Genetic and environmental determinants of cardiometabolic health

Bos, M.M.

בס"ד

GENETIC AND ENVIRONMENTAL DETERMINANTS OF CARDIOMETABOLIC HEALTH

Maxime M Bos

Propositions belonging to the thesis

GENETIC AND ENVIRONMENTAL DETERMINANTS OF CARDIOMETABOLIC HEALTH

1. Metabolic alterations are already present in persons without diabetes mellitus and their detection can serve to better understand biological mechanisms of the disease as well as to identify persons at risk. (this thesis)

2. Sleep-associated cardiovascular risk is influenced by obesity and sleep apnea. (this thesis)

3. The underlying biological mechanisms of short- and long-sleep-associated cardiovascular risk are different. (this thesis)

4. Individuals with a high genetic risk of cardiovascular disease may still benefit from a healthier lifestyle. (this thesis)

5. The focus of metabolomic studies has shifted from cataloguing chemical structures to finding biological stories. (Baker, Nat Methods, 2011)

6. Mendelian randomization studies may prevent unnecessary clinical trials. (Ference, JAMA, 2017)

7. It is important to understand which 'at-risk' individuals are most likely to progress to overt disease. (Wang et al., Nat Med, 2011)

8. Comprehensive integration of multidimensional omics data can effectively capture a holistic view of pathogenic mechanisms. (Arneson, Front Cardiovasc Med, 2017)

9. The key to success is to never stop learning: "The more I live, the more I learn. The more I learn, the more I realize, the less I know." (Michel Legrand, 1954)

10. When performing research, always follow the data: "A man should look for what is, and not for what he thinks should be." (Albert Einstein, 1879-1955)

GENETIC AND ENVIRONMENTAL DETERMINANTS OF CARDIOMETABOLIC HEALTH

Maxime M Bos

Genetic and environmental determinants of cardiometabolic health

Cover design and layout: © evelienjagtman.com
Printing: Optima Grafische Communicatie

ISBN: 978-94-6361-461-0

The work described in this thesis was funded by two European Commision Projects, THYRAGE (Horizon 2020 research and innovation programme, 666869) and HUMAN (EU FP7 no 602757). Financial support from the Netherlands Association for the Study of Obesity (NASO) is greatly acknowledged. Financial support by the Dutch Heart Foundation for the publication of this thesis is gratefully acknowledged.

GENETIC AND ENVIRONMENTAL DETERMINANTS OF CARDIOMETABOLIC HEALTH

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. C.J.J.M. Stolker,
door besluit van het College voor Promoties,
te verdedigen op donderdag 1 oktober 2020,
klokke 11.15 uur

door Maxime M Bos
geboren te Purmerend
in 1991

Promotor
Prof. dr. G.J. Blauw

Co-promotores
Dr. ir. D. van Heemst
Dr. R. Noordam

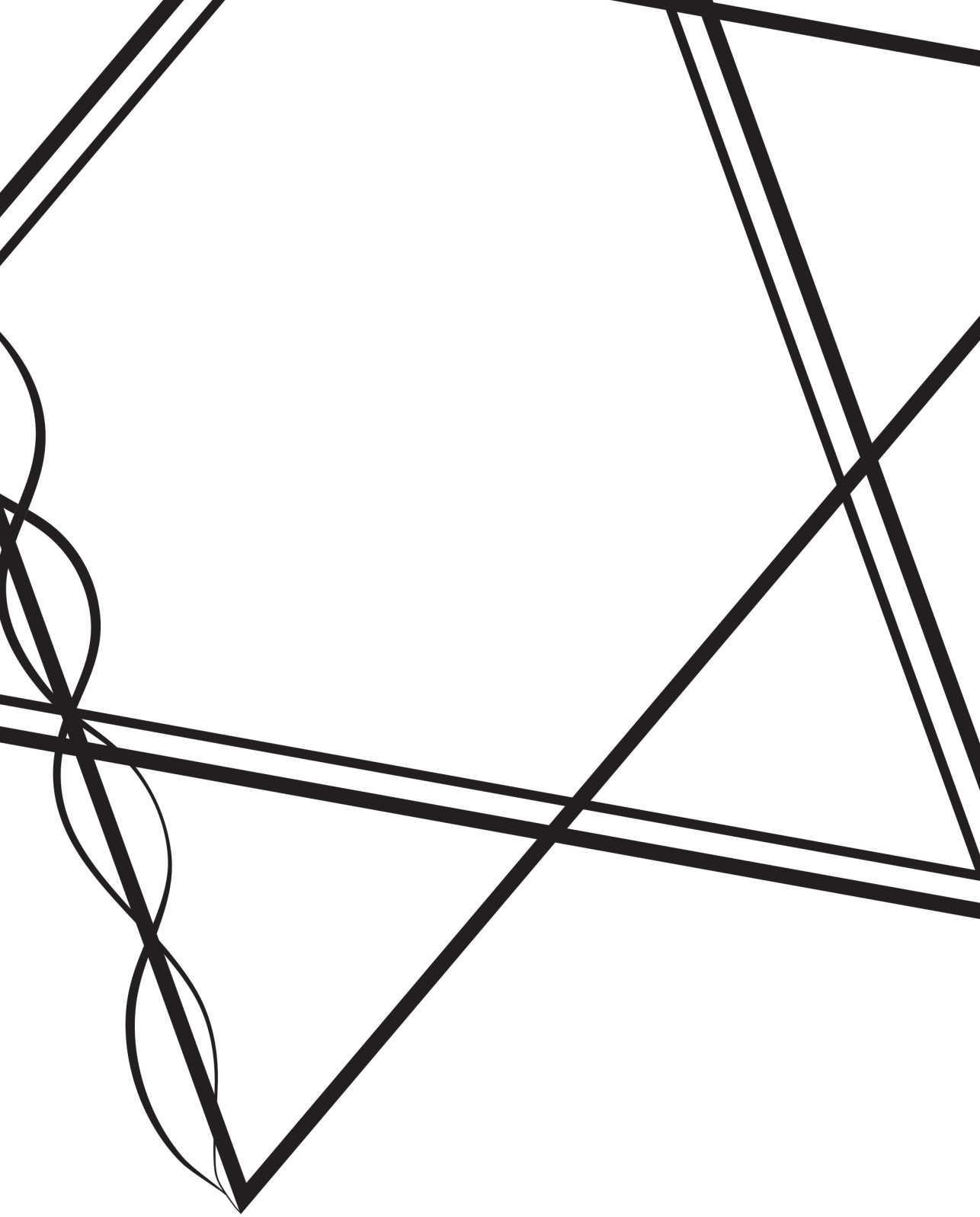
Leden promotiecommissie
Prof. dr. P.C.N. Rensen
Prof. P. Parini (Karolinska Institutet, Stockholm, Sweden)
Dr. F. Rutters (VU University Medical Center, Amsterdam, the Netherlands)

Dit proefschrift draag ik op aan mijn lieve vader,

KANSEN KRIJG JE NIET, DIE MOET JE CREËREN

Nog duizenden knuffels.

TABLE OF CONTENTS



CHAPTER 1

General introduction and thesis outline

Maxime M Bos

INTRODUCTION

The science of medicine is one of the oldest areas of science and has been proven to be very successful. Because of better hygiene, improved treatment options and vaccinations, numerous early deaths have been prevented, thereby greatly increasing our overall survival. As a consequence, our life expectancy has greatly increased. However, because of this increased life expectancy, the world's population is ageing and virtually every country experiences a growth in the proportion of older individuals in their population. In line, the prevalence of age-related diseases has greatly increased and is expected to increase as the population keeps on aging[1]. Although improved cardiovascular risk management (serum cholesterol, blood pressure, smoking, diet) has reduced early deaths due to cardiovascular disease (CVD), CVD remains the leading cause of death of older adults in many developed as well as developing countries and accounted for a total of 17.9 million deaths in 2015[2, 3]. Moreover, a steep increase is seen in the prevalence of type 2 diabetes mellitus. The global prevalence of diabetes mellitus has risen from 108 million individuals (4.7%) in 1980 to 422 million individuals (8.5%) in 2014[4]. Diabetes is one of the major causes for cardiovascular diseases, stroke and kidney failure. Cardiometabolic health and cardiometabolic diseases are terms to describe cardiovascular and metabolic diseases, such as type 2 diabetes mellitus and the metabolic syndrome. Given the public health significance of understanding cardiometabolic diseases, research focusing on a better understanding of causal determinants is of importance.

Obesity is a key risk factor for cardiometabolic diseases. In a few decades, the prevalence of obesity has doubled and this number is expected to further increase. Obesity is just one of the many causes that may lead to cardiometabolic disorders. Many other factors, including age, sex, hormonal-, genetic- and environmental factors jointly determine the individual risk of developing cardiometabolic disorders[5]. For example, non-modifiable factors such as family history and ethnicity all determine metabolic health. Next to these non-modifiable risk factors, modifiable factors such as smoking, physical activity and nutrition all have their own effect on metabolic health as well. Moreover, nearly all diseases result from a complex interaction between an individual's genetic make-up and the environmental factors that one is exposed to. Given the public health significance of understanding cardiometabolic diseases, research focusing on a better understanding of causal determinants is pivotal. Moreover, a better understanding of the interrelations of these risk factors may improve research focused on preventive and treatment strategies for cardiometabolic disorders. Therefore, the aim of this thesis is to study the interplay of non-modifiable (genetic) risk and modifiable lifestyle factors (e.g. sleep, nutrition, physical activity) on cardiometabolic health. In this thesis, several

possible and established risk factors for cardiometabolic disorders will be studied and the causal effects of these factors will be explored. This thesis thereby aims to contribute to a better understanding of determinants of cardiometabolic health in order to decrease the burden of cardiometabolic disorders and age-related diseases on the patient and the society.

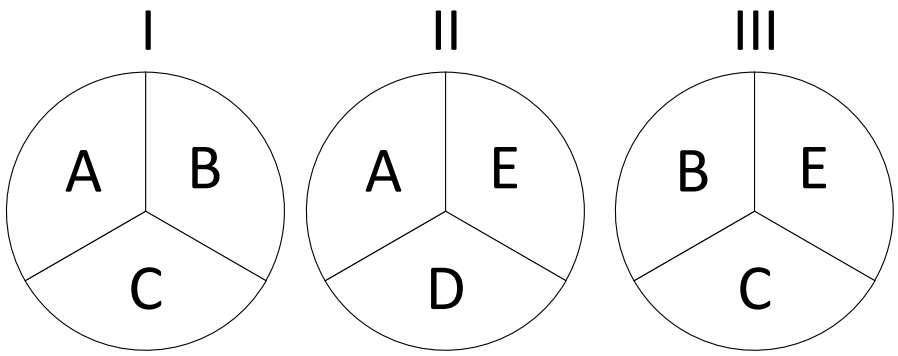


Figure 1. *The causal pie model:* Component causes A-E add up to the sufficient causes I-III. Every sufficient cause consists of different component causes. If and only if all the component causes that constitute the causal pie of a sufficient cause are present, does the sufficient cause exist and does the outcome occur. Hence, the effect of a component cause depends on the presence of its complementary component causes, that is, its complementary set. I, II, and III can be sufficient causes for the same outcome, or for different outcomes, in which case the outcomes are correlated through the component causes. Adapted from Wensink (2014)[6].

Epidemiology and the causal pie model

The study of epidemiology is mainly focussed on determining causes of specific health outcomes and diseases in defined populations. Epidemiology is the cornerstone of public health and is important in shaping policy decisions and evidence-based practices. This field is aimed at identifying risk factors for disease and targets for preventive or curative healthcare. In order to have a better grasp of the idea of causality, the 'causal pie model' of Rothman is a general model widely used in epidemiology[6]. To cite Rothman: "the lights at home shine because they each have a light bulb, there is wire to the light bulbs, the switches are on, there is a power grid, and there is a power source. Take any of these factors away, and there is no light: The system contains 500% causality, for all five factors are 100% causative for the shining of the light. There is no limit to the sum of causes for some outcome"[7]. In **Figure 1**, the causal pie model of Rothman is depicted. In this model, a sufficient cause is a constellation of component causes. The combination of these component causes reflects the causal pie, which leads to a specific outcome. Each component cause can be part of more than one causal pie. In

order for the outcome to occur, all the components that make up the causal pie have to be present. Therefore, the effect of a component cause depends on the presence of the other component causes, which are called complementary component causes. In the case of cardiometabolic health, a sufficient cause can be something as the occurrence of a cardiovascular event. Several component causes make up the causal pie for this outcome, for example, presence of obesity and a high cholesterol level. Moreover, these component causes can be part of other causal pies, such as one for type 2 diabetes mellitus onset (e.g. obesity). In the light of the causal pie model, it is of interest to study which component causes make up a causal pie for a specific cardiometabolic health related outcome.

Lifestyle factors

Lifestyle is a major contributor to risk for the development of cardiometabolic diseases. Besides smoking, the most well-known and established risk factors for cardiometabolic disease are excess caloric intake due to overeating, and a lack of physical activity[5]. As a result of this disbalance between the intake of energy and the expenditure of energy, we eventually will gain weight[5, 8]. During the past decade, sleep has emerged as another lifestyle factor that may contribute to the risk of cardiometabolic disease onset. Sleep is an essential homeostatistically regulated state in which there is decreased activity and alertness[9]. Sleep, or a sleep-like state, is observed across different animal species[9, 10]. Despite extensive studies, the exact mechanistic aspects of sleep are not very well understood. Even less well understood is the association of different aspects of sleep with cardiometabolic health. In several studies, it has been observed that both short and long total sleep duration were associated with a higher risk of obesity, insulin resistance, and diabetes mellitus [11-14]. However, other studies reported only short sleep duration and not long sleep duration to be associated with a higher risk of obesity and metabolic syndrome [15, 16]. Therefore, further study in the field of sleep and cardiometabolic diseases is warranted. Furthermore, insights in the potential biological mechanisms linking disturbances in habitual sleep and cardiometabolic outcomes are warranted to promote further research in prevention and treatment strategies in individuals with disturbed sleep patterns.

Thyroid hormone metabolism

Next to lifestyle factors and genetic predisposition, it is increasingly recognized that endocrine disorders affect cardiometabolic health. Among these are (subclinical) hypothyroidism and hyperthyroidism, which are characterised by a deficit or excess of thyroid hormones. Thyroid hormones act on nearly every cell in the body and thereby exert a wide range of functions. The concentration of thyroid hormones in the circulation is tightly regulated by the hypothalamic pituitary-thyroid axis. Thyrotropin-releasing hormone, secreted by the hypothalamus, regulates synthesis and release of thyroid-

stimulating hormone (TSH) from the pituitary gland, which stimulates the production and secretion of the thyroid hormones by the thyroid gland. Via a classical feedback loop, thyroid hormones inhibit the production of hypothalamic thyrotropin-releasing hormone and pituitary TSH. In target tissues, type 1 and 2 deiodinases convert the prohormone thyroxine into the active hormone triiodothyronine, while type 3 deiodinase converts thyroxine into inactive reverse triiodothyronine[17, 18]. Increasing evidence suggests the existence of an association between thyroid function and cardiometabolic diseases, including atrial and ventricular arrhythmias, atherosclerotic vascular disease, dyslipidemia, and heart failure. In addition, several studies have reported that subclinical hypothyroidism is related to insulin resistance and type 2 diabetes mellitus[19]. However, evidence regarding the causal effect of thyroid hormones on cardiometabolic diseases is lacking.

Genetic epidemiology

The field of genetic epidemiology focusses on the role of genetic factors, and the interplay with environmental factors, in determining health and disease status in populations. Genetic epidemiology thereby seeks to derive a statistical and quantitative analysis on how genetics work in large populations. Genome-wide association studies have been performed to discover genetic variations, single nucleotide polymorphisms (SNPs), without a predefined hypothesis, that are associated with disease traits. A considerable part of the risk of cardiometabolic diseases has a genetic basis. In relation to cardiometabolic disease and human longevity, the top associated SNPs are in the *APOE* gene. Candidate gene studies and GWAS found that genetic variation in *APOE* is robustly associated with multiple cardiometabolic diseases and age-related phenotypes[20, 21]. Apolipoprotein E (ApoE) is an apolipoprotein that plays an important role in triglyceride and cholesterol metabolism. Especially the ApoE ε4 isoform has been associated with mortality, Alzheimer's disease and cardiovascular disease, as compared to the 'neutral' ApoE ε3 isoform[22]. There are, however, large differences in the deleterious effects of the ApoE ε4 isoform between ancestries and populations. One explanation may be the differences in environmental and lifestyle factors between these populations. Nearly all diseases result from a complex interaction between an individual's genetic make-up and the environmental factors that one is exposed to. However, differences in the response to a certain environmental factor may occur in the presence of different genotypes. As a result, some individuals may possess a low risk for developing a certain disease though the exposure to an environmental factor, while other may be more vulnerable. An example of a so called gene-environment interaction, is the higher incidence of skin cancer among fair-skinned individuals as compared to darker skinned individuals as a result to sunlight exposure. In short, two different genotypes respond to an environmental factor in a different way. In relation to the causal pie model, it can therefore be stated that the size of a specific component is not fixed and can differ based on the presence of other component causes.

In the case of a gene-environment interaction, for example, the size of a component (e.g. genetic) can alter based on the presence of a certain environmental factor thereby more rapidly filling the causal pie. In this respect, poor nutrition and physical inactivity are two important lifestyle factors that have been associated with increased risks for Alzheimer's disease and cardiometabolic diseases. It has been hypothesized, that these lifestyle factors may interact with genetic variation in the *APOE* gene thereby altering the risk for disease onset. Therefore, studies focussing on genetic variation in *APOE* and the interaction with lifestyle factors on the risk of cardiometabolic diseases are of importance.

Causal inference and Mendelian Randomization

In general, the study of epidemiology is viewed as a collection of statistical tools used to elucidate associations of exposures to health and disease outcomes. However, a deeper understanding of this science is that of discovering of causal relationships between exposures and outcomes. Therefore, a key term is causal inference. An association or correlation between two variables may be necessary, but is not sufficient for causal inference, notably that one variable causes the other. If a causal variable can be identified and controlled, the disease outcome can be avoided. One method used to ascertain causality of observational associations, free of confounding and reverse causality, is Mendelian randomization [23, 24](**Figure 2**). Mendelian randomization is based on Mendel's second law, the law of independent assortment, which states that germline genetic variation is subject to the random allocation of alleles at conception. This method uses genetic variants, SNPs, as instrumental variable for the exposure of interest[25]. This instrumental variable is associated with the exposure, but not with the outcome of interest, except through its association with the exposure. Since the association between the genetic variants and the clinical outcome of interest are generally independent of environmental or behavioural factors, these SNPs can be used to avoid possible confounding or reverse causality and thereby enable to study the causal effect of an exposure of interest on an outcome.

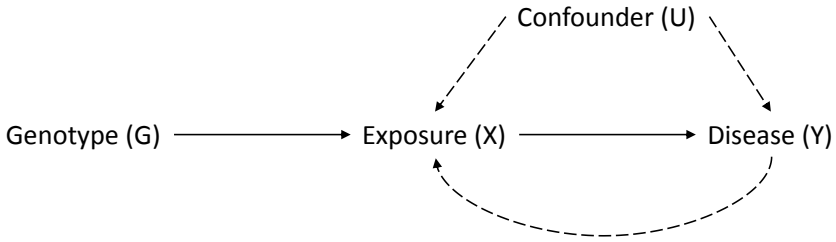


Figure 2. *The design of Mendelian randomization:* the genotype is associated with the exposure, the genotype is independent of confounders and the genotype is associated with the outcome, but only through the exposure of interest.

OUTLINE OF THIS THESIS

In this thesis, we will focus on the interplay between non-modifiable factors (genetics) and modifiable lifestyle factors (e.g. sleep, nutrition, physical activity) with cardiometabolic health. In **Chapter 2**, we aimed to identify early metabolic biomarkers that associate with insulin resistance in non-diabetic individuals and to what extend those biomarkers were associated with diabetes mellitus. This study has been performed in the Leiden Longevity Study, a study of long-lived families of Caucasian origin and the findings were validated in the Netherlands Epidemiology of Obesity (NEO) study. The NEO study is a population-based cohort study in overweight and obese middle-aged individuals from the Leiden region in The Netherlands. Next, we will focus on lifestyle factors. Since sleep is an important lifestyle factor in regard to cardiometabolic health, the following three chapters will be focused around sleep. The first two studies of this section have been performed in the NEO study. First, we will address the association between sleep duration and sleep quality with blood lipid levels and hepatic lipid content in **Chapter 3.1**. In **Chapter 3.2**, the focus will be on sleep duration and sleep quality with insulin resistance. From this chapter on, we will also focus on the genetic aspects of cardiometabolic health, since this study will also include a part that assesses the association between genetically determined sleep duration and insulin resistance. **Chapter 3.3** will focus on the identification of novel lipid loci if we take into account interaction with sleep duration in a large multi-ancestry analysis. This study is part of the CHARGE consortium. The Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium was formed to facilitate genome-wide association study meta-analyses and replication opportunities among multiple large and well-phenotyped longitudinal cohort studies. We will move from sleep duration to food intake and physical activity in **Chapter 4.1**. This chapter is a review in which we discuss the effect of physical activity, oily fish intake and omega-3 fatty acid intake on the risk of diseases associated with genetic variation in the *APOE* gene. Next, we will study whether the risk of cardiovascular disease indeed may be modified by physical activity and food intake in a large cohort study. This is done in **Chapter 4.2**, in which we will describe the associations between genetic variation in *APOE* and lifestyle factors in the UK Biobank. The UK Biobank is a large long-term biobank study with deep genetic, physical and health data collected on ~500,000 individuals in the United Kingdom. In this study, we aimed to investigate the presence of a gene-lifestyle interaction in relation to coronary artery disease incidence. The final two chapters of this thesis, **Chapters 5.1** and **5.2**, will focus on the association between alterations in hormone metabolism and cardiometabolic health. Specifically, we will assess the association between genetically determined thyroid hormone levels and the risk of type 2 diabetes and measures of glucose

homeostasis using Mendelian Randomization in **Chapter 5.1**. In **Chapter 5.2**, the causal of effect of thyroid hormone status on diabetes mellitus and the effect of BMI will be studied. This study will be performed in the UK Biobank. In the last part of this thesis (**Chapter 6**), the study findings together with their implications for future research are discussed.

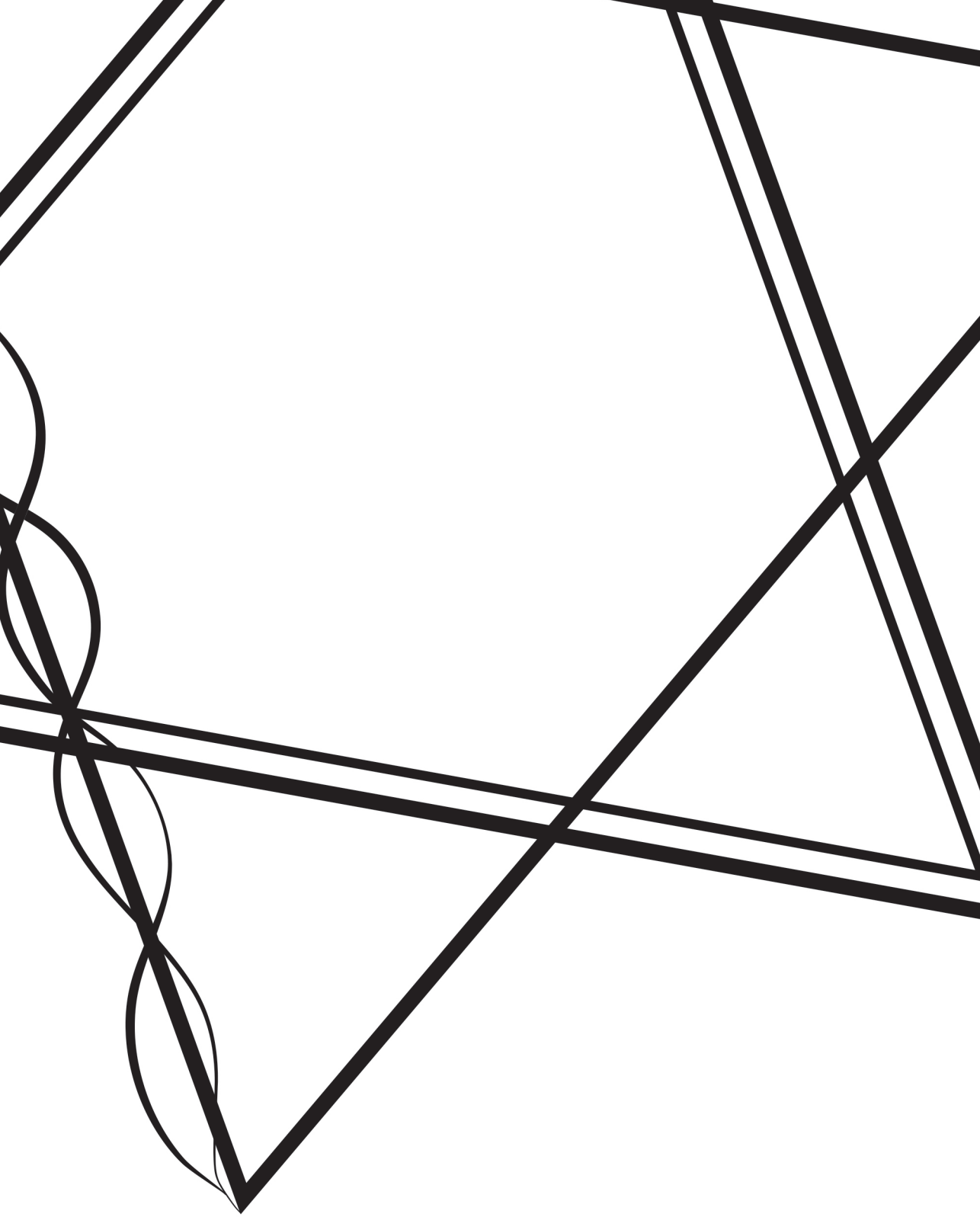
REFERENCES

1. Omran AR. The epidemiologic transition. A theory of the epidemiology of population change. Milbank Mem Fund Q. 1971;49(4):509-38.
2. Afshin A, Reitsma MB, Murray CJL. Health Effects of Overweight and Obesity in 195 Countries. N Engl J Med. 2017;377(15):1496-7.
3. Roth GA, Johnson C, Abajobir A, Abd-Allah F, Abera SF, Abyu G, et al. Global, Regional, and National Burden of Cardiovascular Diseases for 10 Causes, 1990 to 2015. J Am Coll Cardiol. 2017;70(1):1-25.
4. Emerging Risk Factors C, Sarwar N, Gao P, Seshasai SR, Gobin R, Kaptoge S, et al. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. Lancet. 2010;375(9733):2215-22.
5. Medicine Io. Genes, Behavior, and the Social Environment: Moving Beyond the Nature/Nurture Debate. Hernandez LM, Blazer DG, editors. Washington, DC: The National Academies Press; 2006. 384 p.
6. Wensink M, Westendorp RG, Baudisch A. The causal pie model: an epidemiological method applied to evolutionary biology and ecology. Ecol Evol. 2014;4(10):1924-30.
7. Rothman KJ. Causes. Am J Epidemiol. 1976;104(6):587-92.
8. Levine JA, Eberhardt NL, Jensen MD. Role of nonexercise activity thermogenesis in resistance to fat gain in humans. Science. 1999;283(5399):212-4.
9. Zielinski MR, McKenna JT, McCarley RW. Functions and Mechanisms of Sleep. AIMS Neurosci. 2016;3(1):67-104.
10. Allada R, Siegel JM. Unearthing the phylogenetic roots of sleep. Curr Biol. 2008;18(15):R670-R9.
11. Wong PM, Manuck SB, DiNardo MM, Korytkowski M, Muldoon MF. Shorter sleep duration is associated with decreased insulin sensitivity in healthy white men. Sleep. 2015;38(2):223-31.
12. Reutrakul S, Van Cauter E. Interactions between sleep, circadian function, and glucose metabolism: implications for risk and severity of diabetes. Annals of the New York Academy of Sciences. 2014;1311:151-73.
13. Cappuccio FP, Taggart FM, Kandala NB, Currie A, Peile E, Stranges S, et al. Meta-analysis of short sleep duration and obesity in children and adults. Sleep. 2008;31(5):619-26.
14. Cappuccio FP, D'Elia L, Strazzullo P, Miller MA. Quantity and quality of sleep and incidence of type 2 diabetes: a systematic review and meta-analysis. Diabetes Care. 2010;33(2):414-20.
15. Wu Y, Zhai L, Zhang D. Sleep duration and obesity among adults: a meta-analysis of prospective studies. Sleep Med. 2014;15(12):1456-62.
16. Xi B, He D, Zhang M, Xue J, Zhou D. Short sleep duration predicts risk of metabolic syndrome: a systematic review and meta-analysis. Sleep Med Rev. 2014;18(4):293-7.
17. Bianco AC, Larsen PR. Cellular and structural biology of the deiodinases. Thyroid. 2005;15(8):777-86.
18. Bianco AC, Salvatore D, Gereben B, Berry MJ, Larsen PR. Biochemistry, cellular and molecular biology, and physiological roles of the iodothyronine selenodeiodinases. Endocr Rev. 2002;23(1):38-89.
19. Maratou E, Hadjidakis DJ, Kollias A, Tsegka K, Peppa M, Alevizaki M, et al. Studies of insulin resistance in patients with clinical and subclinical hypothyroidism. Eur J Endocrinol. 2009;160(5):785-90.

20. Deelen J, Beekman M, Uh HW, Broer L, Ayers KL, Tan Q, et al. Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. Hum Mol Genet. 2014;23(16):4420-32.
21. Pilling LC, Atkins JL, Bowman K, Jones SE, Tyrrell J, Beaumont RN, et al. Human longevity is influenced by many genetic variants: evidence from 75,000 UK Biobank participants. Aging (Albany NY). 2016;8(3):547-60.
22. Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. JAMA. 1997;278(16):1349-56.
23. Katan MB. Apolipoprotein E isoforms, serum cholesterol, and cancer. Lancet. 1986;1(8479):507-8.
24. Trompet S, Jukema JW, Katan MB, Blauw GJ, Sattar N, Buckley B, et al. Apolipoprotein e genotype, plasma cholesterol, and cancer: a Mendelian randomization study. Am J Epidemiol. 2009;170(11):1415-21.
25. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med. 2008;27(8):1133-63.

PART I

Metabolomics and glucose metabolism



CHAPTER 2

Metabolomics analyses in non-diabetic middle-aged individuals reveal metabolites impacting early glucose disturbances and insulin sensitivity

Maxime M Bos, Raymond Noordam, Kate Bennett, Marian Beekman, Dennis O Mook-Kanamori, Ko Willems van Dijk, P Eline Slagboom, Torbjörn Lundstedt, Izabella Surowiec, Diana van Heemst

Metabolomics 2020;16(3):35

ABSTRACT

Several plasma metabolites have been associated with insulin resistance and type 2 diabetes mellitus. We aimed to identify plasma metabolites associated with different indices of early disturbances in glucose metabolism and insulin sensitivity. This cross-sectional study was conducted in a subsample of the Leiden Longevity Study comprising individuals without a history of diabetes mellitus (n = 233) with a mean age of 63.3 ± 6.7 years of which 48.1% were men. We tested for associations of fasting glucose, fasting insulin, HOMA-IR, Matsuda Index, Insulinogenic Index and glycated hemoglobin with metabolites (Swedish Metabolomics Platform) using linear regression analysis adjusted for age, sex and BMI. Results were validated internally using an independent metabolomics platform (Biocrates platform) and replicated externally in the independent Netherlands Epidemiology of Obesity (NEO) study (Metabolon platform) (n = 545, mean age of 55.8 ± 6.0 years of which 48.6% were men). Moreover, in the NEO study, we replicated our analyses in individuals with diabetes mellitus (cases: n = 36; controls = 561). Out of the 34 metabolites, a total of 12 plasma metabolites were associated with different indices of disturbances in glucose metabolism and insulin sensitivity in individuals without diabetes mellitus. These findings were validated using a different metabolomics platform as well as in an independent cohort of non-diabetics. Moreover, tyrosine, alanine, valine, tryptophan and alpha-ketoglutaric acid levels were higher in individuals with diabetes mellitus. We found several plasma metabolites that are associated with early disturbances in glucose metabolism and insulin sensitivity of which five were also higher in individuals with diabetes mellitus.

INTRODUCTION

Over the past decades, the incidence of type 2 diabetes mellitus (T2D) has increased, partly due to the ever increasing prevalence of obesity [1, 2]. T2D is preceded by several disturbances in glucose metabolism, which can be recognized in the pre-disease state. A novel approach that is being increasingly used to gain additional insight in the disturbances in glucose metabolism before the development of T2D is high-throughput metabolomics. Metabolomics offers the possibility to comprehensively measure a broad range of metabolites in tissues and biological fluids. Multiple observational and causal association studies revealed metabolites such as phospholipids, triacylglycerols, ketone bodies, sphingomyelins, acyl-carnitines and organic acids to be linked to future risk of T2D onset [3-13].

To recognize early stages of T2D one needs increased understanding of the mechanisms contributing to disturbances in glucose metabolism and insulin resistance among individuals without T2D. To quantify early disturbances in glucose metabolism, several indices have been developed based on an oral glucose tolerance test or based on fasting/postprandial plasma and insulin levels. For example, the homeostatic model assessments can be used to quantify insulin resistance (HOMA-IR) based on fasting glucose and insulin levels [14-17]. If both fasting and post-prandial measures are available, indices such as the Matsuda Index or Insulinogenic index can be used to assess insulin resistance and β-cell function respectively[15]. The Matsuda Index reflects both hepatic and peripheral tissue insulin sensitivity and is therefore considered to be an index of whole-body insulin sensitivity[14, 15]. These and other indices all reflect different yet partly overlapping aspects of glucose tolerance and insulin sensitivity that play a role in T2D onset.

In this study, we investigated which metabolites measured by a GC-MS assay are associated with early indices of disturbances in glucose metabolism and insulin sensitivity in individuals that do not use glucose lowering drugs and do not have a history of diabetes mellitus. To this end, we performed association analyses of plasma metabolites with fasting glucose, fasting insulin, the insulinogenic index, two indices of insulin resistance (HOMA-IR and Matsuda) and HbA1c. Subsequently, we replicated the significant findings internally using a different metabolomics platform as well as in an independent cohort in both non-diabetics and individuals with diabetes mellitus.

METHODS

Study design of the Leiden Longevity Study

The main analyses for the present study were embedded in the Leiden Longevity Study, which aims to investigate biomarkers associated with familial longevity and healthy ageing. A more detailed description of the study design and recruitment strategy has been described elsewhere [18]. In short, between 2003 and 2006, a total of 421 long-lived families were recruited, without selection based on health condition or demographics. Families were included when at least two long-lived siblings were still alive and fulfilled the age criteria of being at least 89 years for men and 91 years for women. Of these long-lived families, we also recruited 1,671 of their offspring and 744 partners thereof as controls resembling the general Dutch population at middle age.

For the present study, we used fasting and postprandial blood samples collected between 2006 and 2008 from a subpopulation (N = 280) of the Leiden Longevity Study who lived in close approximation (<45 min by car) from the research center, as described previously [19]. Glucose tolerance was assessed according to a 2-hour oral glucose tolerance test, conducted with a standard loading dose of 75g of glucose per 300mL of water and venous blood samples were drawn at time points of 0, 30, 60, and 120 minutes after glucose loading. We excluded participants that used glucose lowering drugs, had a history of diabetes mellitus, who were not fasted before taking glucose tolerance test, had no fasting glucose or insulin measures, or had incomplete/ unreliable postprandial data. In the present study, we therefore included a total of 233 participants.

The Leiden Longevity Study was approved by the medical ethics committee of the Leiden University Medical Center. All participants provided written informed consent.

Measures of insulin resistance

Fasting plasma glucose concentrations and glycated hemoglobin levels were determined by enzymatic and colorimetric methods (Roche Modular Analytics P800, Roche Diagnostics, Mannheim, Germany; CV < 5%) and serum insulin concentrations were determined by an immunometric method (Siemens Immulite 2500, Siemens Healthcare Diagnostics, Breda, The Netherlands; CV < 5%). All analyses were performed in the central clinical chemical laboratory of the Leiden University Medical Center. Fasting glucose and insulin levels were used to calculate the Homeostatic Model Assessment of Insulin Resistance (HOMA-IR) index as a marker for hepatic insulin resistance. The HOMA-IR was calculated using (fasting insulin * fasting glucose) / 22.5.

The Insulinogenic Index was calculated as (insulin concentration$_{30min}$ – fasting insulin concentration) / (glucose concentration$_{30min}$ – fasting glucose concentration) and the Matsuda Index as 1000 / (squared root (fasting glucose * fasting insulin) * (average glucose * average insulin))[15].

