimura, K. D., Tissenbaum, H. A., Liu, Y. & Ruvkun, G. daf-2, an insulin receptor-like gene that regulates longevity and diapause in Caenorhabditis elegans. *Science* **277**, 942-946, doi:10.1126/science.277.5328.942 (1997).

62 Coschigano, K. T., Clemmons, D., Bellush, L. L. & Kopchick, J. J. Assessment of growth parameters and life span of GHR/BP gene-disrupted mice. *Endocrinology* **141**, 2608-2613, doi:10.1210/endo.141.7.7586 (2000).

63 Clancy, D. J. *et al.* Extension of life-span by loss of CHICO, a Drosophila insulin receptor substrate protein. *Science* **292**, 104-106, doi:10.1126/science.1057991 (2001).

64 Libina, N., Berman, J. R. & Kenyon, C. Tissue-specific activities of C. elegans DAF-16 in the regulation of lifespan. *Cell* **115**, 489-502, doi:10.1016/s0092-8674(03)00889-4 (2003).

65 Vitale, G., Barbieri, M., Kamenetskaya, M. & Paolisso, G. GH/IGF-I/insulin system in centenarians. *Mech Ageing Dev* **165**, 107-114, doi:10.1016/j.mad.2016.12.001 (2017).

66 Reddy, S. S. K. & Chaiban, J. T. The Endocrinology of Aging: A Key to Longevity "Great Expectations". *Endocr Pract* **23**, 1107-1116, doi:10.4158/EP171793.RA (2017).

67 Junnila, R. K., List, E. O., Berryman, D. E., Murrey, J. W. & Kopchick, J. J. The GH/IGF-1 axis in ageing and longevity. *Nat Rev Endocrinol* **9**, 366-376, doi:10.1038/nrendo.2013.67 (2013).

68 Broer, L. *et al.* GWAS of longevity in CHARGE consortium confirms APOE and FOXO3 candidacy. *J Gerontol A Biol Sci Med Sci* **70**, 110-118, doi:10.1093/gerona/glu166 (2015).

69 Sebastiani, P. *et al.* Genetic signatures of exceptional longevity in humans. *PLoS One* **7**, e29848, doi:10.1371/journal.pone.0029848 (2012).

70 Deelen, J. *et al.* A meta-analysis of genome-wide association studies identifies multiple longevity genes. *Nat Commun* **10**, 3669, doi:10.1038/s41467-019-11558-2 (2019).

71 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, doi:10.1016/S1474-4422(18)30053-X (2018).

72 Taudorf, L. *et al.* Declining incidence of dementia: A national registry-based study over 20 years. *Alzheimers Dement* **15**, 1383-1391, doi:10.1016/j.jalz.2019.07.006 (2019).

73 Gatz, M. *et al.* Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry* **63**, 168-174, doi:10.1001/archpsyc.63.2.168 (2006).

74 de Bruijn, R. F. *et al.* Insulin-like growth factor-I receptor stimulating activity is associated with dementia. *J Alzheimers Dis* **42**, 137-142, doi:10.3233/JAD-140186 (2014).

75 Bonafe, M. & Olivieri, F. Genetic polymorphism in long-lived people: cues for the presence of an insulin/IGF-pathway-dependent network affecting human longevity. *Mol. Cell. Endocrinol.* **299**, 118-123, doi:10.1016/j.mce.2008.10.038 (2009).

76 van Heemst, D. *et al.* Reduced insulin/IGF-1 signalling and human longevity. *Aging Cell* **4**, 79-85, doi:10.1111/j.1474-9728.2005.00148.x (2005).

77 Mao, K. *et al.* Late-life targeting of the IGF-1 receptor improves healthspan and lifespan in female mice. *Nat Commun* **9**, 2394, doi:10.1038/s41467-018-04805-5 (2018).

78 Djernes, J. K. Prevalence and predictors of depression in populations of elderly: a review. *Acta Psychiatr Scand* **113**, 372-387, doi:10.1111/j.1600-0447.2006.00770.x (2006).

79 Luijendijk, H. J. *et al.* Incidence and recurrence of late-life depression. *Arch Gen Psychiatry* **65**, 1394-1401, doi:10.1001/archpsyc.65.12.1394 (2008).

80 Copeland, J. R. *et al.* Depression among older people in Europe: the EURODEP studies. *World Psychiatry* **3**, 45-49 (2004).

81 Otte, C. *et al.* Major depressive disorder. *Nat Rev Dis Primers* **2**, 16065, doi:10.1038/nrdp.2016.65 (2016).

82 MacDonald, K. *et al.* Biomarkers for major depressive and bipolar disorders using metabolomics: A systematic review. *Am J Med Genet B Neuropsychiatr Genet* **180**, 122-137, doi:10.1002/ajmg.b.32680 (2019).

83 Dunlop, B. W. *et al.* Predictors of remission in depression to individual and combined treatments (PReDICT): study protocol for a randomized controlled trial. *Trials* **13**, 106, doi:10.1186/1745-6215-13-106 (2012).

84 Williams, H. R. *et al.* Differences in gut microbial metabolism are responsible for reduced hippurate synthesis in Crohn's disease. *BMC Gastroenterol* **10**, 108, doi:10.1186/1471-230X-10-108 (2010).

85 Walsh, M. C. *et al.* Influence of acute phytochemical intake on human urinary metabolomic profiles. *Am J Clin Nutr* **86**, 1687-1693, doi:10.1093/ajcn/86.5.1687 (2007).

86 Pallister, T. *et al.* Hippurate as a metabolomic marker of gut microbiome diversity: Modulation by diet and relationship to metabolic syndrome. *Sci Rep* **7**, 13670, doi:10.1038/s41598-017-13722-4 (2017).

87 Lane, M. A. & Bailey, S. J. Role of retinoid signalling in the adult brain. *Prog Neurobiol* **75**, 275-293, doi:10.1016/j.pneurobio.2005.03.002 (2005).

88 Mey, J. & McCaffery, P. Retinoic acid signaling in the nervous system of adult vertebrates. *Neuroscientist* **10**, 409-421, doi:10.1177/1073858404263520 (2004).

89 O'Reilly, K., Bailey, S. J. & Lane, M. A. Retinoid-mediated regulation of mood: possible cellular mechanisms. *Exp Biol Med (Maywood)* **233**, 251-258, doi:10.3181/0706-MR-158 (2008).

90 de Oliveira, M. R., da Rocha, R. F., Pasquali, M. A. & Moreira, J. C. The effects of vitamin A supplementation for 3 months on adult rat nigrostriatal axis: increased monoamine oxidase enzyme activity, mitochondrial redox dysfunction, increased beta-amyloid(1-40) peptide and TNF-alpha contents, and susceptibility of mitochondria to an in vitro H2O2 challenge. *Brain Res Bull* **87**, 432-444, doi:10.1016/j.brainresbull.2012.01.005 (2012).

91 Bremner, J. D., Shearer, K. D. & McCaffery, P. J. Retinoic acid and affective disorders: the evidence for an association. *J Clin Psychiatry* **73**, 37-50, doi:10.4088/JCP.10r05993 (2012).

92 Ludot, M., Mouchabac, S. & Ferreri, F. Inter-relationships between isotretinoin treatment and psychiatric disorders: Depression, bipolar disorder, anxiety, psychosis and suicide risks. *World J Psychiatry* **5**, 222-227, doi:10.5498/wjp.v5.i2.222 (2015).

93 Lončar, M. *et al.* Coumarins in Food and Methods of Their Determination. *Foods* **9**, doi:10.3390/foods9050645 (2020).

94 Gulbins, E. *et al.* Acid sphingomyelinase-ceramide system mediates effects of antidepressant drugs. *Nat Med* **19**, 934-938, doi:10.1038/nm.3214 (2013).

95 Muller, C. P. *et al.* Brain membrane lipids in major depression and anxiety disorders. *Biochim Biophys Acta* **1851**, 1052-1065, doi:10.1016/j.bbalip.2014.12.014 (2015).

96 Kunutsor, S. K., Seidu, S. & Khunti, K. Depression, antidepressant use, and risk of venous thromboembolism: systematic review and meta-analysis of published observational evidence. *Ann Med* **50**, 529-537, doi:10.1080/07853890.2018.1500703 (2018).

97 Lee, C. W. *et al.* Depression and risk of venous thromboembolism: a population-based retrospective cohort study. *Psychosom Med* **77**, 591-598, doi:10.1097/PSY.0000000000000193 (2015).

98 Loffelholz, K., Klein, J. & Koppen, A. Choline, a precursor of acetylcholine and phospholipids in the brain. *Prog Brain Res* **98**, 197-200, doi:10.1016/s0079-6123(08)62399-7 (1993).

99 Hirsch, M. J. & Wurtman, R. J. Lecithin consumption increases acetylcholine concentrations in rat brain and adrenal gland. *Science* **202**, 223-225, doi:10.1126/science.694529 (1978).

100 Magil, S. G., Zeisel, S. H. & Wurtman, R. J. Effects of ingesting soy or egg lecithins on serum choline, brain choline and brain acetylcholine. *J Nutr* **111**, 166-170, doi:10.1093/jn/111.1.166 (1981).

101 Janowsky, D. S., el-Yousef, M. K. & Davis, J. M. Acetylcholine and depression. *Psychosom Med* **36**, 248-257, doi:10.1097/00006842-197405000-00008 (1974).

102 Zumbe, A., Lee, A. & Storey, D. Polyols in confectionery: the route to sugar-free, reduced sugar and reduced calorie confectionery. *Br J Nutr* **85 Suppl 1**, S31-45, doi:10.1079/bjn2000260 (2001).

103 Muir, J. G. *et al.* Measurement of short-chain carbohydrates in common Australian vegetables and fruits by high-performance liquid chromatography (HPLC). *J Agric Food Chem* **57**, 554-565, doi:10.1021/jf802700e (2009).

104 Ledochowski, M., Widner, B., Bair, H., Probst, T. & Fuchs, D. Fructose- and sorbitol-reduced diet improves mood and gastrointestinal disturbances in fructose malabsorbers. *Scand J Gastroenterol* **35**, 1048-1052, doi:10.1080/003655200451162 (2000).

105 Cheung, S. G. *et al.* Systematic Review of Gut Microbiota and Major Depression. *Front Psychiatry* **10**, 34, doi:10.3389/fpsyt.2019.00034 (2019).

106 Hess, S. *et al.* Decreased serum L-arginine and L-citrulline levels in major depression. *Psychopharmacology (Berl)* **234**, 3241-3247, doi:10.1007/s00213-017-4712-8 (2017).

107 Saldanha, D., Kumar, N., Ryali, V., Srivastava, K. & Pawar, A. A. Serum Serotonin Abnormality in Depression. *Med J Armed Forces India* **65**, 108-112, doi:10.1016/S0377-1237(09)80120-2 (2009).

108 Liu, J. *et al.* Integration of epidemiologic, pharmacologic, genetic and gut microbiome data in a drug-metabolite atlas. *Nat Med* **26**, 110-117, doi:10.1038/s41591-019-0722-x (2020).

109 Schwarz, E. & Bahn, S. The utility of biomarker discovery approaches for the detection of disease mechanisms in psychiatric disorders. *Br J Pharmacol* **153 Suppl 1**, S133-136, doi:10.1038/sj.bjp.0707658 (2008).