Metabolomic profiling

Metabolomic profiling has been performed in three separate analytical batches analysed on three different days, of which the first batch consisted of cases with different levels of metabolic syndrome score as described previously[20]. The samples of the other participants were separated based on the gender of the participants, resulting in a batch of only women (Batch 2) or only men (Batch 3). These three batches were used for the analyses in this study. Fasting plasma samples from the participants were thawed on ice; 630 µL of extraction mixture (H_2O:methanol (1:9, v/v)) was added to 70 µL of plasma. Extraction of the metabolites from the sample was then carried out using a MM301 vibration Mill (Retsch GmbH & Co. KG, Haan, Germany) at a frequency of 30 Hz for 2 min. Samples were stored on ice for 2 hours to allow protein precipitation, after which they were centrifuged at a relative central force of 18,000 for 10 min at 4˚C. An aliquot (200 µL) of the resulting supernatant was transferred to a glass vial and evaporated to dryness at room temperature in a miVac QUATTRO concentrator (Genevac LTD, Ipswich, UK). Gas chromatography-mass spectrometry (GC-MS) (Batch 1, 2 and 3) analyses was performed after metabolite derivatization as described before [21].

Non-processed files from GC-MS were exported in NetCDF format to a MATLAB-based in-house script where all data pre-treatment procedures such as baseline correction, chromatogram alignment, and peak deconvolution were performed. Metabolite identification, was implemented within the script and was based on the retention index (RI) values and MS spectra from the in-house mass spectra library established by the Swedish Metabolomics Centre (Umeå, Sweden) and consisting of 585 compounds (Level 1 identification according to the Metabolomics Standards Initiative [22]). However, since the library covers a wide range of compound classes, and includes a significant number of compounds present in samples from other species than humans (for example plants, bacteria, etc.), as expected, not all metabolites present in the library could be identified in our samples. In the three analytical batches, respectively 105, 48 and 57 compounds were identified, with 34 compounds present in all three batches. In order to strengthen our findings, in calculations we only included those metabolites that were confidently measured in all three batches.

Internal validation

For internal validation, we used metabolomic data measured in the Leiden Longevity Study using the commercially available AbsoluteIDQ p180 (Biocrates Life Sciences, AG, Innsbruck, Austria) mass spectrometry (MS)-based assay kit. Fasting plasma samples were collected from study participants and were stored at −80 °C. A different (unique) aliquot of the sample was used, thus avoiding potential bias caused by introducing a freeze/thaw step when using the same sample for discovery and validation. In this study, we included only those metabolites that overlapped with those that were identified using the GC-MS approach as being significantly associated with indices of disturbances in glucose metabolism and insulin sensitivity after correction for multiple testing, note that only 5 out of the 12 significant metabolites were also present on the Biocrates AbsoluteIDQ p180 kit, namely: phenylalanine, proline, tryptophan, tyrosine and valine.

External validation

For external replication, we used data from the independent Netherlands Epidemiology of Obesity (NEO) study, which had no overlap with the LLS study in regard of study participants. A detailed description of the NEO study design and rationale has been provided elsewhere [23]. We included participants with available data on the same indices of glucose metabolism (e.g. fasting glucose and insulin, HOMA-IR, Matsuda Index, Insulinogenic Index and glycated hemoglobin) as available in the LLS and an LC-MS and GC-MS based metabolic profile (Metabolon, Inc.), who were fasted and drank a complete liquid mixed meal (n= 545). Moreover, we assessed the associations in participants with diabetes mellitus (cases: n = 36; controls = 561). Presence of diabetes mellitus was defined as the usage of glucose lowering medication and/or having a history of diabetes mellitus and/or having fasting glucose levels of ≥7.0 mmol/L.

Metabolomic profiling was performed using ultrahigh-performance liquid-phase chromatography and gas chromatography separation, coupled with tandem mass spectrometry at Metabolon, Inc. using established procedures (12). Fasting plasma samples were collected from study participants and were stored at −80 °C. Here, we only analyzed the identified metabolites in the LLS association study (those metabolites that were significant after correction for multiple testing) and did not include all of the metabolites measured using this platform.

Statistical analyses

All analyses were performed using the R statistical environment[24]. Before analyses, we removed those observations that could be considered outliers based on biological plausible reference values for these measurements. To gain insight in the interrelations of the metabolites included in the present study, we estimated a correlation network

using Pearson correlations for the 34 scaled metabolites measured using the GC-MS approach in the LLS study. The Pearson correlations for each batch were obtained separately and we took the average correlation of these batches taking into account the number of individuals included in each batch. The qgraph package[25], using an absolute weight of edges of >0.5 to be shown, was used to obtain the plot.

We conducted univariate analyses on the metabolites using linear regression analysis adjusted for age, sex and body mass index. All measures of glucose metabolism were log-transformed, independent of normality of the variables in order to improve comparability among the measurements. Peak areas (GC-MS) and metabolite concentrations (Biocrates LC-MS platform) were log-transformed and subsequently standardized using scaling to approach a standard normal distribution (mean = 0, s.d. = 1). Hence, results can be interpreted as the difference in standard deviation per unit increase in measured index. Because of the applied distribution of the individuals in the different batches rather than random allocation, proper correction for batch effect was not possible. Instead, we conducted all univariate analyses separately for the different batches; derived effect estimates were subsequently meta-analysed using a fixed-effect model, assuming a similar direction of effect among batches, by using the rmeta package[26]. Meta-analysis was only performed for those metabolites that were measured in all three batches, which resulted in a total of 34 metabolites. To correct for multiple testing, we calculated the number of independent metabolites based on the methodology described by Li *et al*. [27], and corrected our threshold for statistical significance accordingly. We obtained a total of 24 independent metabolites and used a p-value of 0.05/24 (~0.0021) as a threshold for significance. Analyses were visualized using the ggplot2 package [28].

For internal validation, we repeated our analyses using significant metabolites from our discovery analyses. Of these metabolites, only five metabolites were overlapping with the Biocrates AbsoluteIDQ p180 kit. We performed the same analyses using the same statistics and covariables as in the discovery analyses, however, we used a p-value of <0.05 as level of statistical significance. For external replication, analyses were repeated with the metabolites that were identified in the LLS study and could also be measured by the GC-MS Metabolon platform, comprising twelve metabolites present in the NEO study and Metabolon data. We used the same statistical methods as described previously, however, since we tested hypothesis generated in the discovery phase, we considered a two-sided p-value of <0.05 as statistically significant.

RESULTS

Population characteristics

A total of 233 participants were included in the main study with a mean age of 63.3 ± 6.7 years of which 48.1% were men. The population characteristics of this study are shown in **Table 1**, separately for each batch and combined for all batches. Characteristics of the study population are presented as means with accompanying standard deviations for normally distributed variables, median with interquartile range for non-normal distributed variables and number with percentages for proportional variables. Participants in Batch 3 used slightly more antihypertensive and lipid lowering medication as compared to the other batches.

Network analysis of metabolites

In order to have a better understanding of the interrelations of metabolites in this study, we calculated and visualized the correlations between the different metabolites. In **Figure 1**, a correlation network is shown of the primary metabolite set tested in the LLS study comprising 34 metabolites. Several amino acids were correlated, such as tryptophan, tyrosine, phenylalanine and valine. The strongest correlations were observed between fatty acids, such as myristic acid, dodecanoic acid and octadecanoic acid.

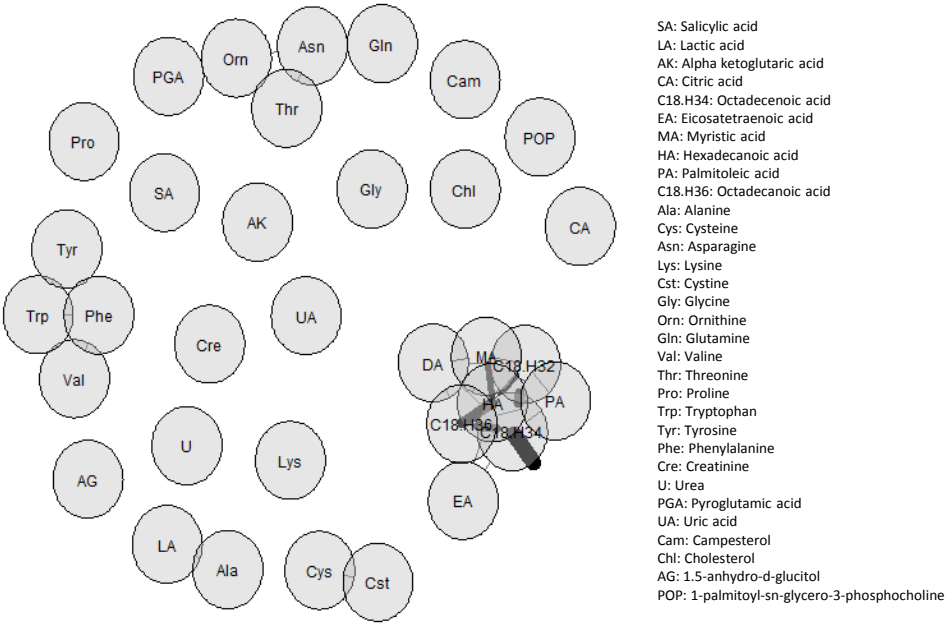


Figure 1 *Correlation network estimated using Pearson correlations.* Each node represents one metabolite. The edges indicate the strength of the correlation. For this plot, only edges with an absolute weight of >0.5 are shown.

Associations between different indices of glucose metabolism with the metabolomic profile
We first tested for association of the 34 single metabolites that were measured in all three batches with six variables of glucose metabolism and insulin sensitivity in the Leiden Longevity Study adjusted for age, sex and BMI. Twelve metabolites were identified to be associated (p-value < 0.05/24) with markers of glucose metabolism, as shown in **Figure 2** and **Online Supplementary Table 1**. Tyrosine, hexadecanoic acid, lysine and alpha-ketoglutaric acid levels were associated with fasting glucose concentrations. Levels of alanine, tyrosine, valine, phenylalanine, tryptophan, proline and uric acid were positively associated with fasting insulin and HOMA-IR and negatively associated with the Matsuda index. Alpha-ketoglutaric acid was positively associated with HOMA-IR and negatively with the Matsuda Index, however, not with fasting insulin. Moreover, lysine levels were positively associated with HOMA-IR, however, not with fasting insulin concentrations or the Matsuda Index. Levels of hexadecanoic acid, myristic acid and octadecanoic acid were associated with Insulinogenic Index. None of these metabolites were associated with either fasting insulin, HOMA-IR or the Matsuda Index. However, hexadecanoic acid levels positively associated with fasting glucose concentrations. Finally, glycated hemoglobin was not associated with any of the metabolites.

Internal validation of main results
Next, we validated the significant observations using the Biocrates platform in order for us to test the robustness of these findings when using a different metabolomics platform. Note that only 5 out of the 12 significant metabolites overlapped with the Biocrates platform (phenylalanine, proline, tryptophan, tyrosine and valine). In **Online Supplementary Table 2**, effect estimates for the association of measures of glucose metabolism with these metabolites are shown. Effect estimates pointed in the same direction and were similar to those observed using the Swedish Metabolomics Platform. For fasting glucose, only tyrosine was present in the Biocrates platform and we were able to validate this result. Proline, tryptophan, tyrosine and valine were positively associated with fasting insulin, HOMA-IR and negatively with the Matsuda Index, thereby validating the previous results. Phenylalanine negatively associated with the Matsuda Index, however, not with fasting insulin and the HOMA-IR, thereby only validating the result for the Matsuda Index. The metabolites that associated with the Insulinogenic Index were not present in the Biocrates platform and therefore, we did not validate these findings. Moreover, for glycated hemoglobin we did not identify any significant associations and therefore we did not validate any finding.

Table 1. *Characteristics of study population stratified by batches and combined.*

	Batch 1 (N=82)
Age in years, mean (SD)	63.2 (6.2)
Men, N (%)	39 (47.6)
BMI in kg/m2, mean (SD)	26.3 (4.9)
Body fat %, mean (SD)	31.3 (9.4)
Fasting glucose in mmol/L, mean (SD)	5.2 (0.7)
Fasting insulin in mU/L, median (IQR)	7.0 (4.0-11.8)
HOMA-IR, median (IQR)	1.6 (0.8-2.7)
Insulinogenic Index, median (IQR)	0.6 (0.4-1.1)
Matsuda Index, median (IQR)	24.6 (12.7-45.3)
HbA1c in %, mean (SD)	5.1 (0.4)
ASAT in U/L, median (IQR)	22.1 (7.2)
ALAT in U/L, median (IQR)	16.0 (13.0-19.8)
GGT in U/L, median (IQR)	22.0 (15.0-35.0)
CRP in mg/dL, median (IQR)	1.2 (0.7-2.5)
Total cholesterol in mmol/L, mean (SD)	5.3 (1.1)
HDL-cholesterol in mmol/L, mean (SD)	1.5 (0.4)
LDL-cholesterol in mmol/L, mean (SD)	3.2 (0.9)
Triglycerides in mmol/L, mean (SD)	1.5 (0.8)
Hypertension, N (%)	20 (24.4)
Statin use, N (%)	7 (8.50

Abbreviations: BMI, Body Mass Index; ALAT, alanine transaminase; ASAT, aspartate transaminase; GGT, gamma-glutamyltransferase; HbA1c, glycated haemoglobin; HDL, high-density lipoprotein; HOMA-IR, Homeostatic Model Assessment of Insulin Resistance; hsCRP, high-sensitivity C-reactive protein; IQR, interquartile range; LDL, low-density lipoprotein; N, number of participants; SD, standard deviation.

	Batch 2 (N=78)	Batch 3 (N=73)	Combined (N=233)
	61.5 (7.0)	65.4 (6.4)	63.3 (6.7)
	0 (0.0)	73 (100.0)	112 (48.1)
	25.9 (3.2)	26.9 (3.0)	26.5 (3.8)
	35.6 (6.5)	26.3 (5.6)	31.2 (8.3)
	5 (0.4)	5.1 (0.5)	5.1 (0.5)
	6.0 (4.0-9.0)	7.0 (4.0-10.0)	6.0 (4.0-10.0)
	1.2 (0.9-2.0)	1.5 (1.0-2.2)	1.5 (0.9-2.2)
	0.8 (0.5-1.2)	0.7 (0.4-1.2)	0.7 (0.4-1.1)
	26.9 (18.7-42.7)	25.1 (17.4-35.7)	26.3 (17.1-41.2)
	5.1 (0.4)	5.0 (0.5)	5.1 (0.4)
	22.5 (8.6)	21.5 (4.6)	22.1 (7.0)
	15.0 (12.0-19.8)	17.0 (15.0-21.0)	16.0 (13.0-20.0)
	18.0 (13.0-24.8)	29.0 (21.0-40.0)	22.0 (15.0-33.0)
	1.1 (0.7-1.9)	1.0 (0.6-2.0)	1.1 (0.6-2.3)
	5.6 (1.0)	5.1 (1.0)	5.3 (1.0)
	1.7 (0.5)	1.4 (0.4)	1.5 (0.4)
	3.3 (0.9)	3.1 (0.9)	3.2 (0.9)
	1.3 (0.5)	1.3 (0.5)	1.4 (0.6)
	19 (24.4)	25 (34.2)	64 (27.5)
	10 (12.8)	10 (13.7)	27 (11.6)

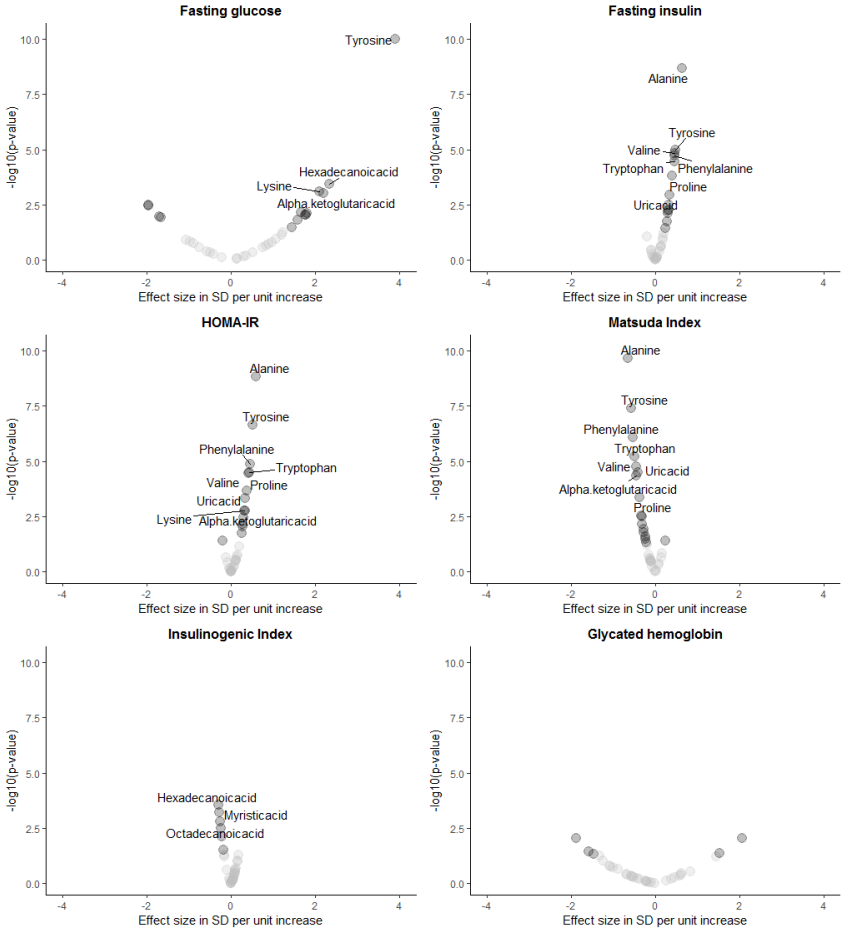


Figure 2. *Associations of plasma metabolites with measures of glucose metabolism*. Analyses can be interpreted as the difference in metabolite level in standard deviation in relation to unit increase of the exposure. The difference in exposure (in standard deviation) is presented on the x-axis; the -log(p-value) of the comparison is presented on the y-axis. Metabolites that were labelled in the figures were those that remained significant after correction for multiple testing ($p < 2.1e-3$); compounds with a p-value < 0.05 are presented as solid black dots.

External validation of main results in individuals with and without diabetes mellitus
We validated our results for the identified metabolites in an independent subset of the NEO cohort in individuals without diagnosis of diabetes mellitus that did not use glucose lowering medication using the same indices of glucose metabolism and insulin sensitivity. As can be seen in **Table 2** and **Table 3**, tyrosine, hexadecanoic acid and alpha-ketoglutaric acid were positively associated with levels of fasting glucose, thereby replicating the associations as observed in the LLS study, except for lysine. Levels of tyrosine, alanine, valine and phenylalanine were positively associated with fasting insulin and HOMA-IR levels in both the LLS and the NEO study. Additionaly, alpha-ketoglutaric acid significantly associated with HOMA-IR. However, proline, tryptophan (fasting insulin and HOMA-IR) and uric acid (HOMA-IR) did not associate in the NEO study. Tyrosine, alanine, valine, phenylalanine, alpha-ketoglutaric acid, hexadecanoic acid and uric acid were negatively associated with the Matsuda Index, similar as in the LLS study. However, tryptophan was not associated with the Matsuda Index in the NEO study. None of the metabolites that associated with the Insulinogenic Index in the LLS study were replicated in the NEO study.

In addition, we studied which of the twelve metabolites associated with diabetes mellitus (cases = 36; controls = 561). As compared to individuals without diabetes mellitus, individuals with diabetes mellitus had higher levels of tyrosine, alanine, valine, tryptophan and alpha-ketoglutaric acid (**Table 3**).

Table 2. *Associations of blood metabolites with measures of glucose metabolism in NEO using Metabolon.*

	Fasting glucose			Fasting insulin		
	Beta	SE	P-value	Beta	SE	P-value
Tyrosine	1.29	0.34	<0.001	0.26	0.07	<0.001
Lysine	-0.45	0.36	0.211			
Alanine				0.36	0.07	<0.001
Valine				0.14	0.06	0.025
Proline				0.14	0.07	0.054
Phenylalanine				0.30	0.07	<0.001
Tryptophan				0.13	0.07	0.079
Hexadecanoic acid	0.84	0.36	0.021			
Alpha-ketoglutaric acid	1.38	0.35	<0.001			
Myristic acid						
Octadecanoic acid						
Uric acid						

Data presented as beta's with accompanying standard errors (SE) and p-values. Results can be interpreted as the difference in metabolite level in standard deviation in relation to unit increase of the glycemic trait. Analyses are adjusted for age, sex and body mass index.

	HOMA-IR			Matsuda Index			Insulinogenic Index		
	Beta	SE	P-value	Beta	SE	P-value	Beta	SE	P-value
	0.27	0.07	<0.001	-0.26	0.08	0.001			
	-0.11	0.07	0.122						
	0.37	0.07	<0.001	-0.48	0.08	<0.001			
	0.15	0.06	0.015	-0.22	0.07	0.002			
	0.11	0.07	0.092	-0.09	0.08	0.239			
	0.27	0.07	<0.001	-0.35	0.08	<0.001			
	0.11	0.07	0.111	-0.12	0.08	0.118			
							-0.07	0.05	0.123
	0.45	0.07	<0.001	-0.54	0.08	<0.001			
							-0.04	0.05	0.431
							-0.04	0.05	0.354
	0.09	0.06	0.148	-0.16	0.07	0.027			

Table 3. *Associations of blood metabolites with measures of glucose metabolism in NEO using Metabolon.*

	Diabetes Mellitus		
	Beta	Se	P-value
Tyrosine	1,93	(1,58 – 2,28)	<0,001
Lysine	1,31	(0,96 – 1,66)	0,134
Alanine	1,59	(1,24 – 1,94)	0,010
Valine	1,94	(1,59 – 2,29)	<0,001
Proline	1,30	(0,94 – 1,65)	0,150
Phenylalanine	1,35	(1,00 – 1,70)	0,097
Tryptophan	1,47	(1,12 – 1,82	0,033
Hexadecanoic acid	0,98	(0,63 – 1,33)	0,902
Alpha-ketoglutaric acid	1,71	(1,36 – 2,06)	0,003
Myristic acid	1,16	(0,81 – 1,51)	0,409
Octadecanoic acid	0,84	(0,48 – 1,19)	0,319
Uric acid	1,14	(0,78 – 1,49)	0,476

Data presented as odd ratio's (OR) with accompanying 95% confidence interval (95%CI) and p-values. Analyses are adjusted for age, sex and body mass index

DISCUSSION

We identified twelve metabolites to be associated with different indices of glucose metabolism and insulin sensitivity, in individuals without diabetes mellitus from the general population. These results were largely validated and externally replicated in an independent cohort. Moreover, five of the twelve metabolites (tyrosine, alanine, valine, tryptophan and alpha-ketoglutaric acid) were associated with diabetes mellitus. These results indicate that specific early alterations in the metabolic profile are already present in individuals without diabetes mellitus and these findings may therefore improve the understanding of mechanisms involved in diabetes mellitus etiology.

To date, several amino acids, sugar metabolites, and lipids have been associated with T2D risk in observational studies and causality has been investigated using Mendelian Randomization [3-13, 29, 30]. The present study replicates these observational findings in participants without diabetes mellitus, indicating that alterations in plasma metabolites levels in relation to perturbed glucose metabolism are already present in the non-diabetic population. For example, alanine, valine, tyrosine and phenylalanine have been consistently associated with the risk of developing type 2 diabetes mellitus in different studies[10]. A better understanding of the molecular mechanisms by which amino acids may impact insulin resistance may aid the identification of novel targets for future diabetes therapies. One of these pathways is via mitochondrial metabolism and the exocytosis of insulin granules. Next to ATP, the main factor in insulin secretion, other factors such as nucleotides, amino acids, enzymes or transporters and alanine aminotransferase have been identified to mediate insulin secretion. Alanine, our strongest association, has previously been described to directly affect β-cell function and insulin secretion[30]. In line, in this study we observed these metabolites to be associated with insulin sensitivity in individuals without diabetes mellitus, as well as in individuals with diabetes mellitus.

Some of the identified associations can be explained by correlations between the metabolites. For example, tyrosine, valine and phenylalanine were correlated which were also observed together in the analyses addressing the association with fasting insulin, HOMA-IR and the Matsuda Index. However, alpha-ketoglutaric acid was not correlated with any of the other metabolites and may therefore reflect an independent pathway that may contribute to disturbances in glucose metabolism and insulin sensitivity. For example, alpha-ketoglutaric acid is a key intermediate in the citric acid cycle and is important for amino acid formation and the urea cycle. Alpha-ketoglutaric acid may be a marker of protein degradation and gluconeogenesis. Hepatic glutamate dehydrogenase catalyzes the reversible oxidative deamination of glutamate to α-ketoglutarate and ammonia, bridging amino acid-to-glucose pathways. In the current study, we observed

higher levels of alpha-ketoglutaric acid to be associated with higher fasting glucose and HOMA-IR and a lower Matsuda Index in two independent cohorts of individuals without diabetes mellitus. Moreover, this metabolite was higher in individuals with diabetes mellitus. Since alpha-ketoglutaric acid was not associated with glycated hemoglobin, but only with measures of short-term insulin sensitivity, this metabolite may be more reflective of short-term glucose control instead of long-term glucose control.

A broad range of studies assessed the association between several lipid classes and type 2 diabetes mellitus[3-13]. Of these classes, for example, plasma phospholipids, triglycerides and sphingolipids were found to be associated with insulin resistance and type 2 diabetes mellitus onset [3-13]. In individuals with impaired fasting glucose and type 2 diabetes mellitus, higher levels of several saturated acids (e.g. palmitic and stearic) have been observed[10, 31]. Our results are in agreement with these findings as we have identified palmitic (hexadecanoic acid) acid to be associated with higher fasting glucose levels.

We observed distinct metabolites to be associated with different measures of glucose metabolism. Of interest is that the metabolites that were associated with fasting insulin and HOMA-IR, which are based on fasting measures of glucose metabolism, were the same metabolites that associated with the Matsuda Index, which also takes into account postprandial measures of glucose metabolism. Since the metabolites that we found to be associated with fasting insulin, HOMA-IR and Matsuda Index were mainly overlapping, these measures may reflect similar biological pathways in the development of type 2 diabetes mellitus. Interestingly, only saturated fatty acids were associated with the Insulinogenic Index in the LLS, which is specific for β-cell function. We did not observe any association between glycated hemoglobin and any of the investigated metabolites. Glycated hemoglobin is the only index that we included that is able to assess glucose metabolism over a longer time period. Since none of the metabolites associated with glycated hemoglobin, the present observations may be more reflective of short-term insulin resistance and non-pathological, merely normal physiology.

The main strength of this study is the internal and external validation of our results in an independent cohort. The same associations were observed, thereby strengthening our results. The participants in the Leiden Longevity Study received a glucose drink and the participants of the NEO study received a mixed meal as a challenge, emphasizing the robustness of the observed associations. However, some limitations of our study have to be acknowledged. Because of the selection of individuals in the three batches (no random selection), we were not able to harmonize for batch effects using statistical techniques. Instead, we performed the analyses separately for the three batches using

standardized metabolite data (mean = 0, s.d. = 1) and performed subsequent meta-analyses to combine the results. As a consequence, we minimized potential bias caused by the batch effects. In this study, we only observed higher levels of the identified metabolites to be associated with higher insulin resistance. High metabolite levels that are associated with lower insulin resistance might (1) have a smaller effect size that would require larger sample sizes, and (2) may only be measured on different platforms. Moreover, due to the observational nature of the study, we could not establish causality and address how the observed metabolites can affect insulin resistance or how insulin resistance may affect the metabolic profile, this will require dedicated prospective studies in the future.

Taken together, in a population of individuals without diabetes mellitus, we observed distinct metabolomic profiles to be associated with different measures of glucose metabolism. In total, we identified 12 metabolites to be associated with indices of glucose metabolism in individuals without diabetes mellitus. Most of these findings could be internally validated and externally replicated. Moreover, five of these metabolites associated with prevalent diabetes mellitus. Our results may improve the understanding of the mechanisms involved in disease etiology and thereby may contribute to improved diagnostics of the early metabolic disturbances preceding type 2 diabetes mellitus.

Author Conflict of Interest statement

Funding: The Leiden Longevity Study was funded by the Innovation Oriented research Program on Genomics (SenterNovem; IGE01014 and IGE5007), the Centre for Medical Systems Biology (CMSB), the Netherlands Genomics Initiative/Netherlands Organization for scientific research (NGI/NOW; 05040202 and 050-060-810, NCHA), Unilever PLC and the EU funded Network of Excellence Lifespan (FP6 036894) and the European Union's Seventh Framework Programme (FP7/2007-2011) under grant agreement number 259679. MMB was supported by the Leiden University fund/Piso-Kuperus Fonds. RN was supported by the EFSD Albert Renold Travel Fellowship Programme. The NEO study is supported by the participating Departments, the Division and the Board of Directors of the Leiden University Medical Centre, and by the Leiden University, Research Profile Area 'Vascular and Regenerative Medicine' This study received funding from the European Commission funded project HUMAN (Health-2013-INNOVATION-1-602757).

Conflict of Interest: DOMK is a part-time research consultant with Metabolon, Inc. All other authors declare to have no conflict of interest.

Acknowledgements

We would like to thank all participants, the secretary staff, Meriam H.G.F. van der Star and Ellen H.M. Bemer-Oorschot for their contribution to this study. The Swedish Metabolomics Centre is acknowledged for support with the GC-MS analyses. We express our gratitude to all individuals who participate in the Netherlands Epidemiology in Obesity study. We are grateful to all participating general practitioners for inviting eligible participants. We furthermore thank P. van Beelen and all research nurses for collecting the data and P. Noordijk and her team for sample handling and storage and I. de Jonge, MSc for data management of the NEO study.

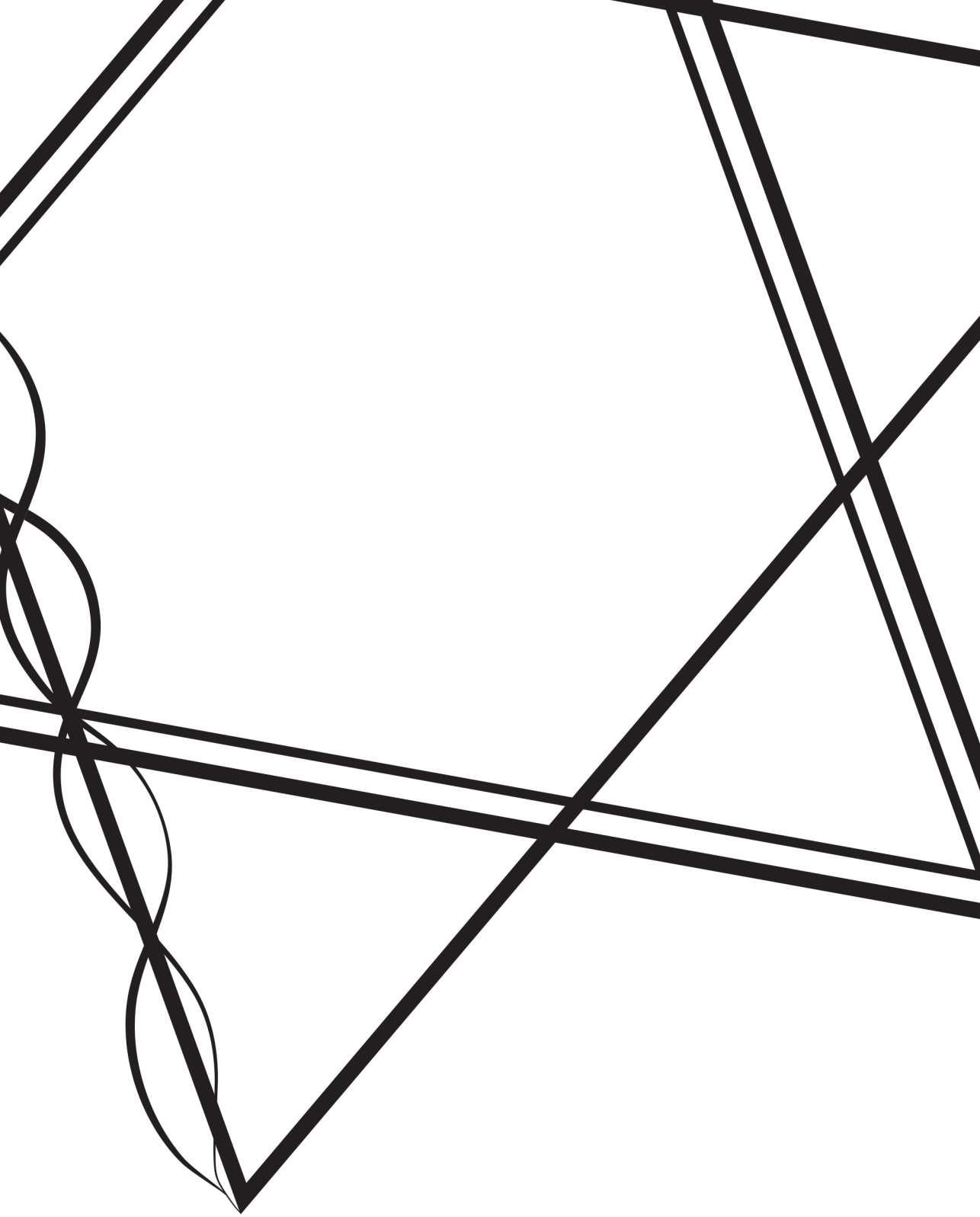
REFERENCES

1. Ogden CL, Carroll MD, Flegal KM. Prevalence of obesity in the United States. Jama. 2014;312(2):189-90.
2. Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of childhood and adult obesity in the United States, 2011-2012. Jama. 2014;311(8):806-14.
3. Mahendran Y, Cederberg H, Vangipurapu J, Kangas AJ, Soininen P, Kuusisto J, et al. Glycerol and fatty acids in serum predict the development of hyperglycemia and type 2 diabetes in Finnish men. Diabetes care. 2013;36(11):3732-8.
4. Mahendran Y, Vangipurapu J, Cederberg H, Stancakova A, Pihlajamaki J, Soininen P, et al. Association of ketone body levels with hyperglycemia and type 2 diabetes in 9,398 Finnish men. Diabetes. 2013;62(10):3618-26.
5. Tillin T, Hughes AD, Wang Q, Wurtz P, Ala-Korpela M, Sattar N, et al. Diabetes risk and amino acid profiles: cross-sectional and prospective analyses of ethnicity, amino acids and diabetes in a South Asian and European cohort from the SABRE (Southall And Brent REvisited) Study. Diabetologia. 2015;58(5):968-79.
6. Wurtz P, Soininen P, Kangas AJ, Ronnemaa T, Lehtimaki T, Kahonen M, et al. Branched-chain and aromatic amino acids are predictors of insulin resistance in young adults. Diabetes care. 2013;36(3):648-55.
7. Wurtz P, Tiainen M, Makinen VP, Kangas AJ, Soininen P, Saltevo J, et al. Circulating metabolite predictors of glycemia in middle-aged men and women. Diabetes care. 2012;35(8):1749-56.
8. Mook-Kanamori DO, Selim MM, Takiddin AH, Al-Homsi H, Al-Mahmoud KA, Al-Obaidli A, et al. 1,5-Anhydroglucitol in saliva is a noninvasive marker of short-term glycemic control. J Clin Endocrinol Metab. 2014;99(3):E479-83.
9. t Hart LM, Vogelzangs N, Mook-Kanamori DO, Brahimaj A, Nano J, van der Heijden A, et al. Blood Metabolomic Measures Associate With Present and Future Glycemic Control in Type 2 Diabetes. J Clin Endocrinol Metab. 2018;103(12):4569-79.
10. Guasch-Ferre M, Hruby A, Toledo E, Clish CB, Martinez-Gonzalez MA, Salas-Salvado J, et al. Metabolomics in Prediabetes and Diabetes: A Systematic Review and Meta-analysis. Diabetes care. 2016;39(5):833-46.
11. Liu J, van Klinken JB, Semiz S, van Dijk KW, Verhoeven A, Hankemeier T, et al. A Mendelian Randomization Study of Metabolite Profiles, Fasting Glucose, and Type 2 Diabetes. Diabetes. 2017;66(11):2915-26.
12. Wang Q, Holmes MV, Davey Smith G, Ala-Korpela M. Genetic Support for a Causal Role of Insulin Resistance on Circulating Branched-Chain Amino Acids and Inflammation. Diabetes care. 2017;40(12):1779-86.
13. Fall T, Salihovic S, Brandmaier S, Nowak C, Ganna A, Gustafsson S, et al. Non-targeted metabolomics combined with genetic analyses identifies bile acid synthesis and phospholipid metabolism as being associated with incident type 2 diabetes. Diabetologia. 2016;59(10):2114-24.
14. Retnakaran R, Shen S, Hanley AJ, Vuksan V, Hamilton JK, Zinman B. Hyperbolic relationship between insulin secretion and sensitivity on oral glucose tolerance test. Obesity (Silver Spring). 2008;16(8):1901-7.
15. Singh B, Saxena A. Surrogate markers of insulin resistance: A review. World J Diabetes. 2010;1(2):36-47.

16. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia. 1985;28(7):412-9.
17. Wallace TM, Levy JC, Matthews DR. Use and abuse of HOMA modeling. Diabetes care. 2004;27(6):1487-95.
18. Schoenmaker M, de Craen AJ, de Meijer PH, Beekman M, Blauw GJ, Slagboom PE, et al. Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. Eur J Hum Genet. 2006;14(1):79-84.
19. Rozing MP, Westendorp RG, de Craen AJ, Frolich M, de Goeij MC, Heijmans BT, et al. Favorable glucose tolerance and lower prevalence of metabolic syndrome in offspring without diabetes mellitus of nonagenarian siblings: the Leiden longevity study. J Am Geriatr Soc. 2010;58(3):564-9.
20. Surowiec I, Noordam R, Bennett K, Beekman M, Slagboom PE, Lundstedt T, et al. Metabolomic and lipidomic assessment of the metabolic syndrome in Dutch middle-aged individuals reveals novel biological signatures separating health and disease. Metabolomics. 2019;15(2):23.
21. A J, Trygg J, Gullberg J, Johansson AI, Jonsson P, Antti H, et al. Extraction and GC/MS analysis of the human blood plasma metabolome. Anal Chem. 2005;77(24):8086-94.
22. Salek RM, Steinbeck C, Viant MR, Goodacre R, Dunn WB. The role of reporting standards for metabolite annotation and identification in metabolomic studies. Gigascience. 2013;2(1):13.
23. de Mutsert R, den Heijer M, Rabelink TJ, Smit JW, Romijn JA, Jukema JW, et al. The Netherlands Epidemiology of Obesity (NEO) study: study design and data collection. Eur J Epidemiol. 2013;28(6):513-23.
24. Team RDC. R: A Language and Environment for Statistical Computing. 2008.
25. Epskamp S, Cramer AOJ, Waldorp LJ, Schmittmann VD, Borsboom D. qgraph: Network Visualizations of Relationships in Psychometric Data. J Stat Softw. 2012;48(4):1-18.
26. Lumley T. rmeta: Meta-Analysis. 2018.
27. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity. 2005;95(3):221-7.
28. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. 2009.
29. Rawat A, Misra G, Saxena M, Tripathi S, Dubey D, Saxena S, et al. (1)H NMR based serum metabolic profiling reveals differentiating biomarkers in patients with diabetes and diabetes-related complication. Diabetes Metab Syndr. 2019;13(1):290-8.
30. Newsholme P, Brennan L, Rubi B, Maechler P. New insights into amino acid metabolism, beta-cell function and diabetes. Clin Sci (Lond). 2005;108(3):185-94.
31. Palomer X, Pizarro-Delgado J, Barroso E, Vazquez-Carrera M. Palmitic and Oleic Acid: The Yin and Yang of Fatty Acids in Type 2 Diabetes Mellitus. Trends Endocrinol Metab. 2018;29(3):178-90.

PART II

Sleep



CHAPTER 3.1

Associations between sleep duration and quality with serum and hepatic lipids: the Netherlands Epidemiology of Obesity Study.

Maxime M Bos, Raymond Noordam, Rosa van den Berg, Renée de Mutsert, Frits R Rosendaal, Gerard Jan Blauw, Patrick CN Rensen, Nienke R Biermasz, Diana van Heemst

J Sleep Res. 2019; 28:e12776

ABSTRACT

Short and long sleep duration and poor sleep quality may affect serum and hepatic lipid content, but available evidence is inconsistent. Therefore, we aimed to investigate the associations between sleep duration and quality with serum and hepatic lipid content in a large population-based cohort of middle-aged individuals. The present cross-sectional study was embedded in the Netherlands Epidemiology of Obesity (NEO) study and consisted of 4,260 participants (mean age: 55 years, proportion men: 46%) not using lipid-lowering agents. Self-reported sleep duration and quality were assessed using the Pittsburgh Sleep Quality Index questionnaire (PSQI). Outcomes of this study were fasting lipid profile (total cholesterol, LDL-cholesterol, HDL-cholesterol, triglycerides), postprandial triglyceride (response) levels, and hepatic triglyceride content as measured with magnetic resonance spectroscopy. We performed multivariable linear regression analyses, adjusted for confounders, and additionally for measures that link to adiposity (e.g., BMI, sleep apnea). We observed that relative to the group with median sleep duration (≈7.0 hours of sleep), the group with shortest sleep (≈5.0 hours of sleep) had 1.5 fold higher hepatic triglyceride content (95% confidence interval (CI): 1.0;2.2). The group with PSQI score ≥10 had a 1.1 (95%CI: 1.0;1.2) fold higher serum triglyceride level compared with the group with PSQI ≤5. However, these associations disappeared after adjustment for BMI and sleep apnea. Therefore, we concluded that previously observed associations between shorter sleep duration and poorer sleep quality with an adverse lipid profile, may be explained by BMI and sleep apnea, rather than by a direct effect of sleep on the lipid profile.

INTRODUCTION

Cardiovascular disease (CVD) is one of the leading causes of death worldwide and is responsible for 17.7 million deaths in 2015[1]. Several lifestyle factors have been associated with a deleterious CVD risk profile and a higher risk of cardiovascular mortality, for example smoking and alcohol intake[2]. Multiple epidemiological studies, have shown that sleep duration is another important risk factor for the development of CVD[3-11]. However, in some studies both extremes (long and short sleep duration) are associated with an increased risk of CVD[3, 4], while in other studies only short sleep duration or long sleep duration has been associated with an increased risk of CVD and with CVD risk factors (notably obesity and metabolic syndrome)[5-13].

Besides sleep duration, poor sleep quality has also been associated with an increased risk of metabolic syndrome[14, 15], yet not in all studies[11]. Several factors could contribute to the discrepant findings regarding the associations between sleep duration and sleep quality with cardiovascular risk factors. Body mass index (BMI) and sleep apnea are associated with alterations in sleep and with high circulating lipids and incidence of coronary heart disease[16, 17]. In the study of Petrov *et al.* (2013) poor sleep quality was associated with a poor lipid profile, however, after adjustment for covariates including BMI and obstructive sleep apnea (OSA) risk, this association disappeared[11]. Importantly, previous studies generally adjusted for BMI, however, most studies did not adjust for sleep apnea[5-13]. Therefore, the question remains to what extent the previously described associations between sleep duration and sleep quality with CVD were confounded by BMI and also OSA. Moreover, both BMI and OSA are risk factors for non-alcoholic liver disease (NAFLD)[18-20], which is a risk factor for myocardial dysfunction[21]. To the best of our knowledge hepatic triglyceride content (HTGC) has not been studied in relation to sleep duration and/or sleep quality.

Based on previous studies, we hypothesized that both short and long sleep duration and poor sleep quality are associated with an adverse serum and hepatic lipid profile. However, we hypothesize that after adjustment for BMI and the risk of sleep apnea, these associations may decrease. In the present study we aim to assess these associations in a large population-based cohort of middle-aged adults from the Netherlands, considering all important confounding factors (including BMI and sleep apnea).

METHODS

Study design and study population

The present study is a cross-sectional analysis of baseline measurements of the Netherlands Epidemiology of Obesity (NEO) study, a cohort of 6,671 individuals with an oversampling of individuals with overweight or obesity. Between September 2008 and September 2012, men and women aged between 45 and 65 years with a self-reported body mass index (BMI) of 27 kg/m^2 or higher living in the greater area of Leiden were invited to participate in the NEO study. In addition, all inhabitants aged between 45 and 65 years from one municipality (Leiderdorp) were invited irrespective of their BMI, allowing for a reference distribution of BMI. Baseline data were collected at the NEO study center of the Leiden University Medical Center (LUMC). Prior to the NEO study visit, participants completed a questionnaire about demographic and clinical information and fasted for at least 10 hours. Participants came to the research site in the morning to undergo several baseline measurements including anthropometric measurements and fasting and postprandial blood sampling. At the study site, a screening form was completed by all participants asking about anything that might create a health risk or interfere with MRI imaging (most notably metallic devices, claustrophobia, and a body circumference of more than 1.70 m). Of the participants who were eligible for MRI, approximately 35% of the total study population were randomly selected to undergo direct assessment of VAT. A medication inventory was performed to collect data on medication use during the month preceding the visit to the study center. More detailed information on the study design and data collection was described elsewhere[22]. This study was approved by the medical ethics committee of the Leiden University Medical Center (LUMC) (and the NEO board) and all participants gave written informed consent. As demonstrated in **Figure 1**, we excluded participants with missing data on the PSQI questionnaire (N=1,402). The PSQI questionnaire was only added to the baseline questionnaire after July 2009 and therefore participants entering the study before this date have missing data on this questionnaire. Moreover, we excluded participants who used lipid lowering drugs (N=791), had missing baseline characteristics (N=118), missing data of the Berlin questionnaire (N=53), were not in a fasting state during the hospital visit (N=19), or missed data on serum triglycerides (N=27) or cholesterol (N=1). We additionally excluded participants who drank >40 g alcohol per day (N=344) from the analyses on hepatic triglyceride content (HTGC). For the analyses on postprandial triglyceride levels, we excluded participants with missing or incomplete postprandial serum triglyceride concentrations (N=221) or who had no or incomplete liquid meal intake (N=2).

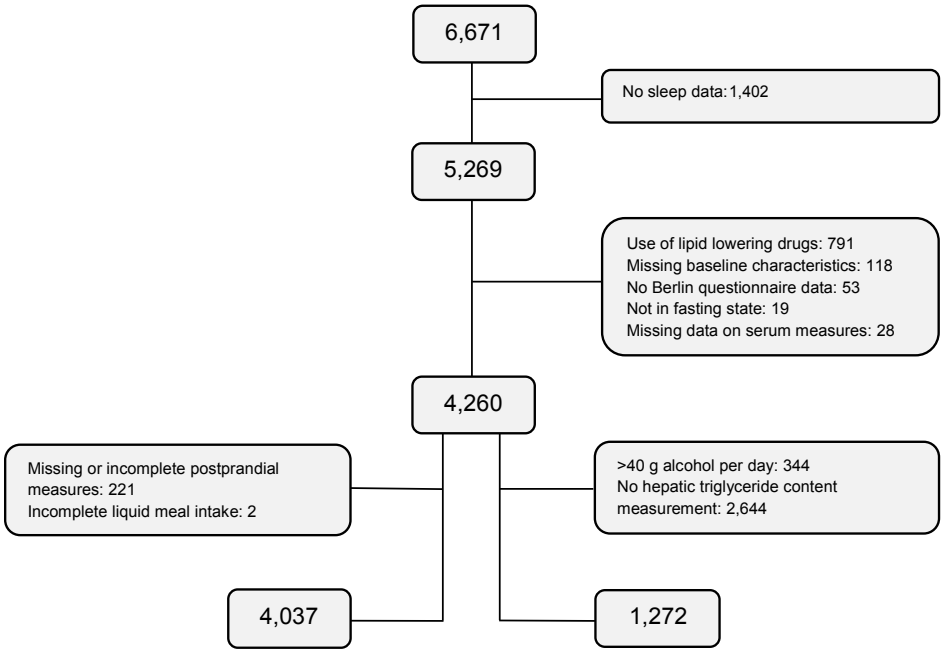


Figure 1. *Flowchart of participant inclusion*

Sleep characteristics

To assess habitual sleep duration and quality, we used the Pittsburgh sleep quality index (PSQI)[23], which is a self-rated questionnaire to retrospectively measure sleep parameters over a one month time interval. Total sleep duration was derived from the question "On an average day, how much sleep do you get?". To obtain a classification of short and long total sleep duration, we calculated the age- and sex- adjusted residuals with linear regression analysis for total sleep duration with age and sex and determined subgroups on the basis of these residuals. We used the 5th lowest percentile of the age- and sex-adjusted residuals to define shortest sleep, the 5th till 20th percentile to define short sleep, the 20th till 80th to define medium sleep, the 80th till 95th to define long sleep and the 95th till 100th percentile to define longest sleep. Sleep quality was assessed using the total score of the PSQI questionnaire. The questionnaire consists of seven components of which an overall score can be calculated. The global score ranges from 0 to 21, in which a higher score indicates a poorer sleep quality[23]. For sleep quality, we formed three sleep quality groups, in which we used the good sleep quality group (PSQI total score ≤5) as a reference group in linear regression analyses. The poor sleep quality group was defined by a PSQI total score between 5 and 10, and worst sleep quality as PSQI total score ≥10.

Serum lipid profile and hepatic triglyceride content

After an overnight fast of at least 10 h, fasting blood samples were taken at the study center. Within 5 min after the first blood sample was taken, participants drank a liquid mixed meal (400 mL) with an energy content of 600 kcal, with 16% of energy derived (En%) from protein, 50 En% from carbohydrates and 34 En% from fat. Postprandial blood samples were taken 30 and 150 minutes after ingestion of the meal. Serum triglyceride concentrations were determined at the 3 time points. Serum total cholesterol and triglyceride concentrations were determined by enzymatic colorimetric methods (Roche Modular Analytics P800, Roche Diagnostics, Mannheim, Germany; CV < 5%) and HDL-cholesterol with homogenous HDLc method, 3rd generation (Roche Modular Analytics P800, Roche Diagnostics, Mannheim, Germany; CV < 5%). Low-density lipoprotein (LDL) cholesterol concentration was estimated using Friedewald's formula[24]. All measures were performed in the central clinical chemistry laboratory of the Leiden University Medical Center. The area under the curve (AUC) for postprandial serum triglyceride levels was calculated using the Trapezoid Rule as (15 * fasting concentration + 75 * $concentration_{30min}$ + 60 * $concentration_{150min}$) / 150[25]. Hepatic ^{1}H magnetic resonance (MR) spectra were obtained in a random subset of 1,207 participants with data on habitual sleep. In short, an 8-mL voxel was positioned in the right lobe of the liver. A point-resolved spectroscopy sequence was used to acquire spectroscopic data during continuous breathing with automated shimming. Spectra were obtained with and without water suppression. Spectral data were fitted by using Java-based MR user interface software (jMRUI, version 3.0; developed by A. van den Boogaart, Katholieke Universiteit Leuven, Leuven, Belgium)[26]. Mean line widths of the spectra were calculated. The resonances that were fitted and used for calculation of the triglycerides were methylene (peak at 1.3 ppm, $[CH_2]_n$) and methyl (peak at 0.9 ppm, CH_3). The HTGC relative to water was calculated with the following formula: (signal amplitude of methylene + methyl)/(signal amplitude of water) × 100.

Covariates

A semi-quantitative food frequency questionnaire (FFQ)[27] questionnaire was used to assess energy intake. Energy intake was estimated from the FFQ with the 2011 version of the Dutch food composition table (NEVO-2011). Participants reported the frequency and duration of their physical activity in leisure time using the Short Questionnaire to Assess Health-enhancing physical activity (SQUASH)[28], which was expressed in hours per week of metabolic equivalents (MET-h/week). Body weight was measured at the study center without shoes and one kilogram (kg) was subtracted to correct for the weight of clothing. BMI was calculated by dividing the weight in kilograms by the height in meters squared. The risk for the presence of obstructive sleep apnea syndrome was assessed using the Berlin questionnaire[29]. This questionnaire consists of 10 questions

that form three categories (snoring (category 1), daytime somnolence (category 2) and hypertension and BMI (category 3)) related to the likelihood of the presence of sleep apnea. Individuals can be classified as either having a high (2 or more categories with a positive score) or low likelihood of sleep apnea (only 1 or no categories with a positive score) .

Statistical Analysis

Because individuals with a BMI of 27 kg/m^2 or higher were oversampled in the NEO study population, adjustments were made to correctly represent associations in the general population[30-32]. This was done by weighting individuals towards to the BMI distribution of participants from the Leiderdorp municipality, whose BMI distribution was similar to the BMI distribution of the general Dutch population[22]. Consequently, all presented results are based on weighted analyses and apply to a population-based study without oversampling of participants with a BMI of 27 kg/m^2 or higher. Characteristics of the study population were expressed as mean (with standard deviation, SD) for normally distributed measures, median with inter-quartile ranges for non-normally distributed measures, and proportions for categorical variables. We performed all statistical analyses using Stata version 12.1 (Stata, College Station, Texas, USA) software.

Not normally distributed outcomes were log transformed to approximate a normal distribution (notably serum triglycerides, HTGC and AUC of serum triglycerides). However, in order to present the results with a similar interpretation, we log transformed normally distributed outcomes (notably serum HDL-cholesterol, LDL-cholesterol, total cholesterol) as well. We performed linear regression analyses using the medium sleep category (characterized by 20th till 80th percentile of sleep duration residuals) as reference group. The subsequent beta regression coefficients were back-transformed and expressed as a ratio with accompanying 95% confidence interval (95% CI), which can be interpreted as the relative change in outcome compared to the reference group. The initial model for linear regression analyses was adjusted for age and sex (Model 1). In addition to age and sex, we adjusted in Model 2 for ethnicity (white/other), education level (high/other), smoking (never/former/current), alcohol consumption, energy intake, physical activity and sleep medication (yes/no). In Model 3 we additionally adjusted for BMI and sleep apnea. In the analyses for sleep quality we did not adjust for sleep medication in Models 2 and 3, as this is a component of the PSQI total score.

RESULTS

Characteristics of the study population

In total, after exclusion of non-eligible participants, this study comprised 4,260 participants with a mean age of 55 (SD 6.0) years, of whom 46% were men. As compared with the medium sleep group (40%), there were more men in both the shortest (45%) and the longest (51%) sleep group (**Table 1**). Less individuals had higher education in both the shortest sleep group (39%) and the longest sleep group (36%) than in the medium sleep group (51%). More participants used sleep medication in the shortest (14%) and longest sleep group (7%) as compared with the medium sleep group (4%). HTGC was higher both in the shortest sleep group (6% [2.5;10.5]) and in the longest sleep group (4% [2.0;7.4]) than the medium sleep group (2%, [1.2;5.1]). All other studied characteristics were similar between the groups.

Associations between sleep duration and fasted and postprandial lipids

In the analyses adjusted for age and sex (Model 1), shortest sleep duration was associated with a 1.52 (95%CI: 1.04-2.24) fold higher HTGC as compared with the medium sleep group (**Figure 2 and Online Supplementary Table 1**). This association persisted after adjustment for potential confounding factors (Model 2). However, the association between shortest sleep and higher HTGC disappeared after we additionally adjusted for BMI and sleep apnea in Model 3 (ratio of 1.00 (95%CI: 0.68-1.45). There were no associations between short and long sleep duration and total cholesterol, LDL-cholesterol, HDL-cholesterol, triglycerides or AUC of triglycerides.

Associations between sleep quality and fasting and postprandial lipids

A poor sleep quality (PSQI total score 5-10; **Figure 3 and Online Supplementary Table 2**) was associated with 1.07 (95%CI: 1.01;1.13) fold increased serum triglyceride level in the analyses (age and sex adjusted) as compared with good sleep quality (PSQI score ≤5). Adjustment for potential confounding factors (Model 2) did not materially change the results (ratio 1.06 (95%CI: 1.00;1.11)), but when we additionally adjusted for BMI and sleep apnea (Model 3), the association between poor sleep quality and serum triglycerides disappeared (ratio 1.04 (95%CI: 0.99;1.09). Poor sleep quality was associated with a 1.24 (95% CI: 1.04;1.49) fold increased HTGC as compared with good sleep quality in Model 1, which persisted in Model 2 (ratio 1.21 (95%CI: 1.01;1.45)). However, after additional adjustment for BMI and sleep apnea (Model 3) the association disappeared (ratio 1.08 (95%CI: 0.91;1.27)). Worst sleep quality (PSQI ≥ 10) was associated with a 1.10 (95% CI: 1.02;1.18) fold increased fasting serum triglyceride level as compared with good sleep quality in Model 1.

This association persisted in Model 2 (ratio 1.08 (95%CI: 1.00;1.16), but the association disappeared in Model 3 (ratio 1.01 (95%CI: 0.99;1.09). There were no associations between sleep quality and total cholesterol, LDL-cholesterol, HDL-cholesterol or AUC of triglycerides.

Table 1. *Characteristics of participants in the Netherlands Epidemiology of Obesity study, stratified by sleep duration (N=4,260)*

Sleep duration	Shortest	Short	Medium	Long	Longest
	0-5%	5-20%	20-80%	80-95%	95-100%
Age (years)	57 (5)	57 (5)	55 (6)	54 (6)	57 (6)
Sex (% men)	45	47	40	41	51
BMI (kg/m²)	27 (5)	26 (5)	26 (4)	26 (4)	26 (5)
Ethnicity (% white)	90	93	96	96	93
Education (% high)	39	43	51	48	36
Smoking (% current)	19	14	16	17	19
Sleep medication (%)	14	9	4	5	7
Alcohol consumption (g/day)	12 (3;22)	10 (3;22)	10 (3;21)	9 (2;21)	9 (0;21)
Physical activity (MET-h/week)	25 (12;44)	30 (16;47)	30 (17;50)	32 (16;52)	30 (15;51)
Sleep duration (h/day)	5 (4;5)	6 (6;6)	7 (7;8)	8 (8;8)	9 (9;9)
PSQI (total score)	11 (9;13)	7 (5;10)	4 (3;6)	3 (2;4)	3 (2;5)
Sleep apnea (%)	33	26	16	16	22
Fasting total cholesterol (mmol/L)	6 (1)	6 (1)	6 (1)	6 (1)	6 (1)
Fasting LDL-cholesterol (mmol/L)	4 (1)	4 (1)	4 (1)	4 (1)	4 (1)
Fasting HDL-cholesterol (mmol/L)	2 (1)	2 (1)	2 (1)	2 (0)	2 (0)
Fasting triglycerides (mmol/L)	1 (1;2)	1 (1;2)	1 (1;1)	1 (1;2)	1 (1;1)
Triglycerides 30 min (mmol/L)	1 (1;2)	1 (1;2)	1 (1;2)	1 (1;2)	1 (1;2)
Triglycerides 120 min (mmol/L)	2 (1;3)	2 (1;2)	2 (1;2)	2 (1;2)	2 (1;2)
AUC Triglycerides±	48 (28;77)	48 (26;73)	46 (26;66)	46 (27;68)	44 (26;68)
Hepatic triglyceride content (%)*	6 (3;11)	3 (2;6)	2 (1;5)	2 (1;6)	4 (2;8)

Abbreviations: AUC, area under the curve; BMI, body mass index; HDL, high-density lipoprotein; kJ, kilojoule; LDL, low-density lipoprotein; MET, metabolic equivalents of task; NEO, Netherlands Epidemiology of Obesity; PSQI, Pittsburgh Sleep Questionnaire Index. Results were based on analyses weighted towards the BMI distribution of the general Dutch population. Data presented as mean ± standard deviation (SD); proportion (%); median (25^{th}-75^{th} percentile). ±, N=4,037; *, N=1,272.

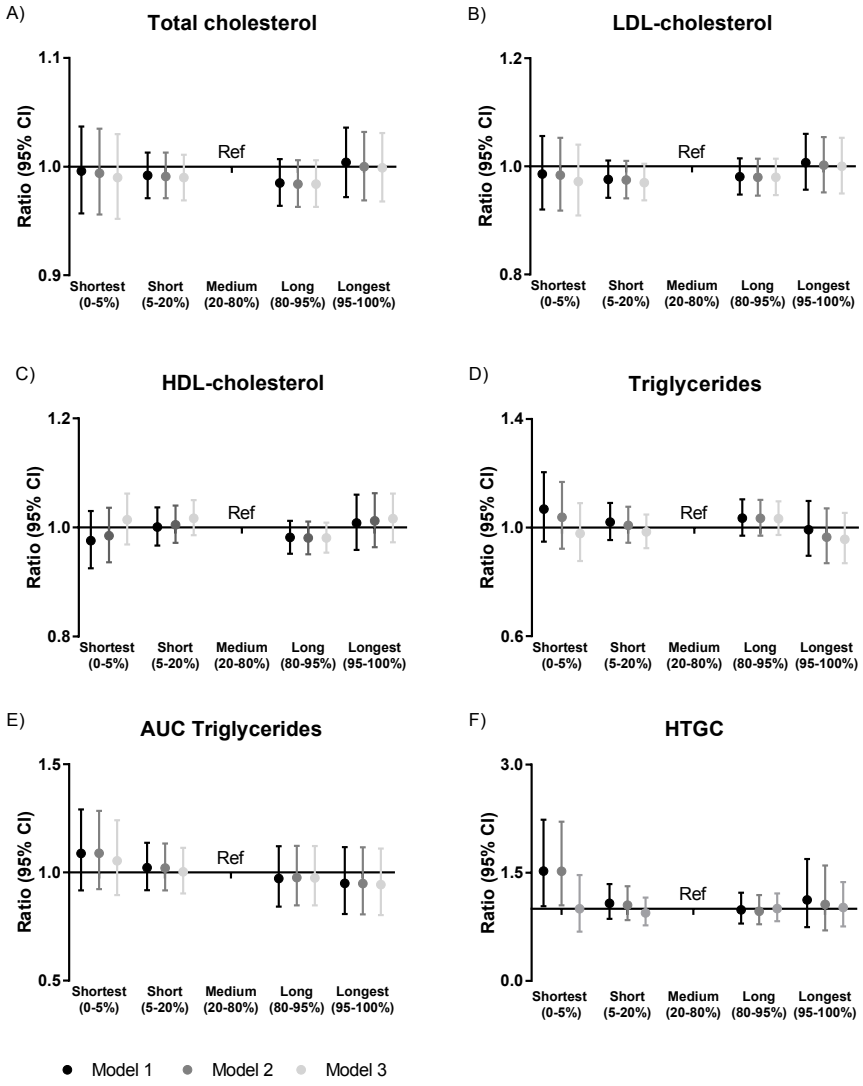


Figure 2. *Associations between sleep duration and A) TC, B) LDL-cholesterol, C) HDL-cholesterol, D) TG, E) AUC of TG and F) hepatic triglyceride content (HTGC).* The medium sleep duration group is used as reference category in linear regression analyses. Results are presented as ratios with accompanying 95% confidence intervals, linear regression coefficients of the log transformed outcomes were back transformed in order to present ratios. The ratio reflects the relative change to provide an indication of the fold change of the outcome as compared to the reference category. Results were based on analyses weighted towards the BMI distribution of the general Dutch population. Model 1: adjusted for age and sex; Model 2: adjusted for age, sex, ethnicity, education level, smoking, alcohol intake, caloric intake and physical activity; Model 3: adjusted for Model 2 + sleep apnea and BMI. Abbreviations: AUC, area under the curve; CI, confidence interval; HDL, high-density lipoprotein; HTGC, hepatic triglyceride content; LDL, low-density lipoprotein; Ref, reference category.

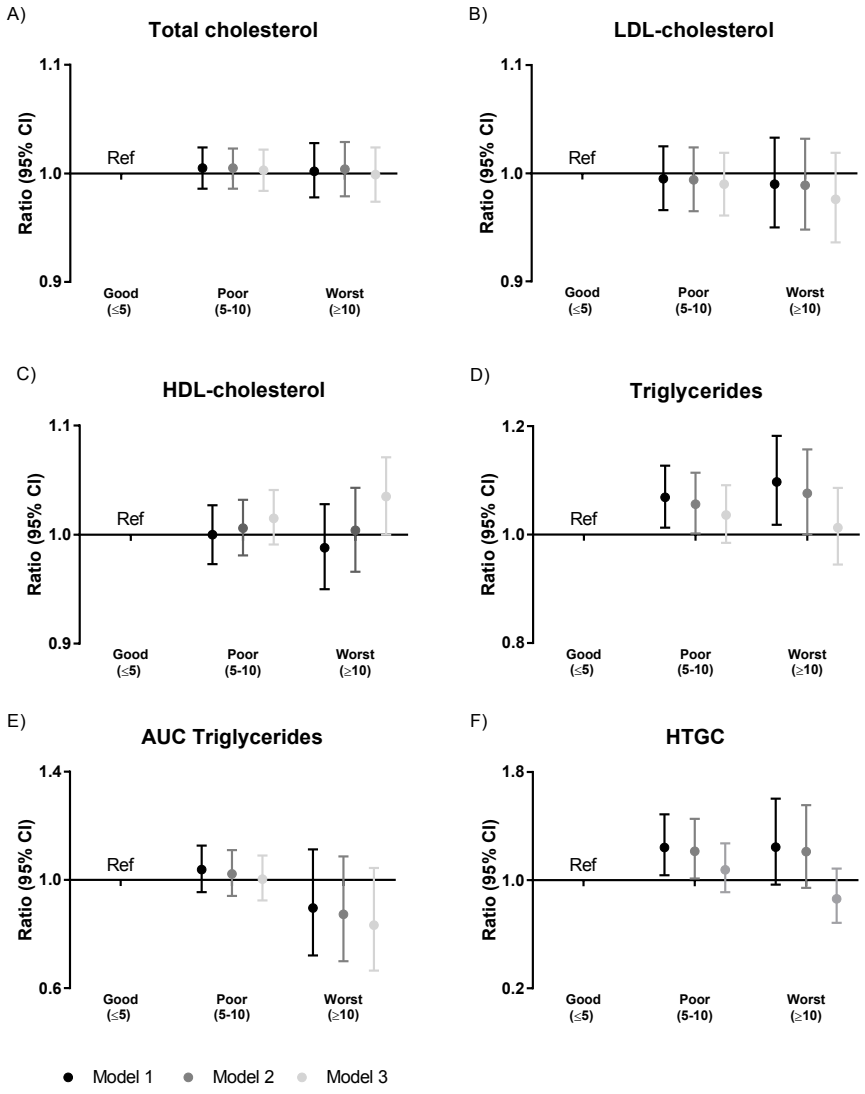


Figure 3. *Associations between sleep quality and A) TC, B) LDL-cholesterol, C) HDL-cholesterol, D) TG, E) AUC of TG and F) HTGC.* The good sleep quality group is used as reference category in linear regression analyses. Results are presented as ratios with accompanying 95% confidence intervals, linear regression coefficients of the log transformed outcomes were back transformed in order to present ratios. The ratio reflects the relative change to provide an indication of the fold change of the outcome as compared to the reference category. Results were based on analyses weighted towards the BMI distribution of the general Dutch population. Model 1: adjusted for age and sex; Model 2: adjusted for age, sex, ethnicity, education level, smoking, alcohol intake, caloric intake and physical activity; Model 3: adjusted for Model 2 + sleep apnea and BMI. AUC, area under the curve; CI, confidence interval; HDL, high-density lipoprotein; HTGC, hepatic triglyceride content; LDL, low-density lipoprotein; Ref, reference category.

DISCUSSION

The present study aimed to address the associations between sleep duration and quality with hepatic triglyceride content (HTGC) and serum lipid levels in a cohort of 4,260 middle-aged individuals. When analyses were adjusted for age and sex, we observed an association between shortest sleep duration with higher HTGC. Moreover, poor sleep quality was associated with higher fasting serum triglyceride levels and higher HTGC than good sleep quality. However, all observed associations disappeared after additional adjustment for BMI and sleep apnea.

In previous research it has been shown that both short and long sleep duration, or only short or long sleep duration were associated with cardiovascular risk and cardiovascular risk factors[3-11]. We hypothesized that these discrepant findings could be explained by differences in the considered confounding factors, which includes the adjustment for the confounding factors BMI and sleep apnea[16, 17]. In agreement, in our study, when adjusted for age and sex, short sleep duration was associated with higher serum triglyceride levels. When we additionally adjusted for ethnicity, education level, smoking, alcohol intake, caloric intake and physical activity, this association remained.. However, this association disappeared after additional adjustment for BMI and sleep apnea. Similar, we observed an association between poor sleep quality and higher serum triglyceride levels and HTGC after adjustment for classical confounders, however, again these associations disappeared after additional adjustment for BMI and sleep apnea. In agreement, in a cohort study comprising 503 adults, it has been observed that there was no association between poor sleep quality and lipid profile after adjustment for covariates including BMI and sleep apnea risk[11].

In contrast to our findings, a study of Anujuo *et al.*[33] did not observe an association between short sleep duration and higher triglyceride levels in neither of their analyses (adjusted for only age and sex, or other confounding factors including BMI) in a population consisting of 2,146 participants of Dutch origin. One of the explanations for these discrepant findings could be the different cut-off point used for determining short sleep duration, which was <7 hours based on the small number of individuals with very short sleep duration. Therefore, the group defined as "short sleep" might not be sleeping sufficiently short to observe clinical relevant associations with lipid levels.

Questions regarding the direction of the observed associations remain to be elucidated. In this study we considered BMI and sleep apnea as potential confounding factors, suggesting that BMI and sleep apnea are a common cause of alterations in sleep and in lipid metabolism. Alternatively, sleep duration might have a causal effect on BMI and

sleep apnea meaning that BMI and sleep apnea may mediate the association between sleep and cardiovascular risk factors. In this case the observed associations between sleep duration/quality and serum and hepatic lipid profile in the present study may be underestimated.

There are several biological pathways that could link sleep duration and quality to CVD. For example, a short sleep duration might have an adverse effect on cardiovascular health via increased cortisol levels and/or inflammatory mediators through higher sympathetic nervous system activity[34]. In addition, a genome-wide association study on total sleep duration suggests that there is also a shared genetic component between insomnia symptoms and a higher BMI, waist circumference and insulin resistance, which have all been associated with a higher risk of developing CVD[35]. Moreover, a shorter sleep duration has been associated with altered levels of leptin and ghrelin and a higher craving for carbohydrate-rich foods[36, 37]. These are possible mechanisms via which alteration in sleep could lead to obesity and obesity-related disorders. Both obesity and sleep apnea are shown to be associated with high circulating lipids and a higher incidence of coronary heart disease[16, 17]. Moreover, obesity is a risk factor for non-alcoholic liver disease (NAFLD)[18]. It was shown that a higher HTGC was associated with a higher risk of myocardial dysfunction, as characterized by a lower diastolic function in the NEO study population, however, only in obese individuals[21]. Also, it was shown that obstructive sleep apnea (OSA) is prevalent in 60% of NAFLD patients, and that OSA may contribute to progression of NAFLD[19, 20]. OSA is thought to exert an effect via different mechanisms (e.g. inflammation, oxidative stress and intermittent hypoxia)[17] and can be treated with continuous positive airway pressure (CPAP)[38]. This supports our hypothesis that sleep apnea and BMI affect the relation between short sleep duration and poor sleep quality with CVD. Our findings support the idea of weight loss and sleep apnea screening in individuals with short sleep duration and poor sleep quality in order to contribute to prevention of cardiovascular disease onset. Nevertheless, the individual contributions of sleep, sleep quality, circadian rhythm, and other lifestyle adaptations and their interrelations are complex and difficult to assess separately and at least require prospective analyses in large populations.

A strength of this study is the use of residuals as a determinant for sleep duration. By using the residuals we obtained a classification of sleep duration which is independent of age and sex. One of the other strengths of this study is the extensive phenotyping of the NEO study which enables us to correct for a broad range of possible confounding factors (e.g. ethnicity, physical activity and risk of sleep apnea). Moreover, previous studies used different questionnaires to assess sleep duration and sleep quality, which may result in inconclusive results. Therefore, we have used the PSQI questionnaire, which is a widely

used and validated too to assess sleep disturbances [39]. However, the present study also has some limitations. First, inherent to the observational cross-sectional design, whereby we are not able to exclude reverse causation or residual confounding in this study. Second, we used subjective sleep questionnaires to assess sleep duration and sleep quality and the risk of sleep apnea. Self-reported data is subject to recall bias, whereby participants may either under- or over-report their sleep duration and their quality of sleep. There may be a measurement error which results in non-differential misclassification for the exposure. However, the Pittsburgh sleep quality index was shown to be a reliable and validated tool to assess sleep dysfunction [39]. Moreover, although the risk of sleep apnea is self-reported using the Berlin questionnaire, this tool has also been validated in several populations and showed the highest specificity to detect mild and severe OSA in patients from sleep clinics, as compared to other OSA screening questionnaires[40].

In conclusion, we observed that shorter sleep duration and poorer sleep quality were associated with an adverse lipid profile. However, all observed associations disappeared after additional adjustment for BMI and sleep apnea, indicating that BMI and risk of sleep apnea, likely confound previously observed associations and should therefore be considered in future studies.

Acknowledgements and funding

We express our gratitude to all individuals who participate in the Netherlands Epidemiology of Obesity study, in addition to all participating general practitioners. We furthermore thank P.R. van Beelen and all research nurses for collecting the data, P.J. Noordijk and her team for sample handling and storage, and I. de Jonge for data management of the NEO study. The NEO study is supported by the participating Departments, the Division and the Board of Directors of the Leiden University Medical Centre, and by the Leiden University, Research Profile Area 'Vascular and Regenerative Medicine'. Nutricia Research, Utrecht, The Netherlands, provided the mixed meal. We acknowledge the support from the Netherlands Cardiovascular Research Initiative: an initiative with support of the Dutch Heart Foundation (CVON2014-02 ENERGISE). DvH was supported by the European Commission funded project HUMAN (Health-2013-INNOVATION-1-602757). NB was supported by the Netherlands Organization for Scientific Research (NWO-VENI 016.136.125).