110 Sethi, S. & Brietzke, E. Omics-Based Biomarkers: Application of Metabolomics in Neuropsychiatric Disorders. *Int J Neuropsychopharmacol* **19**, pyv096, doi:10.1093/ijnp/pyv096 (2015).

111 Fava, M. & Davidson, K. G. Definition and epidemiology of treatment-resistant depression. *Psychiatr Clin North Am* **19**, 179-200, doi:10.1016/s0193-953x(05)70283-5 (1996).

112 Papakostas, G. I. Managing partial response or nonresponse: switching, augmentation, and combination strategies for major depressive disorder. *J Clin Psychiatry* **70 Suppl 6**, 16-25, doi:10.4088/JCP.8133su1c.03 (2009).

113 Espiritu, J. R. Aging-related sleep changes. *Clin Geriatr Med* **24**, 1-14, v, doi:10.1016/j.cger.2007.08.007 (2008).

114 Medic, G., Wille, M. & Hemels, M. E. Short- and long-term health consequences of sleep disruption. *Nat Sci Sleep* **9**, 151-161, doi:10.2147/NSS.S134864 (2017).

115 Partinen, M., Kaprio, J., Koskenvuo, M., Putkonen, P. & Langinvainio, H. Genetic and environmental determination of human sleep. *Sleep* **6**, 179-185, doi:10.1093/sleep/6.3.179 (1983).

116 Heath, A. C., Kendler, K. S., Eaves, L. J. & Martin, N. G. Evidence for genetic influences on sleep disturbance and sleep pattern in twins. *Sleep* **13**, 318-335 (1990).

117 Klei, L. *et al.* Heritability of morningness-eveningness and self-report sleep measures in a family-based sample of 521 hutterites. *Chronobiol Int* **22**, 1041-1054, doi:10.1080/07420520500397959 (2005).

118 Geisler, P. *et al.* The influence of age and sex on sleep latency in the MSLT-30--a normative study. *Sleep* **29**, 687-692, doi:10.1093/sleep/29.5.687 (2006).

119 Rundo, J. V. & Downey, R., 3rd. Polysomnography. *Handb Clin Neurol* **160**, 381-392, doi:10.1016/B978-0-444-64032-1.00025-4 (2019).

120 de Souza, L. *et al.* Further validation of actigraphy for sleep studies. *Sleep* **26**, 81-85, doi:10.1093/sleep/26.1.81 (2003).

121 Tobler, I. & Borbely, A. A. European isolation and confinement study. Twenty-four hour rhythm of rest/activity and sleep/wakefulness: comparison of subjective and objective measures. *Adv Space Biol Med* **3**, 163-183 (1993).

122 Kripke, D. F. *et al.* Genetic variants associated with sleep disorders. *Sleep Med* **16**, 217-224, doi:10.1016/j.sleep.2014.11.003 (2015).

123 Jones, S. E. *et al.* Genetic studies of accelerometer-based sleep measures yield new insights into human sleep behaviour. *Nat Commun* **10**, 1585, doi:10.1038/s41467-019-09576-1 (2019).

124 Mehra, R. *et al.* Prevalence and correlates of sleep-disordered breathing in older men: osteoporotic fractures in men sleep study. *J Am Geriatr Soc* **55**, 1356-1364, doi:10.1111/j.1532-5415.2007.01290.x (2007).
125 Senaratna, C. V. *et al.* Prevalence of obstructive sleep apnea in the general population: A systematic review. *Sleep Med Rev*, doi:10.1016/j.smrv.2016.07.002 (2016).
126 Young, T., Peppard, P. E. & Gottlieb, D. J. Epidemiology of obstructive sleep apnea: a population health perspective. *Am J Respir Crit Care Med* **165**, 1217-1239 (2002).
127 Buysse, D. J., Reynolds, C. F., 3rd, Monk, T. H., Berman, S. R. & Kupfer, D. J. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res* **28**, 193-213, doi:0165-1781(89)90047-4 [pii] (1989).
128 Palmer, L. J. *et al.* A whole-genome scan for obstructive sleep apnea and obesity. *Am J Hum Genet* **72**, 340-350, doi:10.1086/346064 (2003).
129 Stelzer, G. *et al.* The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr Protoc Bioinformatics* **54**, 1 30 31-31 30 33, doi:10.1002/cpbi.5 (2016).
130 Ryan, S., Taylor, C. T. & McNicholas, W. T. Selective activation of inflammatory pathways by intermittent hypoxia in obstructive sleep apnea syndrome. *Circulation* **112**, 2660-2667, doi:10.1161/CIRCULATIONAHA.105.556746 (2005).
131 Garvey, J. F., Taylor, C. T. & McNicholas, W. T. Cardiovascular disease in obstructive sleep apnoea syndrome: the role of intermittent hypoxia and inflammation. *Eur Respir J* **33**, 1195-1205, doi:10.1183/09031936.00111208 (2009).
132 Mukherjee, S., Saxena, R. & Palmer, L. J. The genetics of obstructive sleep apnoea. *Respirology* **23**, 18-27, doi:10.1111/resp.13212 (2018).
133 Iglesias, A. I. *et al.* Haplotype reference consortium panel: Practical implications of imputations with large reference panels. *Hum Mutat* **38**, 1025-1032, doi:10.1002/humu.23247 (2017).
134 Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *bioRxiv*, 563866, doi:10.1101/563866 (2019).
135 Kowalski, M. H. *et al.* Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. *bioRxiv*, 683201, doi:10.1101/683201 (2019).
136 Liu, D. J. *et al.* Meta-analysis of gene-level tests for rare variant association. *Nat Genet* **46**, 200-204, doi:10.1038/ng.2852 (2014).
137 Li, B. & Leal, S. M. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet* **83**, 311-321, doi:10.1016/j.ajhg.2008.06.024 (2008).
138 Cawthon, R. M. Telomere measurement by quantitative PCR. *Nucleic Acids Res* **30**, e47 (2002).
139 Lai, T. P., Wright, W. E. & Shay, J. W. Comparison of telomere length measurement methods. *Philos Trans R Soc Lond B Biol Sci* **373**, doi:10.1098/rstb.2016.0451 (2018).
140 Demanelis, K. *et al.* Determinants of telomere length across human tissues. *Science* **369**, doi:10.1126/science.aaz6876 (2020).
141 Lin, J., Smith, D. L., Esteves, K. & Drury, S. Telomere length measurement by qPCR - Summary of critical factors and recommendations for assay design. *Psychoneuroendocrinology* **99**, 271-278, doi:10.1016/j.psyneuen.2018.10.005 (2019).
142 Broer, L. *et al.* Meta-analysis of telomere length in 19,713 subjects reveals high heritability, stronger maternal inheritance and a paternal age effect. *Eur J Hum Genet* **21**, 1163-1168, doi:10.1038/ejhg.2012.303 (2013).
143 Lewinsohn, P. M., Seeley, J. R., Roberts, R. E. & Allen, N. B. Center for Epidemiologic Studies Depression Scale (CES-D) as a screening instrument for depression among community-residing older adults. *Psychol Aging* **12**, 277-287 (1997).
144 Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci* **22**, 343-352, doi:10.1038/s41593-018-0326-7 (2019).
145 Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826, doi:10.1038/s41467-017-01261-5 (2017).
146 Lawlor, D. A., Harbord, R. M., Sterne, J. A., Timpson, N. & Davey Smith, G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med* **27**, 1133-1163, doi:10.1002/sim.3034 (2008).

147 Story Jovanova, O. *et al.* DNA Methylation Signatures of Depressive Symptoms in Middle-aged and Elderly Persons: Meta-analysis of Multiethnic Epigenome-wide Studies. *JAMA psychiatry* **75**, 949-959, doi:10.1001/jamapsychiatry.2018.1725 (2018).
148 Carlson, C. S. *et al.* Generalization and dilution of association results from European GWAS in populations of non-European ancestry: the PAGE study. *PLoS Biol* **11**, e1001661, doi:10.1371/journal.pbio.1001661 (2013).
149 Ikram, M. A. *et al.* Objectives, design and main findings until 2020 from the Rotterdam Study. *Eur J Epidemiol* **35**, 483-517, doi:10.1007/s10654-020-00640-5 (2020).
150 VanderWeele, T. J., Jackson, J. W. & Li, S. Causal inference and longitudinal data: a case study of religion and mental health. *Soc Psychiatry Psychiatr Epidemiol* **51**, 1457-1466, doi:10.1007/s00127-016-1281-9 (2016).
151 Liu, J. *et al.* A Mendelian Randomization Study of Metabolite Profiles, Fasting Glucose, and Type 2 Diabetes. *Diabetes* **66**, 2915-2926, doi:10.2337/db17-0199 (2017).
152 Telomeres Mendelian Randomization, C. *et al.* Association Between Telomere Length and Risk of Cancer and Non-Neoplastic Diseases: A Mendelian Randomization Study. *JAMA Oncol* **3**, 636-651, doi:10.1001/jamaoncol.2016.5945 (2017).
153 Greenberg, P. E., Fournier, A. A., Sisitsky, T., Pike, C. T. & Kessler, R. C. The economic burden of adults with major depressive disorder in the United States (2005 and 2010). *J Clin Psychiatry* **76**, 155-162, doi:10.4088/JCP.14m09298 (2015).

CHAPTER 5

Summary/Samenvatting

Summary

The world's population is aging, and age is the largest risk factor for the development of many common diseases such as cardiovascular disease, type 2 diabetes, various cancers, and neurodegenerative disorders like dementia and Alzheimer's disease (AD). The aging process is complex and includes both calendar age and biological age. Although there are many biological and genetic processes involved, the molecular mechanisms underlying aging remain largely unknown. Therefore, in this thesis, I have used two omics approaches to study genetic variants and circulating metabolites in blood (i.e., genomics and metabolomics) and their association with age-related outcomes using data from multiple cohorts across the world. In this chapter, I will provide a summary of the findings of this thesis. Ultimately, when proven, interventions targeting underlying molecular mechanisms may lead to healthier aging.