Conflict of interest

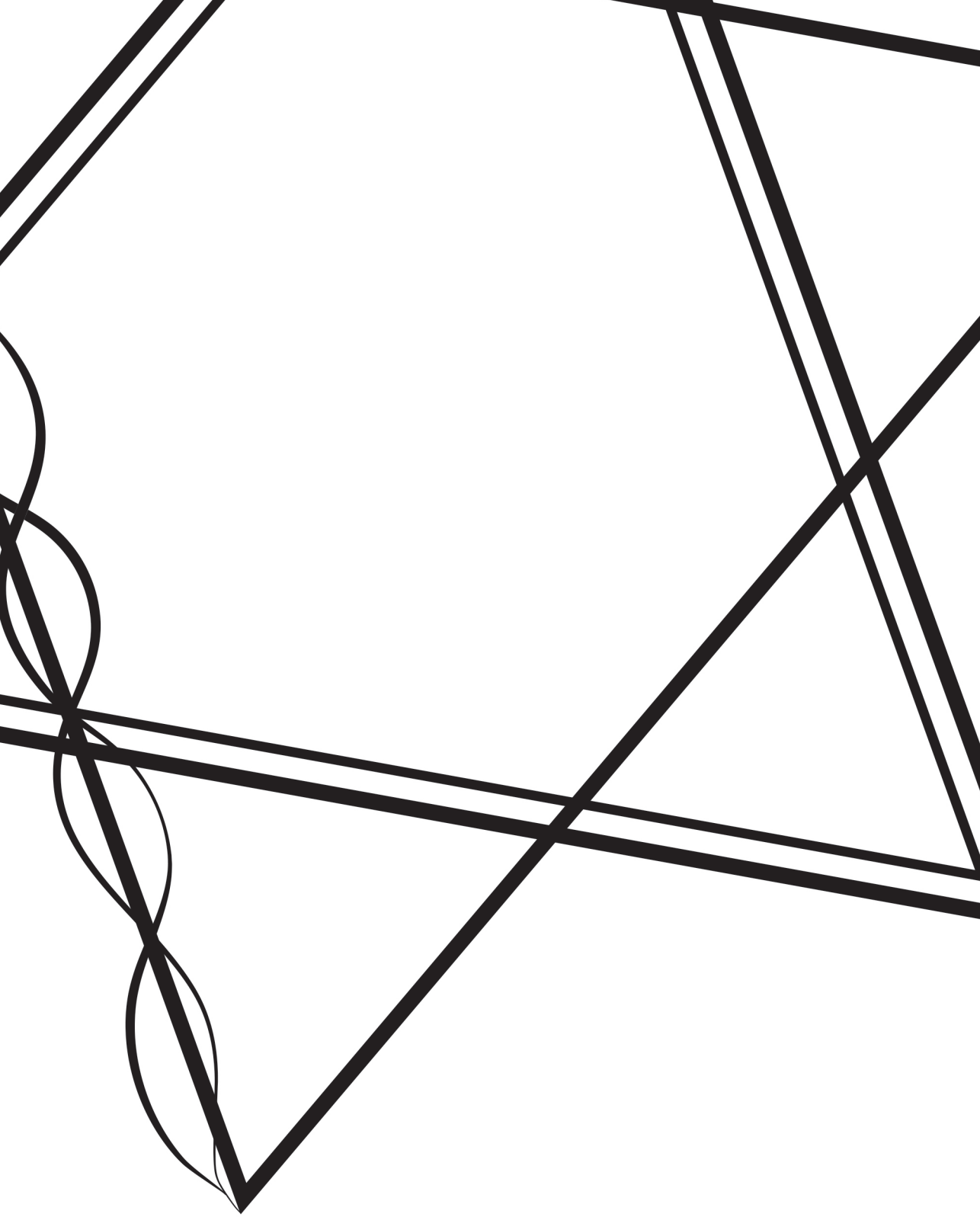
All authors declare to have no conflict of interest.

REFERENCES

1. Fact sheet Cardiovascular diseases World Health Organization2017 [updated May 2017. Available from: http://www.who.int/mediacentre/factsheets/fs317/en/.
2. Jha P, Ramasundarahettige C, Landsman V, Rostron B, Thun M, Anderson RN, et al. 21st-century hazards of smoking and benefits of cessation in the United States. N Engl J Med. 2013;368(4):341-50.
3. Ford ES. Habitual sleep duration and predicted 10-year cardiovascular risk using the pooled cohort risk equations among US adults. J Am Heart Assoc. 2014;3(6):e001454.
4. Aggarwal S, Loomba RS, Arora RR, Molnar J. Associations between sleep duration and prevalence of cardiovascular events. Clin Cardiol. 2013;36(11):671-6.
5. Wu Y, Zhai L, Zhang D. Sleep duration and obesity among adults: a meta-analysis of prospective studies. Sleep Med. 2014;15(12):1456-62.
6. Xi B, He D, Zhang M, Xue J, Zhou D. Short sleep duration predicts risk of metabolic syndrome: a systematic review and meta-analysis. Sleep Med Rev. 2014;18(4):293-7.
7. Cappuccio FP, Taggart FM, Kandala NB, Currie A, Peile E, Stranges S, et al. Meta-analysis of short sleep duration and obesity in children and adults. Sleep. 2008;31(5):619-26.
8. Cappuccio FP, Cooper D, D'Elia L, Strazzullo P, Miller MA. Sleep duration predicts cardiovascular outcomes: a systematic review and meta-analysis of prospective studies. Eur Heart J. 2011;32(12):1484-92.
9. Lee JA, Park HS. Relation between sleep duration, overweight, and metabolic syndrome in Korean adolescents. Nutr Metab Cardiovas. 2014;24(1):65-71.
10. van den Berg JF, Miedema HM, Tulen JH, Neven AK, Hofman A, Witteman JC, et al. Long sleep duration is associated with serum cholesterol in the elderly: the Rotterdam Study. Psychosom Med. 2008;70(9):1005-11.
11. Petrov ME, Kim Y, Lauderdale D, Lewis CE, Reis JP, Carnethon MR, et al. Longitudinal associations between objective sleep and lipids: the CARDIA study. Sleep. 2013;36(11):1587-95.
12. Zhan Y, Chen R, Yu J. Sleep duration and abnormal serum lipids: the China Health and Nutrition Survey. Sleep Med. 2014;15(7):833-9.
13. Shin HY, Kang G, Kim SW, Kim JM, Yoon JS, Shin IS. Associations between sleep duration and abnormal serum lipid levels: data from the Korean National Health and Nutrition Examination Survey (KNHANES). Sleep Med. 2016;24:119-23.
14. Koren D, Dumin M, Gozal D. Role of sleep quality in the metabolic syndrome. Diabetes Metab Syndr Obes. 2016;9:281-310.
15. Mesas AE, Guallar-Castillon P, Lopez-Garcia E, Leon-Munoz LM, Graciani A, Banegas JR, et al. Sleep quality and the metabolic syndrome: the role of sleep duration and lifestyle. Diabetes Metab Res Rev. 2014;30(3):222-31.
16. Dale CE, Fatemifar G, Palmer TM, White J, Prieto-Merino D, Zabaneh D, et al. Causal Associations of Adiposity and Body Fat Distribution With Coronary Heart Disease, Stroke Subtypes, and Type 2 Diabetes Mellitus: A Mendelian Randomization Analysis. Circulation. 2017;135(24):2373-88.
17. Javaheri S, Barbe F, Campos-Rodriguez F, Dempsey JA, Khayat R, Javaheri S, et al. Sleep Apnea: Types, Mechanisms, and Clinical Cardiovascular Consequences. J Am Coll Cardiol. 2017;69(7):841-58.
18. Bray GA. Medical consequences of obesity. J Clin Endocrinol Metab. 2004;89(6):2583-9.

19. Sundaram SS, Sokol RJ, Capocelli KE, Pan Z, Sullivan JS, Robbins K, et al. Obstructive sleep apnea and hypoxemia are associated with advanced liver histology in pediatric nonalcoholic fatty liver disease. J Pediatr. 2014;164(4):699-706 e1.
20. Mirrakhimov AE, Polotsky VY. Obstructive sleep apnea and non-alcoholic Fatty liver disease: is the liver another target? Front Neurol. 2012;3:149.
21. Widya RL, de Mutsert R, den Heijer M, le Cessie S, Rosendaal FR, Jukema JW, et al. Association between Hepatic Triglyceride Content and Left Ventricular Diastolic Function in a Population-based Cohort: The Netherlands Epidemiology of Obesity Study. Radiology. 2016;279(2):443-50.
22. de Mutsert R, den Heijer M, Rabelink TJ, Smit JW, Romijn JA, Jukema JW, et al. The Netherlands Epidemiology of Obesity (NEO) study: study design and data collection. Eur J Epidemiol. 2013;28(6):513-23.
23. Buysse DJ, Reynolds CF, 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res. 1989;28(2):193-213.
24. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin Chem. 1972;18(6):499-502.
25. Retnakaran R, Shen S, Hanley AJ, Vuksan V, Hamilton JK, Zinman B. Hyperbolic relationship between insulin secretion and sensitivity on oral glucose tolerance test. Obesity (Silver Spring). 2008;16(8):1901-7.
26. Naressi A, Couturier C, Devos JM, Janssen M, Mangeat C, de Beer R, et al. Java-based graphical user interface for the MRUI quantitation package. MAGMA. 2001;12(2-3):141-52.
27. Feunekes GI, Van Staveren WA, De Vries JH, Burema J, Hautvast JG. Relative and biomarker-based validity of a food-frequency questionnaire estimating intake of fats and cholesterol. The American journal of clinical nutrition. 1993;58(4):489-96.
28. Wendel-Vos GC, Schuit AJ, Saris WH, Kromhout D. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. Journal of clinical epidemiology. 2003;56(12):1163-9.
29. Netzer NC, Stoohs RA, Netzer CM, Clark K, Strohl KP. Using the Berlin Questionnaire to identify patients at risk for the sleep apnea syndrome. Annals of internal medicine. 1999;131(7):485-91.
30. Korn EL, Graubard BI. Epidemiologic studies utilizing surveys: accounting for the sampling design. Am J Public Health. 1991;81(9):1166-73.
31. Lumley T. Analysis of complex survey samples [Available from: http://www.jstatsoft.org/v09/i08/paper.
32. VWS Mv. Hoveel mensen hebben overgewicht? [Available from: http://www.rivm.nl/Onderwerpen/N/Nederland_de_Maat_Genomen.
33. Anujuo K, Stronks K, Snijder MB, Jean-Louis G, Rutters F, van den Born BJ, et al. Relationship between short sleep duration and cardiovascular risk factors in a multi-ethnic cohort - the helius study. Sleep Med. 2015;16(12):1482-8.
34. Lucassen EA, Rother KI, Cizza G. Interacting epidemics? Sleep curtailment, insulin resistance, and obesity. Annals of the New York Academy of Sciences. 2012;1264:110-34.
35. Lane JM, Liang J, Vlasac I, Anderson SG, Bechtold DA, Bowden J, et al. Genome-wide association analyses of sleep disturbance traits identify new loci and highlight shared genetics with neuropsychiatric and metabolic traits. Nature genetics. 2017;49(2):274-81.
36. Taheri S, Lin L, Austin D, Young T, Mignot E. Short sleep duration is associated with reduced leptin, elevated ghrelin, and increased body mass index. PLoS Med. 2004;1(3):e62.

37. Spiegel K, Tasali E, Penev P, Van Cauter E. Brief communication: Sleep curtailment in healthy young men is associated with decreased leptin levels, elevated ghrelin levels, and increased hunger and appetite. Annals of internal medicine. 2004;141(11):846-50.
38. Gottlieb DJ, Punjabi NM, Mehra R, Patel SR, Quan SF, Babineau DC, et al. CPAP versus oxygen in obstructive sleep apnea. N Engl J Med. 2014;370(24):2276-85.
39. Mollayeva T, Thurairajah P, Burton K, Mollayeva S, Shapiro CM, Colantonio A. The Pittsburgh sleep quality index as a screening tool for sleep dysfunction in clinical and non-clinical samples: A systematic review and meta-analysis. Sleep Med Rev. 2016;25:52-73.
40. Amra B, Rahmati B, Soltaninejad F, Feizi A. Screening Questionnaires for Obstructive Sleep Apnea: An Updated Systematic Review. Oman Med J. 2018;33(3):184-92.



CHAPTER 3.2

The association between habitual sleep duration and sleep quality with glycemic traits: assessment by cross-sectional and Mendelian Randomization analyses

Maxime M Bos, Diana van Heemst, Esther Donga, Renée de Mutsert, Frits R Rosendaal, Gerard Jan Blauw, Patrick C N Rensen, Nienke R Biermasz, Raymond Noordam

J Clin Med 2019;8(5):682

ABSTRACT

Evidence whether habitual sleep duration and sleep quality are associated with insulin resistance is inconsistent. Therefore, we aimed to investigate the associations between different measures of habitual sleep with glycemic traits by cross-sectional as well as Mendelian Randomization analyses. Using a cross-sectional study design in 4,672 middle-aged (45-65 years; 48% men) non-diabetic participants of the Netherlands Epidemiology of Obesity (NEO) study, we assessed the associations of sleep duration and sleep quality with glycemic traits using multivariable linear regression models adjusted for demographics, anthropometrics and lifestyle factors. Genetic variants for total, short and long sleep duration were used as instrumental variables in two-sample Mendelian Randomization analyses using summary-level data of genome wide association studies on glycemic traits in individuals without diabetes in a cohort of European ancestry (MAGIC; n=58,074). Compared with medium sleepers (median 7.0 hours of sleep per night), the shortest sleepers (median 5.0 hours of sleep per night) had 14.5% (95% confidence interval (CI): 2.0-28.6%) higher fasting insulin level and 16.3% (95% CI: 2.7-31.7%) higher HOMA-β. Bad sleep quality as compared with good sleep quality was associated with higher insulin resistance (e.g., 14.3% (95% CI: 4.7-24.9%) higher HOMA-IR). All these associations disappeared after adjustment for BMI and the risk of sleep apnea. Mendelian Randomization analyses did not indicate a causal association between total, short or long sleep duration and glycemic traits. Therefore, we conclude that measures of habitual sleep duration and sleep quality as applied in our study are unlikely to causally associate with insulin resistance.

INTRODUCTION

During the past decades obesity and conditions that reflect disturbances in metabolism have increased[1]. Multiple studies assessed the association between measures of habitual sleep and metabolic diseases. In several studies, it has been observed that both short and long total sleep duration were associated with a higher risk of obesity, insulin resistance, diabetes mellitus and a higher body weight[2-8]. However, other studies reported only short sleep duration and not long sleep duration to be associated with a higher risk of obesity and metabolic syndrome [9-11]. As total sleep duration has decreased over the past decades[12], studies in this area are of increasing importance.

Several factors could contribute to the discrepancy in the observed associations between sleep duration and insulin resistance. For example, most studies investigated only short sleep duration [2, 11], or short sleep duration under artificial circumstances[2, 13], which might not resemble habitual short sleep duration. Another explanation might be adjustments for confounding factors. Body mass index (BMI) is known on one hand to be one of the largest risk factors for insulin resistance and diabetes while on the other hand it is associated with shorter sleep duration[14-16]. In a study in adolescents, an association between shorter sleep duration and insulin resistance was observed, however, this association disappeared after adjustment for BMI[17]. Although related to BMI, another potential confounding factor is the presence of obstructive sleep apnea (OSA). The presence of OSA was demonstrated to be associated with a higher risk of insulin resistance and diabetes[18]. However, most of the previous performed studies did not adjust for OSA[4, 9, 10, 19-21].

The question remains to what extent the previously described associations between sleep duration with insulin resistance were confounded by BMI and OSA. Besides sleep duration, poor sleep quality is associated with higher presence of obesity, metabolic syndrome and diabetes[22, 23]. However, studies examining poor sleep quality have mainly been performed in patients with diabetes[5, 23, 24]. If assessed in non-diabetic individuals, artificially altered sleep quality was studied by e.g. suppressing slow-wave sleep[25], which does not resemble habitual poor sleep quality. Moreover, methods to assess sleep quality differed between studies. For example, some studies used only a single question about difficulty of initiating or maintaining sleep[5], while others used the Pittsburgh Sleep Quality Index (PSQI) questionnaire[24] or other questionnaires[26, 27]. These different methods to assess sleep quality complicate comparability among these studies.

Based on earlier studies we hypothesized that both short sleep duration and bad sleep quality are associated with higher insulin resistance, but this association is likely dependent on anthropometric traits like BMI and the risk of sleep apnea. We aimed to examine this hypothesis in a middle-aged non-diabetic population embedded in the Netherlands Epidemiology of Obesity (NEO) study[28]. Furthermore, we extended this study with a two-sample Mendelian Randomization (MR) analysis to provide evidence whether the association between sleep duration and insulin resistance is causal using data of the Meta-Analyses of Glucose and Insulin-Related Traits Consortium (MAGIC).

METHODS

Study design and study population

The Netherlands Epidemiology of Obesity (NEO) study is a population-based cohort study. Participants were recruited from September 2008 until September 2012, resulting in a cohort of 6,671 individuals with an oversampling of individuals with overweight or obesity. Men and women aged between 45 and 65 years with a self-reported body mass index (BMI) of 27 kg/m^2 or higher living in the greater area of Leiden were eligible to participate in the NEO study. In addition, all inhabitants aged between 45 and 65 years from one municipality (Leiderdorp) were invited irrespective of their BMI, allowing for a reference distribution of BMI. Baseline data were collected at the NEO study center of the Leiden University Medical Center (LUMC). Prior to the NEO study visit, participants completed a questionnaire about demographic and clinical information and fasted for at least 10 hours. Participants came to the research site in the morning to undergo several baseline measurements including anthropometric measurements and blood sampling. The participants drank a liquid mixed meal where after postprandial blood sampling was performed. All medication used in the month preceding the visit to the study center was recorded by research nurses. More detailed information on the study design and data collection was described elsewhere [28]. The NEO study was approved by the medical ethics committee of the LUMC and all participants gave written informed consent.

In the present study, we excluded participants with missing data on the PSQI questionnaire (N=1,402), which was collected in participants enrolled in NEO after July 2009. Moreover, we excluded participants that used glucose lowering medication (N=263), had a medical history of diabetes mellitus (N=77), had missing baseline characteristics (N=132), had missing glucose or insulin concentrations (N=48), had missing data on the Berlin questionnaire (N=57) or were not fasted (N=20). In the analyses with the postprandial measures, we additionally excluded participants with incomplete or no liquid meal intake (N=8) and participants with either missing postprandial glucose (N=153) or postprandial insulin concentrations (N=78).

Sleep characteristics

To assess habitual sleep duration and sleep quality, we used data collected with the PSQI [29], which is a self-rated questionnaire, to retrospectively measure sleep parameters over a one month time period. Total sleep duration was derived from the question "On an average night, how much sleep do you get?". To obtain a classification of short and long total sleep duration, we calculated the age- and sex- adjusted residuals with linear regression analysis for total sleep duration and determined subgroups on the basis of these residuals. We used the 5th lowest percentile to define shortest sleep, the 5th till

20th percentile to define short sleep, the 20th till 80th to define medium sleep, the 80th till 95th to define long sleep and the 95th till 100th percentile to define longest sleep of the adjusted residuals. Sleep quality was assessed using the total score of the PSQI questionnaire. The questionnaire consists of seven components based on which an overall score can be calculated ranging from 0 to 21, in which a higher score indicates a poorer sleep quality [29]. In the sleep quality analyses, the good sleep quality group (PSQI total score ≤5) was used as a reference group in linear regression analyses with two groups with either a PSQI total score of >5 or ≥10, the latter in order to investigate a more extreme disturbed habitual sleep.

Glycemic traits

After an overnight fast of at least 10 hours, fasting blood samples were taken at the study center. Within 5 minutes after the first blood sample was taken, participants drank a liquid mixed meal. This meal (400 mL) contained 600 kcal, with 16% of energy derived (En%) from protein, 50% En% from carbohydrates and 34 En% from fat. Two postprandial blood samples were taken at 30 and 150 minutes after ingestion of the mixed meal. Serum and plasma was collected during each of the three blood draws and concentrations of glucose and insulin were determined. Fasting plasma glucose concentrations were determined by enzymatic and colorimetric methods (Roche Modular Analytics P800, Roche Diagnostics, Mannheim, Germany; CV < 5%) and serum insulin concentrations were determined by an immunometric method (Siemens Immulite 2500, Siemens Healthcare Diagnostics, Breda, The Netherlands; CV < 5%). All analyses were performed in the central clinical chemistry laboratory of the Leiden University Medical Center. Fasting glucose and insulin levels were used to calculate the Homeostatic Model Assessment of Insulin Resistance (HOMA-IR) index as a marker for hepatic insulin resistance. The HOMA-IR was calculated using (fasting insulin * fasting glucose) / 22.5. HOMA of β-cell function (HOMA-β) was used as a marker that indicates basal insulin release, and calculated as 20 * (fasting insulin / fasting glucose) - 3.5 [30, 31]. The area under the curve (AUC) for postprandial overall glucose and insulin levels was calculated using the Trapezoid Rule as (15 * fasting concentration + 75 * concentration$_{30min}$ + 60 * concentration$_{150min}$) / 150 [32].

Covariates

Level of education was reported in 10 categories according to the Dutch education system and grouped into high versus low education [28]. A semi-quantitative food frequency questionnaire (FFQ) [33] questionnaire was used to assess daily total energy intake. Energy intake was estimated from the FFQ with the 2011 version of the Dutch food composition table (NEVO-2011). Participants reported the frequency and duration of their physical activity in leisure time using the Short Questionnaire to Assess Health-enhancing

physical activity (SQUASH) [34], which was expressed in hours per week of metabolic equivalents (MET-h/week). Body weight was measured without shoes and one kilogram (kg) was subtracted to correct for the weight of clothing. BMI was calculated by dividing the weight in kilograms by the height in meters squared. The Berlin questionnaire was used to assess the risk for the presence of obstructive sleep apnea syndrome [35]. This questionnaire consists of 10 questions that form three categories (snoring (category 1), daytime somnolence (category 2) and hypertension and BMI (category 3)) related to the likelihood of the presence of sleep apnea. Individuals can be classified as either having a high (2 or more categories with a positive score) or low likelihood of the presence of sleep apnea (only 1 or no categories with a positive score).

Statistical Analysis

Because individuals with a BMI of 27 kg/m^2 or higher were oversampled in the NEO study, adjustments were made to correctly represent associations in the general population [36]. This was done by weighting individuals towards the BMI distribution of participants from the Leiderdorp municipality [37], whose BMI distribution was similar to the BMI distribution of the general Dutch population. All presented results are based on weighted analyses [38]. Consequently, the results apply to a population-based study without oversampling of participants with a BMI of 27 kg/m^2 or higher. We performed all statistical analyses of the NEO cohort using Stata version 12.1 (Stata, College Station, Texas, USA).

Baseline characteristics were expressed as means (with standard deviation, SD) for normally distributed measures, median with inter-quartile ranges for non-normally distributed measures, and proportions for categorical variables. Not normally distributed outcomes were log-transformed to approximate a normal distribution (notably fasting insulin, HOMA-IR, HOMA-β, AUC insulin). In order to present the results with a similar interpretation, normally distributed outcomes (notably fasting glucose and AUC glucose) were log-transformed as well. Linear regression analyses using the medium sleep category (characterized by 20th till 80th percentile of sleep duration residuals) as reference group were performed. The subsequent beta regression coefficients are expressed as percentages with accompanying 95% confidence interval (95% CI), which can be interpreted as the percentage change in outcome with respect to the reference group. The initial model in linear regression analyses was adjusted for age and sex (Model 1). In addition to age and sex, we adjusted in Model 2 for ethnicity (white/other), education level (high/other), smoking (never/former/current), alcohol consumption (g/day), energy intake (kJ/day), physical activity (MET/h/week) and sleep medication (yes/no). In Model 3 we additionally adjusted for sleep apnea (high risk/low risk) and BMI (kg/m^2). In the analyses for sleep quality we did not adjust for sleep medication in Model 2 and Model 3, as this is a component of the PSQI total score.

Mendelian Randomization analysis

For the MR analysis, we selected 78 single nucleotide polymorphisms (SNPs) that have been shown to associate with self-reported total sleep duration (p-value < $5e^{-8}$) as genetic instruments from the largest genome-wide association study (GWAS; source: UK Biobank, 446,118 unrelated European-ancestry individuals [39]). Moreover, we selected 27 SNPs associated with short sleep duration (<7 hours; n = 106,192 cases) and 8 SNPs associated with long sleep duration (≥9 hours; n = 34,184 cases) relative to 7-8 hours of sleep duration (n = 305,742) [39].

Summary statistics data of GWAS on glycaemic traits were used as outcomes in a two-sample MR approach. We used data from MAGIC, which comprised of a meta-analysis of European ancestry studies that investigated genetic variants associated with glycemic traits. For fasting glucose and fasting insulin, we used the GWAS of Manning *et al*. (2012)[40] which included a total of 58,074 and 51,570 individuals without diabetes mellitus, respectively. In addition, we used HOMA-IR and HOMA-B as measures of insulin resistance (n = 46,186) [41].

For the two-sample MR analyses, we used similar methodology as have been described previously [42]. In short, we combined effects of the individual genetic instruments using the inverse variance weighted (IVW) approach as our main analysis method with MRCIEU/TwoSampleMR package in R [43]. The resulting estimate (presented with accompanying 95% CI) can be interpreted as a weighted mean effect of genetically determined increase in total sleep duration (per hour) and a higher risk for short or long sleep duration on our study outcomes. In order to formally test for potential pleiotropic effects of the genetic variants, in which the genetic variants have pleiotropic effects that influence glycemic traits via alternative pathways, we conducted MR-Egger regression as sensitivity analysis [44]. We furthermore conducted weighted median estimator (WME) analyses; similarity between the IVW and WME effect estimates are indicative of robustness of the results [45].

RESULTS

Baseline characteristics

In the present study we included a total of 4,672 participants with a mean age of 56 (SD 6.0) years, of whom 48% was men. In **Table 1** the population characteristics are presented for the study population, stratified by the sleep duration groups. As compared with the medium sleep group (sleep duration = 7 h/day), participants in the shortest sleep group (sleep duration = 5 h/day) were more of non-white ancestry (96% vs. 90%), had a lower education (50% high vs. 39% high), higher BMI (26 vs. 27 kg/m^2), smoked more (16% vs. 18%), were less physically active (31 vs. 25 MET/h/week), used more sleep medication (4% vs. 14%), had a higher PSQI total score (4 vs. 11) and had a higher risk of having sleep apnea (17% vs. 35%). In the shortest sleep group fasting insulin, HOMA-β and AUC of insulin were higher than in the medium sleep group. As compared with the medium group, the longest sleep duration group consisted of more men (42% vs. 53%), more current smokers (16% vs. 20%), less high-educated participants (50% vs. 36%), more frequent use of sleep medication (4% vs. 7%) and more participants with a higher risk of having sleep apnea (17% vs. 24%) as compared with the medium sleep group. All other characteristics were similar between the groups.

Sleep duration and glycemic traits

In the analyses adjusted for age and sex (Model 1), shortest sleep was associated with a 17.9% (95%CI: 4.8;32.5%) higher fasting insulin, a 18.4% (95%CI: 3.9;34.9%) higher HOMA-IR, a 18.8% (95%CI: 4.9;34.7%) higher HOMA-β and a 17.5% (95%CI: 7.9;28.0%) higher AUC for insulin as compared with the medium sleep group (**Figure 1** and **Online Supplementary Table 1**). When we adjusted for potential confounding factors (Model 2), shortest sleep duration was associated with a 14.5% (95%CI: 2.0;28.6%) higher fasting insulin, a 14.9% (95%CI: 1.0;30.7%) higher HOMA-IR, a 16.3% (95%CI: 2.7;31.7%) higher HOMA-β and a 14.5% (95%CI: 4.9;24.9%) higher AUC of insulin. However, the associations between shortest sleep duration and HOMA-IR attenuated when analyses were adjusted for considered potential confounding factors (Model 2). When we additionally adjusted for BMI and risk of sleep apnea (Model 3) the other associations observed for shortest sleep duration disappeared. However, in Model 3, longest sleep was associated with a -16.3% (95%CI: -29.6;-0.4%) lower HOMA-IR as compared with the medium sleep duration group. The same attenuation of associations was observed when we adjusted separately for either BMI or the risk of sleep apnea (**Online Supplementary Table 1**: Model 3 and 4).

Table 1. *Characteristics of participants in the Netherlands Epidemiology of Obesity study, stratified by sleep duration (N=4,672).*

Sleep duration	Shortest	Short	Medium	Long	Longest
	0-5%	5-20%	20-80%	80-95%	95-100%
Age (years)	57 [5]	58 [5]	55 [6]	54 [6]	58 [6]
Sex (% men)	45	49	42	43	53
BMI (kg/m^2)	27 [5]	26 [4]	26 [4]	26 [4]	26 [5]
Ethnicity (% white)	90	94	96	96	94
Education (% high)	39	41	50	48	36
Smoking (%current)	18	14	16	16	20
Sleep medication (%)	14	10	4	5	7
Alcohol consumption (g/day)	12 [3;22]	11 [3;22]	10 [3;21]	9 [2;21]	11 [1;23]
Physical activity (MET/h/week)	25 [12;44]	30 [16;50]	31 [17;51]	32 [15;52]	30 [16;49]
Sleep duration (h/day)	5 [4;5]	6 [6;6]	7 [7;8]	8 [8;8]	9 [9;9]
PSQI (total score)	11 [9;13]	7 [5;9]	4 [3;6]	3 [1;4]	3 [1;5]
Sleep apnea (%)	35	25	17	17	24
Fasting glucose (mmol/L)	6 [1]	6 [2]	6 [1]	6 [1]	6 [2]
Fasting insulin (mmol/L)	9 [6;14]	8 [6;12]	7 [5;11]	7 [5;11]	7 [4;12]
HOMA-IR	2 [1;4]	2 [1;3]	2 [1;3]	2 [1;3]	2 [1;3]
HOMA-β	28 [18;42]	26 [18;40]	24 [16;37]	25 [15;36]	25 [13;37]
AUC Glucose*	6 [1]	6 [1]	6 [1]	6 [1]	6 [1]
AUC Insulin#	47 [34;62]	41 [30;57]	38 [29;53]	38 [30;54]	41 [26;61]

Abbreviations: AUC, area under the curve; BMI, body mass index; HOMA-β, Homeostatic Model of Assessment β-cell function; HOMA-IR, Homeostatic Model of Assessment Insulin Resistance; kJ, kilojoule; MET, metabolic equivalents of task; NEO, Netherlands Epidemiology of Obesity; PSQI, Pittsburgh Sleep Questionnaire Index. Results were based on analyses weighted towards the BMI distribution of the general Dutch population. Data presented as mean ± standard deviation (SD); proportion (%); median (25^{th}-75^{th} percentile). *, N=4,511; # N=4,586.

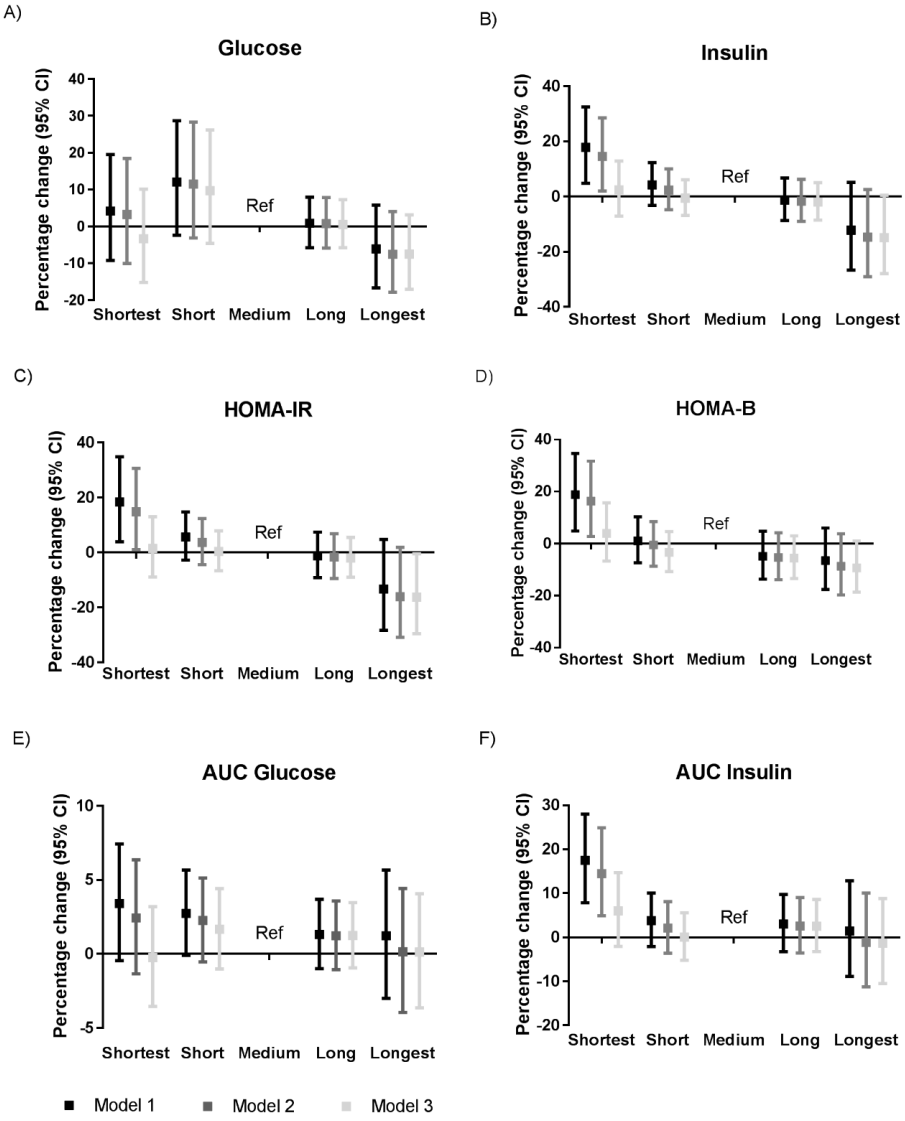


Figure 1. *Associations between sleep duration and A) fasting glucose, B) fasting insulin, C) HOMA-IR, D) HOMA-β, E) AUC of glucose and F) AUC of insulin.* Results were based on analyses weighted towards the BMI distribution of the general Dutch population (N=4,672), and were derived from regression coefficients with 95% confidence intervals from linear regression analyses and expressed as percentage change in outcome measure, as compared with the medium sleep duration group as reference category. Model 1: adjusted for age and sex; Model 2: adjusted for age, sex, ethnicity, education level, smoking, alcohol intake, caloric intake and physical activity; Model 3: adjusted for Model 2 + risk of sleep apnea and BMI. Abbreviations: AUC, area under the curve; CI, confidence interval; HOMA-IR, Homeostatic Model of Insulin Resistance; HOMA-β, Homeostatic Model of β cell function; Ref, reference category. N=4,511 for the analyses with AUC Glucose and N=4,586 for the analyses with AUC Insulin.

Sleep quality and glycemic traits

The associations between sleep quality and glycemic traits are visualized in **Figure 2** and **Online Supplementary Table 2**. We did find evidence for an association between poor sleep quality, as defined by a PSQI score of >5 and a 6.5% (95%CI: 0.0;13.5%) higher HOMA- β in Model 1 as compared with the good sleep quality group (PSQI score ≤5). However, when we adjusted for potential confounding factors in Model 2, this association disappeared (percentage change 5.1% (95%CI: -1.1-11.8%). In the age and sex adjusted analyses (Model 1), a PSQI score ≥10 was associated with a 16.9% (95%CI; 7.6;27.1%) higher fasting insulin, a 17.3% (95%CI; 7.0;28.5%) higher HOMA-IR, a 18.6% (95%CI: 8.4;29.7%) higher HOMA-β and a 7.7% (95%CI: 0.7;15.2%) higher AUC of insulin as compared with the good sleep quality group. When we additionally adjusted for potential confounding factors in Model 2, we observed a 14.1% (95%CI: 5.3;23.7%) higher fasting insulin, a 14.3% (95%CI: 4.7;24.9%) higher HOMA-IR, and a 15.7% (95%CI: 6.0;26.4%) higher HOMA-β in the group with a PSQI total score of ≥10 as compared with the good sleep quality group. However, all associations attenuated when we additionally adjusted for BMI and the risk of sleep apnea. These associations attenuated in a similar manner when we adjusted the model for only BMI or the risk of sleep apnea (**Online Supplementary Table 2**: Model 3 and 4).

Genetically-determined habitual sleep duration and glycemic traits

In order to assess whether the association observed between sleep duration and glycemic traits (Model 1 and 2) is causal, we performed an MR study. We found no evidence for an association between genetically determined total sleep duration, short sleep duration and long sleep duration with glycemic traits using IVW-analyses (**Table 2** and **Online Supplementary Table 3**). The WME analyses were consistent with these results (**Online Supplementary Table 3**). However, we did find evidence for an association between total sleep duration and a higher HOMA- β (IVW-estimate: 0.08 [95%CI: 0.01;0.014]). The estimate from WME analysis was consistent and MR-Egger did not indicate the presence of directional pleiotropy in this analysis (**Online Supplementary Table 3**).