In the Introduction (**Chapter 1)**, I have provided the background of this PhD thesis. The studies provided in **Chapter 2** show the associations of genetic variants and circulating metabolites with telomere length, a biomarker for biological aging. The aim of these studies was to provide a better understanding of the molecular mechanisms involved in telomere length regulation. In addition, I have studied the relationship between the insulin-like growth factor I (IGF-I), a biomarker for aging, and the development of dementia and AD. In **Chapter 2.1,** I investigated the association of rare genetic variants with telomere length. I found nine rare genetic variants associated with shorter telomere length in the family-based Erasmus Rucphen Family (ERF) study, of which eight were located in a region on chromosome 11 that contains the *ATM* gene. *ATM* is known to be involved in the regulation of processes involved in telomere maintenance. These rare variants seem to be specific for the ERF study, as I was not able to replicate these variants in two other cohorts. Moreover, the segregation analysis showed that all variants segregate with shorter telomere length in one specific ERF family. Additionally, by performing a meta-analysis including all three studies I found nominal significant associations of rare variants in the *RPL8* and *RTEL1* genes with telomere length, which, together with *ATM*, require follow-up in future studies. In **Chapter 2.2,** I studied the association of metabolites with telomere length using data of seven cohorts that are part of the ENGAGE (European Network for Genetic and Genomic Epidemiology) consortium. I identified six metabolites (i.e., lysophosphatidylcholine acyl C17:0, methionine, tyrosine, phosphatidylcholine diacyl C32:1, hydroxypropionylcarnitine, and phosphatidylcholine acyl-alkyl C38:4) significantly associated with telomere length. These metabolites are involved in inflammation, oxidative stress, and lipid metabolism and have been implicated in cardiovascular disease and in diabetes. Pathway analysis additionally revealed the involvement of the homocysteine metabolism in telomere biology. In **Chapter 2.3,** I showed the results of a second study investigating the association of metabolites with telomere length. In this study, other metabolites were investigated than those in **Chapter 2.2** by using a different platform for metabolite quantification. This study included data from six cohorts which are part of the BBMRI-NL (Biobanking and Biomolecular Research Infrastructure, the

Netherlands) consortium. I found four metabolite ratios positively associated with telomere length: two cholesterol to lipid ratios and two omega-6 fatty acid ratios. I replicated the ratio of omega-6 fatty acids to total fatty acids (FAw6/FA) in an independent dataset. As these ratios have not been the focus of previous research, this study provided important novel knowledge. Finally, in **Chapter 2.4**, I examined the robustness of the previously found association of IGF-I receptor stimulating activity (i.e., a measure of biological availability of IGF-I) with dementia and AD in the prospective Rotterdam Study cohort. To this end, I conducted an analysis extending the follow-up period of the Rotterdam Study participants from 12 to 16 years. In this analysis, the significant association of IGF-I receptor stimulating activity with dementia was no longer observed. However, an increase of IGF-I receptor stimulating activity was associated with dementia and AD in carriers of the apolipoprotein ε4 (*APOE*- ε4) allele, but not in non-carriers. The *APOE*-ε4 genotype is known to increase risk of AD. We also found that *APOE*-ε44 carriers had the lowest IGF-I receptor stimulating activity, supporting the hypothesis that low IGF-I activity increases the risk of dementia. However, the findings of this study require replication in other cohorts.

In **Chapter 3** of this thesis, I studied the association of metabolites with late-onset depression and the association of genetic variants with two sleep outcomes: sleep latency and symptoms of sleep apnea. In **Chapter 3.1,** I provided novel insights into the association of circulating metabolites with depression. In this study, we investigated data from five independent cohorts and 806 metabolites. We found that 53 metabolites were significantly associated with depression and we showed that the associations of most metabolites were explained by the use of antidepressants, including previously found metabolites in the monoamine and neurotransmitter pathways (serotonin, glutamate, kynurenine, and leucine). Eight metabolites including retinol (vitamin A), 1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1), mannitol/sorbitol, hippurate, 4-hydroxycoumarin, 2-aminooctanoate, 10-undecenoate (11:1n1), and 1-linoleoyl-GPA (18:2) were associated with depression, independently of medication use, smoking and body mass index. Additionally, further analyses showed that hippurate may be causally related with depression. Besides new important insides on the involvement of metabolites in depression pathophysiology, this study suggests that dietary interventions may prevent depression and this hypothesis should be further tested in clinical trials. In **Chapter 3.2**, I studied the association of common genetic variants with sleep latency (i.e., the time it takes to fall asleep) using data from seven European cohorts. I identified a cluster of three highly correlated variants in the *RBFOX3* gene associated with sleep latency, which was replicated in an independent dataset. A gene network analysis showed expression of *RBFOX3* in plausible tissues such as the brain and nervous system. This analysis provided further evidence for the expression of *RBFOX3* in brain regions that are important for sleep regulation. These findings suggest *RBFOX3* as a candidate gene for sleep latency. Although this finding should be studied more thoroughly, larger studies have implicated *RBFOX3* in alcohol dependence, a trait known to be associated with sleep. Lastly, in **Chapter 3.3**, I studied the association of rare genetic variants with symptoms of sleep apnea. By performing a meta-analysis, I found a rare genetic variant positively associated with symptoms of sleep

apnea and subsequently replicated this finding in an independent dataset. This rare variant is located in the overlapping 3' untranslated regions, which are not protein coding but involved in gene expression regulation, of the *ERCC1* and *CD3EAP* genes. Both genes are likely candidate genes for sleep apnea based on their expression in tissues such as the tongue, muscles, cartilage and trachea and the biological pathways they are involved in, such as immune-related pathways.

In the last chapter of this thesis (**Chapter 4**), I presented an overview, discussed the findings of this thesis and I provided directions for future research. To conclude, the findings of this thesis contribute to a better understanding of the molecular mechanisms involved in age-related outcomes. When replicated, the presented findings will provide targets for earlier diagnosis, new treatments and new prevention strategies for age-related diseases to improve healthy aging.

Samenvatting

De bevolking is aan het verouderen en veroudering is de grootste risicofactor voor de ontwikkeling van een scala aan veel voorkomende ziekten zoals hart- en vaatziekten, diabetes type 2, kanker, en neurologische aandoeningen waaronder dementie en de ziekte van Alzheimer. Veroudering is een complex begrip en veel breder dan alleen de kalenderleeftijd. Er liggen verschillende biologische en genetische processen aan ten grondslag, maar de onderliggende moleculaire mechanismen zijn nog onvoldoende bekend. Daarom heb ik in dit proefschrift de associatie van genetische varianten en metabolieten in het bloed met leeftijd-gerelateerde uitkomstmaten onderzocht. Hiervoor heb ik twee omics methodes toegepast, genomics en metabolomics, en gegevens uit studies van over de hele wereld gebruikt. In dit hoofdstuk zal ik een samenvatting geven van de bevindingen van de onderzoeken die beschreven zijn in dit proefschrift. Uiteindelijk kunnen interventies gericht op de betrokken moleculaire mechanismen mogelijk leiden tot gezondere veroudering.

In de inleiding (**Hoofdstuk 1**) heb ik de achtergrond van dit proefschrift beschreven. In de studies in **Hoofdstuk 2** onderzocht ik genetische varianten en metabolieten in bloed die samenhangen met telomeerlengte, een biomarker voor biologische veroudering. Het doel van deze studie was om meer inzicht te verkrijgen in de moleculaire mechanismen die betrokken zijn bij de regulatie van telomeerlengte. Daarnaast bestudeerde ik de relatie tussen IGF-I (insulin-like growth factor I), een biomarker voor veroudering, en het ontstaan van dementie en de ziekte van Alzheimer. In **Hoofdstuk 2.1** onderzocht ik de samenhang tussen zeldzame genetische varianten en telomeerlengte. In de Erasmus Rucphen Familie (ERF) studie, een familiestudie, vond ik negen zeldzame genetische varianten die samenhingen met kortere telomeerlengte, een risicofactor voor veel voorkomende verouderingsziekten. Acht van deze varianten liggen in een gebied op chromosoom 11 dat het *ATM*-gen bevat. *ATM* staat bekend om zijn rol in de regulatie van processen die een effect hebben op het onderhoud van telomeren. Omdat ik deze associaties niet terugvond in twee andere studies, lijken deze varianten specifiek een effect te hebben in deelnemers van de ERF studie. Vervolgens toonde ik aan dat overerving van deze genetische varianten samen met kortere telomeerlengte specifiek is voor één ERF familie. In een meta-analyse van alle drie de studies samen vond ik daarnaast nominaal significante associaties van twee zeldzame genetische varianten in de *RPL8* en *RTEL1* genen met telomeerlengte. Samen met *ATM* zouden deze genen nader onderzocht moeten worden in vervolgstudies. In **Hoofdstuk 2.2** bestudeerde ik het verband tussen metabolieten en telomeerlengte door gebruik te maken van zeven cohortstudies binnen het ENGAGE (European Network for Genetic and Genomic Epidemiology) consortium. Ik vond significant bewijs voor de associatie van zes metabolieten (lysophosphatidylcholine acyl C17:0, methionine, tyrosine, phosphatidylcholine diacyl C32:1, hydroxypropionylcarnitine, en phosphatidylcholine acyl-alkyl C38:4) met telomeerlengte. Deze metabolieten zijn betrokken bij ontstekingsreacties, oxidatieve stress en de vetstofwisseling; processen die een rol spelen bij onder andere hart- en vaatziekten en

diabetes. Verdere analyse van de biologische processen waar deze metabolieten onderdeel van uitmaken liet zien dat de stofwisseling van homocysteïne betrokken is bij de biologische processen die telomeren reguleren. Ook in **Hoofdstuk 2.3** onderzocht ik deze relatie, maar met andere metabolieten die gemeten zijn met een ander metabolieten platform dan in **Hoofdstuk 2.2**. In deze studie analyseerde ik de gegevens van zes studies binnen het BBMRI-NL (Biobanking and Biomolecular Research Infrastructure, the Netherlands) consortium. Vier metaboliet ratio's hadden een significante positieve associatie met telomeerlengte: twee ratio's van cholesterol in vetten en twee ratio's van omega-6 vetzuren. In een onafhankelijke studie kon ik de associatie van de ratio van omega-6 vetzuren tot alle vetzuren (FAw6/FA) in het bloed repliceren. De resultaten van deze studie gaven belangrijke nieuwe kennis over de betrokkenheid van metabolieten en hun verhoudingen bij de biologie van telomeren. Deze nieuwe bevindingen vereisen de aandacht van toekomstige studies. Tot slot heb ik in **Hoofdstuk 2.4** de eerder geobserveerde associatie tussen de IGF-I receptor stimulerende activiteit, een maat van biologische beschikbaarheid van IGF-I, en dementie waaronder de ziekte van Alzheimer onderzocht. Deze eerdere studie was uitgevoerd in de Rotterdam Studie en in **Hoofdstuk 2.4** heb ik de follow-up periode verlengd van 12 tot 16 jaar. Ondanks dat we de significante associatie niet meer zagen, toonden wij wel aan dat een verhoging van de IGF-I receptor stimulerende activiteit leidde tot een verhoogd risico op het krijgen van dementie en de ziekte van Alzheimer. Dit was echter enkel in dragers van het Apolipoprotein ε4 (*APOE*-ε4) allel, wat bekend staat voor een verhoogd risico op het krijgen van de ziekte van Alzheimer. We vonden ook de laagste IGF-I receptor stimulerende activiteit in *APOE*-ε44 dragers. Dit ondersteund de hypothese dat lage IGF-I activiteit het risico op dementie verhoogd. Deze bevindingen moeten echter nog wel gerepliceerd worden in andere cohorten.

In **Hoofdstuk 3** van dit proefschrift onderzocht ik de relatie tussen metabolieten en depressie op latere leeftijd en de associatie van genetische varianten met twee slaap uitkomsten: slaaplatentie en slaap apnoe. In **Hoofdstuk 3.1** bracht ik nieuwe inzichten in de relatie tussen circulerende metabolieten en depressie aan het licht. In deze studie onderzochten we de gegevens van deelnemers van vijf cohortstudies en 806 metabolieten. We vonden een associatie tussen 53 metabolieten en depressie en we toonden aan dat de meeste associaties gedreven werden door het gebruik van antidepressiva. Dit gold ook voor de eerder gevonden metabolieten in de monoamine en neurotransmitter netwerken zoals serotonine, glutamaat, kynurenine, en leucine. Er waren acht metabolieten waarvan de associatie met depressie niet beïnvloed werd door antidepressiva, waaronder retinol (vitamine A), 1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1), mannitol/sorbitol, hippuraat, 4-hydroxycoumarine, 2-aminooctanoate, 10-undecenoate (11:1n1), en 1-linoleoyl-GPA (18:2). Daarnaast werden de associaties van deze metabolieten met depressie ook niet beïnvloed door medicatie, roken en body mass index. Naast belangrijke kennis over de betrokkenheid van metabolieten bij het ontstaan van depressie hebben deze metabolieten mogelijk ook een voorspellende functie. We vonden nieuwe aanknopingspunten voor de preventie van depressie die getoetst kunnen worden in klinisch experimenteel onderzoek. In **Hoofdstuk**

3.2 bestudeerde ik de associatie van veel voorkomende genetische varianten met slaaplatentie, de tijd die mensen nodig hebben om in slaap te vallen. Voor deze studie gebruikte ik gegevens van deelnemers uit zeven Europese cohorten. Ik vond een associatie van een cluster van drie gecorreleerde genetische varianten in het *RBFOX3* gen met slaaplatentie, wat gerepliceerd werd in een onafhankelijke dataset. Een gen netwerkanalyse liet expressie van het *RBFOX3* gen zien in biologisch aannemelijke weefsels zoals het brein en zenuwstelsel. Ook gaf dit bewijs voor de betrokkenheid van gebieden van het brein, waar het *RBFOX3* gen tot expressie komt, bij processen die belangrijk zijn voor slaapregulatie. Deze bevindingen suggereren dat *RBFOX3* een kandidaat gen is voor slaaplatentie. Vervolgonderzoek laat zien dat het gen een rol speelt in alcoholverslaving, een determinant van slaaplatentie. In **Hoofdstuk 3.3**, bestudeerde ik de associatie van zeldzame genetische varianten met symptomen van slaap apnoe in een meta-analyse. We vonden een zeldzame genetische variant die positief geassocieerd was met symptomen van slaap apnoe. Deze bevinding was gerepliceerd in een onafhankelijk cohort. De zeldzame variant bevond zich in de overlappende 3' untranslated gebieden van de *ERCC1* en *CD3EAP* genen. Deze gebieden coderen niet voor een eiwit maar zijn wel betrokken bij de regulatie van de expressie van genen. Beide zijn waarschijnlijke kandidaat genen voor slaap apnoe vanwege hun expressie in biologisch gezien relevante weefsels zoals de tong, spieren, kraakbeen en luchtpijp, maar ook vanwege de biologische netwerken waarbij ze betrokken zijn zoals bijvoorbeeld immuungerelateerde netwerken.

In het laatste hoofdstuk van dit proefschrift **(Hoofdstuk 4**) gaf ik een overzicht en bediscussieerde ik de bevindingen, sterke punten en uitdagingen van dit proefschrift en gaf ik aanwijzingen voor toekomstig onderzoek. De bevindingen van dit proefschrift dragen bij aan het beter begrijpen van de moleculaire mechanismen die betrokken zijn bij leeftijdsafhankelijke aandoeningen. De bevindingen geven nieuwe aanknopingspunten voor het verbeteren van de diagnostiek, nieuwe behandelingen en nieuwe preventiestrategieën van leeftijdsafhankelijke aandoeningen om gezonder ouder te worden.

CHAPTER 6

Appendix

CHAPTER 6.1

Acknowledgements/Dankwoord

Ik heb er jaren naar uitgekeken om dit stuk van mijn proefschrift te schrijven, en nu is het moment daar! Ik had mijn onderzoeken en dit proefschrift niet af kunnen ronden zonder de hulp en steun van vele mensen. Ik wil iedereen die hierbij betrokken is geweest hartelijk bedanken voor hun bijdrage. Ik hoop oprecht dat ik in dit dankwoord niemand vergeet te noemen (anders sorry!).

Allereerst wil ik mijn promotor **Prof. Cornelia van Duijn** bedanken. Beste Cornelia, bedankt voor alle kansen die u mij heeft gegeven tijdens mijn promotietraject. Ik heb me altijd verbaasd over uw energie en enthousiasme en dit heeft me gemotiveerd om hard te werken aan mijn projecten. Ik heb veel van u geleerd en ik ben u dankbaar voor uw steun tijdens de afronding van mijn proefschrift. Next, I would like to thank my co-promotor **Dr. Najaf Amin**. Dear Najaf, I am grateful for your support and guidance during my PhD. You always took time to help me with the, often statistical, questions I had. I learned so much from you. Cornelia and Najaf, I wish you and your families all the best in Oxford.

Daarnaast wil ik alle leden van de leescommissie bedanken. Ik waardeer het zeer dat jullie mijn proefschrift hebben beoordeeld. **Prof. Kamran Ikram**, we hebben niet direct samengewerkt tijdens mijn PhD, maar ik vind het een eer dat u plaats wilt nemen in de commissie. **Prof. Eline Slagboom**, ik vond het fijn om met u en uw team samen te werken in verschillende CHARGE longevity en BBMRI-NL metabolomics projecten. Bedankt dat u lid wilt zijn van de commissie! **Prof. André Uitterlinden**, ik heb mijn allereerste stage bij u in de groep gelopen (van mijn HLO opleiding) en dit heeft mij geïnspireerd om door te gaan in de genetica. Ik heb veel van u en uw groep geleerd, bijvoorbeeld tijdens de NIHES en MOLMED-cursussen en tijdens meetings, en ik vind het een mooie afsluiting dat u nu plaats wilt nemen in mijn commissie. Bedankt!

Ik wil ook graag de overige leden van de commissie bedanken. Ik waardeer het enorm dat jullie als opponenten willen deelnemen aan mijn verdediging. **Prof. Dorret Boomsma**, bedankt voor de prettige samenwerking in verschillende projecten. **Dr. Annemarie Luik,** bedankt voor de samenwerking en de fijne gesprekken die we gehad hebben tijdens mijn PhD. **Prof. Eric Sijbrands**, bedankt dat u plaats wilt nemen in de commissie.

Mijn paranimfen, Ester en Ivana, wat ben ik blij dat jullie allebei gelijk enthousiast waren toen ik jullie vroeg om paranimf te zijn bij mijn promotie. Ik had dit echt niet zonder jullie gekund! **Ester**, wat een geluk dat wij elkaar ontmoet hebben tijdens de NIHES summer school (op de eerste dag!). Je bent een hele goede vriendin en ik ben je onwijs dankbaar voor alle gesprekken, alle gezellige etentjes en drankjes en de leuke dingen die we gedaan hebben. Ik kijk uit naar alle leuke dingen die we nog gaan doen! **Ivana**, my roommate since day one from the master. We shared all parts of the PhD life, the fun and difficult parts. I have learned so much from you and I loved our talks about differences between Serbia and the Netherlands. It was amazing to see our families grow during our PhD trajectories and I hope we will have many more coffee and dinner dates in the future.

A special thanks to my colleagues from the GenEpi group, life would not have been the same without you: Adriana, Andrea, Andy, Ayse, Bernadette, Dina, Dream, Elena, Elisa, Jeannette, Hata, Maaike, Sara (Galle), Shahzad, and Sven. **Adriana**, onwijs bedankt voor alle koffiemomentjes en lunches tijdens en na mijn PhD. Ik verbaas me altijd over jouw energie, kennis en motivatie. Succes met je opleiding tot klinisch genetisch laboratoriumspecialist! **Ayse**, thank you for your help and support with the metabolomics studies and for the nice conversations we had about our families and our shared interest in watersports. **Dina**, thank you for all your help, advice, and nice (daily) conversations. I have learned so much from you and I am very grateful that you asked me to be your paranymph! Dear **Dream** (Jun), I am so happy to have worked with you. Traveling together to conferences and Ivana's wedding was so much fun. I am also grateful for the time we spent together in Barcelona and I hope we will stay in touch in the future. **Elena**, you are an amazing friend and you were always there to help and to have nice conversations. I loved the time we spent together in Split and I hope we will meet again soon. **Elisa**, ik vond het fijn om samen te werken en een kamer te delen. Jij hebt me altijd van adviezen voorzien waar ik nog steeds gebruik van maak ;) Ik vind het heel fijn dat we soms nog afspreken om bij te kletsen. **Hata**, nadat Sven weggegaan is, heb jij een deel van zijn werk overgenomen. Hieronder viel ook de dataverzameling voor de Erasmus Rucphen Familie (ERF) studie, waarvoor we samen veel huisartsen hebben bezocht en aardig wat uren in de auto hebben doorgebracht. Dankjewel voor je gezelligheid, ook bij de conferenties. Ik wens je het allerbeste met het afronden van je PhD en je toekomst als arts! **Maaike**, bedankt voor alles! Je hebt me vaak geholpen en stond altijd voor mij en de andere GenEpi's klaar. **Sara**, bedankt voor je gezelligheid en de samenwerking in Rucphen. Succes met de afronding van je PhD. **Shahzad**, we were roommates for a long time and I really enjoyed talking to you. A special thanks for your help with my scripts and bioinformatics analyses ;) I really appreciate your support during the finalization of my PhD. **Sven**, bedankt voor alle steun en hulp! Ik moet gelijk weer denken aan onze bezoekjes aan de huisartsen voor de dataverzameling van de ERF studie. We hebben heel wat dagen samen doorgebracht en ik ben blij dat ik dit met jou heb kunnen doen. Ook fijn dat jij net zo gek bent op koffie als ik ;) Jouw kijk op dingen was heel leerzaam en de schop die jij me soms onder mijn kont gaf heeft me echt geholpen om door te gaan.

Natuurlijk wil ik ook iedereen van het lab bedanken voor jullie hulp, steun en de gezelligheid! Ik vond het jammer dat we elkaar zo weinig zagen, maar gelukkig waren er de wekelijkse meetings en zagen we elkaar bij borrels en andere gelegenheden: **Andrea**, wat was het altijd gezellig samen. Ik vond het zo leuk om te horen hoe trots je bent op je kleinkind, lekker genieten! **Andy**, wij kennen elkaar al langer van mijn stage bij Inwendige Geneeskunde, het was altijd gezellig om je te zien en bij te praten. **Bernadette**, ik vond het altijd gezellig om je te spreken. Geniet lekker van je vrije tijd en je familie! **Jeannette**, enorm bedankt voor je hulp bij alles wat met de ERF studie te maken had. Zonder jouw hulp waren we zeker niet zo ver gekomen met de dataverzameling. Het was ook altijd gezellig om even bij te praten. Also, a special thanks to the GenEpi's I have only worked with for a short period: **Aaron**, **Constanza**, **Fizzah**, **Henriët**, **Maarten**, and **Sara**. I wish all the GenEpi's and their families all the best and I hope we will stay in touch!