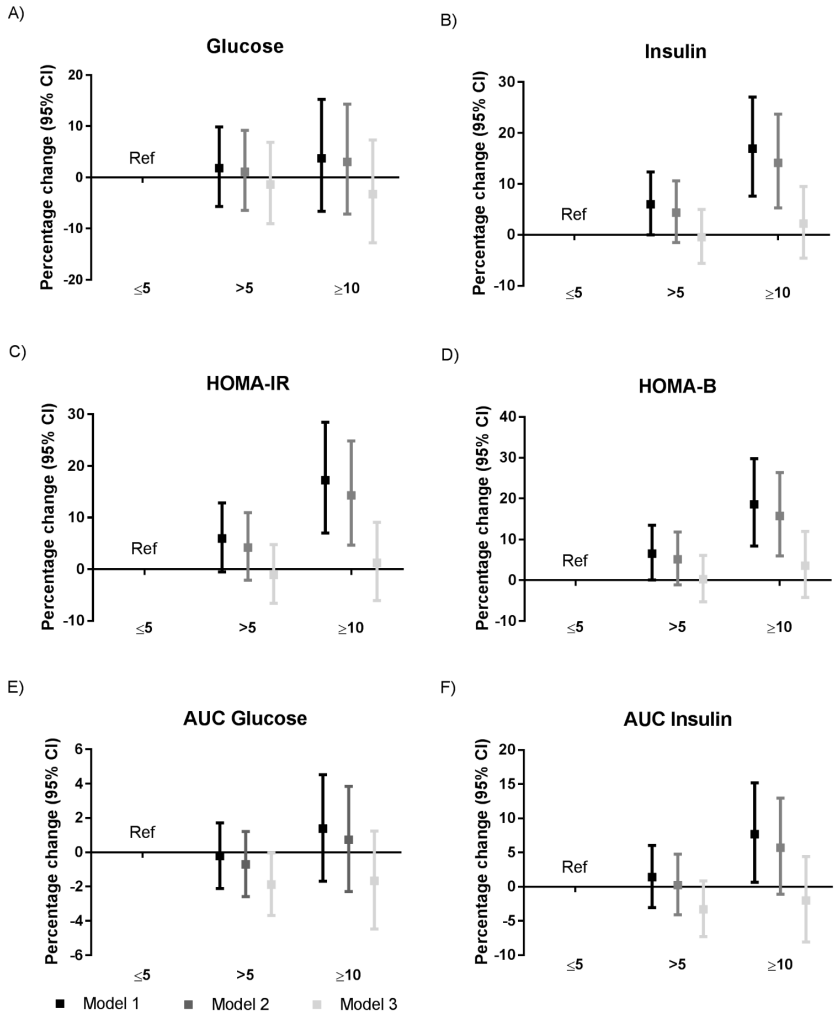


Figure 2. *Associations between sleep quality and A) fasting glucose, B) fasting insulin, C) HOMA-IR, D) HOMA-β, E) AUC of glucose and F) AUC of insulin.* Results were based on analyses weighted towards the BMI distribution of the general Dutch population (N=4,672), and were derived from regression coefficients with 95% confidence intervals from linear regression analyses, using either a group with a PSQI total score of >5 or ≥10, and expressed as percentage change in outcome measure, as compared with the good sleep group as reference category. Model 1: adjusted for age and sex; Model 2: adjusted for age, sex, ethnicity, education level, smoking, alcohol intake, caloric intake and physical activity; Model 3: adjusted for Model 2 + risk of sleep apnea and BMI. Abbreviations: AUC, area under the curve; CI, confidence interval; HOMA-IR, Homeostatic Model of Insulin Resistance; HOMA-β, Homeostatic Model of β cell function; Ref, reference category. N=4,511 for the analyses with AUC Glucose and N=4,586 for the analyses with AUC Insulin.

Table 2. *Inverse-variance weighted estimates for sleep duration on glycemic traits.*

		Total sleep duration		Short sleep duration		Long sleep duration
	SNPs	Estimate [95%CI]	SNPs	Estimate [95%CI]	SNPs	Estimate [95%CI]
Fasting glucose	45	-0.03 [-0.11;0.06]	20	-0.09 [-0.19;0.01]	5	-0.05 [-0.24;0.13]
Fasting Insulin	45	0.01 [-0.04;0.07]	20	-0.01 [-0.11;0.08]	5	-0.06 [-0.36;0.23]
HOMA-IR	53	0.07 [-0.01;0.15]	20	0.04 [-0.08;0.15]	5	-0.01 [-0.47;0.45]
HOMA-β	53	0.08 [0.01;0.14]	20	0.09 [-0.01;0.19]	5	0.09 [-0.26;0.45]

Data presented as odd ratios (type 2 diabetes) or beta coefficients with 95% confidence interval per standard deviation increase in exposure.

DISCUSSION

In the present study, we performed a cross-sectional analysis in a middle-aged population of 4,519 non-diabetic participants to determine the associations between habitual sleep duration and sleep quality with glycemic traits. When analyses were adjusted for age and sex, shortest sleep duration was associated with higher fasting insulin, higher HOMA-IR, higher HOMA-β and higher AUC of insulin. Poorer sleep quality was associated with higher fasting insulin, higher HOMA-IR, higher HOMA-β and higher AUC of insulin. However, all these associations disappeared after additional adjustment for BMI and the risk of sleep apnea. Furthermore, in the MR analyses, we, overall, did not observe associations between genetically-determined total, short and long sleep duration and glycemic traits. Of interest, we found some evidence that longer genetically-determined sleep duration was associated with a higher HOMA-β.

Previous studies showed that both short and long sleep duration, or only short sleep duration were associated with a higher risk of obesity, body weight, insulin resistance, and diabetes mellitus [2-5, 9, 10]. We hypothesized that these differences in findings may be explained by confounding factors such as BMI and sleep apnea. In agreement, in the age- and sex-adjusted analyses we observed associations between shorter sleep duration and higher insulin resistance which disappeared when we additionally adjusted for BMI and the presence of sleep apnea. Our findings are in agreement with a study of Javaheri *et al.* who also observed an association between short sleep duration and insulin resistance, which disappeared after adjustment for obesity [17]. In line, our MR study, genetically-determined sleep duration was, in general, not associated with glycemic traits, indicating that there is no direct effect of sleep on glycemic traits. However, it has been observed that there is a shared genetic component between sleep and anthropometrics [46]. Moreover, adiposity has been shown to be causally associated with an increased risk of diabetes and altered glycemic traits [15]. Therefore, we hypothesize that excess adiposity is a common cause for alterations in habitual sleep and insulin resistance, instead of being in the causal path of this association.

An alternative explanation for these observations is that BMI/risk of sleep apnea is in the causal pathway between shorter sleep duration and increased insulin resistance. It was shown that sleep restriction reduces leptin levels and enhances ghrelin levels and thereby increases cravings for carbohydrate-rich foods which may result in weight gain [47]. Another mechanism that links short sleep duration with insulin resistance might be via obstructive sleep apnea (OSA). It was demonstrated that patients with OSA showed increased insulin resistance compared with controls without OSA, independent of obesity [48]. These might indicate that, although obesity is known to effect insulin resistance

[15], OSA may have an independent additional effect on insulin resistance possibly via nocturnal hypoxemia [49]. A GWAS of Lane *et al.* on total sleep duration suggested that there is a shared genetic component between sleep duration and energy metabolism [46], which demonstrates that the genetic background of an individual may also interfere with altered sleep duration and energy metabolism.

One of the strengths of this study is the use of residuals to determine the sleep duration. As no clear cut-off for short and long sleep exists in literature, using residuals is a good way to harmonize grouping of sleep duration among studies. Moreover, heterogeneity of future meta-analyses and systematic reviews might benefit from using this methodology. Another strength of this study is the extensive phenotyping of the NEO study. This enables us to correct for a broad range of possible confounding factors, such as ethnicity, physical activity, BMI, and risk of sleep apnea. In addition, we have used the PSQI questionnaire, which is a widely used validated tool to assess sleep disturbances. A limitation of the present study is the cross-sectional design, therefore, we are not able to exclude reverse causation or residual confounding in this study. Another limitation of this study is the use of subjective sleep questionnaires to assess sleep duration, sleep quality and the risk of sleep apnea. One of the drawbacks of self-reported data is the fact that this data is subject to recall bias. Moreover, there may be a measurement error resulting in non-differential misclassification for the exposure. Therefore, the use of a sleep diary may improve reliability of the measures of habitual sleep. However, the PSQI questionnaire has been shown to be a reliable and validated tool to assess sleep dysfunction and the Berlin questionnaire, used to assess the risk of sleep apnea, has been shown to be validated to detect mild and severe OSA in several populations [50, 51]. The MR study has the advantage of being a highly efficient method that allows for usage of large sample sizes, however, although we used the largest GWAS to date, we cannot rule out the use of weak instruments. The 78 loci used in the current study explained 0.69% of the variance in sleep duration. Moreover, the amount of genetic instruments that were present in this current study, especially for long sleep duration, was minimal. Finally, the identification of the genetic variants as well as the MR study were both performed in cohorts of European ancestry, which hampers the generalizability to non-European populations.

In summary, shorter sleep duration and poorer sleep quality were associated with higher insulin resistance, but these associations were dependent on BMI and the risk of sleep apnea. Both BMI and the risk of sleep apnea thereby likely explain previous observed associations between adverse habitual sleep and an increased risk of insulin resistance, suggesting that these factors should therefore be considered in future studies.

Acknowledgements and funding

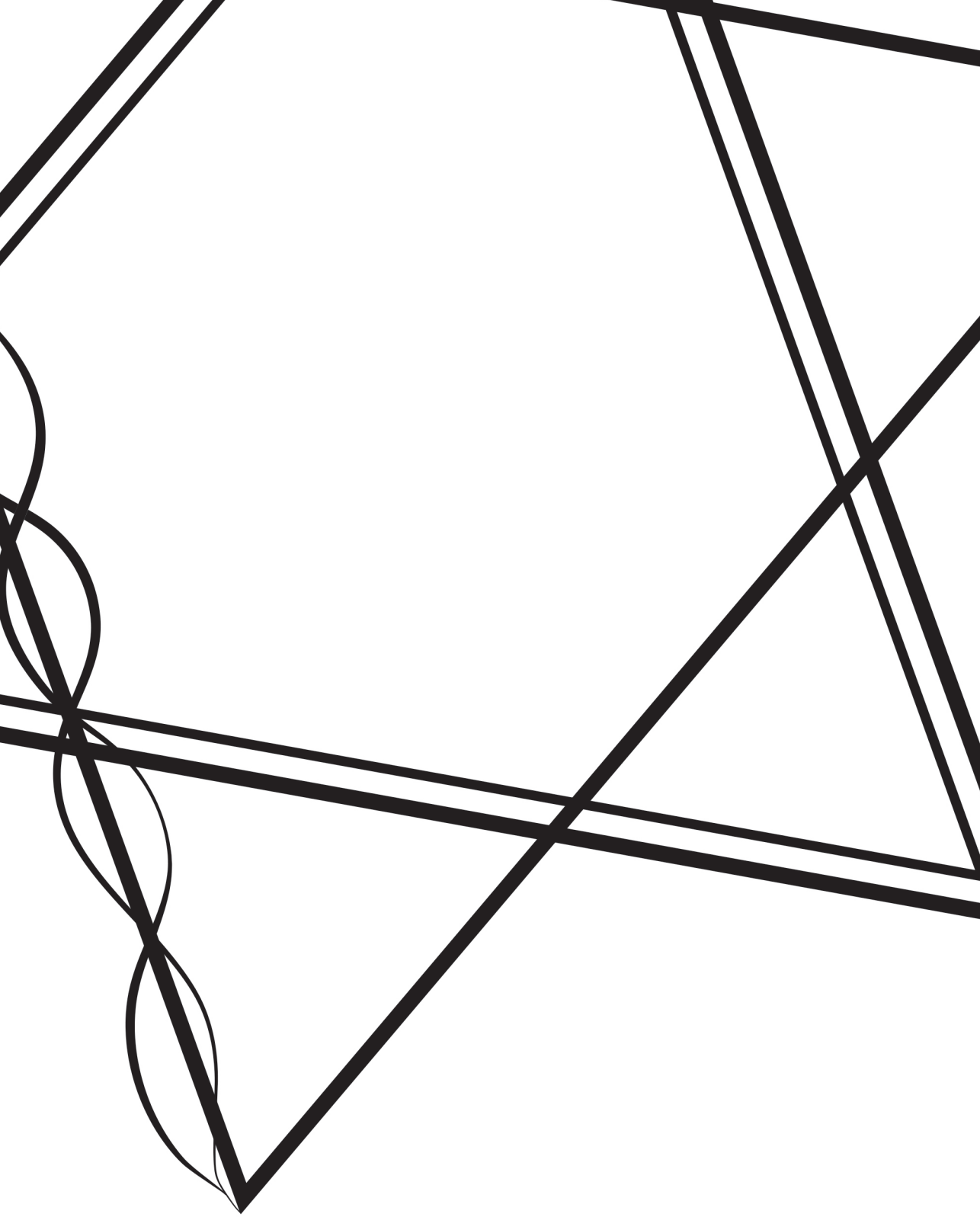
We express our gratitude to all individuals who participate in the Netherlands Epidemiology of Obesity study. We are grateful to all participating general practitioners for inviting eligible participants. We furthermore thank P.R. van Beelen and all research nurses for collecting the data, P.J. Noordijk and her team for sample handling and storage, and I. de Jonge for all data management of the NEO study. The NEO study is supported by the participating Departments, the Division and the Board of Directors of the Leiden University Medical Centre, and by the Leiden University, Research Profile Area 'Vascular and Regenerative Medicine'. Nutricia Research, Utrecht, The Netherlands, provided the mixed meal. We acknowledge the support from the Netherlands Cardiovascular Research Initiative: an initiative with support of the Dutch Heart Foundation (CVON2014-02 ENERGISE). DvH was supported by the European Commission funded project HUMAN (Health-2013-INNOVATION-1-602757). NRB was supported by the Netherlands Organization for Scientific Research (NWO-VENI 016.136.125).

REFERENCES

1. Gregg EW, Shaw JE. Global Health Effects of Overweight and Obesity. N Engl J Med. 2017;377(1):80-1.
2. Wong PM, Manuck SB, DiNardo MM, Korytkowski M, Muldoon MF. Shorter sleep duration is associated with decreased insulin sensitivity in healthy white men. Sleep. 2015;38(2):223-31.
3. Reutrakul S, Van Cauter E. Interactions between sleep, circadian function, and glucose metabolism: implications for risk and severity of diabetes. Annals of the New York Academy of Sciences. 2014;1311:151-73.
4. Cappuccio FP, Taggart FM, Kandala NB, Currie A, Peile E, Stranges S, et al. Meta-analysis of short sleep duration and obesity in children and adults. Sleep. 2008;31(5):619-26.
5. Cappuccio FP, D'Elia L, Strazzullo P, Miller MA. Quantity and quality of sleep and incidence of type 2 diabetes: a systematic review and meta-analysis. Diabetes Care. 2010;33(2):414-20.
6. Brady EM, Bodicoat DH, Hall AP, Khunti K, Yates T, Edwardson C, et al. Sleep duration, obesity and insulin resistance in a multi-ethnic UK population at high risk of diabetes. Diabetes Res Clin Pract. 2018;139:195-202.
7. Rutters F, Besson H, Walker M, Mari A, Konrad T, Nilsson PM, et al. The Association Between Sleep Duration, Insulin Sensitivity, and beta-Cell Function: The EGIR-RISC Study. J Clin Endocrinol Metab. 2016;101(9):3272-80.
8. van Dijk D, Balkau B, Segrestin B, Gottsater M, Gabriel R, Hatunic M, et al. Associations between sleep duration and sleep debt with insulin sensitivity and insulin secretion in the EGIR-RISC Study. Diabetes Metab. 2018.
9. Wu Y, Zhai L, Zhang D. Sleep duration and obesity among adults: a meta-analysis of prospective studies. Sleep Med. 2014;15(12):1456-62.
10. Xi B, He D, Zhang M, Xue J, Zhou D. Short sleep duration predicts risk of metabolic syndrome: a systematic review and meta-analysis. Sleep Med Rev. 2014;18(4):293-7.
11. Liu A, Kushida CA, Reaven GM. Habitual shortened sleep and insulin resistance: an independent relationship in obese individuals. Metabolism. 2013;62(11):1553-6.
12. Schoenborn CA, Adams PF, Peregoy JA. Health behaviors of adults: United States, 2008-2010. Vital Health Stat 10. 2013(257):1-184.
13. Donga E, van Dijk M, van Dijk JG, Biermasz NR, Lammers GJ, van Kralingen KW, et al. A single night of partial sleep deprivation induces insulin resistance in multiple metabolic pathways in healthy subjects. J Clin Endocrinol Metab. 2010;95(6):2963-8.
14. Kahn SE, Hull RL, Utzschneider KM. Mechanisms linking obesity to insulin resistance and type 2 diabetes. Nature. 2006;444(7121):840-6.
15. Dale CE, Fatemifar G, Palmer TM, White J, Prieto-Merino D, Zabaneh D, et al. Causal Associations of Adiposity and Body Fat Distribution With Coronary Heart Disease, Stroke Subtypes, and Type 2 Diabetes Mellitus: A Mendelian Randomization Analysis. Circulation. 2017;135(24):2373-88.
16. Dekker SA, Noordam R, Biermasz NR, de Roos A, Lamb HJ, Rosendaal FR, et al. Habitual Sleep Measures are Associated with Overall Body Fat, and not Specifically with Visceral Fat, in Men and Women. Obesity (Silver Spring). 2018;26(10):1651-8.
17. Javaheri S, Storfer-Isser A, Rosen CL, Redline S. Association of short and long sleep durations with insulin sensitivity in adolescents. J Pediatr. 2011;158(4):617-23.

18. Morgenstern M, Wang J, Beatty N, Batemarco T, Sica AL, Greenberg H. Obstructive sleep apnea: an unexpected cause of insulin resistance and diabetes. Endocrinol Metab Clin North Am. 2014;43(1):187-204.
19. Cappuccio FP, Cooper D, D'Elia L, Strazzullo P, Miller MA. Sleep duration predicts cardiovascular outcomes: a systematic review and meta-analysis of prospective studies. Eur Heart J. 2011;32(12):1484-92.
20. van den Berg JF, Miedema HM, Tulen JH, Neven AK, Hofman A, Witteman JC, et al. Long sleep duration is associated with serum cholesterol in the elderly: the Rotterdam Study. Psychosom Med. 2008;70(9):1005-11.
21. Shin HY, Kang G, Kim SW, Kim JM, Yoon JS, Shin IS. Associations between sleep duration and abnormal serum lipid levels: data from the Korean National Health and Nutrition Examination Survey (KNHANES). Sleep Med. 2016;24:119-23.
22. Jennings JR, Muldoon MF, Hall M, Buysse DJ, Manuck SB. Self-reported sleep quality is associated with the metabolic syndrome. Sleep. 2007;30(2):219-23.
23. Koren D, Dumin M, Gozal D. Role of sleep quality in the metabolic syndrome. Diabetes Metab Syndr Obes. 2016;9:281-310.
24. Tang Y, Meng L, Li D, Yang M, Zhu Y, Li C, et al. Interaction of sleep quality and sleep duration on glycemic control in patients with type 2 diabetes mellitus. Chin Med J (Engl). 2014;127(20):3543-7.
25. Tasali E, Leproult R, Ehrmann DA, Van Cauter E. Slow-wave sleep and the risk of type 2 diabetes in humans. Proc Natl Acad Sci U S A. 2008;105(3):1044-9.
26. Byberg S, Hansen AL, Christensen DL, Vistisen D, Aadahl M, Linneberg A, et al. Sleep duration and sleep quality are associated differently with alterations of glucose homeostasis. Diabet Med. 2012;29(9):e354-60.
27. Engeda J, Mezuk B, Ratliff S, Ning Y. Association between duration and quality of sleep and the risk of pre-diabetes: evidence from NHANES. Diabet Med. 2013;30(6):676-80.
28. de Mutsert R, den Heijer M, Rabelink TJ, Smit JW, Romijn JA, Jukema JW, et al. The Netherlands Epidemiology of Obesity (NEO) study: study design and data collection. Eur J Epidemiol. 2013;28(6):513-23.
29. Buysse DJ, Reynolds CF, 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res. 1989;28(2):193-213.
30. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia. 1985;28(7):412-9.
31. Wallace TM, Levy JC, Matthews DR. Use and abuse of HOMA modeling. Diabetes Care. 2004;27(6):1487-95.
32. Retnakaran R, Shen S, Hanley AJ, Vuksan V, Hamilton JK, Zinman B. Hyperbolic relationship between insulin secretion and sensitivity on oral glucose tolerance test. Obesity (Silver Spring). 2008;16(8):1901-7.
33. Feunekes GI, Van Staveren WA, De Vries JH, Burema J, Hautvast JG. Relative and biomarker-based validity of a food-frequency questionnaire estimating intake of fats and cholesterol. The American journal of clinical nutrition. 1993;58(4):489-96.
34. Wendel-Vos GC, Schuit AJ, Saris WH, Kromhout D. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. Journal of clinical epidemiology. 2003;56(12):1163-9.

35. Netzer NC, Stoohs RA, Netzer CM, Clark K, Strohl KP. Using the Berlin Questionnaire to identify patients at risk for the sleep apnea syndrome. Annals of internal medicine. 1999;131(7):485-91.
36. Korn EL, Graubard BI. Epidemiologic studies utilizing surveys: accounting for the sampling design. Am J Public Health. 1991;81(9):1166-73.
37. Lumley T. Analysis of complex survey samples [Available from: http://www.jstatsoft.org/v09/i08/paper.
38. VWS Mv. Hoveel mensen hebben overgewicht? [Available from: http://www.rivm.nl/Onderwerpen/N/Nederland_de_Maat_Genomen.
39. Dashti H, Jones SE, Wood AR, Lane J, van Hees VT, Wang H, et al. GWAS in 446,118 European adults identifies 78 genetic loci for self-reported habitual sleep duration supported by accelerometer-derived estimates. bioRxiv. 2018:274977.
40. Manning AK, Hivert MF, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat Genet. 2012;44(6):659-69.
41. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet. 2010;42(2):105-16.
42. Bos MM, Smit RAJ, Trompet S, van Heemst D, Noordam R. Thyroid Signaling, Insulin Resistance, and 2 Diabetes Mellitus: A Mendelian Randomization Study. J Clin Endocrinol Metab. 2017;102(6):1960-70.
43. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018;7.
44. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44(2):512-25.
45. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol. 2016;40(4):304-14.
46. Lane JM, Liang J, Vlasac I, Anderson SG, Bechtold DA, Bowden J, et al. Genome-wide association analyses of sleep disturbance traits identify new loci and highlight shared genetics with neuropsychiatric and metabolic traits. Nature genetics. 2017;49(2):274-81.
47. Spiegel K, Tasali E, Penev P, Van Cauter E. Brief communication: Sleep curtailment in healthy young men is associated with decreased leptin levels, elevated ghrelin levels, and increased hunger and appetite. Annals of internal medicine. 2004;141(11):846-50.
48. Tassone F, Lanfranco F, Gianotti L, Pivetti S, Navone F, Rossetto R, et al. Obstructive sleep apnoea syndrome impairs insulin sensitivity independently of anthropometric variables. Clin Endocrinol (Oxf). 2003;59(3):374-9.
49. Dewan NA, Nieto FJ, Somers VK. Intermittent hypoxemia and OSA: implications for comorbidities. Chest. 2015;147(1):266-74.
50. Mollayeva T, Thurairajah P, Burton K, Mollayeva S, Shapiro CM, Colantonio A. The Pittsburgh sleep quality index as a screening tool for sleep dysfunction in clinical and non-clinical samples: A systematic review and meta-analysis. Sleep Med Rev. 2016;25:52-73.
51. Amra B, Rahmati B, Soltaninejad F, Feizi A. Screening Questionnaires for Obstructive Sleep Apnea: An Updated Systematic Review. Oman Med J. 2018;33(3):184-92.



CHAPTER 3.3

Multi-ancestry sleep-by-SNP interaction analysis in 126,926 individuals reveals lipid loci stratified by sleep duration.

Raymond Noordam, Maxime M Bos, Heming Wang, *et al.*

Nat Commun. 2019;10(1):5121.

ABSTRACT

Both short and long sleep are associated with an adverse lipid profile, likely through different biological pathways. To provide new insights in the biology of sleep-associated adverse lipid profile, we conducted multi-ancestry genome-wide sleep-SNP interaction analyses on three lipid traits (HDL-c, LDL-c and triglycerides). In the total study sample (discovery + replication) of 126,926 individuals from 5 different ancestry groups, when considering either long or short total sleep time interactions in joint analyses, we identified 49 novel lipid loci, and 10 additional novel lipid loci in a restricted sample of European-ancestry cohorts. In addition, we identified new gene-sleep interactions for known lipid loci such as *LPL* and *PCSK9*. The novel gene-sleep interactions had a modest explained variance in lipid levels: most notable, gene-short-sleep interactions explained 4.25% of the variance in triglyceride concentration. Collectively, these findings contribute to our understanding of the biological mechanisms involved in sleep-associated adverse lipid profiles.

INTRODUCTION

Dyslipidemia is defined as abnormalities in one or more types of lipids, such as high blood LDL-cholesterol (LDL-c) and triglyceride (TG) concentrations and a low HDL-cholesterol (HDL-c) concentration. High LDL-c and TG are well-established modifiable causal risk factors for cardiovascular disease[1-3], and therefore are a primary focus for preventive and therapeutic interventions. Over 300 genetic loci have been identified to be associated with blood lipid concentrations[4-10]. Recent studies showed that only 12.3% of the total variance in lipid concentration is explained by common single nucleotide polymorphisms (SNPs), suggesting additional lipid loci could be uncovered [10]. Some of the unexplained heritability may be due to the presence of gene-environment and gene-gene interactions. Recently, high levels of physical activity were shown to modify the effects of four genetic loci on lipid levels[11], an additional 18 novel lipid loci were identified when considering interactions with high alcohol consumption[12], and 13 novel lipid loci were identified when considering interaction with smoking status[13], suggesting that behavioral factors may interact with genetic loci to influence lipid levels.

Sleep is increasingly recognized as a fundamental behavior that influences a wide range of physiological processes[14]. A large volume of epidemiological research implicates disturbed sleep in the pathogenesis of atherosclerosis[15], and specifically, both a long and short sleep duration are associated with an adverse blood lipid profile[16-26]. However, it is unknown whether sleep duration modifies genetic risk factors for adverse blood lipid profiles. We hypothesize that short and long habitual sleep duration may modify genetic associations with blood lipid levels. The identification of SNPs involved in such interactions will facilitate our understanding of the biological background of sleep-associated adverse lipid profiles.

We investigated gene-sleep duration interaction effects on blood lipid levels as part of the Gene-Lifestyle Interactions Working Group within the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium[27, 28]. To permit the detection of both such sleep-duration-SNP interactions and lipid-SNP associations accounting for total sleep duration, a 2 degree of freedom (2df) test that jointly tests the SNP-main and SNP-interaction effect was applied[29]. Given that there are differences among ancestry groups in sleep behaviours and lipid levels, analysis of data from cohorts of varying ancestries may facilitate the discovery of robust interactions between genetic loci and sleep traits. We focused on short total sleep time (STST; defined as the lower 20% of age- and sex-adjusted sleep duration residuals) and long total sleep time (LTST; defined as the upper 20% of age- and

sex-adjusted sleep duration residuals) as exposures compared to the remaining individuals in the study population, given that each extreme sleep trait has been associated with multiple metabolic and health outcomes[15-26, 30-34]. Within the present study, we report multi-ancestry sleep-by-SNP interaction analyses for blood lipid levels that successfully identified several novel loci for blood lipid traits.

METHODS

Details regarding motivation and methodology of this and other projects of the CHARGE Gene-Lifestyle Interactions Working Group are available elsewhere[35].

Participants

Analyses were performed locally by the different participating studies. Discovery and replication analyses comprised men and women between the age of 18 and 80 years, and were conducted separately for the different contributing (self-defined) ancestry groups, including: European, African, Asian, Hispanic, and Brazilian (discovery analysis only). Descriptions of the different participating studies are described in detail in the **Online Supplementary Materials**, and study-specific characteristics (sizes, trait distribution and data preparation) are presented in **Online Supplementary Tables 1-6**. Every effort was made to include as many studies as possible. The present work was approved by the Institutional Review Board of Washington University in St. Louis and complies with all relevant ethical regulations. Each participating study obtained written informed consent from all participants and received approval from the appropriate local institutional review boards.

Lipid traits

We conducted all analyses on the following lipid traits: HDL-c, LDL-c, and TG. TG and LDL-c concentrations were measured in samples from individuals who had fasted for at least 8 hours. LDL-c could be either directly assayed or derived using the Friedewald equation[36] (the latter being restricted to those with TG ≤ 400 mg/dL). We furthermore corrected LDL-c for the use of lipid-lowering drugs, defined as any use of a statin drug or any unspecified lipid-lowering drug after the year 1994 (when statin use became common in general practice). If LDL-c was directly assayed, the concentration of LDL-c was corrected by dividing the LDL-c concentration by 0.7. If LDL-c was derived using the Friedewald equation, we first divided the concentration of total cholesterol by 0.8 before LDL-c was calculated by the Friedewald equation. Due to the skewed distribution of HDL-c and TG, we ln-transformed the concentration prior to the analyses; no transformation for LDL-c was required. When an individual cohort measured the lipid traits during multiple visits, the visit with the largest available sample and concurrent availability of the sleep questions was selected.

Nocturnal total sleep time

Contributing cohorts collected information on the habitual sleep duration using either a single question such as "on an average night, how long do you sleep?" or as part of a standardized sleep questionnaire (e.g., the Pittsburgh Sleep Quality Index questionnaire[37]). For the present project, we defined both STST and LTST. For the

present project, we defined both STST and LTST. To harmonize the sleep duration data across cohorts from different countries, cultures and participants with different physical characteristics, in whom sleep duration was assessed using various questions, we defined STST and LTST using cohort- specific residuals, adjusting for age and sex. An exception was for AGES and HANDLS cohorts, we used a cohort-specific definition due to limited response categories in relationship to the available question on sleep duration. Instead, we defined STST or LTST based on expert input. Exposure to STST was defined as the lowest 20% of the sex- and age-adjusted sleep-time residuals (coded as "1"). Exposure to LTST was defined as the highest 20% of the sex- and age-adjusted sleep-time residuals (coded as "1"). For both sleep-time definitions, we considered the remaining 80% of the population as being unexposed to either STST or LTST (coded as "0").

Genotype Data

Genotyping was performed by each participating study locally using genotyping arrays from either Illumina (San Diego, CA, USA) or Affymetrix (Santa Clara, CA, USA). Each study conducted imputation using various software programs and with local cleaning thresholds for call rates (usually > 98%) and Hardy-Weinberg equilibrium (usually p-value $<1e^{-5}$). The cosmopolitan reference panel from the 1000 Genomes Project Phase I Integrated Release Version 3 Haplotypes (2010-11 data freeze, 2012-03-14 haplotypes) was specified for imputation. Only SNPs on the autosomal chromosomes with a minor allele frequency of at least 0.01 were considered in the analyses. Specific details of each participating study's genotyping platform and imputation software are described (**Online Supplementary Tables 3** and **6)**.

Stage 1 Analysis (discovery phase)

The discovery phase of the present project included 21 cohorts contributing data from 28 study/ancestry groups, and included up to 62,457 participants of EUR, AFR, ASN, HISP and BR ancestry (**Online Supplementary Tables 1-3**). All cohorts ran statistical models according to a standardized analysis protocol. The main model for this project examined the SNP-main effect and the multiplicative interaction term between the SNP and either LTST or STST:

$$E(Y) = \beta_0 + \beta_E E + \beta_G SNP + \beta_{GE} E * SNP + \beta_C C$$

In which E is the sleep exposure variable (LTST/STST) and C are the (study-specific) covariates, which was similar to what we have done in previous studies [4, 11, 12]. In addition, we examined the SNP-main effect (without incorporating LTST/STST) and the SNP-main effect stratified by the exposure:

$E(Y) = \beta_o + \beta_G . SNP + \beta_C . C$

All models were performed for each lipid trait and separately for the different ancestry groups. Consequently, per ancestry group, we requested a total of 7 GWA analyses per lipid trait. All models were adjusted for age, sex, field center (if required), and the first principal components to correct for population stratification. The number of principal components included in the model was chosen according to cohort-specific preferences (ranging from 0 to 10). All studies were asked to provide the effect estimates (SNP-main and -interaction effect) with accompanying robust estimates of the standard error for all requested models. A robust estimate of the covariance between the main and interaction effects was also provided. To obtain robust estimates of covariance matrices and standard errors, studies with unrelated participants used R packages such as either sandwich[38, 39] or ProbABEL[40]. Studies including related individuals used either generalized estimating equations (R package geepack[41]) or linear mixed models (GenABEL[42], MMAP, or R package sandwich[38, 39]). Sample code provided to studies to generate these data has been previously published (see **Online Supplementary Materials** [35]).

Upon completion of the analyses by local institution, all summary data were stored centrally for further processing and meta-analyses. We performed estimative quality control (QC) using the R-based package EasyQC[43] (www.genepi-regensburg.de/easyqc) at the study level (examining the results of each study individually), and subsequently at the ancestry level (after combining all ancestry-specific cohorts using meta-analyses). Study-level QC consisted of excluding all SNPs with MAF < 0.01, harmonization of alleles, comparison of allele frequencies with ancestry-appropriate 1000 Genomes reference data, and harmonization of all SNPids to a standardized nomenclature according to chromosome and position. Ancestry-level QC included the compilation of summary statistics on all effect estimates, standard errors and p-values across studies to identify potential outliers, and production of SE-N and QQ plots to identify analytical problems (such as improper trait transformations)[44].

Prior to the ancestry-specific meta-analyses, we excluded the following SNPs from the cohort-level data files: all SNPs with an imputation quality < 0.5, and all SNPs with a minor allele count in the exposed group (LTST or STST equals "1") x imputation quality of less than 20. SNPs in the European-ancestry and multi-ancestry analyses had to be present in at least 3 cohorts and 5000 participants. Due to the limited sample size of the non-European ancestries (either discovery or replication), we did not take into account this filter in those ancestry-level meta-analyses.

Meta-analyses were conducted for all models using the inverse variance-weighted fixed effects method as implemented in METAL[45] (http://genome.sph.umich.edu/wiki/METAL). We evaluated both a 1 degree (1df) of freedom test of interaction effect and a 2 degree of freedom (2df) joint test of main and interaction effects, following previously published methods[29]. A 1df Wald test was used to evaluate the 1df interaction, as well as the main effect in models without an interaction term. A 2df Chi-squared test was used to jointly test the effects of both the variant and the variant x LTST/STST interaction[46]. Meta-analyses were conducted within each ancestry separately. Multi-ancestry meta-analyses were conducted on all ancestry-specific meta-analyses. Genomic control correction was applied on all cohorts incorporated in the ancestry-level meta-analyses as well as on the final meta-analyses for the publication. From this effort, we selected all SNPs associated with any of the lipid traits with $p \leq 5x10^{-7}$ for replication in the Stage 2 analysis. This cut-off was selected to minimize false-negative results.

Stage 2 Analysis (replication phase)
All SNPs selected in Stage 1 for replication were evaluated in the interaction model in up to 18 cohorts contributing data from 20 study groups totalling up to 64,469 individuals (**Online Supplementary Tables 4-6**). As we had a limited number of individuals from non-European ancestry in the replication analyses, we did not consider an the non-European ancestries separately and only focussed on a European-ancestry and multi-ancestry analysis.

Study- and ancestry-level QC was carried out as in stage 1. In contrast to stage 1, no additional filters were included for the number of studies or individuals contributing data to stage 2 meta-analyses, as these filters were implemented to reduce the probability of false positives, and were less relevant in stage 2. Stage 2 SNPs were evaluated in all ancestry groups and for all traits, no matter what specific meta-analysis met the p-value threshold in the stage 1 analysis. We did not apply genomic control to any of the Stage 2 analyses given the expectation of association.

An additional meta-analysis was performed combining the Stage 1 and 2 meta-analyses. SNPs (irrespective of being known or novel) were considered to be replicated when Stage 1 p-values $<5x10^{-7}$, Stage 2 p-value <0.05 with a similar direction of effect as in the discovery meta-analysis, and Stage 1+2 p-value $<5x10^{-8}$. Replicated SNPs were subsequently used in different bioinformatics tools for further processing. In addition, 1 df p-values (SNP-sleep interaction effect only) of the lead SNPs of both the replicated known and novel genetic loci were calculated to explore whether genetic variant were specifically driven by SNP-main or SNP-interaction effects. Based on the total number of lead SNPs across all analyses, we performed correction using the false-discovery rate to quantify statistical significance [47].

Bioinformatics

Replicated SNPs were first processed using the online tool FUMA[48] to identify independent lead SNPs and to perform gene mapping. From the SNP that has a $P_{joint} < 5x10^{-8}$, we determined lead SNPs that were independent from each other at $R^2 < 0.1$ using the 1000G Phase 3 EUR as a reference panel population. Independent lead SNPs with a physical distance >1 mB from a known locus were considered as novel. Regional plots of the novel loci from the European- and multi-ancestry meta-analyses were made using the online LocusZoom tool [49]. The explained variance of the newly identified genetic independent variants was calculated based on the summary statistics of the combined analysis of Stage 1 and 2 using the R-based VarExp package, which has been previously validated to provide similar results to individual participant data [50]. This package calculates the variance explained on the basis of the combined (joint) SNP-main and SNP-interaction effect. Differential expression analyses of the lead SNPs in the newly identified genetic loci was performed using GTEx (https://gtexportal.org/home/) [51, 52].