Ik wil ook alle **deelnemers van de ERF studie en de Rotterdam Study/ERGO** bedanken, zonder jullie had dit onderzoek niet uitgevoerd kunnen worden. Daarnaast wil ik ook alle huisartsen in Rucphen en omstreken bedanken voor de hulp bij de dataverzameling van de ERF studie. Jullie hebben ons heel fijn ontvangen! Ik wil twee huisartsen in het bijzonder bedanken: **Pieter Snijders** en (in memoriam) **Peter van Wouw**.

Everybody that I met during the Master Genetic Epidemiology and from the department of Epidemiology: thank you for your support, the interesting meetings, the laughs and nice lunches. I can't name all of you but I will give it a try: **Annemarie**, **Chantal**, **Desi**, **Elena**, **Emmely**, **Fjorda**, **Frank**, **Gennady**, **Giorgia**, **Gwen**, **Henning**, **Hieab**, **Hoyan**, **Irene**, **Janine**, **Jelena**, **Kate**, **Kim**, **Laurence**, **Leila**, **Loes**, **Marieke**, **Marten**, **Milly**, **Mohsen**, **Natalie**, **Ruchika**, **Ryan**, **Sirwan**, **Strahinja**, and **Trudy**. A special thanks to **Prof. Albert Hofman** for setting up the Rotterdam Study and the amazing courses, it was very valuable to get the opportunity to learn from you. And to **Prof. Arfan Ikram**, it was an honor to work with you and I appreciate the nice conversations we had. I would also like to thank **Frank van Rooij** for taking care of the Rotterdam Study data, **Nano** for always being there for all the computer-related questions, and **Kris** for the talks and good coffee.

Also special thanks to the Internal Medicine group: **Carolina**, **Cindy**, **Djawad**, **Fernando**, **Jeroen**, **Joyce**, **Linda**, **Lisette**, **Michael**, **Pooja**, and **Robert**. **Linda**, bedankt voor je support en fijne gesprekken over de studies. Ik ging verder met wat jij gestart bent in de groep en het was fijn om te kunnen overleggen of om gewoon even te kletsen. **Lisette**, bedankt voor jouw steun tijdens en na mijn stage. Ik weet niet of ik over een PhD was gaan nadenken zonder jou! **Jeroen**, bedankt voor alle fijne gesprekken tijdens en na de cursussen die we samen gevolgd hebben. Ik hoop dat we nog af en toe koffie gaan drinken.

A special thanks to everybody I collaborated with from the **Cohorts for Heart Aging Research in Genomic Epidemiology (CHARGE) consortium**. A special thanks to the members of the longevity workgroup and **Prof. Joanne Murabito**: thank you for the nice collaboration during the past years! It was a pleasure to meet everybody during the annual CHARGE meetings and I am grateful for all opportunities that were provided and for everything that I learned from you all. The same is true for everybody that I collaborated with from the **Biobanking for Medical Research Infrastructure of the Netherlands (BBMRI-NL) consortium**. I really enjoyed the meetings and I am grateful that I had the opportunity to be involved in multiple metabolomics projects. It was truly a pleasure to work with you! **Dr. Sophie Warner**, **Dr. Veryan Codd**, **Dr. Christopher Nelson**, and **Prof. Sir Nilesh Samani**: thank you for the collaboration and for the opportunity to work on the telomere length project focusing on rare genetic variants in Leicester. **Sophie**, thank you for helping me and for the daily talks during my visit in Leicester. My Spanish friends at **Linkcare Health Sciences**, thank you for the wonderful and inspiring time in Barcelona. I fell in love with the beautiful city and the amazing food!

Mijn collega's bij **SkylineDx**, ik waardeer het dat jullie altijd geïnteresseerd zijn in mijn PhD en ik ben blij dat ik de kans heb om met jullie te werken, van jullie te leren en met jullie te lachen. In het bijzonder wil ik **Dennie**, **Deniz** en **Daniëlle** bedanken voor jullie steun tijdens het afronden van mijn proefschrift. **Dennie**, ik waardeer onze gesprekken en ik vind het heel fijn dat je altijd bereid was om even met me mee te denken! Ook wil ik graag mijn oud-collega's van de afdeling Immunologie bedanken voor jullie steun toen ik besloten had om de masteropleiding genetische epidemiologie te gaan volgen. Vanuit de afdeling Immunologie is er een vriendengroep ontstaan en helaas zien we elkaar veel te weinig: **Dennis & Claudia**, **Joyce & Stefan** en **Youri & Leoni**, bedankt voor jullie gezelligheid en het delen van vele mooie momenten. **Rianne**, ik vind onze vriendschap heel waardevol en ik hoop dat we elkaar regelmatig blijven zien.

Wouter & Myriam, **Niek & Stephania**, **Marcel & Miruna**, bedankt voor jullie vriendschap, de gezelligheid en steun! Ik waardeer onze vriendschap enorm. Ik hoop dat we snel weer lekker een weekendje weg gaan met z'n allen! **Eva** en **Else**, we zijn samen de masteropleiding doorgekomen en sindsdien hebben we nog steeds vaak contact. Ik ben heel erg blij met jullie vriendschap en ondanks dat we allemaal een andere kant op zijn gegaan ben ik ervan overtuigd dat we elkaar zullen blijven zien en onze kindjes lekker samen kunnen spelen. Bedankt voor jullie steun en adviezen. Tot snel! **Debora**, super bedankt voor jouw steun en peptalks. We kennen elkaar al vanaf onze tijd bij de Bijenkorf, maar onze vriendschap is daarna alleen maar hechter geworden. Ik ben trots op wat jij allemaal doet en ik wens jou het allerbeste in San Antonio. Ik hoop natuurlijk stiekem wel dat je uiteindelijk weer wat dichterbij komt wonen ;) **Myrte**, helaas zien we elkaar niet zoveel meer vanwege onze drukke schema's. Maar ik heb altijd genoten van onze gezellige koffiepauzes en ik vind het onwijs knap hoe jij alles aanpakt. Mijn vrienden, en ondertussen familie, van de **Kralingsche zeilclub**, bedankt voor alle afleiding, gezelligheid en leuke weekendjes! Ik hoop na het afronden van mijn PhD meer tijd te hebben om lekker te zeilen en te varen.

Mijn lieve familie, ik had dit niet kunnen doen zonder jullie steun. Lieve **papa**, jij bent altijd mijn steun en toeverlaat geweest. Je bent altijd geïnteresseerd geweest en je bent me altijd blijven steunen (met de nodige peptalks) en volgens mij ben jij een van de weinigen die dit boekje ook echt gaat lezen. **Corina**, bedankt voor de gesprekken en wat boffen wij met jou. Jullie zijn allebei zo lief voor Milan en Emma en staan altijd voor ons klaar! Lieve **mama**, helaas heb je dit gedeelte van mijn leven niet meer mee kunnen maken. Maar ik weet zeker dat het je niet zou verbazen dat ik dit ben gaan doen. Jij stond altijd voor mij klaar en hebt me altijd gemotiveerd om te doen wat ik wilde doen. Ik mis je veel meer dan ik hier kan schrijven en ik hoop dat je trots bent! Mijn "kleine" zusje **Jemaine**, tot hilariteit van mijn collega's belde je vaak tijdens mijn werk met de vraag "Ben je aan het werk?". Ik ben trots op wat je hebt bereikt en het is jammer dat jij en **Ramon** niet dichtbij wonen, want ik zou jullie en **Lizzy** graag vaker zien. Dikke knuffel voor jullie! Mijn lieve schoonfamilie: **Bram en Gerda**, wat een fijne schoonouders zijn jullie. Ondanks dat de onderzoekswereld ver van jullie afstond, wilden jullie wel altijd luisteren naar mijn verhalen en meedenken als ik het

niet meer zag zitten. Ik vind het ook heel fijn dat Milan vaak spontaan naar jullie toe kon zodat ik nog even kon werken tijdens de afronding van mijn PhD. **Sarah en René**, bedankt voor jullie gezelligheid en de afleiding. Ik kijk uit naar alle familie-uitjes die nog gaan komen en een dikke kus voor **Benthe** en **Linde**! Onze **opa's** en **oma's**: bedankt voor jullie steun en de fijne gesprekken. Jullie zijn er altijd voor ons! **Brenda**, **Richard**, en **Kai**: bedankt voor jullie interesse in mijn onderzoek en de gezelligheid in de weekenden en avonden op de boot.

En als allerlaatste, mijn lieve gezin. Lieve **Joris**, ik kan me geen leven zonder jou voorstellen. Jij bent er altijd en we hebben heel veel samen meegemaakt, zowel leuke als hele zware tijden. Zonder jou had ik deze PhD niet afgemaakt. Wat ben ik trots dat wij samen een prachtige zoon en dochter hebben, dat hebben wij goed gedaan ;) Ik hou van jou! De geboorte van Milan en Emma maakte het afronden van mijn promotietraject een uitdaging, maar ik zou dit natuurlijk voor geen goud hebben willen missen. Lieve lieve lieve **Milan** en **Emma**, wat zou ik zonder jullie moeten! Ik hoop dat jullie je hart gaan volgen en gaan doen wat jullie leuk vinden. Ik zal jullie altijd steunen en er altijd voor jullie zijn. Jullie zullen het later niet meer weten, maar ik had jullie knuffels en kusjes nodig om me te kunnen focussen op het afronden van dit boekje. Ik hou zoveel van jullie en ik kijk uit naar alles wat de toekomst gaat brengen!

CHAPTER 6.2

PhD portfolio

Name PhD student:	Ashley van der Spek
Erasmus MC Department:	Epidemiology
Research School:	NIHES
PhD period:	2015 – 2019
Promotor:	Prof.dr. C.M. van Duijn
Supervisor:	Dr. N. Amin

1. PhD training

	Year	Workload (ECTS)*
General courses		
- Biomedical English Writing and Communication	2017	1.5
- Research Integrity	2017	0.3
- Photoshop and Indesign	2018	0.3
- Employability outside academia	2019	1
Seminars and workshops		
- Seminars and 2020 meetings at the department of Epidemiology	2015-2019	2
- Seminars Genetic Epidemiology unit, department of Epidemiology	2015-2018	1.0
- Molecular Epidemiology research meetings	2015-2018	1.0
- Symposium Biolink NL: gegevenskoppeling in Nederland	2015	0.6
- KNAW symposium "Het Alzheimermysterie"	2016	0.1
- BBMRI metabolomics workshop	2017	0.6
- PhD day	2017	0.3
- PhD day	2018	0.3
- Workshop Career Development for PhD candidates	2019	0.15
- Innovation for Health	2019	0.3
Presentations		
- Genetic Epidemiology unit: "*ERF2015: Follow-up of the ERF study*" (oral)	2016	1
- Genetic Epidemiology unit: "*Exome-wide association analysis of telomere length in the ERF study*" (oral)	2016	1
- 2020 meeting at the department of Epidemiology "*The Genetics of Ageing*" (oral)	2016	1
- Molecular Epidemiology meeting "*Exome-wide meta-analysis of sleep apnea symptoms*" (oral)	2017	1
- Genetic Epidemiology unit: "*Exome sequence meta-analysis of telomere length in two Dutch studies*" (oral)	2018	1
- CHARGE Investigator meeting, Baltimore, USA: "Exome chip association meta-analyses of longevity in Europeans and African-Americans" (poster)	2018	1

- Health Sciences Research Day: "Exome-sequencing analysis identifies rare variants in ATM locus associated with shorter telomere length in Dutch family study" (poster)	2019	1
- Dutch Epidemiology Conference (WEON): "Exome sequencing analysis identifies rare variants in ATM locus associated with shorter telomere length in Dutch family study" (oral)	2019	1

(Inter)national conferences

- BBMRI Metabolomics Consortium meetings	2016-2018	1.0
- CHARGE Investigator meeting, Boston, USA	2017	1.1
- CHARGE Investigator meeting, Baltimore, USA	2018	1.1
- Health Sciences Research Day – poster presentation	2019	0.3
- Dutch Epidemiology Conference (WEON)	2019	0.7

Other

- Organisation ERF Follow-up meeting	2016	1.5
- Organisation ERF meeting with VUMC "Bewegen, niet vergeten"	2016	1.5
- Research fellow at Linkcare Health Services, as part of PRECeDI, a Marie Sklodowska-Curie Research and Innovation Staff Exchange (MCSA-RISE) program	2017	3 months

2. Teaching

Teaching assistant

- Teaching assistant for the NIHES course "Biostatistical Methods I: basis principles"	2015	0.5
- Teaching assistant for the NIHES course "Biostatistical Methods I: basis principles"	2016	0.8
- Teaching assistant for the NIHES Summer Programme course "Principles of Genetic Epidemiology"	2016	1
- Teaching assistant for the NIHES Summer Programme course "Principles of Genetic Epidemiology"	2017	1
- Teaching assistant for the NIHES Summer Programme course "Principles of Genetic Epidemiology"	2018	1

Supervising Master's thesis

- Supervision with Sven van der Lee: Ilse Geraedts Bachelor student Biomedische Wetenschappen	2016	1

Other

- Organisation weekly seminars department of Epidemiology	2015-2016	2
- Peer review of manuscripts	2018-2019	0.5

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

CHAPTER 6.3

List of publications

1. Prokić I, Lahousse L, de Vries M, Liu J, Kalaoja M, Vonk JM, van der Plaat DA, van Diemen CC, **van der Spek A**, Zhernakova A, Fu J, Ghanbari M, Ala-Korpela M, Kettunen J, Havulinna AS, Perola M, Salomaa V, Lind L, Ärnlöv J, Stricker BHC, Brusselle GG, Boezen HM, van Duijn CM, Amin N. **A cross-omics integrative study of metabolic signatures of chronic obstructive pulmonary disease.** *BMC Pulm Med*. 2020 Jul 16;20(1):193. doi: 10.1186/s12890-020-01222-7.

2. **van der Spek A**, Warner SC, Broer L, Nelson CP, Vojinovic D, Ahmad S, Arp PP, Brouwer RWW, Denniff M, van den Hout MCGN, van Rooij JGJ, Kraaij R, van IJcken WFJ, Samani NJ, Ikram MA, Uitterlinden AG, Codd V, Amin N, van Duijn CM. **Exome sequencing analysis identifies rare variants in *ATM* and *RPL8* that are associated with shorter telomere length.** *Front Genet*. 2020 Apr 30;11:337. doi: 10.3389/fgene.2020.00337. eCollection 2020.

3. Li C, Stoma S, Lotta LA, Warner S, Albrecht E, Allione A, Arp PP, Broer L, Buxton JL, Da Silva Couto Alves A, Deelen J, Fedko IO, Gordon SD, Jiang T, Karlsson R, Kerrison N, Loe TK, Mangino M, Milaneschi Y, Miraglio B, Pervjakova N, Russo A, Surakka I, **van der Spek A**, Verhoeven JE, Amin N, Beekman M, Blakemore AI, Canzian F, Hamby SE, Hottenga JJ, Jones PD, Jousilahti P, Mägi R, Medland SE, Montgomery GW, Nyholt DR, Perola M, Pietiläinen KH, Salomaa V, Sillanpää E, Suchiman HE, van Heemst D, Willemsen G, Agudo A, Boeing H, Boomsma DI, Chirlaque MD, Fagherazzi G, Ferrari P, Franks P, Gieger C, Eriksson JG, Gunter M, Hägg S, Hovatta I, Imaz L, Kaprio J, Kaaks R, Key T, Krogh V, Martin NG, Melander O, Metspalu A, Moreno C, Onland-Moret NC, Nilsson P, Ong KK, Overvad K, Palli D, Panico S, Pedersen NL, Penninx BWJH, Quirós JR, Jarvelin MR, Rodríguez-Barranco M, Scott RA, Severi G, Slagboom PE, Spector TD, Tjonneland A, Trichopoulou A, Tumino R, Uitterlinden AG, van der Schouw YT, van Duijn CM, Weiderpass E, Denchi EL, Matullo G, Butterworth AS, Danesh J, Samani NJ, Wareham NJ, Nelson CP, Langenberg C, Codd V. **Genome-wide association analysis in humans links nucleotide metabolism to leukocyte telomere length.** *Am J Hum Genet*. 2020 Mar 5;106(3):389-404. doi: 10.1016/j.ajhg.2020.02.006.

4. Fedko IO, Hottenga JJ, Helmer Q, Mbarek H, Huider F, Amin N, Beulens JW, Bremmer MA, Elders PJ, Galesloot TE, Kiemeney LA, van Loo HM, Picavet HSJ, Rutters F, **van der Spek A**, van de Wiel AM, van Duijn C, de Geus EJC, Feskens EJM, Hartman CA, Oldehinkel AJ, Smit JH, Verschuren WMM, Penninx BWJH, Boomsma DI, Bot M. **Measurement and genetic architecture of lifetime depression in the Netherlands as assessed by LIDAS (Lifetime Depression Assessment Self-report).** *Psychol Med*. 2020 Feb 27:1-10. doi: 10.1017/S0033291720000100.

5. Noordam R, Bos MM, Wang H, Winkler TW, Bentley AR, Kilpeläinen TO, de Vries PS, Sung YJ, Schwander K, Cade BE, Manning A, Aschard H, Brown MR, Chen H, Franceschini N, Musani SK, Richard M, Vojinovic D, Aslibekyan S, Bartz TM, de Las Fuentes L, Feitosa M,

Horimoto AR, Ilkov M, Kho M, Kraja A, Li C, Lim E, Liu Y, Mook-Kanamori DO, Rankinen T, Tajuddin SM, **van der Spek A**, Wang Z, Marten J, Laville V, Alver M, Evangelou E, Graff ME, He M, Kühnel B, Lyytikäinen LP, Marques-Vidal P, Nolte IM, Palmer ND, Rauramaa R, Shu XO, Snieder H, Weiss S, Wen W, Yanek LR, Adolfo C, Ballantyne C, Bielak L, Biermasz NR, Boerwinkle E, Dimou N, Eiriksdottir G, Gao C, Gharib SA, Gottlieb DJ, Haba-Rubio J, Harris TB, Heikkinen S, Heinzer R, Hixson JE, Homuth G, Ikram MA, Komulainen P, Krieger JE, Lee J, Liu J, Lohman KK, Luik AI, Mägi R, Martin LW, Meitinger T, Metspalu A, Milaneschi Y, Nalls MA, O'Connell J, Peters A, Peyser P, Raitakari OT, Reiner AP, Rensen PCN, Rice TK, Rich SS, Roenneberg T, Rotter JI, Schreiner PJ, Shikany J, Sidney SS, Sims M, Sitlani CM, Sofer T, Strauch K, Swertz MA, Taylor KD, Uitterlinden AG, van Duijn CM, Völzke H, Waldenberger M, Wallance RB, van Dijk KW, Yu C, Zonderman AB, Becker DM, Elliott P, Esko T, Gieger C, Grabe HJ, Lakka TA, Lehtimäki T, North KE, Penninx BWJH, Vollenweider P, Wagenknecht LE, Wu T, Xiang YB, Zheng W, Arnett DK, Bouchard C, Evans MK, Gudnason V, Kardia S, Kelly TN, Kritchevsky SB, Loos RJF, Pereira AC, Province M, Psaty BM, Rotimi C, Zhu X, Amin N, Cupples LA, Fornage M, Fox EF, Guo X, Gauderman WJ, Rice K, Kooperberg C, Munroe PB, Liu CT, Morrison AC, Rao DC, van Heemst D, Redline S. **Multi-ancestry sleep-by-SNP interaction analysis in 126,926 individuals reveals lipid loci stratified by sleep duration.** *Nat Commun*. 2019 Nov 12;10(1):5121. doi: 10.1038/s41467-019-12958-0.

6. Clark DW, Okada Y, Moore KHS, Mason D, Pirastu N, Gandin I, Mattsson H, Barnes CLK, Lin K, Zhao JH, Deelen P, Rohde R, Schurmann C, Guo X, Giulianini F, Zhang W, Medina-Gomez C, Karlsson R, Bao Y, Bartz TM, Baumbach C, Biino G, Bixley MJ, Brumat M, Chai JF, Corre T, Cousminer DL, Dekker AM, Eccles DA, van Eijk KR, Fuchsberger C, Gao H, Germain M, Gordon SD, de Haan HG, Harris SE, Hofer E, Huerta-Chagoya A, Igartua C, Jansen IE, Jia Y, Kacprowski T, Karlsson T, Kleber ME, Li SA, Li-Gao R, Mahajan A, Matsuda K, Meidtner K, Meng W, Montasser ME, van der Most PJ, Munz M, Nutile T, Palviainen T, Prasad G, Prasad RB, Priyanka TDS, Rizzi F, Salvi E, Sapkota BR, Shriner D, Skotte L, Smart MC, Smith AV, **van der Spek A**, Spracklen CN, Strawbridge RJ, Tajuddin SM, Trompet S, Turman C, Verweij N, Viberti C, Wang L, Warren HR, Wootton RE, Yanek LR, Yao J, Yousri NA, Zhao W, Adeyemo AA, Afaq S, Aguilar-Salinas CA, Akiyama M, Albert ML, Allison MA, Alver M, Aung T, Azizi F, Bentley AR, Boeing H, Boerwinkle E, Borja JB, de Borst GJ, Bottinger EP, Broer L, Campbell H, Chanock S, Chee ML, Chen G, Chen YI, Chen Z, Chiu YF, Cocca M, Collins FS, Concas MP, Corley J, Cugliari G, van Dam RM, Damulina A, Daneshpour MS, Day FR, Delgado GE, Dhana K, Doney ASF, Dörr M, Doumatey AP, Dzimiri N, Ebenesersdóttir SS, Elliott J, Elliott P, Ewert R, Felix JF, Fischer K, Freedman BI, Girotto G, Goel A, Gögele M, Goodarzi MO, Graff M, Granot-Hershkovitz E, Grodstein F, Guarrera S, Gudbjartsson DF, Guity K, Gunnarsson B, Guo Y, Hagenaars SP, Haiman CA, Halevy A, Harris TB, Hedayati M, van Heel DA, Hirata M, Höfer I, Hsiung CA, Huang J, Hung YJ, Ikram MA, Jagadeesan A, Jousilahti P, Kamatani Y, Kanai M, Kerrison ND, Kessler T, Khaw KT, Khor CC, de Kleijn DPV, Koh WP, Kolcic I, Kraft P, Krämer BK, Kutalik Z, Kuusisto J, Langenberg C, Launer LJ, Lawlor DA, Lee IT, Lee WJ, Lerch MM, Li L, Liu J, Loh