Look-ups of novel loci in publicly available databases

Newly identified genetic loci for the three lipid traits were further explored in the GWAS catalogue (https://www.ebi.ac.uk/gwas/) to investigate the role of the newly identified mapped genes in other traits. Furthermore, we extracted the lead SNPs from the novel genetic loci from publically available GWAS data from the UK Biobank (http://www.nealelab.is/uk-biobank/) for different questionnaire-based sleep phenotypes, notably "daytime snoozing/sleeping (narcolepsy)", "getting up in the morning", "morning/evening person (chronotype)", "nap during the day", "sleep duration", "sleeplessness/insomnia", and "snoring". Analyses on these phenotypes were generally done using continuous outcomes; the variable "sleep duration" was expressed in hours of total sleep per day. GWAS in the UK Biobank were done in European-ancestry individuals only (N up to 337,074). We furthermore extracted the newly identified lead SNPs from the GWAS analyses done on accelerometer-based sleep variables, which was done in European-ancestry individuals from the UK Biobank (N = 85,670) [53]. In addition, we extracted the newly identified lead SNPs from publically available summary-statistics data on coronary artery disease of the CARDIoGRAMplusC4D consortium, which included 60,801 cases of coronary artery disease and 123,504 controls [54].

RESULTS

Study population

Discovery analyses were performed in up to 62,457 individuals (40,041 European-ancestry, 14,908 African-ancestry, 4,460 Hispanic-ancestry, 2,379 Asian-ancestry, and 669 Brazilian/mixed-ancestry individuals) from 21 studies spanning 5 different ancestry groups (**Online Supplementary Tables 1-3**). Of the total discovery analysis, 13,046 (20.9%) individuals were classified as short sleepers and 12,317 (19.7%) individuals as long sleepers. Replication analyses were performed in up to 64,469 individuals (47,612 European-ancestry, 12,578 Hispanic-ancestry, 3,133 Asian-ancestry, and 1,146 African-ancestry individuals) from 19 studies spanning 4 different ancestry groups (**Online Supplementary Tables 4-6**). Of the total replication analysis, 12,952 (20.1%) individuals were classified as short sleepers and 12,834 (19.9%) individuals as long sleepers.

Identification of novel loci for lipid traits when considering potential interaction with long or short total sleep time

An overview of the multi-ancestry analyses process for both STST and LTST is presented in **Figure 1**. QQ plots of the combined multi-ancestry and European meta-analysis of the discovery and replication analysis are presented in **Online Supplementary Figures 1 and 2**. Lambda values ranged between 1.023 and 1.055 (trans-ancestry meta-analysis) before the second genomic control and were all 1 after second genomic control correction. In the combined discovery and replication meta-analyses comprising all contributing ancestry groups, we found that many SNPs replicated for the lipid traits (P_{joint} in replication < 0.05 with similar directions of effect as in the discovery analyses and P_{joint} in combined discovery and replication analysis < $5x10^{-8}$). Notably, we replicated 2,395 and 2,576 SNPs for HDL-c, 2,012 and 2,074 SNPs for LDL-c, and 2,643 and 2,734 SNPs for TG in the joint model with LTST and STST respectively.

Most of the replicated SNPs were mapped to known loci (**Online Supplementary Tables 7 and 8**). We looked at the 427 known lipid SNPs (**Online Supplementary Table 9**), but these did not reveal significant 1df interactions with either LTST or STST. In addition, we identified lead SNPs mapping to novel regions when considering the joint model with potential interaction for either STST or LTST (>1 Mb distance from known locus). Ultimately, in the multi-ancestry analysis, we identified 14 novel loci for HDL-c, 12 novel loci for LDL-c, and 23 novel loci for TG (R^2 < 0.1; **Figure 2**). Of these, 7 loci for HDL-c, 4 loci for LDL-c and 7 loci for TG were identified after considering an interaction with LTST (**Online Supplementary Table 10**). Furthermore, 7 loci for HDL-c, 8 loci for LDL-c and 16 loci for TG were identified when considering an interaction with STST (**Online Supplementary Table 11**). Importantly, none of the novel loci for the three lipid traits identified through

LTST were identified in the analyses with STST, and *vice versa*. Furthermore, the novel lipid loci were specific to a single lipid trait. Regional plots of the newly identified loci from the multi-ancestry analyses are presented in **Online Supplementary Figures 3-8**. Some of the novel SNPs identified through modelling a short or long sleep duration interaction (1 df) also showed suggestive evidence of association with lipid levels in the joint model (2 df test). However, this pattern suggested a main effect that appeared once sleep duration was adjusted for rather than an effect due to an interaction between sleep and the novel SNPs (**Online Supplementary Table 10-11**).

Using the R-based VarExp package[50], we calculated the explained variance based on the summary statistics of the combined discovery and replication analysis. Collectively, novel lead SNPs identified with LTST explained 0.97% of the total HDL-c variation, 0.13% of the total LDL-c variation, and 1.51% of the total TG variation. In addition, novel lead SNPs identified with STST explained 1.00% of the total HDL-c variation, 0.38% of the total LDL-c variation, and 4.25% of the total TG variation.

In the analyses restricted to European-ancestry individuals (overview **Online Supplementary Figure 9)**, we identified 10 additional novel loci (7 novel loci with LTST and 3 novel loci with STST; **Online Supplementary Figure 10**), which were not identified in the multi-ancestry analyses. Of these, we identified 4 loci for HDL-c, 2 loci for LDL-c, and 1 locus for TG with LTST (**Online Supplementary Table 12**). In addition, we identified 1 locus for HDL-c and 2 for TG with STST (**Online Supplementary Table 13**). Again, we observed no overlapping findings between the two sleep exposures and the three lipid traits. Regional plots of the identified novel loci were presented in **Online Supplementary Figures 11-15**.

Sleep × SNP interactions in identified novel and known lipid loci in the combined sample of discovery and replication studies

Based on a total of 402 lead SNPs in known and novel regions for both exposures and the three lipid traits that were identified using the joint test in the combined sample of discovery and replication studies, we subsequently explored the extent the effects were driven by 1df interaction with the sleep exposure trait being tested[29]. We corrected the 1df interaction p-value for multiple testing using the false-discovery rate [47] considering all 402 lead SNPs for the present investigation, which was equivalent in our study to a 1df interaction p-value $<5x10^{-4}$. Overall, in the multi-ancestry meta-analyses, the novel lipid loci show clearly stronger interaction with either LTST or STST than the loci defined as known (**Figure 3**). The majority of the newly identified lead variants were generally common, with minor allele frequencies (MAF) mostly > 0.2, and SNP × sleep interaction effects were not specifically identified in lower frequency SNPs (e.g., MAF<0.05).

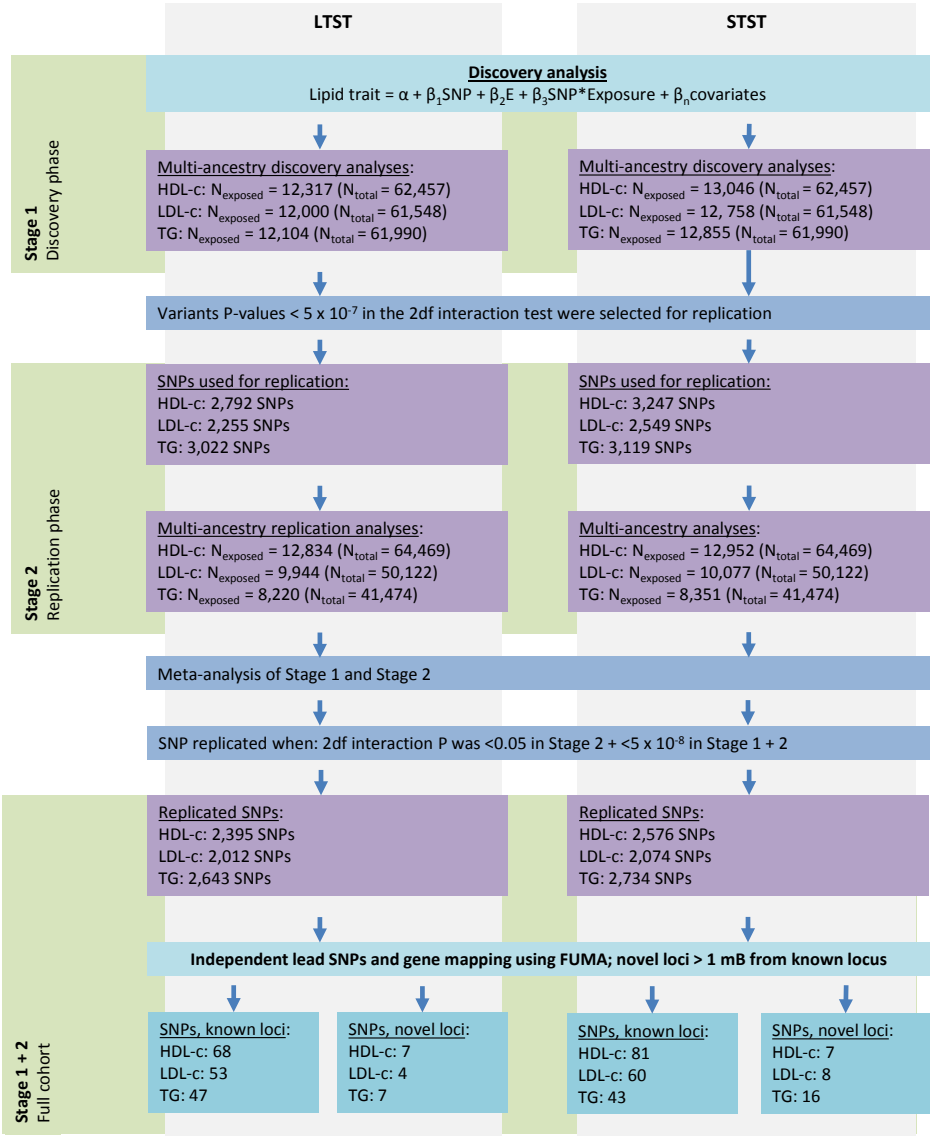


Figure 1. *Project overview and SNP selection in the multi-ancestry analyses.* Project overview of the multi-ancestry analyses of how the new lipid loci were identified in the present project. Replicated variants had to have a $P_{joint} < 5x10^{-7}$ in stage 1, $P_{joint} < 0.05$ with similar direction of effect in stage 2, and $P_{joint} < 5x10^{-8}$ in the combined stage 1 + 2.

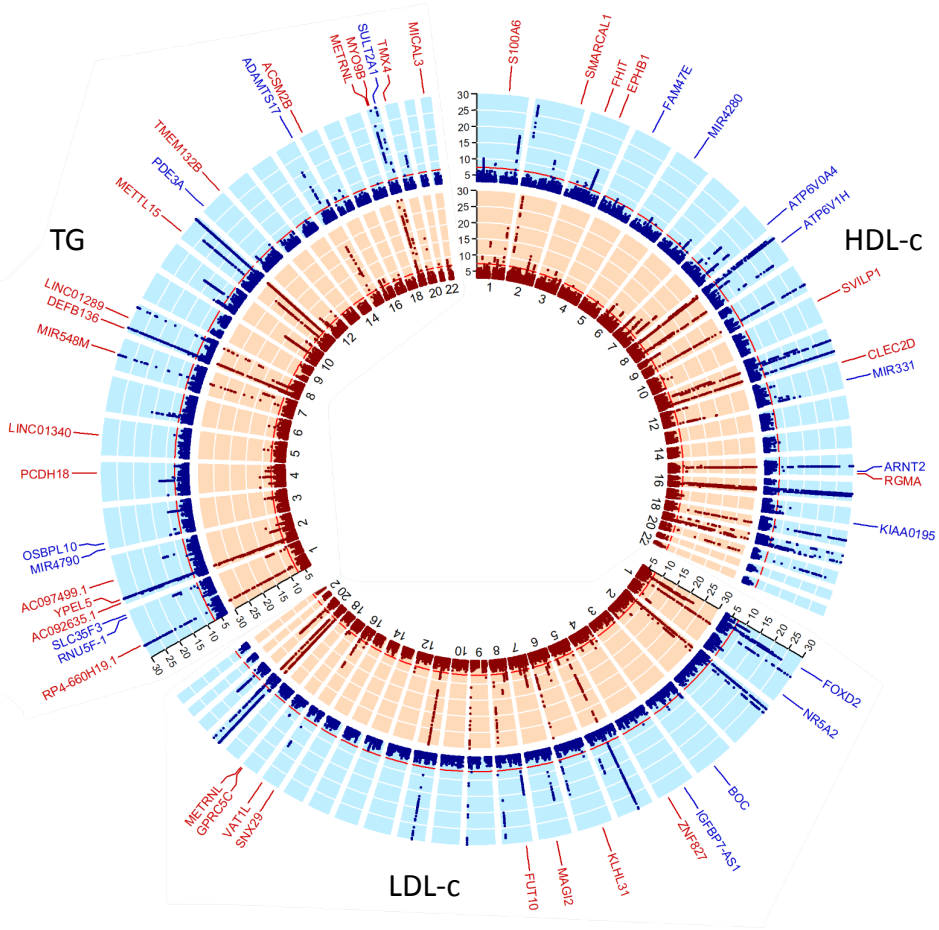


Figure 2. *Circular –log(p-value) plots of the multi-ancestry sleep-SNP interactions analyses for the three lipid traits.* Plot visualizes the –log(p_{joints}) for HDL-c, LDL-c and TG per chromosome. In red (inner circle) are the –log(p-value) plots for the analyses taking into account potential interaction with short total sleep time. In blue (outer circle) are the –log(p-value plots for the analyses taking into account potential interaction with long total sleep time. Loci defined as novel and replicated are labeled. Replicated variants had to have a $P_{joint} < 5x10^{-7}$ in stage 1, $P_{joint} < 0.05$ in stage 2, and $P_{joint} < 5x10^{-8}$ in the combined stage 1 + 2. Labeled gene names in red were identified in the STST analysis; Labeled gene names in blue were identified in the LTST analysis. All –log(p_{joints}) > 30 were truncated to 30 for visualization purposes only. The unlabeled regions with $P_{joint} < 5x10^{-8}$ were in known loci. Figure prepared using the R package circlize[55].

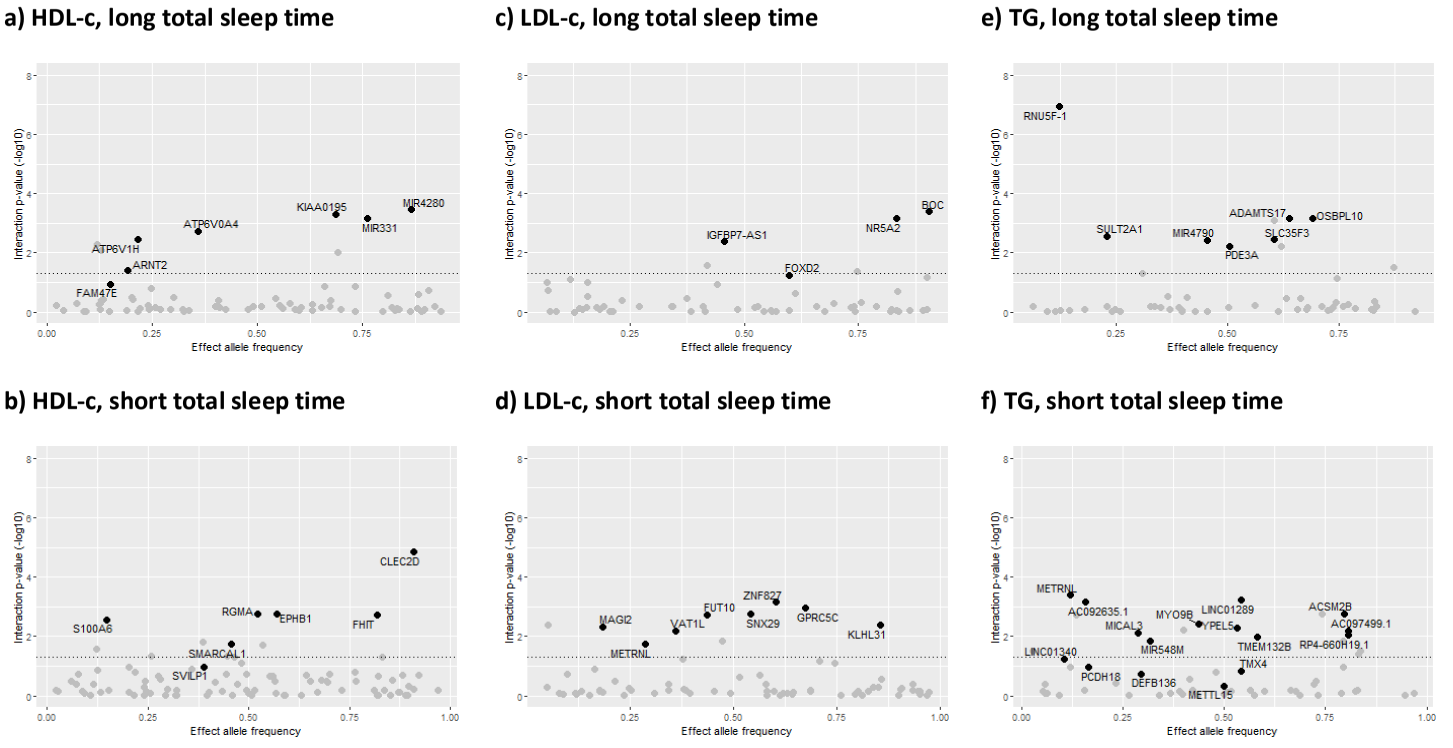


Figure 3. *Sleep-interactions on lipid traits in novel and known loci in the multi-ancestry meta-analyses.* Plot displaying the –log(p-value) of the 1df interaction between the SNP and either LTST or STST on the lipid trait after correction for multiple testing using false-discovery rate against the allele frequency of the effect allele. Dotted horizontal line resembles the cut-off for the 1df interaction p-value$_{FDR}$ <0.05 after correction for multiple testing using false-discovery rate. In black are the novel loci for lipid traits; in grey are the identified lead SNPs mapped within a 1 Mb physical distance from a known lipid locus. Visualization of the plots was performed using the R package ggplot2 [56].

Out of the 7 novel HDL-c loci identified in the joint model with LTST, 6 had a 1df interaction p-value$_{FDR}$ < 0.05, notably lead SNPs mapped to *ATP6V1H*, *ARTN2, ATP6V0A4*, *KIAA0195*, *MIR331*, and *MIR4280.* Based on exposure-stratified analyses in the meta-analysis of the discovery cohorts, we further explored the effect sizes per exposure group. The lead SNPs that showed significant sleep × SNP interaction also showed effect estimates that modestly differed between LTST exposure groups (**Online Supplementary Table 14**). Interestingly, two lead SNPs near known HDL-c loci showed a 1df interaction p-value$_{FDR}$ < 0.05, including SNPs near *CETP* and *LIPC* (**Online Supplementary Table 7**). Out of the 7 novel HDL-c loci identified in the joint model with STST, we found 6 loci with a 1df interaction p-value$_{FDR}$ < 0.05, notably lead SNPs mapped to *S1000A6*, *SMARCAL1, RGMA*, *EPHB1*, *FHIT* and *CLEC2D.* Again, their effect estimates differed between the exposure groups in the discovery multi-ancestry meta-analysis (**Online Supplementary Table 15; Figure 4**). Some lead SNPs near known HDL-c loci showed evidence of a 1df interaction with STST (e.g., *MADD* and *LPL;* p-value$_{FDR}$ < 0.05).

For all four novel lead SNPs associated with LDL-c when considering LTST, we observed a 1df interaction p-value$_{FDR}$ < 0.05; notably, lead SNPs mapped to *IGFBP7-AS1*, *FOXD2*, *NR5A2* and *BOC*. One locus that mapped within a 1 Mb physical distance from known LDL-c locus (*PCSK9*) showed 1df interaction with LTST (**Online Supplementary Table 7**). Similarly, all 8 independent novel lead SNPs associated with LDL-c when considering STST, had a 1df interaction p-value$_{FDR}$ <0.05; notably, lead SNPs mapped to *MAGI2*, *METRNL*, *VAT1L*, *FUT10*, *SNX29*, *ZNF827*, *GPRC5C* and *KLHL31*. In addition, of the known LDL-c loci, lead SNPs mapped within a physical distance of 1 Mb of *APOB* and *SLC22A1* showed a 1df interaction p-value$_{FDR}$ < 0.05 (**Online Supplementary Table 8**). For both analyses, we observed that effect estimates differed between the LTST and STST exposure groups in the multi-ancestry discovery analysis (**Online Supplementary Table 14 and 15; Figure 4**).

All 7 independent novel lead SNPs associated with TG when considering LTST, had a 1df interaction p-value$_{FDR}$ <0.05; notably, lead SNPs mapped to *RNU5F-1*, *SULT2A1*, *MIR4790*, *PDE3A, SLC35F3*, *ADAMTS17* and *OSBPL10*. In addition, we found some evidence for long sleep-SNP interaction in lead SNPs near known TG loci, including lead SNPs near *AKR1C4* and *NAT2* (**Online Supplementary Table 7**). Of the 16 novel lead SNPs associated with TG when considering STST, we observed 12 lead SNPs with a 1df interaction p-value <5x10^{-4} (p-value$_{FDR}$ < 0.05), including lead SNPs mapped to *LINC0140, METRNL*, *AC092635.1*, *MICAL3*, *MIR548M*, *MYO9B*, *YPEL5*, *LINC01289*, *TMEM132B*, *ACSM2B, AC097499.1* and *RP4-660H19.1.* In addition, we observed some lead SNPs within 1 Mb physical distance from known TG loci, such as *MMP3* and *NECTIN2* (**Online Supplementary Table 8**). For both LTST and STST analyses, we again observed differing effects dependent on the exposure group in the discovery meta-analyses (**Online Supplementary Table 14 and 15; Figure 4**).

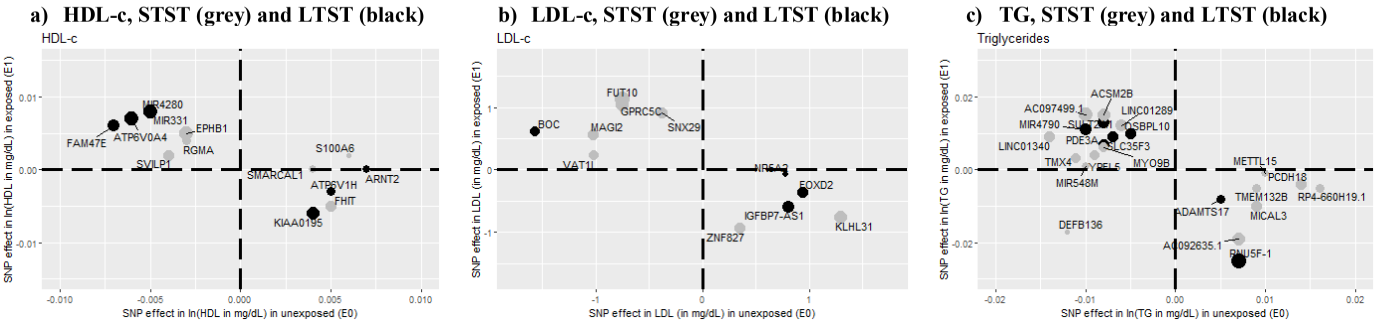


Figure 4. *Comparison of SNP-main effects stratified by exposure in the multi-ancestry discovery meta-analyses.* X-axis displays the effect sizes of the novel lead SNPs as observed in the meta-analyses of the unexposed individuals (LTST = "0", STST = "0"). Y-axis displays the effect sizes of the novel lead SNPs as observed in the meta-analyses of the exposed individuals (LTST = "1", STST = "1"). In black are the novel lead SNPs identified with LTST; in grey are the novel lead SNPs identified with STST. Sizes of the dots were weighted to the difference observed between exposed and unexposed. Visualization of the plots was performed using the R package ggplot2 [56].

Look-ups and bioinformatics analyses

Based on the lead SNPs mapped to novel loci, we conducted a look-up in GWAS summary statistics data on different questionnaire-based sleep phenotypes from up to 337,074 European-ancestry individuals of the UK Biobank (**Online Supplementary Table 15**). We only observed the TG-identified rs7924896 (*METTL15*) to be associated with snoring (p-value = $1e^{-5}$) after correction for a total of 343 explored SNP-sleep associations (7 sleep phenotypes × 49 genes; 10 SNPs were unavailable; threshold for significance = $1.46e^{-4}$). Furthermore, we did not observe that any of the newly identified SNPs was associated with accelerometer-based sleep traits (**Online Supplementary Table 17**). In general, we did not find substantial evidence that the identified novel lead SNPs were associated with coronary artery disease in the CARGIoGRAMplusC4D consortium (**Online Supplementary Table 18**).

Newly identified lipid loci were further explored in the GWAS catalogue (**Online Supplementary Table 19**). Several of the mapped genes of our novel lead SNPs have previously been identified with multiple other traits, such as body mass index (*FHIT*, *KLH31*, *ADAMTS17*, *MAGI2*), mental health (*FHIT* [autism/schizophrenia, depression], *SNX13* [cognition]), gamma-glutamyltransferase (*ZNF827*, *MICAL3*), and inflammatory processes (*ZNF827*, *NR5A2*).

We additionally investigated differential expression of the novel lead SNPs using data from multiple tissues from the GTEx consortium [51, 52] (**Online Supplementary Table 20**). Lead SNPs were frequently associated with mRNA expression levels of the mapped gene and with trans-eQTLs. For example, rs429921 (mapped to *VAT1L*) was associated with differential mRNA expression levels of *CLEC3A* and *WWOX,* which are located more upstream on chromosome 16 (**Online Supplementary Figure 6**). rs3826692 (mapped to *MYO9B*) was specifically associated with differential expression of the nearby *USE1* gene. Identified SNPs were frequently associated with differential expression in the arteries. For example, rs6501801 (*KIAA0195*) was associated with differential expression in arteries at different locations. Several of the other identified SNPs showed differential expression in multiple tissues, including the gastrointestinal tract, (subcutaneous/visceral) adipose tissue, brain, heart, muscle, lung, liver, nervous system, skin, spleen, testis, thyroid and whole blood.

DISCUSSION

We investigated SNP-sleep interactions in a large, multi-ancestry, meta-analysis of blood lipid levels. Given the growing evidence that sleep influences metabolism [57-62], at least in part through effects on gene expression, we hypothesized that short/long habitual sleep duration may modify the effects of genetic loci on lipid levels. In a total study population of 126,926 individuals from 5 different ancestry groups, we identified 49 novel lipid loci when considering either long or short total sleep time in the analyses. An additional 10 novel lipid loci were identified in analyses in Europeans only. Of the newly identified loci, most loci at least in part were driven by differing effects in short/long sleepers compared to the rest of the study population. Multiple of the novel genes identified by our efforts have been previously identified in relation to adiposity, hepatic function, inflammation or psychosocial traits, collectively contributing to potential biological mechanisms involved in sleep-associated adverse lipid profile.

In addition to the over 300 genetic loci that already have been identified in relation to blood lipid concentrations in different efforts [4-10], we identified 49 additional loci associated with either HDL-c, LDL-c or TG in our multi-ancestry analysis. While for some of the novel SNPs had no neighboring SNPs in high LD (e.g., rs7799249; mapped to *ATP6V0A4*), our applied filters (e.g., imputation quality > 0.5) would suggest that the chance of invalidity of the findings is negligible. Furthermore, in the case of rs7799249, no SNPs in high LD are known in individuals from different ancestries [63]. Considering the novel TG loci identified by considering interactions with total sleep duration explain an additional 4.25% and 1.51% of the total variation in TG concentrations, for STST and LTST, respectively. While the additionally explained variance for LDL-c (0.38% and 0.13%) and HDL-c (1.00% and 0.97%) was low/modest, the novel lead SNPs identified include genes known to be associated with adiposity, inflammatory disorders, cognition, and liver function, thus identifying pathways by which sleep disturbances may influence lipid biology.

Across multiple populations, both short and long sleep duration have been associated with cardiovascular disease and diabetes [64]. There are numerous likely mechanisms for these associations. Experimental sleep loss results in inflammation, cellular stress in brain and peripheral tissues, and altered expression of genes associated with oxidative stress [65, 66]. The impact of long sleep on metabolism is less well understood than the effect of short sleep, and multiple of the associations seem to overlap with short sleep as well. Long sleep duration is associated with decreased energy expenditure, increased sedentary time, depressed mood, and obesity-related factors associated with inflammation and a pro-thrombotic state [67], as well as with higher C-reactive

protein and interleukin-6 concentrations [68]. However, studies that adjusted for multiple confounders, including obesity, depression, and physical activity, showed that long sleep remained a significant predictor of adverse cardiovascular outcomes [64, 69]. Therefore, the adverse effects of long sleep also may partly reflect altered sleep-wake rhythms and chronodisruption resulting from misalignment between the internal biological clock with timing of sleep and other behaviours that track with sleep, such as timing of food intake, activity, and light exposure [70]. Altered sleep-wake and circadian rhythms influence glucocorticoid signalling and autonomic nervous system excitation patterns across the day [59], which can influence the phase of gene expression. These inputs appear to be particularly relevant for genes controlling lipid biosynthesis, absorption and degradation, many of which are rhythmically regulated and under circadian control [71]. Moreover, the molecular circadian clock acts as a rate limiting step in cholesterol and bile synthesis, supporting the potential importance of circadian disruption in lipid biology [72]. Collectively, these data suggest different biological mechanisms involved in short and long sleep-associated adverse lipid profiles.

Consistent with different hypothesized physiological effects of short and long sleep, we observed no overlap in the novel loci that were identified by modelling interactions with short or long sleep duration. Novel lipid loci that were identified after considering STST include *FHIT*, *MAGI2* and *KLH3,* which have been previously associated with body mass index (BMI) [73-79]. Interestingly, although not genome-wide significant, variation in *MAGI2* has been associated with sleep duration [80], however, we did not find evidence for an association with rs10244093 in *MAGI2* with any sleep phenotype in the UK Biobank sample. Variants in *MICAL3* and *ZNF827,* that were also identified after considering STST, have been associated with serum liver enzymes gamma-glutamyl transferase measurement and/or aspartate aminotransferase levels [81, 82], which have been implicated in cardiometabolic disturbances [83-86] and associated with prolonged work hours (which often results in short or irregular sleep) [87]. Other loci identified through interactions with STST were in genes previously associated with neurocognitive and neuropsychiatric conditions, possibly reflecting associations mediated by heightened levels of cortisol and sympathetic activity that frequently accompany short sleep.

In relation to LTST, the novel genes identified have been previously related to inflammation-driven diseases of the intestine, blood pressure and blood count measurements, including traits influenced by circadian rhythms [88, 89]. However, no novel identified lipid loci with LTST directly interacted with genes involved in the central circadian clock (e.g., *PER2*, *CRY2* and *CLOCK*) in the KEGG pathways database [90]. The novel loci *NR5A2* and *SLC35F3* have been associated with inflammation-driven diseases of the intestine [91, 92]. Ulcerative colitis, an inflammatory bowel disease, has been associated with

both longer sleep duration [93] and circadian disruption [88]. *ARNT2*, also identified via a LTST interaction, heterodimerizes with transcriptional factors implicated in homeostasis and environmental stress responses [94, 95]. A linkage association study has reported nominal association of this gene with lipids in a Caribbean Hispanic population [96].

We identified a number of additional genetic lead SNPs in the meta-analyses performed in European-Americans only. For example, we identified rs3938236 mapped to *SPRED1* to be associated with HDL-c after accounting for potential interaction with LTST. Interestingly, this gene has been previously associated with hypersomnia in Caucasian and Japanese populations [97], but was not identified in our larger multi-ancestry analysis, possibly due to cultural differences in sleep behaviours [98].

We additionally found evidence, amongst others, in the known lipid loci *APOB*, *PCSK9* and *LPL* for interaction with either short or long sleep. Associations have been observed previously between short sleep and ApoB concentrations, have been observed previously [99]. LPL expression has been shown to follows a diurnal rhythm in several metabolic organs [61, 100], and disturbing sleeping pattern by altered light exposure can lower LPL activity, at least in brown adipose tissue [61]. Similar effects of sleep on hepatic secretion of ApoB and PCSK9 may be expected. Indeed, in humans PCSK9 has a diurnal rhythm synchronous with hepatic cholesterol synthesis [101]. Although the interaction effects we observed were rather weak, the supporting evidence from the literature suggests that sleep potentially modifies the effect of some of the well-known lipid regulators that are also targets for therapeutic interventions.

Some of the novel lipid loci have been previously associated with traits related to sleep. For example, *MAGI2* and *MYO9B* [80] have been suggestively associated with sleep duration and quality, respectively. Genetic variation in *TMEM132B* has been associated with excessive daytime sleepiness [102], and *EPHB1* has been associated with self-reported chronotype [103]. These findings suggest some shared genetic component of lipid regulation and sleep biology. However, with the exception of the *METTL15*-mapped rs7924896 variant in relation to snoring, none of the lead SNPs mapped to the new lipid loci were associated with any of the investigated sleep phenotypes in the UK Biobank population, suggesting no or minimal shared component in sleep and lipid biology but rather that sleep duration specifically modifies the effect of the variant on the lipid traits.

The present study was predominantly comprised of individuals of European ancestry, despite our efforts to include as many studies of diverse ancestries as possible. For this reason, additional efforts are required to specifically study gene-sleep interactions in those of African, Asian and Hispanic ancestry once more data becomes available.

In line, we identified several loci that were identified only in the European-ancestry analysis, and not in the multi-ancestry analysis, suggestion that there might be ancestry-specific effects. The multi-ancestry analysis highlighted the genetic regions that are more likely to play a role in sleep-associated adverse lipid profiles across ancestries. In addition, our study used questionnaire-based data on sleep duration. Although the use of questionnaires likely increased measurement error and decreased statistical power, questionnaire-based assessments of sleep duration have provided important epidemiological data, including the identification of genetic variants for sleep traits in genome-wide association studies[102]. Identified variants for sleep traits have been recently successfully validated using accelerometer data [104], although the overall genetic correlation with accelerometer-based sleep duration was shown to be low [53]. Moreover, observational studies showed only a modest correlation between the phenotypes [105], which suggest that each approach characterizes somewhat different phenotypes. At this time, we did not have sufficient data to evaluate other measures of sleep duration such as polysomnography or accelerometery; however, a more comprehensive characterization and circadian traits likely will refine our understanding of the interaction of these fundamental phenotypes and lipid biology.

In summary, the gene-sleep interaction efforts described in the present multi-ancestry study identified many novel genetic loci associated with either HDL-c, LDL-c or triglycerides levels. Multiple of the novel genetic loci were driven by interactions with either short or long sleep duration, and were mapped to genes also associated with adiposity, inflammatory or neuropsychiatric traits. Collectively, the results highlight the interactions between extreme sleep-wake exposures and lipid biology.

Acknowledgments

This project was supported by a grant from the US National Heart, Lung, and Blood Institute (NHLBI) of the National Institutes of Health (R01HL118305). This research was supported in part by the Intramural Research Program of the National Human Genome Research Institute, National Institutes of Health. Tuomas O. Kilpeläinen was supported by the Danish Council for Independent Research (DFF – 6110-00183) and the Novo Nordisk Foundation (NNF18CC0034900, NNF17OC0026848 and NNF15CC0018486). Diana van Heemst was supported by the European Commission funded project HUMAN (Health-2013-INNOVATION-1-602757). Susan Redline was supported in part by NIH R35HL135818 and HL11338. Study-specific acknowledgments can be found in the **Supplementary Materials**. Data on coronary artery disease have been contributed by the Myocardial Infarction Genetics and CARDIoGRAM investigators and have been downloaded from www.CARDIOGRAMPLUSC4D.ORG.

Conflict of interest statement

DOMK is a part-time research consultant for Metabolon, Inc. HJG has received travel grants and speakers honoraria from Fresenius Medical Care, Neuraxpharm and Janssen Cilag. HJG has received research funding from the German Research Foundation (DFG), the German Ministry of Education and Research (BMBF), the DAMP Foundation, Fresenius Medical Care, the EU "Joint Programme Neurodegenerative Disorders (JPND) and the European Social Fund (ESF)". SA reports employment and stock options with 23andMe, Inc.