M, London SJ, Loomis S, Lu Y, Luan J, Mägi R, Manichaikul AW, Manunta P, Másson G, Matoba N, Mei XW, Meisinger C, Meitinger T, Mezzavilla M, Milani L, Millwood IY, Momozawa Y, Moore A, Morange PE, Moreno-Macías H, Mori TA, Morrison AC, Muka T, Murakami Y, Murray AD, de Mutsert R, Mychaleckyj JC, Nalls MA, Nauck M, Neville MJ, Nolte IM, Ong KK, Orozco L, Padmanabhan S, Pálsson G, Pankow JS, Pattaro C, Pattie A, Polasek O, Poulter N, Pramstaller PP, Quintana-Murci L, Räikkönen K, Ralhan S, Rao DC, van Rheenen W, Rich SS, Ridker PM, Rietveld CA, Robino A, van Rooij FJA, Ruggiero D, Saba Y, Sabanayagam C, Sabater-Lleal M, Sala CF, Salomaa V, Sandow K, Schmidt H, Scott LJ, Scott WR, Sedaghati-Khayat B, Sennblad B, van Setten J, Sever PJ, Sheu WH, Shi Y, Shrestha S, Shukla SR, Sigurdsson JK, Sikka TT, Singh JR, Smith BH, Stančáková A, Stanton A, Starr JM, Stefansdottir L, Straker L, Sulem P, Sveinbjornsson G, Swertz MA, Taylor AM, Taylor KD, Terzikhan N, Tham YC, Thorleifsson G, Thorsteinsdottir U, Tillander A, Tracy RP, Tusié-Luna T, Tzoulaki I, Vaccargiu S, Vangipurapu J, Veldink JH, Vitart V, Völker U, Vuoksimaa E, Wakil SM, Waldenberger M, Wander GS, Wang YX, Wareham NJ, Wild S, Yajnik CS, Yuan JM, Zeng L, Zhang L, Zhou J, Amin N, Asselbergs FW, Bakker SJL, Becker DM, Lehne B, Bennett DA, van den Berg LH, Berndt SI, Bharadwaj D, Bielak LF, Bochud M, Boehnke M, Bouchard C, Bradfield JP, Brody JA, Campbell A, Carmi S, Caulfield MJ, Cesarini D, Chambers JC, Chandak GR, Cheng CY, Ciullo M, Cornelis M, Cusi D, Smith GD, Deary IJ, Dorajoo R, van Duijn CM, Ellinghaus D, Erdmann J, Eriksson JG, Evangelou E, Evans MK, Faul JD, Feenstra B, Feitosa M, Foisy S, Franke A, Friedlander Y, Gasparini P, Gieger C, Gonzalez C, Goyette P, Grant SFA, Griffiths LR, Groop L, Gudnason V, Gyllensten U, Hakonarson H, Hamsten A, van der Harst P, Heng CK, Hicks AA, Hochner H, Huikuri H, Hunt SC, Jaddoe VWV, De Jager PL, Johannesson M, Johansson Å, Jonas JB, Jukema JW, Junttila J, Kaprio J, Kardia SLR, Karpe F, Kumari M, Laakso M, van der Laan SW, Lahti J, Laudes M, Lea RA, Lieb W, Lumley T, Martin NG, März W, Matullo G, McCarthy MI, Medland SE, Merriman TR, Metspalu A, Meyer BF, Mohlke KL, Montgomery GW, Mook-Kanamori D, Munroe PB, North KE, Nyholt DR, O'connell JR, Ober C, Oldehinkel AJ, Palmas W, Palmer C, Pasterkamp GG, Patin E, Pennell CE, Perusse L, Peyser PA, Pirastu M, Polderman TJC, Porteous DJ, Posthuma D, Psaty BM, Rioux JD, Rivadeneira F, Rotimi C, Rotter JI, Rudan I, Den Ruijter HM, Sanghera DK, Sattar N, Schmidt R, Schulze MB, Schunkert H, Scott RA, Shuldiner AR, Sim X, Small N, Smith JA, Sotoodehnia N, Tai ES, Teumer A, Timpson NJ, Toniolo D, Tregouet DA, Tuomi T, Vollenweider P, Wang CA, Weir DR, Whitfield JB, Wijmenga C, Wong TY, Wright J, Yang J, Yu L, Zemel BS, Zonderman AB, Perola M, Magnusson PKE, Uitterlinden AG, Kooner JS, Chasman DI, Loos RJF, Franceschini N, Franke L, Haley CS, Hayward C, Walters RG, Perry JRB, Esko T, Helgason A, Stefansson K, Joshi PK, Kubo M, Wilson JF**. Associations of autozygosity with a broad range of human phenotypes.** *Nat Commun*. 2019 Oct 31;10(1):4957. doi: 10.1038/s41467-019-12283-6.

7. Deelen J, Kettunen J, Fischer K, **van der Spek A**, Trompet S, Kastenmüller G, Boyd A, Zierer J, van den Akker EB, Ala-Korpela M, Amin N, Demirkan A, Ghanbari M, van Heemst D, Ikram MA, van Klinken JB, Mooijaart SP, Peters A, Salomaa V, Sattar N, Spector TD,

Tiemeier H, Verhoeven A, Waldenberger M, Würtz P, Davey Smith G, Metspalu A, Perola M, Menni C, Geleijnse JM, Drenos F, Beekman M, Jukema JW, van Duijn CM, Slagboom PE. **A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals.** *Nat Commun*. 2019 Aug 20;10(1):3346. doi: 10.1038/s41467-019-11311-9.

8. Deelen J, Evans DS, Arking DE, Tesi N, Nygaard M, Liu X, Wojczynski MK, Biggs ML, **van der Spek A**, Atzmon G, Ware EB, Sarnowski C, Smith AV, Seppälä I, Cordell HJ, Dose J, Amin N, Arnold AM, Ayers KL, Barzilai N, Becker EJ, Beekman M, Blanché H, Christensen K, Christiansen L, Collerton JC, Cubaynes S, Cummings SR, Davies K, Debrabant B, Deleuze JF, Duncan R, Faul JD, Franceschi C, Galan P, Gudnason V, Harris TB, Huisman M, Hurme MA, Jagger C, Jansen I, Jylhä M, Kähönen M, Karasik D, Kardia SLR, Kingston A, Kirkwood TBL, Launer LJ, Lehtimäki T, Lieb W, Lyytikäinen LP, Martin-Ruiz C, Min J, Nebel A, Newman AB, Nie C, Nohr EA, Orwoll ES, Perls TT, Province MA, Psaty BM, Raitakari OT, Reinders MJT, Robine JM, Rotter JI, Sebastiani P, Smith J, Sørensen TIA, Taylor KD, Uitterlinden AG, van der Flier W, van der Lee SJ, van Duijn CM, van Heemst D, Vaupel JW, Weir D, Ye K, Zeng Y, Zheng W, Holstege H, Kiel DP, Lunetta KL, Slagboom PE, Murabito JM. **A meta-analysis of genome-wide association studies identifies multiple longevity genes.** *Nat Commun*. 2019 Aug 14;10(1):3669. doi: 10.1038/s41467-019-11558-2.

9. **van der Spek A**, Broer L, Draisma HHM, Pool R, Albrecht E, Beekman M, Mangino M, Raag M, Nyholt DR, Dharuri HK, Codd V, Amin N, de Geus EJC, Deelen J, Demirkan A, Yet I, Fischer K, Haller T, Henders AK, Isaacs A, Medland SE, Montgomery GW, Mooijaart SP, Strauch K, Suchiman HED, Vaarhorst AAM, van Heemst D, Wang-Sattler R, Whitfield JB, Willemsen G, Wright MJ, Martin NG, Samani NJ, Metspalu A, Eline Slagboom P, Spector TD, Boomsma DI, van Duijn CM, Gieger C. **Metabolomics reveals a link between homocysteine and lipid metabolism and leukocyte telomere length: the ENGAGE consortium.** *Sci Rep*. 2019 Aug 12;9(1):11623. doi: 10.1038/s41598-019-47282-6.

10. Jones SE, van Hees VT, Mazzotti DR, Marques-Vidal P, Sabia S, **van der Spek A**, Dashti HS, Engmann J, Kocevska D, Tyrrell J, Beaumont RN, Hillsdon M, Ruth KS, Tuke MA, Yaghootkar H, Sharp SA, Ji Y, Harrison JW, Freathy RM, Murray A, Luik AI, Amin N, Lane JM, Saxena R, Rutter MK, Tiemeier H, Kutalik Z, Kumari M, Frayling TM, Weedon MN, Gehrman PR, Wood AR. **Genetic studies of accelerometer-based sleep measures yield new insights into human sleep behaviour.** *Nat Commun*. 2019 Apr 5;10(1):1585. doi: 10.1038/s41467-019-09576-1.

11. Galle SA, **van der Spek A**, Drent ML, Brugts MP, Scherder EJA, Janssen JAMJL, Ikram MA, van Duijn CM. **Revisiting the role of insulin-like growth factor-I receptor stimulating activity and the apolipoprotein E in Alzheimer's disease.** *Front Aging Neurosci*. 2019 Feb 12;11:20. doi: 10.3389/fnagi.2019.00020. eCollection 2019.

12. van Dongen J, Bonder MJ, Dekkers KF, Nivard MG, van Iterson M, Willemsen G, Beekman M, **van der Spek A**, van Meurs JBJ, Franke L, Heijmans BT, van Duijn CM, Slagboom PE, Boomsma DI; BIOS consortium. **DNA methylation signatures of educational attainment.** *NPJ Sci Learn*. 2018 Mar 23;3:7. doi: 10.1038/s41539-018-0020-2.

13. **van der Spek A**, Luik AI, Kocevska D, Liu C, Brouwer RWW, van Rooij JGJ, van den Hout MCGN, Kraaij R, Hofman A, Uitterlinden AG, van IJcken WFJ, Gottlieb DJ, Tiemeier H, van Duijn CM, Amin N. **Exome-wide meta-analysis identifies rare 3'-UTR variant in *ERCC1/CD3EAP* associated with symptoms of sleep apnea.** *Front Genet*. 2017 Oct 18;8:151. doi: 10.3389/fgene.2017.00151.

14. Joshi PK, Pirastu N, Kentistou KA, Fischer K, Hofer E, Schraut KE, Clark DW, Nutile T, Barnes CLK, Timmers PRHJ, Shen X, Gandin I, McDaid AF, Hansen TF, Gordon SD, Giulianini F, Boutin TS, Abdellaoui A, Zhao W, Medina-Gomez C, Bartz TM, Trompet S, Lange LA, Raffield L, **van der Spek A**, Galesloot TE, Proitsi P, Yanek LR, Bielak LF, Payton A, Murgia F, Concas MP, Biino G, Tajuddin SM, Seppälä I, Amin N, Boerwinkle E, Børglum AD, Campbell A, Demerath EW, Demuth I, Faul JD, Ford I, Gialluisi A, Gögele M, Graff M, Hingorani A, Hottenga JJ, Hougaard DM, Hurme MA, Ikram MA, Jylhä M, Kuh D, Ligthart L, Lill CM, Lindenberger U, Lumley T, Mägi R, Marques-Vidal P, Medland SE, Milani L, Nagy R, Ollier WER, Peyser PA, Pramstaller PP, Ridker PM, Rivadeneira F, Ruggiero D, Saba Y, Schmidt R, Schmidt H, Slagboom PE, Smith BH, Smith JA, Sotoodehnia N, Steinhagen-Thiessen E, van Rooij FJA, Verbeek AL, Vermeulen SH, Vollenweider P, Wang Y, Werge T, Whitfield JB, Zonderman AB, Lehtimäki T, Evans MK, Pirastu M, Fuchsberger C, Bertram L, Pendleton N, Kardia SLR, Ciullo M, Becker DM, Wong A, Psaty BM, van Duijn CM, Wilson JG, Jukema JW, Kiemeney L, Uitterlinden AG, Franceschini N, North KE, Weir DR, Metspalu A, Boomsma DI, Hayward C, Chasman D, Martin NG, Sattar N, Campbell H, Esko T, Kutalik Z, Wilson JF. **Genome-wide meta-analysis associates HLA-DQA1/DRB1 and LPA and lifestyle factors with human longevity.** *Nat Commun*. 2017 Oct 13;8(1):910. doi: 10.1038/s41467-017-00934-5.

15. Liu J, Semiz S, van der Lee SJ, **van der Spek A**, Verhoeven A, van Klinken JB, Sijbrands E, Harms AC, Hankemeier T, van Dijk KW, van Duijn CM, Demirkan A. **Metabolomics based markers predict type 2 diabetes in a 14-year follow-up study.** *Metabolomics*. 2017;13(9):104. doi: 10.1007/s11306-017-1239-2.

16. Pirastu N, Kooyman M, Robino A, **van der Spek A**, Navarini L, Amin N, Karssen LC, Van Duijn CM, Gasparini P. **Non-additive genome-wide association scan reveals a new gene associated with habitual coffee consumption.** *Sci Rep*. 2016 Aug 25;6:31590. doi: 10.1038/srep31590.

17. Amin N, Allebrandt KV, **van der Spek A**, Müller-Myhsok B, Hek K, Teder-Laving M, Hayward C, Esko T, van Mill JG, Mbarek H, Watson NF, Melville SA, Del Greco FM, Byrne

EM, Oole E, Kolcic I, Chen TH, Evans DS, Coresh J, Vogelzangs N, Karjalainen J, Willemsen G, Gharib SA, Zgaga L, Mihailov E, Stone KL, Campbell H, Brouwer RW, Demirkan A, Isaacs A, Dogas Z, Marciante KD, Campbell S, Borovecki F, Luik AI, Li M, Hottenga JJ, Huffman JE, van den Hout MC, Cummings SR, Aulchenko YS, Gehrman PR, Uitterlinden AG, Wichmann HE, Müller-Nurasyid M, Fehrmann RS, Montgomery GW, Hofman A, Kao WH, Oostra BA, Wright AF, Vink JM, Wilson JF, Pramstaller PP, Hicks AA, Polasek O, Punjabi NM, Redline S, Psaty BM, Heath AC, Merrow M, Tranah GJ, Gottlieb DJ, Boomsma DI, Martin NG, Rudan I, Tiemeier H, van IJcken WF, Penninx BW, Metspalu A, Meitinger T, Franke L, Roenneberg T, van Duijn CM. **Genetic variants in *RBFOX3* are associated with sleep latency.** *Eur J Hum Genet*. 2016 Oct;24(10):1488-95. doi: 10.1038/ejhg.2016.31. Epub 2016 May 4.

18. Joshi PK, Esko T, Mattsson H, Eklund N, Gandin I, Nutile T, Jackson AU, Schurmann C, Smith AV, Zhang W, Okada Y, Stančáková A, Faul JD, Zhao W, Bartz TM, Concas MP, Franceschini N, Enroth S, Vitart V, Trompet S, Guo X, Chasman DI, O'Connel JR, Corre T, Nongmaithem SS, Chen Y, Mangino M, Ruggiero D, Traglia M, Farmaki AE, Kacprowski T, Bjonnes A, **van der Spek A**, Wu Y, Giri AK, Yanek LR, Wang L, Hofer E, Rietveld CA, McLeod O, Cornelis MC, Pattaro C, Verweij N, Baumbach C, Abdellaoui A, Warren HR, Vuckovic D, Mei H, Bouchard C, Perry JRB, Cappellani S, Mirza SS, Benton MC, Broeckel U, Medland SE, Lind PA, Malerba G, Drong A, Yengo L, Bielak LF, Zhi D, van der Most PJ, Shriner D, Mägi R, Hemani G, Karaderi T, Wang Z, Liu T, Demuth I, Zhao JH, Meng W, Lataniotis L, van der Laan SW, Bradfield JP, Wood AR, Bonnefond A, Ahluwalia TS, Hall LM, Salvi E, Yazar S, Carstensen L, de Haan HG, Abney M, Afzal U, Allison MA, Amin N, Asselbergs FW, Bakker SJL, Barr RG, Baumeister SE, Benjamin DJ, Bergmann S, Boerwinkle E, Bottinger EP, Campbell A, Chakravarti A, Chan Y, Chanock SJ, Chen C, Chen YI, Collins FS, Connell J, Correa A, Cupples LA, Smith GD, Davies G, Dörr M, Ehret G, Ellis SB, Feenstra B, Feitosa MF, Ford I, Fox CS, Frayling TM, Friedrich N, Geller F, Scotland G, Gillham-Nasenya I, Gottesman O, Graff M, Grodstein F, Gu C, Haley C, Hammond CJ, Harris SE, Harris TB, Hastie ND, Heard-Costa NL, Heikkilä K, Hocking LJ, Homuth G, Hottenga JJ, Huang J, Huffman JE, Hysi PG, Ikram MA, Ingelsson E, Joensuu A, Johansson Å, Jousilahti P, Jukema JW, Kähönen M, Kamatani Y, Kanoni S, Kerr SM, Khan NM, Koellinger P, Koistinen HA, Kooner MK, Kubo M, Kuusisto J, Lahti J, Launer LJ, Lea RA, Lehne B, Lehtimäki T, Liewald DCM, Lind L, Loh M, Lokki ML, London SJ, Loomis SJ, Loukola A, Lu Y, Lumley T, Lundqvist A, Männistö S, Marques-Vidal P, Masciullo C, Matchan A, Mathias RA, Matsuda K, Meigs JB, Meisinger C, Meitinger T, Menni C, Mentch FD, Mihailov E, Milani L, Montasser ME, Montgomery GW, Morrison A, Myers RH, Nadukuru R, Navarro P, Nelis M, Nieminen MS, Nolte IM, O'Connor GT, Ogunniyi A, Padmanabhan S, Palmas WR, Pankow JS, Patarcic I, Pavani F, Peyser PA, Pietilainen K, Poulter N, Prokopenko I, Ralhan S, Redmond P, Rich SS, Rissanen H, Robino A, Rose LM, Rose R, Sala C, Salako B, Salomaa V, Sarin AP, Saxena R, Schmidt H, Scott LJ, Scott WR, Sennblad B, Seshadri S, Sever P, Shrestha S, Smith BH, Smith JA, Soranzo N, Sotoodehnia N, Southam L, Stanton AV, Stathopoulou MG, Strauch K, Strawbridge RJ, Suderman MJ, Tandon N, Tang ST,

Taylor KD, Tayo BO, Töglhofer AM, Tomaszewski M, Tšernikova N, Tuomilehto J, Uitterlinden AG, Vaidya D, van Hylckama Vlieg A, van Setten J, Vasankari T, Vedantam S, Vlachopoulou E, Vozzi D, Vuoksimaa E, Waldenberger M, Ware EB, Wentworth-Shields W, Whitfield JB, Wild S, Willemsen G, Yajnik CS, Yao J, Zaza G, Zhu X, Project TBJ, Salem RM, Melbye M, Bisgaard H, Samani NJ, Cusi D, Mackey DA, Cooper RS, Froguel P, Pasterkamp G, Grant SFA, Hakonarson H, Ferrucci L, Scott RA, Morris AD, Palmer CNA, Dedoussis G, Deloukas P, Bertram L, Lindenberger U, Berndt SI, Lindgren CM, Timpson NJ, Tönjes A, Munroe PB, Sørensen TIA, Rotimi CN, Arnett DK, Oldehinkel AJ, Kardia SLR, Balkau B, Gambaro G, Morris AP, Eriksson JG, Wright MJ, Martin NG, Hunt SC, Starr JM, Deary IJ, Griffiths LR, Tiemeier H, Pirastu N, Kaprio J, Wareham NJ, Pérusse L, Wilson JG, Girotto G, Caulfield MJ, Raitakari O, Boomsma DI, Gieger C, van der Harst P, Hicks AA, Kraft P, Sinisalo J, Knekt P, Johannesson M, Magnusson PKE, Hamsten A, Schmidt R, Borecki IB, Vartiainen E, Becker DM, Bharadwaj D, Mohlke KL, Boehnke M, van Duijn CM, Sanghera DK, Teumer A, Zeggini E, Metspalu A, Gasparini P, Ulivi S, Ober C, Toniolo D, Rudan I, Porteous DJ, Ciullo M, Spector TD, Hayward C, Dupuis J, Loos RJF, Wright AF, Chandak GR, Vollenweider P, Shuldiner A, Ridker PM, Rotter JI, Sattar N, Gyllensten U, North KE, Pirastu M, Psaty BM, Weir DR, Laakso M, Gudnason V, Takahashi A, Chambers JC, Kooner JS, Strachan DP, Campbell H, Hirschhorn JN, Perola M, Polašek O, Wilson JF. **Directional dominance on stature and cognition in diverse human populations.** *Nature*. 2015 Jul 23;523(7561):459-462. doi: 10.1038/nature14618.

CHAPTER 6.4

About the author

Ashley van der Spek was born in Rotterdam, The Netherlands, on July 13th 1987. Ashley grew up in Spijkenisse, where she completed her Higher General Secondary Education (HAVO) in 2005. After obtaining her bachelor degree in Applied Sciences at the Rotterdam University of Applied Sciences (Hogeschool Rotterdam), Rotterdam, she started her career as technician at the Department of Immunology, Erasmus MC, in 2009. Ashley had always been eager to learn and because of her interest in genetics, she decided to continue her education at the Netherlands Institute of Health Sciences (NIHES), Rotterdam. In 2015, Ashley received her Master of Science degree in Genetic Epidemiology. During the master program, Ashley worked on a research project at the Genetic Epidemiology unit, Department of Epidemiology, Erasmus MC, under supervision of Prof. Cornelia van Duijn and Dr. Najaf Amin. After graduation, Ashley continued her research career as PhD candidate in the same group in 2015. As part of the Marie Sklodowska-Curie Research and Innovation Staff Exchange (MCSA-RISE) scholarship, Ashley worked three months at Linkcare Health Services, Barcelona, Spain. During this period, she worked on the development of healthcare protocols to improve the diagnosis, treatment, and prevention of chronic diseases. Since September 2019, Ashley works as Study Coordinator at SkylineDx B.V., Rotterdam.