REFERENCES

1. Holmes MV, Asselbergs FW, Palmer TM, Drenos F, Lanktree MB, Nelson CP, et al. Mendelian randomization of blood lipids for coronary heart disease. Eur Heart J. 2015;36(9):539-50.
2. Ference BA, Yoo W, Alesh I, Mahajan N, Mirowska KK, Mewada A, et al. Effect of long-term exposure to lower low-density lipoprotein cholesterol beginning early in life on the risk of coronary heart disease: a Mendelian randomization analysis. J Am Coll Cardiol. 2012;60(25):2631-9.
3. Voight BF, Peloso GM, Orho-Melander M, Frikke-Schmidt R, Barbalic M, Jensen MK, et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. Lancet. 2012;380(9841):572-80.
4. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. Nature genetics. 2013;45(11):1274-83.
5. Do R, Willer CJ, Schmidt EM, Sengupta S, Gao C, Peloso GM, et al. Common variants associated with plasma triglycerides and risk for coronary artery disease. Nature genetics. 2013;45(11):1345-52.
6. Peloso GM, Auer PL, Bis JC, Voorman A, Morrison AC, Stitziel NO, et al. Association of low-frequency and rare coding-sequence variants with blood lipids and coronary heart disease in 56,000 whites and blacks. Am J Hum Genet. 2014;94(2):223-32.
7. Spracklen CN, Chen P, Kim YJ, Wang X, Cai H, Li S, et al. Association analyses of East Asian individuals and trans-ancestry analyses with European individuals reveal new loci associated with cholesterol and triglyceride levels. Human Molecular Genetics. 2017;26(9):1770-84.
8. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature. 2010;466(7307):707-13.
9. Kathiresan S, Melander O, Anevski D, Guiducci C, Burtt NP, Roos C, et al. Polymorphisms associated with cholesterol and risk of cardiovascular events. N Engl J Med. 2008;358(12):1240-9.
10. Klarin D, Damrauer SM, Cho K, Sun YV, Teslovich TM, Honerlaw J, et al. Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. Nature genetics. 2018;50(11):1514-23.
11. Kilpelainen TO, Bentley AR, Noordam R, Sung YJ, Schwander K, Winkler TW, et al. Multi-ancestry study of blood lipid levels identifies four loci interacting with physical activity. Nat Commun. 2019;10(1):376.
12. de Vries PS, Brown MR, Bentley AR, Sung YJ, Winkler TW, Ntalla I, et al. Multi-Ancestry Genome-Wide Association Study of Lipid Levels Incorporating Gene-Alcohol Interactions. Am J Epidemiol. 2019.
13. Bentley AR, Sung YJ, Brown MJ, Winkler TW, Kraja AT, Ntalla I, et al. Multi-ancestry genome-wide smoking interaction study of 387,272 individuals identifies novel lipid loci. Nature Genetics (accepted for publication). 2019.
14. Tobaldini E, Costantino G, Solbiati M, Cogliati C, Kara T, Nobili L, et al. Sleep, sleep deprivation, autonomic nervous system and cardiovascular diseases. Neurosci Biobehav Rev. 2017;74(Pt B):321-9.
15. Tobaldini E, Pecis M, Montano N. Effects of acute and chronic sleep deprivation on cardiovascular regulation. Arch Ital Biol. 2014;152(2-3):103-10.

16. Ford ES. Habitual sleep duration and predicted 10-year cardiovascular risk using the pooled cohort risk equations among US adults. J Am Heart Assoc. 2014;3(6):e001454.
17. Aggarwal S, Loomba RS, Arora RR, Molnar J. Associations between sleep duration and prevalence of cardiovascular events. Clin Cardiol. 2013;36(11):671-6.
18. Wu Y, Zhai L, Zhang D. Sleep duration and obesity among adults: a meta-analysis of prospective studies. Sleep Med. 2014;15(12):1456-62.
19. Xi B, He D, Zhang M, Xue J, Zhou D. Short sleep duration predicts risk of metabolic syndrome: a systematic review and meta-analysis. Sleep Med Rev. 2014;18(4):293-7.
20. Cappuccio FP, Taggart FM, Kandala NB, Currie A, Peile E, Stranges S, et al. Meta-analysis of short sleep duration and obesity in children and adults. Sleep. 2008;31(5):619-26.
21. Cappuccio FP, Cooper D, D'Elia L, Strazzullo P, Miller MA. Sleep duration predicts cardiovascular outcomes: a systematic review and meta-analysis of prospective studies. Eur Heart J. 2011;32(12):1484-92.
22. Lee JA, Park HS. Relation between sleep duration, overweight, and metabolic syndrome in Korean adolescents. Nutr Metab Cardiovas. 2014;24(1):65-71.
23. van den Berg JF, Miedema HM, Tulen JH, Neven AK, Hofman A, Witteman JC, et al. Long sleep duration is associated with serum cholesterol in the elderly: the Rotterdam Study. Psychosom Med. 2008;70(9):1005-11.
24. Petrov ME, Kim Y, Lauderdale D, Lewis CE, Reis JP, Carnethon MR, et al. Longitudinal associations between objective sleep and lipids: the CARDIA study. Sleep. 2013;36(11):1587-95.
25. Bos MM, Noordam R, van den Berg R, de Mutsert R, Rosendaal FR, Blauw GJ, et al. Associations of sleep duration and quality with serum and hepatic lipids: The Netherlands Epidemiology of Obesity Study. J Sleep Res. 2018:e12776.
26. Kaneita Y, Uchiyama M, Yoshiike N, Ohida T. Associations of usual sleep duration with serum lipid and lipoprotein levels. Sleep. 2008;31(5):645-52.
27. Psaty BM, O'Donnell CJ, Gudnason V, Lunetta KL, Folsom AR, Rotter JI, et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from five cohorts. Circulation Cardiovascular genetics. 2009;2(1):73-80.
28. Rao DC, Sung YJ, Winkler TW, Schwander K, Borecki I, Cupples LA, et al. Multiancestry Study of Gene-Lifestyle Interactions for Cardiovascular Traits in 610 475 Individuals From 124 Cohorts: Design and Rationale. Circ Cardiovasc Genet. 2017;10(3).
29. Manning AK, LaValley M, Liu C-T, Rice K, An P, Liu Y, et al. Meta-analysis of Gene-Environment interaction: joint estimation of SNP and SNP×Environment regression coefficients. Genetic Epidemiology. 2011;35(1):11-8.
30. Lopez-Garcia E, Faubel R, Leon-Munoz L, Zuluaga MC, Banegas JR, Rodriguez-Artalejo F. Sleep duration, general and abdominal obesity, and weight change among the older adult population of Spain. The American journal of clinical nutrition. 2008;87(2):310-6.
31. van den Berg JF, Knvistingh Neven A, Tulen JH, Hofman A, Witteman JC, Miedema HM, et al. Actigraphic sleep duration and fragmentation are related to obesity in the elderly: the Rotterdam Study. International journal of obesity (2005). 2008;32(7):1083-90.
32. Wong PM, Manuck SB, DiNardo MM, Korytkowski M, Muldoon MF. Shorter sleep duration is associated with decreased insulin sensitivity in healthy white men. Sleep. 2015;38(2):223-31.
33. Reutrakul S, Van Cauter E. Interactions between sleep, circadian function, and glucose metabolism: implications for risk and severity of diabetes. Annals of the New York Academy of Sciences. 2014;1311:151-73.

34. Cappuccio FP, D'Elia L, Strazzullo P, Miller MA. Quantity and quality of sleep and incidence of type 2 diabetes: a systematic review and meta-analysis. Diabetes Care. 2010;33(2):414-20.
35. Rao DC, Sung YJ, Winkler TW, Schwander K, Borecki I, Cupples LA, et al. Multiancestry Study of Gene–Lifestyle Interactions for Cardiovascular Traits in 610 475 Individuals From 124 Cohorts. Design and Rationale. 2017;10(3).
36. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin Chem. 1972;18(6):499-502.
37. Buysse DJ, Reynolds CF, 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry research. 1989;28(2):193-213.
38. Zeileis A. Object-oriented Computation of Sandwich Estimators. 2006. 2006;16(9):16.
39. Zeileis A. Econometric COmputing with HC and HAC Covariance Matrix Estimators. J of Stat Software. 2004;11(10).
40. Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. BMC Bioinformatics. 2010;11:134.
41. Halekoh U, Højsgaard S, Yan J. The R package geepack for generalized estimating equations. Journal of Statistical Software. 2006;15(2):1-11.
42. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. Bioinformatics. 2007;23(10):1294-6.
43. Winkler TW, Day FR, Croteau-Chonka DC, Wood AR, Locke AE, Magi R, et al. Quality control and conduct of genome-wide association meta-analyses. Nat Protoc. 2014;9(5):1192-212.
44. Winkler TW, Kutalik Z, Gorski M, Lottaz C, Kronenberg F, Heid IM. EasyStrata: evaluation and visualization of stratified genome-wide association meta-analysis data. Bioinformatics. 2015;31(2):259-61.
45. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26(17):2190-1.
46. Kraft P, Yen YC, Stram DO, Morrison J, Gauderman WJ. Exploiting Gene-Environment Interaction to Detect Genetic Associations. Human Heredity. 2007;63(2):111-9.
47. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B. 1995;57(1):289-300.
48. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017;8(1):1826.
49. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010;26(18):2336-7.
50. Laville V, Bentley AR, Prive F, Zhu X, Gauderman J, Winkler TW, et al. VarExp: estimating variance explained by genome-wide GxE summary statistics. Bioinformatics. 2018;34(19):3412-4.
51. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015;348(6235):648-60.
52. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nature genetics. 2013;45(6):580-5.
53. Jones SE, van Hees VT, Mazzotti DR, Marques-Vidal P, Sabia S, van der Spek A, et al. Genetic studies of accelerometer-based sleep measures yield new insights into human sleep behaviour. Nat Commun. 2019;10(1):1585.

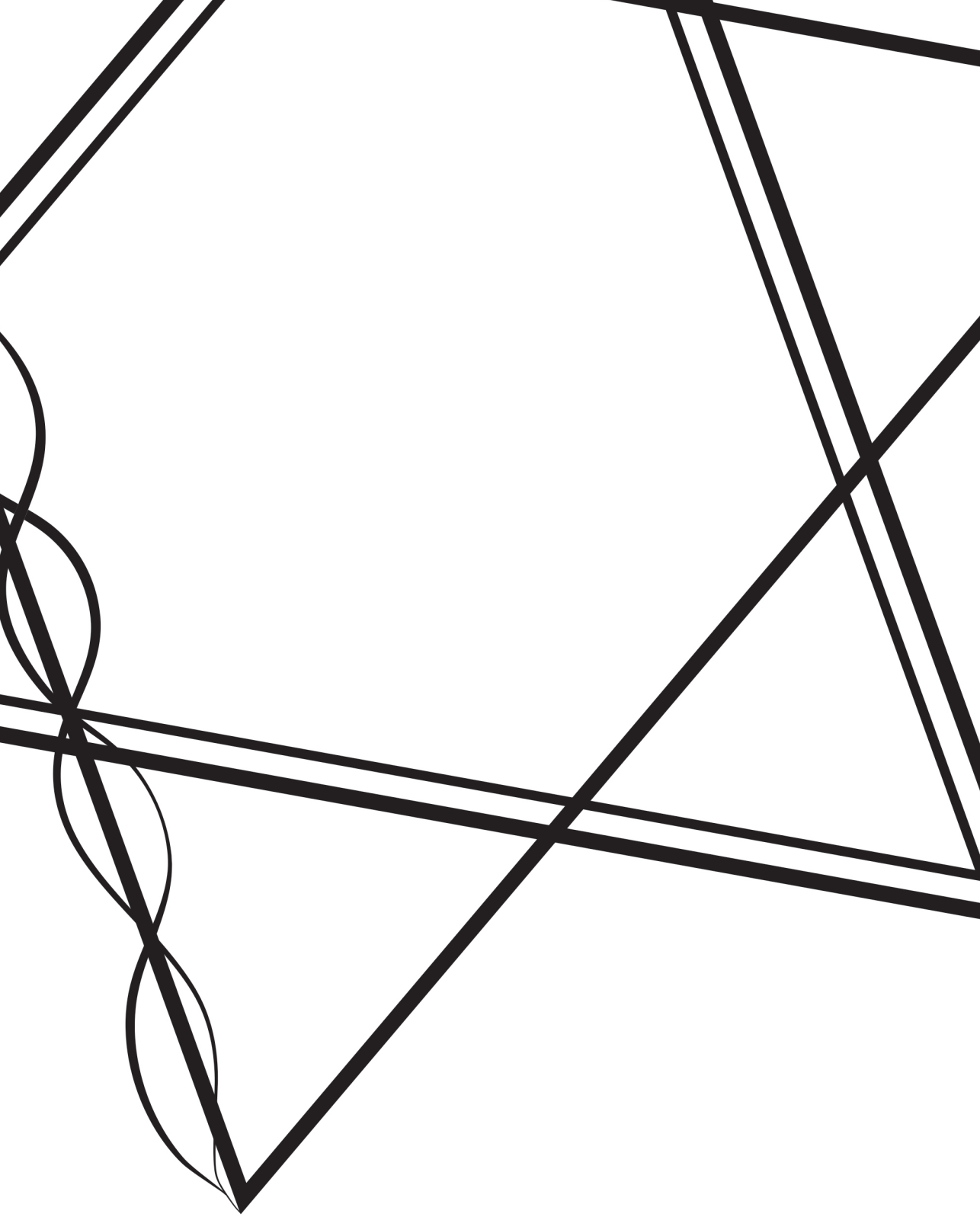
54. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet. 2015;47(10):1121-30.
55. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize Implements and enhances circular visualization in R. Bioinformatics. 2014;30(19):2811-2.
56. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. 2016.
57. Aho V, Ollila HM, Kronholm E, Bondia-Pons I, Soininen P, Kangas AJ, et al. Prolonged sleep restriction induces changes in pathways involved in cholesterol metabolism and inflammatory responses. Sci Rep. 2016;6:24828.
58. Chua EC, Shui G, Cazenave-Gassiot A, Wenk MR, Gooley JJ. Changes in Plasma Lipids during Exposure to Total Sleep Deprivation. Sleep. 2015;38(11):1683-91.
59. Gooley JJ. Circadian regulation of lipid metabolism. Proc Nutr Soc. 2016;75(4):440-50.
60. Huang T, Zeleznik OA, Poole EM, Clish CB, Deik AA, Scott JM, et al. Habitual sleep quality, plasma metabolites and risk of coronary heart disease in post-menopausal women. Int J Epidemiol. 2018.
61. van den Berg R, Kooijman S, Noordam R, Ramkisoensing A, Abreu-Vieira G, Tambyrajah LL, et al. A Diurnal Rhythm in Brown Adipose Tissue Causes Rapid Clearance and Combustion of Plasma Lipids at Wakening. Cell Rep. 2018;22(13):3521-33.
62. van den Berg R, Noordam R, Kooijman S, Jansen SWM, Akintola AA, Slagboom PE, et al. Familial longevity is characterized by high circadian rhythmicity of serum cholesterol in healthy elderly individuals. Aging Cell. 2017;16(2):237-43.
63. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012;40(Database issue):D930-4.
64. Yang L, Yang H, He M, Pan A, Li X, Min X, et al. Longer Sleep Duration and Midday Napping Are Associated with a Higher Risk of CHD Incidence in Middle-Aged and Older Chinese: the Dongfeng-Tongji Cohort Study. Sleep. 2016;39(3):645-52.
65. Anafi RC, Pellegrino R, Shockley KR, Romer M, Tufik S, Pack AI. Sleep is not just for the brain: transcriptional responses to sleep in peripheral tissues. BMC Genomics. 2013;14:362.
66. Moller-Levet CS, Archer SN, Bucca G, Laing EE, Slak A, Kabiljo R, et al. Effects of insufficient sleep on circadian rhythmicity and expression amplitude of the human blood transcriptome. Proc Natl Acad Sci U S A. 2013;110(12):E1132-41.
67. Carson V, Tremblay MS, Chaput JP, Chastin SF. Associations between sleep duration, sedentary time, physical activity, and health indicators among Canadian children and youth using compositional analyses. Appl Physiol Nutr Metab. 2016;41(6 Suppl 3):S294-302.
68. Patel SR, Zhu X, Storfer-Isser A, Mehra R, Jenny NS, Tracy R, et al. Sleep duration and biomarkers of inflammation. Sleep. 2009;32(2):200-4.
69. Ayas NT, White DP, Manson JE, Stampfer MJ, Speizer FE, Malhotra A, et al. A prospective study of sleep duration and coronary heart disease in women. Arch Intern Med. 2003;163(2):205-9.
70. Wefers J, van Moorsel D, Hansen J, Connell NJ, Havekes B, Hoeks J, et al. Circadian misalignment induces fatty acid metabolism gene profiles and compromises insulin sensitivity in human skeletal muscle. Proc Natl Acad Sci U S A. 2018;115(30):7789-94.
71. Adamovich Y, Aviram R, Asher G. The emerging roles of lipids in circadian control. Biochim Biophys Acta. 2015;1851(8):1017-25.
72. Galman C, Angelin B, Rudling M. Bile acid synthesis in humans has a rapid diurnal variation that is asynchronous with cholesterol synthesis. Gastroenterology. 2005;129(5):1445-53.

73. Akiyama M, Okada Y, Kanai M, Takahashi A, Momozawa Y, Ikeda M, et al. Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. Nature genetics. 2017;49(10):1458-67.
74. Winkler TW, Justice AE, Graff M, Barata L, Feitosa MF, Chu S, et al. The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. PLoS Genet. 2015;11(10):e1005378.
75. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015;518(7538):197-206.
76. Justice AE, Winkler TW, Feitosa MF, Graff M, Fisher VA, Young K, et al. Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits. Nat Commun. 2017;8:14977.
77. Hoffmann TJ, Choquet H, Yin J, Banda Y, Kvale MN, Glymour M, et al. A Large Multiethnic Genome-Wide Association Study of Adult Body Mass Index Identifies Novel Loci. Genetics. 2018;210(2):499-515.
78. Graff M, Scott RA, Justice AE, Young KL, Feitosa MF, Barata L, et al. Genome-wide physical activity interactions in adiposity - A meta-analysis of 200,452 adults. PLoS Genet. 2017;13(4):e1006528.
79. Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Magi R, et al. New genetic loci link adipose and insulin biology to body fat distribution. Nature. 2015;518(7538):187-96.
80. Spada J, Scholz M, Kirsten H, Hensch T, Horn K, Jawinski P, et al. Genome-wide association analysis of actigraphic sleep phenotypes in the LIFE Adult Study. J Sleep Res. 2016;25(6):690-701.
81. Chambers JC, Zhang W, Sehmi J, Li X, Wass MN, Van der Harst P, et al. Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. Nature genetics. 2011;43(11):1131-8.
82. Kanai M, Akiyama M, Takahashi A, Matoba N, Momozawa Y, Ikeda M, et al. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. Nature genetics. 2018;50(3):390-400.
83. Kunutsor SK, Abbasi A, Adler AI. Gamma-glutamyl transferase and risk of type II diabetes: an updated systematic review and dose-response meta-analysis. Ann Epidemiol. 2014;24(11):809-16.
84. Kunutsor SK, Apekey TA, Cheung BM. Gamma-glutamyltransferase and risk of hypertension: a systematic review and dose-response meta-analysis of prospective evidence. J Hypertens. 2015;33(12):2373-81.
85. Kunutsor SK, Apekey TA, Seddoh D. Gamma glutamyltransferase and metabolic syndrome risk: a systematic review and dose-response meta-analysis. Int J Clin Pract. 2015;69(1):136-44.
86. Wang J, Zhang D, Huang R, Li X, Huang W. Gamma-glutamyltransferase and risk of cardiovascular mortality: A dose-response meta-analysis of prospective cohort studies. PLoS One. 2017;12(2):e0172631.
87. Park SG, Lee YJ, Ham JO, Jang EC, Kim SW, Park H. Association between long working hours and serum gamma-glutamyltransferase levels in female workers: data from the fifth Korean National Health and Nutrition Examination Survey (2010-2011). Ann Occup Environ Med. 2014;26(1):40.
88. Swanson GR, Burgess HJ, Keshavarzian A. Sleep disturbances and inflammatory bowel disease: a potential trigger for disease flare? Expert Rev Clin Immunol. 2011;7(1):29-36.
89. Giles TD. Circadian rhythm of blood pressure and the relation to cardiovascular events. J Hypertens Suppl. 2006;24(2):S11-6.

90. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27-30.
91. de Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. Nature genetics. 2017;49(2):256-61.
92. Maguire LH, Handelman SK, Du X, Chen Y, Pers TH, Speliotes EK. Genome-wide association analyses identify 39 new susceptibility loci for diverticular disease. Nature genetics. 2018;50(10):1359-65.
93. Ananthakrishnan AN, Khalili H, Konijeti GG, Higuchi LM, de Silva P, Fuchs CS, et al. Sleep duration affects risk for ulcerative colitis: a prospective cohort study. Clin Gastroenterol Hepatol. 2014;12(11):1879-86.
94. Sullivan AE, Raimondo A, Schwab TA, Bruning JB, Froguel P, Farooqi IS, et al. Characterization of human variants in obesity-related SIM1 protein identifies a hot-spot for dimerization with the partner protein ARNT2. Biochem J. 2014;461(3):403-12.
95. Hao N, Bhakti VL, Peet DJ, Whitelaw ML. Reciprocal regulation of the basic helix-loop-helix/Per-Arnt-Sim partner proteins, Arnt and Arnt2, during neuronal differentiation. Nucleic Acids Res. 2013;41(11):5626-38.
96. Dong C, Beecham A, Wang L, Slifer S, Wright CB, Blanton SH, et al. Genetic loci for blood lipid levels identified by linkage and association analyses in Caribbean Hispanics. J Lipid Res. 2011;52(7):1411-9.
97. Khor SS, Miyagawa T, Toyoda H, Yamasaki M, Kawamura Y, Tanii H, et al. Genome-wide association study of HLA-DQB1*06:02 negative essential hypersomnia. PeerJ. 2013;1:e66.
98. Egan KJ, Knutson KL, Pereira AC, von Schantz M. The role of race and ethnicity in sleep, circadian rhythms and cardiovascular health. Sleep Med Rev. 2017;33:70-8.
99. Ren H, Liu Z, Zhou X, Yuan G. Association of sleep duration with apolipoproteins and the apolipoprotein B/A1 ratio: the China health and nutrition survey. Nutr Metab (Lond). 2018;15:1.
100. Tsutsumi K, Inoue Y, Kondo Y. The relationship between lipoprotein lipase activity and respiratory quotient of rats in circadian rhythms. Biol Pharm Bull. 2002;25(10):1360-3.
101. Persson L, Cao G, Stahle L, Sjoberg BG, Troutt JS, Konrad RJ, et al. Circulating proprotein convertase subtilisin kexin type 9 has a diurnal rhythm synchronous with cholesterol synthesis and is reduced by fasting in humans. Arterioscler Thromb Vasc Biol. 2010;30(12):2666-72.
102. Lane JM, Liang J, Vlasac I, Anderson SG, Bechtold DA, Bowden J, et al. Genome-wide association analyses of sleep disturbance traits identify new loci and highlight shared genetics with neuropsychiatric and metabolic traits. Nature genetics. 2017;49(2):274-81.
103. Hu Y, Shmygelska A, Tran D, Eriksson N, Tung JY, Hinds DA. GWAS of 89,283 individuals identifies genetic variants associated with self-reporting of being a morning person. Nat Commun. 2016;7:10448.
104. Dashti H, Jones SE, Wood AR, Lane J, van Hees VT, Wang H, et al. GWAS in 446,118 European adults identifies 78 genetic loci for self-reported habitual sleep duration supported by accelerometer-derived estimates. bioRxiv. 2018:274977.
105. Jackson CL, Patel SR, Jackson WB, 2nd, Lutsey PL, Redline S. Agreement between self-reported and objectively measured sleep duration among white, black, Hispanic, and Chinese adults in the United States: Multi-Ethnic Study of Atherosclerosis. Sleep. 2018;41(6).

PART III

APOE



CHAPTER 4.1

The ApoE ε4 Isoform: Can the Risk of Diseases be Reduced by Environmental Factors?

Maxime M Bos, Raymond Noordam, Gerard Jan Blauw, P Eline Slagboom, Patrick CN Rensen, Diana van Heemst

J Gerontol A Biol Sci Med Sci. 2019;74(1):99–107

ABSTRACT

Candidate gene studies and genome-wide association studies found that genetic variation in *APOE* is robustly associated with multiple age-related diseases and longevity. Apolipoprotein E (ApoE) is an apolipoprotein that plays an important role in triglyceride and cholesterol metabolism. In literature, especially the ApoE ε4 isoform has been associated with an increased risk of mortality and age-related diseases such as Alzheimer's disease (AD), cardiovascular diseases (CVD) , as compared to the 'neutral' ApoE ε3 isoform. There are, however, large differences in the deleterious effects of the ApoE ε4 isoform between ancestries and populations, which might be explained by differences in environmental and lifestyle exposures. In this respect, poor nutrition and physical inactivity are two important lifestyle factors that have been associated with increased risks for AD and CVD. Therefore, in this narrative review we discuss how omega-3 fatty acid intake and physical activity, may modify the impact of ApoE ε4 on AD and CVD risk.

INTRODUCTION

Genetic variation in *APOE* is robustly associated with human longevity[1, 2]. The *APOE* gene, located on chromosome 19, consists of three different isoforms, notably ApoE ε2 (Cys^{112}, Cys^{158}), ApoE ε3 (Cys^{112}, Arg^{158}) and ApoE ε4 (Arg^{112}, Arg^{158}), of which the ApoE ε3 isoform is generally considered the 'neutral' isoform[3]. With respect to longevity, the ApoE ε2 isoform has been associated with an increased survival and with a more beneficial lipid profile[4, 5]. In contrast, compared with ApoE ε3 carriers, carriers of the ApoE ε4 isoform have higher mean total serum cholesterol levels[6]. Moreover, previous research indicated that the ApoE ε4 isoform decreases the efficacy of cholesterol lowering statin therapy[7, 8]. ApoE ε4 is an established risk factor for ageing and various age-related diseases, such as multiple types of dementia (including Alzheimer's disease (AD)) and cardiovascular disease (CVD)[6, 9]. In a study comprising individuals of European ancestry (5,107 AD patients and 6,262 controls), ApoE ε3/ε4 carriers had a 3.2-fold increased risk and ApoE ε4/ε4 carriers had a 14.9-fold increased risk to develop AD compared to ApoE ε3/ε3 carriers[6]. Furthermore, in a meta-analysis of studies from different ancestries (15,492 cases and 32,965 controls), it was shown that both ApoE ε3/ε4 carriers and ApoE ε4/ε4 carriers had a 1.4-fold higher risk to develop coronary artery disease[10]. Most interestingly, the increased risk of disease associated with ApoE ε4 seems to be variable between individuals of different ancestries with Kenyan or Nigerian ancestry individuals having no harmful effects of ApoE ε4[11, 12]. Strikingly, Nigerian ancestry individuals have the highest frequency of the ApoE ε4, but a relatively low incidence of AD[13].

Possibly, the differences observed in risk conferred by ApoE genotype between individuals of different ancestries could be attributable to environmental- and lifestyle-factors. Therefore, for our biological understanding and to be eventually of added value to public health, it is of interest to disentangle the mechanisms resulting in the lower disease risk conferred by ApoE ε4 in certain populations. Lifestyle factors vary between populations, and are associated with increased risks of disease. Because of the broad definition of lifestyle, we will only elaborate on nutritional intake and physical activity in the context of ApoE and age-related disease. Therefore, the primary aim of this narrative review is to discuss potential pathways that might attenuate the effects of the genetic susceptibility for AD and CVD in ApoE ε4 carriers. To the best of our knowledge, this is the first narrative review to discuss the current (biological) evidence of APOE-lifestyle interactions in the pathophysiology of age-related diseases. First, we will provide a short overview of the ApoE protein, the different isoforms and their function. Next, we will focus on how nutrition and physical activity could modify the effect of genetic predisposition of individuals carrying the ApoE ε4 risk allele.

The ApoE protein and isoform prevalence

ApoE is a 299-residue protein, which is predominantly produced by hepatocytes, macrophages and astrocytes[14, 15]. Human ApoE comprises multiple amphipathic α-helices and is known to contain a low-density lipoprotein receptor (LDLR) binding site on the fourth helix[16]. ApoE is a major apolipoprotein found in plasma and is presented on chylomicron remnants, very low-density lipoproteins (VLDLs), intermediate-density lipoproteins (IDL), and high-density lipoproteins (HDL) to mediate their receptor-mediated uptake from the circulation. ApoE is also involved in VLDL assembly and secretion by hepatocytes [17]. Moreover, ApoE is present in the central nervous system[18, 19] where it plays an important role in the transport of cholesterol and in cellular reparative processes (e.g. neuronal repair)[18].

Figure 1 provides a schematic overview of the role of ApoE in lipoprotein metabolism. In the exogenous pathway, dietary triglycerides and cholesterol are absorbed by enterocytes and are used for lipidation of ApoB48 to generate chylomicrons, which enter the blood circulation via the lymphatic system. In the circulation, lipoprotein lipase (LPL) on metabolically active tissues, hydrolyses triglycerides (TG) within these particles to release free fatty acids that are taken up by these tissues (including the heart, skeletal muscles, white adipose tissue and brown adipose tissue). As a consequence of LPL-mediated lipolysis, smaller chylomicron remnant particles are formed, which become enriched with ApoE that is acquired from HDL. Enrichment with ApoE abrogates the lipolysis by LPL and mediates the subsequent uptake of the chylomicron remnants by hepatocytes through receptor-mediated endocytosis via the LDLR, and the LDLR-related protein (LRP)[20]. In addition to these high-affinity receptors, heparan sulfate proteoglycans (HSPG) also play a role in the low affinity/high capacity binding and internalization of chylomicron remnants[21].

In the endogenous pathway, VLDL particles are synthesized in the liver by lipidation of ApoB100 with cholesterol and triglycerides, and serve to deliver endogenous fatty acids as well as cholesterol towards peripheral tissues. After secretion from the liver into the plasma, LPL hydrolyses VLDL similarly to chylomicrons, which results in the formation of VLDL remnants. ApoE mediates the uptake of VLDL remnants (also termed 'intermediate-density lipoproteins; IDL) via hepatic receptors and binding sites in a similar fashion as to chylomicron remnants. VLDL remnants that escape uptake by the liver are completely lipolysed by LPL to generate low-density lipoprotein (LDL) particles that mainly carry cholesteryl esters. The main apolipoprotein in LDL is ApoB100, as all other apolipoproteins including ApoE are lost during lipolysis. LDL particles are taken up via recognition of ApoB100 through the LDLR on the liver and peripheral tissues (e.g. adrenals, testes and ovaria), which need cholesterol for e.g. steroid hormone synthesis.

Both the liver and intestines produce lipid-poor HDL particles containing ApoAI and ApoAII in addition to ApoE. Through these apolipoproteins, HDL can induce the efflux of cholesterol from peripheral tissues via ATP-Binding Cassette Transporter A1 (ABCA1) and ATP-binding Cassette Transporter G1 (ABCG1) and transport the cholesterol to the liver via Scavenger Receptor Class B Member 1 (SR-BI), after which cholesterol can be converted into bile acids. Collectively, this pathway is called reverse cholesterol transport (RCT)[16]. Alternatively, cholesterol from peripheral tissues can reach the liver after transfer of cholesteryl esters from HDL to VLDL via the cholesteryl ester transfer protein (CETP), with subsequent receptor-mediated uptake of remnants by the liver.

Table 1. *ApoE isoforms and their properties*

Isoform	2	3	4
Residue 112	Cysteine	Cysteine	Arginine
Residue 158	Cysteine	Arginine	Arginine
Overall frequency (mean %)	7	79	14
Plasma triglycerides (mmol/L)	Higher	Normal	Higher
Plasma cholesterol (mmol/L)	Lower	Normal	Higher
ApoE stability	Higher	Normal	Lower
Associated disorders	Type III hyper-lipoproteinaemia, PVC, ASCVD	Normal	Hypercholesterolemia, CVD, AD
LDLR binding affinity (%)	1	100	100
Lipid binding ability	Normal	Normal	Higher
Binding preference	HDL	HDL	VLDL

Abbreviations: AD, Alzheimer's disease; ASCVD, atherosclerotic cardiovascular disease; CVD, cardiovascular disease, HDL, high-density lipoprotein, PVC, peripheral vascular disease; VLDL, very-low-density lipoprotein.

APOE isoforms and prevalence

Table 1 provides an overview of the different isoforms and their characteristics. The frequencies of the different ApoE isoforms vary greatly between populations, but ApoE ε3 is most common in all (mean global frequency ≈79%) followed by the ε4 isoform (≈14%) and the ε2 isoform (≈7%)[24, 25]. For example, ApoE ε3 frequency ranges from 54% in African Pygmies to 91% in Mayans[26, 27], while ApoE ε4 frequency ranges from 5% in Sardinians to 41% in African Pygmies[26]. Importantly, among European populations, ApoE ε4 frequency is higher in northern European countries than in Southern countries[28, 29]. Compared to ApoE ε3 (normal plasma cholesterol levels[3]), ApoE ε4 is associated with altered plasma lipid levels and lipoprotein particle

distributions as the resulting ApoE protein has altered binding affinity for either lipoprotein particles or the low-density lipoprotein receptor (LDLR)[30]. The ApoE ε4 isoform is associated with higher total cholesterol levels, higher LDL-cholesterol levels and lower HDL-cholesterol levels[24]. ApoE ε4 has a higher binding affinity for larger TG-rich lipoproteins (such as VLDL and chylomicron remnants)[22, 23]. The ApoE ε2 isoform displays about 1% of the binding affinity to the LDLR compared to ApoE ε3 and ε4[31]. Moreover, ApoE ε2 carriers have a lower hepatic VLDL assembly and VLDL uptake, resulting in diminished clearance from the blood and subsequent type III hyperlipoproteinemia[32, 33]. *In vitro* studies have shown that ApoE ε2 can protect cells from oxidative stress induced cell death. Moreover, ApoE is important in neural injury repair by initiating membrane repair by redistribution of lipids [34]. The molecular stability of ApoE is lower for ApoE ε4 as compared to the ε2 and ε3 isoform[35]. The structural differences of the ApoE isoforms result in different susceptibility to proteolytic cleavage, by which neurotoxic fragments are formed. Proteolytic cleavage is lowest for ApoE ε2[34,36-38] and highest for ApoE ε4, providing a potential mechanism explaining the higher risk of AD in ApoE ε4 carriers[34]. Moreover, it was shown that HDL-induced recycling of ApoE ε4-containing TG-rich lipoproteins is strongly reduced in hepatocytes, resulting in increased intracellular cholesterol levels[39]. As ApoE ε4 has a preference to bind VLDL and chylomicrons, this results in an enhanced uptake of these particles by the hepatocytes thereby competing with the uptake of LDL particles, resulting in increased LDL concentration[19]. Since this review focuses on strategies to alleviate the health risk associated with ApoE ε4, a detailed discussion of the ApoE ε2 isoform is beyond the scope of this review.

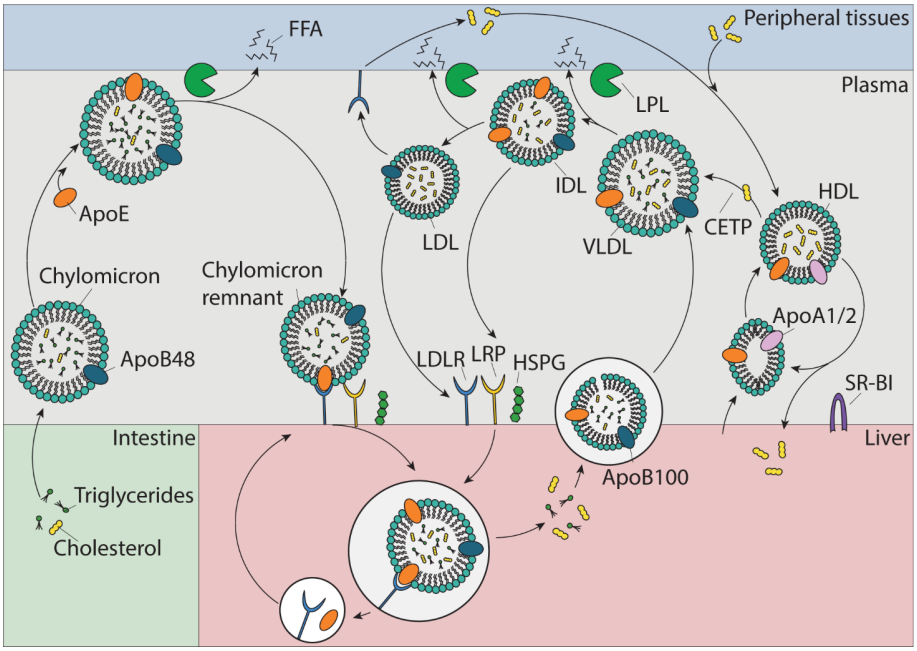


Figure 1. *Schematic overview of lipoprotein metabolism*. In the exogenous pathway, dietary cholesterol and triglycerides (TG) are absorbed by the small intestine and incorporated into chylomicrons within enterocytes. Via the lymphatic system, chylomicrons reach the circulation, where their triglycerides are hydrolysed by lipoprotein lipase (LPL) present on metabolically active tissues, to deliver exogenous fatty acids to these tissues[20]. This results in formation of smaller remnant particles that are enriched with ApoE to mediate subsequent internalisation by hepatocytes via the LDL receptor (LDLR), the LDLR-related protein (LRP), and heparan sulfate proteoglycans (HSPG)[21]. In the endogenous pathway, very-low-density lipoproteins (VLDL) are assembled and secreted by the liver to deliver endogenous fatty acids towards metabolically active tissues. Similar to chylomicrons, VLDL are lipolysed by LPL and their remnants can be taken up by hepatocytes via ApoE. Alternatively, the particles can be further processed by LPL to yield low-density lipoproteins (LDL) as lipolytic end product. LDL lacks ApoE and therefore ApoB100 serves as a ligand to bind exclusively the LDLR. The free fatty acids (FFA) derived from lipolysis are used in the peripheral tissues. High density lipoproteins (HDL) are synthesized as discoidal precursors by the liver and small intestine. Also, surface remnants that are produced during lipolysis of chylomicrons and VLDL can contribute to the HDL pool. HDL can acquire cholesterol from peripheral tissues and transport the cholesterol back to the liver through the direct pathway of reverse cholesterol transport (RCT) via scavenger receptor class B member 1 protein (SR-BI), which selectively takes up cholesteryl esters from HDL, after which liberated cholesterol can be converted into bile acids and secreted into the feces[16]. Reverse cholesterol transport can also occur through the indirect pathway, whereby cholesterol is transported via CETP from HDL to VLDL and LDL. Compared to the ApoE ε3 isoform, the ApoE ε4 isoform binds preferentially to VLDL particles and slows down lipolysis. This results in higher VLDL concentrations and lower HDL concentrations[22, 23].

Effect modifiers of ApoE isoforms

Nutritional intake and physical activity have been hypothesized to modify the metabolic effects of ApoE ε4. Therefore, it is of interest to elaborate more on these effect modifiers in relation to ApoE genotype.

Fatty acids and disease risk

A meta-analysis comprising 11 prospective cohort studies (371,965 participants from general populations and 31,185 death events) showed that higher dietary intake and higher circulating levels of n-3 long-chain polyunsaturated fatty acids were associated with a lower risk for all-cause mortality[40]. Notably, it was found that a 0.3 g daily increase in dietary intake of n-3 long-chain polyunsaturated fatty acids (also called omega-3 fatty acids) was associated with 6% lower risk of all-cause mortality in the general population. Furthermore, a 1% increase in circulating levels of the omega-3 fatty acids eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA), were associated with a 20% and 21% risk reduction of all cause-mortality, respectively[40]. In addition, emerging evidence suggests that dietary factors and cognitive function are related[41]; several epidemiological studies showed that higher fish intake is associated with a lower risk for cognitive decline and dementia during follow-up[42, 43]. In more detail, a study comprising 2,031 Norwegian individuals aged 70-74 years, recruited from the general population, showed that fish intake was associated with a better cognitive function in a dose-dependent manner, with 75 g fish a day as optimum[44]. Omega-3 fatty acids not only exert a beneficial effect on cognition, but also influence cardiovascular risk in the general population. Overall, omega-3 fatty acids are able to reduce CVD mortality by 37%[45, 46]. However, the effects of omega-3 fatty acid intake on disease risk are conflicting with studies indicating positive, null or negative effects[47-50]. For example, no significant difference in triglyceride concentration has been observed upon omega-3 fatty acids supplementation in elderly[48]. However, a linear correlation between higher doses of omega-3 fatty acid intake and a triglyceride lowering effect has also been observed[47]. Therefore, given this heterogeneity in research findings, there are yet no definite conclusions on the role of omega-3 fatty acids in CVD and neurodegenerative disease. Importantly, these studies did not take any specific ApoE isoform into account.

Fatty acids, disease risk and APOE genotype

In the following paragraph, we discuss potential effect modification of the association between fatty acid intake and disease risk by *APOE*. Kariv-Inbal *et al.*[43] described that the detrimental effects of the ApoE ε4 isoform on AD risk could be reduced by a diet enriched with fish oil (DHA) in ApoE ε4-targeted replacement mice. Another study, conducted in humans, determined the association between seafood and n-3 fatty acid intake and cognitive decline in relation to the ApoE ε4 isoform[51]. This longitudinal, community-

based epidemiologic study in 915 elderly participants of Caucasian ancestry (recruited from retirement communities in Illinois, USA), demonstrated that ApoE ε4 carriers had a slower decline in multiple cognitive domains with weekly seafood consumption and moderate to high intake of n-3 fatty acids than ε3 and ε2 carriers consuming the same amount of seafood after an average follow up of 4.9 ± 2.5 years[51]. Intake of vegetable α-linolenic acid, which is used by the body to form long chain n-3 fatty acids, was also associated with slower cognitive decline only in carriers of the ApoE ε4 isoform[51]. These studies indicate that omega-3 fatty acids are beneficial in preventing cognitive decline and suggest that individuals with the ApoE ε4 isoform may especially benefit from higher n-3 fatty acid consumption for the prevention of cognitive decline and AD.

The possible mechanisms of action of unsaturated fatty acids

It is of interest to elaborate more on how polyunsaturated 3-n fatty acids may be beneficial in slowing AD and CVD development related to ApoE ε4, and healthy aging in general, which may be through multiple biological pathways. For example, in the brain, omega-3 fatty acids are incorporated in phospholipids where they replace omega-6 fatty acids, which increases fluidity of membranes of neuronal cells[52, 53]. This increased fluidity allows for better signal transduction between the neuronal cells. Omega-3 fatty acids also improve neurotransmission by increasing receptor binding affinity and increasing the number of receptors of ion channels[54], which therefore counteracts the synaptic deficits associated with ApoE ε4[55].

Another biological mechanism might work via the ability of omega-3 fatty acids to lower the synthesis of new VLDL particles and triglycerides from the liver[56], as illustrated **in Figure 2A** and **2B**. ApoE ε4 isoform carriers have a faster clearance of VLDL particles compared to ApoE ε3 carriers[57], which by competition for the hepatic clearance of LDL raises LDL-cholesterol. We therefore hypothesize that omega-3 fatty acids could possibly lower the synthesis of VLDL particles, whereby competition for hepatic uptake between VLDL remnants and LDL is reduced, and the uptake of LDL particles by the liver is increased. Subsequently, this might lead to lower serum LDL-cholesterol concentrations to be of specific importance to ApoE ε4 carriers.

The positive effects of polyunsaturated fatty acids may also be explained through inflammatory pathways. In relation to ApoE ε4, increased inflammation and oxidative stress has been observed in cell lines, rodents and human volunteers[58]. It was previously reported in animal studies that fish oil has beneficial effects on triglyceride levels and inflammatory factors by downregulation of inflammatory genes and upregulation of peroxisome proliferator-activated receptor-gamma (PPAR-γ)[59]. The PPARs belong to a nuclear receptor group that act as lipid-activated transcription

factors. Increasing evidence suggests a protective role of PPAR-γ signaling in atherosclerosis by decreasing inflammatory cytokine production and mediating lipid metabolism[60, 61]. Moreover, a placebo-controlled study in hyperlipidemic individuals demonstrated that n-3 polyunsaturated fatty acids in combination with plant sterols were able to reduce several inflammatory markers, such as C-reactive protein (CRP), tumor necrosis factor-alpha (TNF-α), interleukin-6 (IL-6) and leukotriene B(4) (LTB(4))[62]. In the same study, the overall CVD risk was reduced, suggesting that higher n-3 polyunsaturated fatty acid intake works in a cardio-protective manner possibly through reduced inflammation[62].

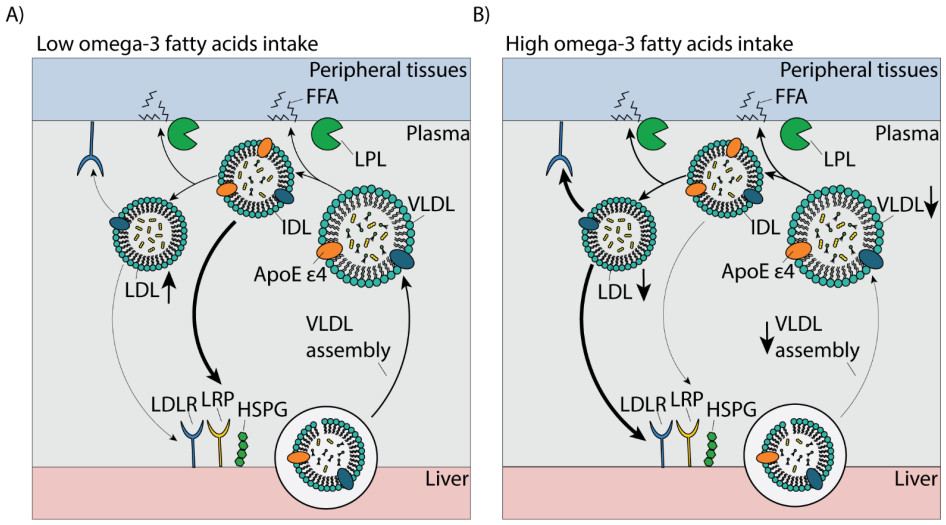


Figure 2. *The effects of omega-3 fatty acid intake on the ApoE ε4 isoform* A) ApoE ε4 has a higher affinity for the LDLR receptor than LDL itself, and therefore binds preferentially to this receptor, resulting in receptor-mediated endocytosis of VLDL instead of LDL. This preferential uptake of VLDL results in higher levels of LDL in the plasma[19]. B) Unsaturated fatty acid intake decreases VLDL assembly, resulting in a lower concentration of VLDL and thereby increases the uptake of LDL and IDL by the liver, leading to lower LDL levels[56].

Physical activity and disease risk

It has been argued that a high level of aerobic exercise can attenuate the process of aging by reducing amyloid plaque formation and increasing overall vascular health[63]. Previous studies in healthy older adults demonstrated that high physical activity is associated with a preservation of both cognitive function and hippocampal volume[64-66]. Moreover, cognitive function and amyloid plaque formation in elderly AD patients benefits from daily exercise[67].

Physical activity, disease risk and APOE genotype

The following paragraph will focus on physical activity in relation to cognitive and cardiovascular health with respect to ApoE ε4. In cognitively healthy adults, a more sedentary lifestyle was associated with higher amyloid deposition[68]. Interestingly, this finding was only observed in ApoE ε4 carriers and not in carriers of the other ApoE genotypes[68]. Moreover, in a study of 78 cognitive healthy older adults with 18 months follow-up, high physical activity was associated with a slower decline in hippocampal volume. Again, this effect was specifically observed only in ApoE ε4 carriers[69]. In addition, aerobic exercise was associated with a slower cognitive decline, and decreased risks of various types of dementia, including AD, but specifically in ApoE *ε4 carriers*[70-72]. In relation to cardiovascular health, it has been shown that age-related changes in cholesterol and LDL-cholesterol were counteracted by life-long endurance exercise in 15 old trained healthy men as compared to 12 old untrained, 10 young trained and 12 young untrained men[73]. Compared to a group of mild-to-moderate physically active men who maintained their physical exercise level, men that increased their exercise during a 1-year follow-up had a more favorable lipid profile[74]. Furthermore, it was shown that HDL-cholesterol levels increased directly after exercise training in 17 overweight men, possibly through reduction in HDL protein catabolism[75]. However, these intervention studies have been generally conducted in small samples and with short follow-up. In a population-based cross-sectional study (N=1,708, aged 35-74 years), higher physical activity was associated with higher HDL cholesterol and lower triglyceride levels in specifically ApoE ε4 carriers[76]. However, these results have not been confirmed in a subsequent study[77]. These discrepancies require further research and might yield insights in other mechanisms associated with physical activity.

The possible mechanisms of action of physical activity

One possible mechanistic explanation underlying the association between physical activity and AD specifically in ApoE ε4 carriers, is based on the finding that neuronal ApoE ε4 has an increased susceptibility for proteolytic cleavage compared to ApoE ε3[36-38]. In brain tissue samples of AD patients, fragments of the ApoE protein are present in much higher concentrations as compared to those of controls[36, 38]. Physical exercise in ApoE ε4 carriers is able to reduce the neuronal level of ApoE ε4 and thereby lower the total amount of ApoE ε4 fragments in the brains of these individuals. Subsequently, the risk of developing AD in these individuals might be reduced. However, this is merely a hypothesis based on a small number of studies[36, 38]. Additional research is warranted to disentangle the protective mechanism of physical exercise on ApoE ε4 concentrations in the brain.

Based on previous studies, we are able to hypothesize the biological mechanism through which physical activity might modify the detrimental effects of ApoE ε4 carriership (visualized in **Figure 3A** and **3B**). Exercise enhances the LPL-dependent flux of triglyceride-derived fatty acids from chylomicrons and VLDL to myocytes, which decreases the level of serum triglycerides[16]. As a consequence, excess surface lipids are released from chylomicrons and VLDL as surface remnants that mainly contain phospholipids and unesterified cholesterol. These surface remnants are precursors of HDL that subsequently accept additional cholesterol from peripheral tissues, thereby increasing total HDL cholesterol levels[78, 79]. In this way, increased LPL activity may decrease serum triglyceride levels in more physically active ApoE ε4 carriers. Indeed, it was shown that physical exercise decreases VLDL particle size, which is consistent with hydrolysis of these particles by LPL[80].

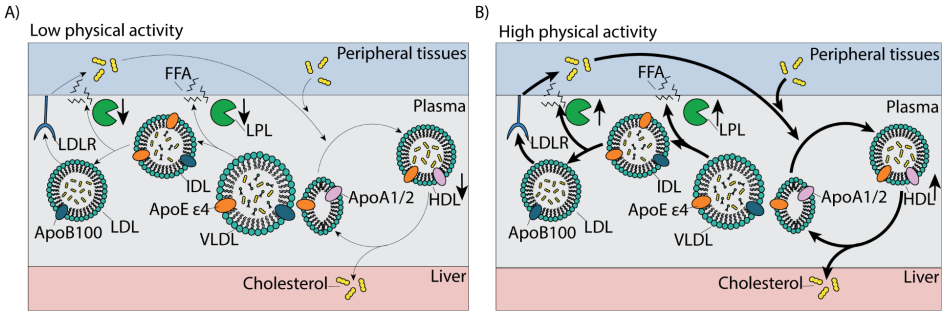


Figure 3. *The effects of physical activity on the ApoE ε4 isoform.* A) The ApoE ε4 isoform binds preferably to large lipoprotein particles, such as VLDL, due to higher lipid binding ability. The increased binding to VLDL slows down lipolysis of these particles[22, 23]. B) Physical activity enhances LPL activity. This increased activity enhances lipolysis of VLDL particles and thereby decreases the VLDL concentration. Generation of more surface remnants increases the level of HDL that can accept cholesterol from peripheral tissues[16, 78, 79].

FUTURE PERSPECTIVES

Due to advances in technology and availability of large datasets, multiple novel genetic determinants of diseases are being identified. ApoE ε4 carriership is the strongest genetic risk factor for multiple age-related diseases, including diseases for which no drug treatments are (currently) available. In the present narrative review, we described several biological mechanisms on how unsaturated fatty acids and physical activity may prevent or delay cognitive decline and CVD, and discuss how these effects extend to and are possibly even stronger in carriers of the ApoE ε4 risk allele.

On the one hand, the general public is becoming increasingly aware of the impact of nutrition and physical activity on their health. However, on the other hand, current consumption of omega-3 fatty acids is low due to modern agriculture and a Western diet[81] and a large part of modern society is now adapted to a sedentary lifestyle whereby the largest proportion of adults does not even meet the proposed physical activity guidelines[82]. In line, a higher incidence of cognitive decline, AD, and other age-related diseases in relation to the Western diet is observed[83-85]. This is especially of importance when populations that still have a high prevalence of the ApoE ε4 isoform (e.g., Nigerian ancestry or Northern European countries) adapt to a more sedentary lifestyle, because an even higher increase in CVD and AD may occur in these at-risk individuals[63]. In agreement with this hypothesis, African populations that move to cities and reduce their physical activity are much more susceptible to acquire CVD and AD than Western populations [86]. Therefore, it seems that individuals carrying the ApoE ε4 isoform could specifically benefit from increasing their physical activity and/or increasing their omega-3 fatty acid intake.

In order to assess if an individual is a carrier of the ApoE ε4 isoform, screening for this genotype has to be implemented. However, screening of ApoE ε4 carriers runs into a vast amount of ethical, methodological, and economic aspects that need to be addressed first in order to make the implementation of these models feasible as well as cost effective. For example, important questions, such as the clinical meaning and implications of such screening and which professional figures should manage the implementation, are only some of many questions that have to be answered first. However, there is an increasing body of evidence suggesting that lifestyle may influence genetic susceptibility to several chronic diseases that may not be left unnoticed. Therefore, in line with the evidence as discussed in this narrative review, next to focusing on the general population to increase their omega-3 fatty acid intake and enhance their physical activity, it may be valuable to specifically focus on at-risk individuals and/or families that have a higher susceptibility to carry ApoE ε4. An example of a at-risk group

may be certain families with a high incidence of AD, in which a higher prevalence of hypertension, pro-inflammatory markers and ApoE ε4 genotype has been observed. These factors may be early risk factors for AD in old age, as those have been observed already at middle-age before the onset of AD[87]. Specifically, increasing awareness of physicians and general practitioners may lead them to stress the importance of adhering to a healthier lifestyle in at-risk individuals. For example, in previous research, it was demonstrated that lifestyle interventions to improve physical activity and/or nutritional habits, even in older adults, seem promising[88, 89]. A 13-weeks lifestyle program already induced metabolic health benefits, which might increase the positive adaption of lifestyle changes in the general population as effects occurred relatively fast[88]. These studies suggest that diminishing the occurrence of non-communicable diseases associated with ApoE ε4 via improving physical activity seems possible. Next to increasing physical activity, we hypothesize that a diet rich in polyunsaturated fatty acids will benefit ApoE ε4 isoform carriers. For example, the Mediterranean diet is a plant-based diet rich of unsaturated fatty acids and antioxidants. The Mediterranean diet is characterized by a high content of olive oil, high intake of fruits and vegetables, moderate-to-high fish and seafood consumption, low intake of dairy products, low meat consumption and a regular intake of red wine[90]. The Mediterranean diet is associated with a lower risk of AD and cognitive decline[91, 92] and has beneficial effects on overall health[93]. In a randomized clinical trial in healthy elderly, Valls-Pedret *et al.*[94] showed that a Mediterranean diet supplemented with olive oil and mixed nuts was able to improve cognitive function. The Mediterranean diet is rich in bioactive phytochemicals that are known to have antioxidant and anti-inflammatory properties. For example, olive oil is rich in phenolic compounds that may counteract oxidative stress processes in the brain and thereby decrease neurodegeneration[94].

There are still many questions remaining to be addressed in future research. For example, it needs to be investigated whether short term or only prolonged physical activity is beneficial in ApoE ε4 carriers and at what age it can still restore the metabolic consequences of ApoE ε4. As only life-long high aerobic exercise exerts a protective effect, the overall health benefits of increasing physical activity at high age might be lower than those in younger individuals[73]. In this review, we described the effects of unsaturated fatty acids intake and physical activity separately. However, further research should also warrant attention to the combined effect of these lifestyle factors to disentangle those mechanisms. For example, a synergistic effect of these two lifestyle-factors on ApoE ε4-related disorders might exist, or one of the two might have a higher impact on these outcomes. Not only omega-3 fatty acids, but also other macro- and micronutrients might be of interest in relation to healthy aging in ApoE ε4 carriers and the general population, here further research is also warranted[19, 95]. Moreover, this review only focused on two

lifestyle-related factors. However, other lifestyle-factors (e.g. sleep), but also cultural- and environmental factors and medication use may modify ApoE ε4 related effects. A recent trial in high-risk individuals investigating the effect of a multidomain lifestyle intervention program on cognition in different *APOE* genotype subgroups did not show specific beneficial effects on cognition in ApoE ε4 carriers[96]. However, sample size and follow-up duration might have been limited. We acknowledge, however, that in general it is very difficult for individuals to alter their lifestyle, and adherence to the intervention might be an issue to longer follow-ups. Alternatively, medication has been suggested to specifically target the mechanisms described in this review. For example, it has been described that drugs like CETP inhibitors and APOC3 antisense may work through similar pathways as described in this narrative review. The efficacy and safety of APOC3 antisense for the treatment of hypertriglyceridemia is currently being tested in phase 3 trials[97]. CETP inhibitors, however, have not been able to demonstrate clinical benefit and were found to have effects that are modest at best in phase 3 clinical trials[97, 98]. An important area of future research comprises assessment of interactions between genes, lifestyle and medication use. Especially in a medical world trying to de-prescribe, these future studies focusing on lifestyle and its interactions may have considerable value.

CONCLUDING REMARKS

In this review, we discussed lifestyle-related factors and their contribution to the effects of genetic variation in the *APOE* gene on age-related diseases. This review provides an overview of the current literature, however, some limitations should be mentioned. First, because of the consistently growing body of evidence regarding this topic we may have missed important results that could influence our conclusions. Moreover, the authors are aware that the discussed epidemiological studies differ in their design and their study population, which may therefore cause the results to not be directly comparable. For example, there may be differences in the methods of administering omega-3 fatty acids (EPA, DHA, fish oil etc.) and in the amount of time exposed to physical activity (long-term, high-intensity, low-intensity etc.). In addition, most studies included in this narrative review did not take into account the effect of *APOE* genotype heterozygosity, which have been found previously to be of importance[6, 10]. However, studies addressing the relation of *APOE* genotype heterozygosity with omega-3 fatty acid intake and physical activity are scarce. Further research should consider this heterozygosity in relation to lifestyle factors and disease risk.

Taken together, an increasing body of evidence suggests a protective role for omega-3 fatty acids and physical activity in carriers of the ε4 allele. The risks associated with the ApoE ε4 isoform consist of several components that jointly contribute to disease onset. By modifying the risk of the ApoE ε4 isoform, disease burden associated with this risk allele might be decreased in the general population. This information is of interest as it now seems that the risks associated with the ApoE ε4 isoform are modifiable which may stimulate risk-reducing behaviors.

REFERENCES

1. Deelen J, Beekman M, Uh HW, Broer L, Ayers KL, Tan Q, et al. Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. Human molecular genetics. 2014;23(16):4420-32.
2. Pilling LC, Atkins JL, Bowman K, Jones SE, Tyrrell J, Beaumont RN, et al. Human longevity is influenced by many genetic variants: evidence from 75,000 UK Biobank participants. Aging. 2016;8(3):547-60.
3. Eisenberg DT, Kuzawa CW, Hayes MG. Worldwide allele frequencies of the human apolipoprotein E gene: climate, local adaptations, and evolutionary history. American journal of physical anthropology. 2010;143(1):100-11.
4. Vaarhorst AA, Beekman M, Suchiman EH, van Heemst D, Houwing-Duistermaat JJ, Westendorp RG, et al. Lipid metabolism in long-lived families: the Leiden Longevity Study. Age (Dordr). 2011;33(2):219-27.
5. Noordam R, Oudt CH, Deelen J, Slagboom PE, Beekman M, van Heemst D. Assessment of the contribution of APOE gene variants to metabolic phenotypes associated with familial longevity at middle age. Aging. 2016;8(8):1790-801.
6. Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. Jama. 1997;278(16):1349-56.
7. Postmus I, Trompet S, Deshmukh HA, Barnes MR, Li X, Warren HR, et al. Pharmacogenetic meta-analysis of genome-wide association studies of LDL cholesterol response to statins. Nat Commun. 2014;5:5068.
8. Smit RA, Postmus I, Trompet S, Barnes MR, Warren H, Arsenault BJ, et al. Rooted in risk: genetic predisposition for low-density lipoprotein cholesterol level associates with diminished low-density lipoprotein cholesterol response to statin treatment. Pharmacogenomics. 2016;17(15):1621-8.
9. Snowdon DA, Greiner LH, Mortimer JA, Riley KP, Greiner PA, Markesbery WR. Brain infarction and the clinical expression of Alzheimer disease. The Nun Study. Jama. 1997;277(10):813-7.
10. Song Y, Stampfer MJ, Liu S. Meta-analysis: apolipoprotein E genotypes and risk for coronary heart disease. Ann Intern Med. 2004;141(2):137-47.
11. Chen CH, Mizuno T, Elston R, Kariuki MM, Hall K, Unverzagt F, et al. A comparative study to screen dementia and APOE genotypes in an ageing East African population. Neurobiol Aging. 2010;31(5):732-40.
12. Gureje O, Ogunniyi A, Baiyewu O, Price B, Unverzagt FW, Evans RM, et al. APOE epsilon4 is not associated with Alzheimer's disease in elderly Nigerians. Annals of neurology. 2006;59(1):182-5.
13. Notkola IL, Sulkava R, Pekkanen J, Erkinjuntti T, Ehnholm C, Kivinen P, et al. Serum total cholesterol, apolipoprotein E epsilon 4 allele, and Alzheimer's disease. Neuroepidemiology. 1998;17(1):14-20.
14. Liu CC, Kanekiyo T, Xu H, Bu G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. Nature reviews Neurology. 2013;9(2):106-18.
15. Segrest JP, Jones MK, De Loof H, Brouillette CG, Venkatachalapathi YV, Anantharamaiah GM. The amphipathic helix in the exchangeable apolipoproteins: a review of secondary structure and function. Journal of lipid research. 1992;33(2):141-66.

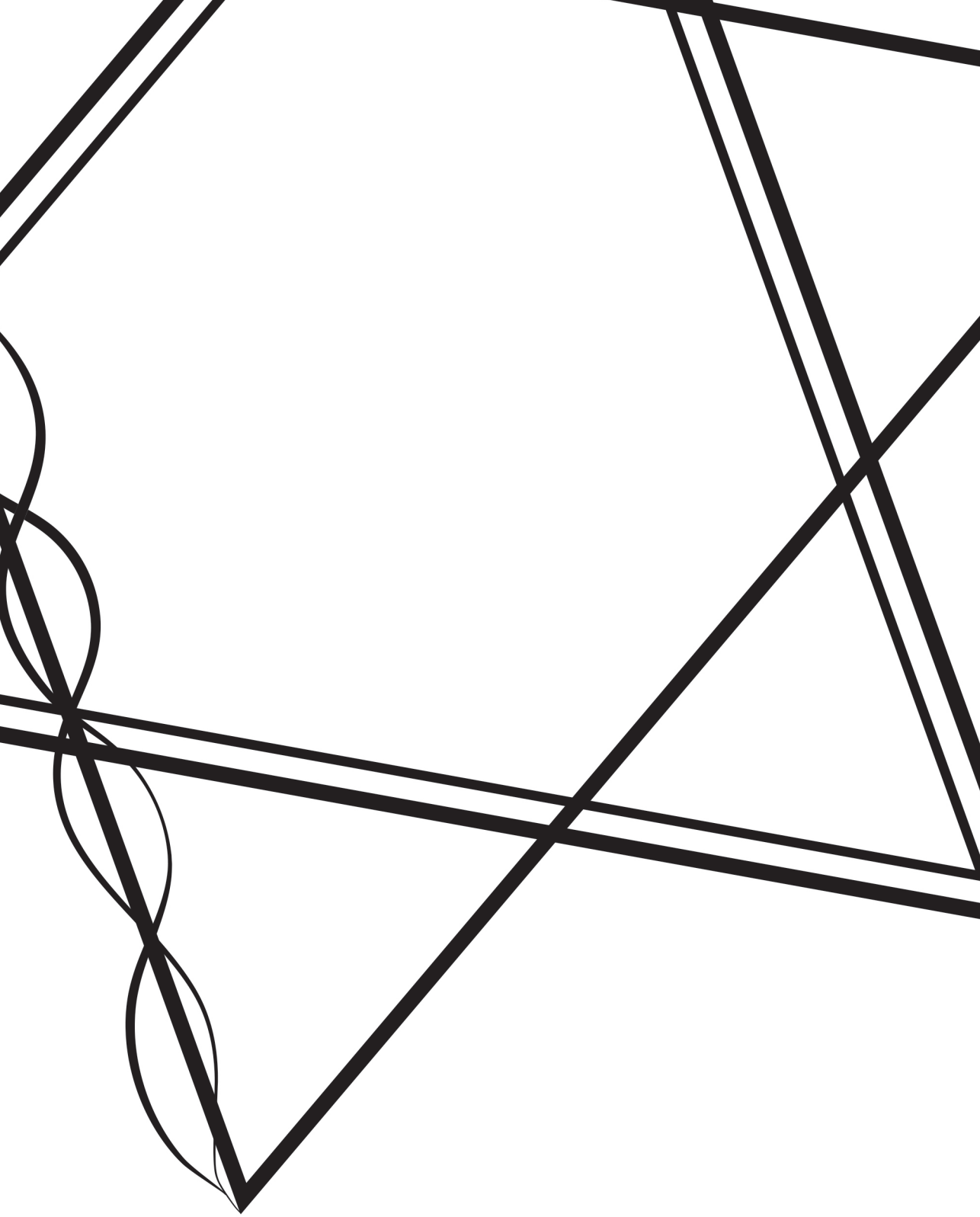
16. Phillips MC. Apolipoprotein E isoforms and lipoprotein metabolism. IUBMB life. 2014;66(9):616-23.
17. Mensenkamp AR, Jong MC, van Goor H, van Luyn MJ, Bloks V, Havinga R, et al. Apolipoprotein E participates in the regulation of very low density lipoprotein-triglyceride secretion by the liver. J Biol Chem. 1999;274(50):35711-8.
18. Mahley RW, Huang Y. Apolipoprotein e sets the stage: response to injury triggers neuropathology. Neuron. 2012;76(5):871-85.
19. Huebbe P, Nebel A, Siegert S, Moehring J, Boesch-Saadatmandi C, Most E, et al. APOE epsilon4 is associated with higher vitamin D levels in targeted replacement mice and humans. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 2011;25(9):3262-70.
20. Rensen PC, van Berkel TJ. Apolipoprotein E effectively inhibits lipoprotein lipase-mediated lipolysis of chylomicron-like triglyceride-rich lipid emulsions in vitro and in vivo. J Biol Chem. 1996;271(25):14791-9.
21. Stanford KI, Bishop JR, Foley EM, Gonzales JC, Niesman IR, Witztum JL, et al. Syndecan-1 is the primary heparan sulfate proteoglycan mediating hepatic clearance of triglyceride-rich lipoproteins in mice. J Clin Invest. 2009;119(11):3236-45.
22. Steinmetz A, Jakobs C, Motzny S, Kaffarnik H. Differential distribution of apolipoprotein E isoforms in human plasma lipoproteins. Arteriosclerosis. 1989;9(3):405-11.
23. Weisgraber KH. Apolipoprotein E distribution among human plasma lipoproteins: role of the cysteine-arginine interchange at residue 112. J Lipid Res. 1990;31(8):1503-11.
24. Mahley RW, Rall SC, Jr. Apolipoprotein E: far more than a lipid transport protein. Annu Rev Genomics Hum Genet. 2000;1:507-37.
25. Davignon J, Gregg RE, Sing CF. Apolipoprotein E polymorphism and atherosclerosis. Arteriosclerosis. 1988;8(1):1-21.
26. Zekraoui L, Lagarde JP, Raisonnier A, Gerard N, Aouizerate A, Lucotte G. High frequency of the apolipoprotein E *4 allele in African pygmies and most of the African populations in sub-Saharan Africa. Human biology. 1997;69(4):575-81.
27. Kamboh MI. Apolipoprotein E polymorphism and susceptibility to Alzheimer's disease. Human biology. 1995;67(2):195-215.
28. Ewbank DC. The APOE gene and differences in life expectancy in Europe. J Gerontol A Biol Sci Med Sci. 2004;59(1):16-20.
29. Crean S, Ward A, Mercaldi CJ, Collins JM, Cook MN, Baker NL, et al. Apolipoprotein E epsilon4 prevalence in Alzheimer's disease patients varies across global populations: a systematic literature review and meta-analysis. Dement Geriatr Cogn Disord. 2011;31(1):20-30.
30. Hui DY, Innerarity TL, Mahley RW. Defective hepatic lipoprotein receptor binding of beta-very low density lipoproteins from type III hyperlipoproteinemic patients. Importance of apolipoprotein E. J Biol Chem. 1984;259(2):860-9.
31. Weisgraber KH, Innerarity TL, Mahley RW. Abnormal lipoprotein receptor-binding activity of the human E apoprotein due to cysteine-arginine interchange at a single site. J Biol Chem. 1982;257(5):2518-21.
32. Eichner JE, Dunn ST, Perveen G, Thompson DM, Stewart KE, Stroehla BC. Apolipoprotein E polymorphism and cardiovascular disease: a HuGE review. American journal of epidemiology. 2002;155(6):487-95.
33. Utermann G, Hees M, Steinmetz A. Polymorphism of apolipoprotein E and occurrence of dysbetalipoproteinaemia in man. Nature. 1977;269(5629):604-7.

34. Suri S, Heise V, Trachtenberg AJ, Mackay CE. The forgotten APOE allele: a review of the evidence and suggested mechanisms for the protective effect of APOE varepsilon2. Neurosci Biobehav Rev. 2013;37(10 Pt 2):2878-86.
35. Acharya P, Segall ML, Zaiou M, Morrow J, Weisgraber KH, Phillips MC, et al. Comparison of the stabilities and unfolding pathways of human apolipoprotein E isoforms by differential scanning calorimetry and circular dichroism. Biochimica et biophysica acta. 2002;1584(1):9-19.
36. Huang Y, Liu XQ, Wyss-Coray T, Brecht WJ, Sanan DA, Mahley RW. Apolipoprotein E fragments present in Alzheimer's disease brains induce neurofibrillary tangle-like intracellular inclusions in neurons. Proceedings of the National Academy of Sciences of the United States of America. 2001;98(15):8838-43.
37. Harris FM, Brecht WJ, Xu Q, Tesseur I, Kekonius L, Wyss-Coray T, et al. Carboxyl-terminal-truncated apolipoprotein E4 causes Alzheimer's disease-like neurodegeneration and behavioral deficits in transgenic mice. Proceedings of the National Academy of Sciences of the United States of America. 2003;100(19):10966-71.
38. Brecht WJ, Harris FM, Chang S, Tesseur I, Yu GQ, Xu Q, et al. Neuron-specific apolipoprotein e4 proteolysis is associated with increased tau phosphorylation in brains of transgenic mice. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2004;24(10):2527-34.
39. Heeren J, Grewal T, Laatsch A, Becker N, Rinninger F, Rye KA, et al. Impaired recycling of apolipoprotein E4 is associated with intracellular cholesterol accumulation. J Biol Chem. 2004;279(53):55483-92.
40. Chen GC, Yang J, Eggersdorfer M, Zhang W, Qin LQ. N-3 long-chain polyunsaturated fatty acids and risk of all-cause mortality among general populations: a meta-analysis. Scientific reports. 2016;6:28165.
41. Gomez-Pinilla F. Brain foods: the effects of nutrients on brain function. Nature reviews Neuroscience. 2008;9(7):568-78.
42. Cole GM, Ma QL, Frautschy SA. Omega-3 fatty acids and dementia. Prostaglandins, leukotrienes, and essential fatty acids. 2009;81(2-3):213-21.
43. Kariv-Inbal Z, Yacobson S, Berkecz R, Peter M, Janaky T, Lutjohann D, et al. The isoform-specific pathological effects of apoE4 in vivo are prevented by a fish oil (DHA) diet and are modified by cholesterol. Journal of Alzheimer's disease : JAD. 2012;28(3):667-83.
44. Nurk E, Drevon CA, Refsum H, Solvoll K, Vollset SE, Nygard O, et al. Cognitive performance among the elderly and dietary fish intake: the Hordaland Health Study. The American journal of clinical nutrition. 2007;86(5):1470-8.
45. Harris WS, Miller M, Tighe AP, Davidson MH, Schaefer EJ. Omega-3 fatty acids and coronary heart disease risk: clinical and mechanistic perspectives. Atherosclerosis. 2008;197(1):12-24.
46. Lee JH, O'Keefe JH, Lavie CJ, Marchioli R, Harris WS. Omega-3 fatty acids for cardioprotection. Mayo Clinic proceedings. 2008;83(3):324-32.
47. Buoite Stella A, Gortan Cappellari G, Barazzoni R, Zanetti M. Update on the Impact of Omega 3 Fatty Acids on Inflammation, Insulin Resistance and Sarcopenia: A Review. Int J Mol Sci. 2018;19(1).
48. Smith GI, Julliand S, Reeds DN, Sinacore DR, Klein S, Mittendorfer B. Fish oil-derived n-3 PUFA therapy increases muscle mass and function in healthy older adults. The American journal of clinical nutrition. 2015;102(1):115-22.
49. Burke MF, Burke FM, Soffer DE. Review of Cardiometabolic Effects of Prescription Omega-3 Fatty Acids. Curr Atheroscler Rep. 2017;19(12):60.

50. Balk EM, Lichtenstein AH. Omega-3 Fatty Acids and Cardiovascular Disease: Summary of the 2016 Agency of Healthcare Research and Quality Evidence Review. Nutrients. 2017;9(8).
51. van de Rest O, Wang Y, Barnes LL, Tangney C, Bennett DA, Morris MC. APOE epsilon4 and the associations of seafood and long-chain omega-3 fatty acids with cognitive decline. Neurology. 2016;86(22):2063-70.
52. Yuen AW, Sander JW, Fluegel D, Patsalos PN, Bell GS, Johnson T, et al. Omega-3 fatty acid supplementation in patients with chronic epilepsy: a randomized trial. Epilepsy & behavior : E&B. 2005;7(2):253-8.
53. Yehuda S, Rabinovitz S, Mostofsky DI. Modulation of learning and neuronal membrane composition in the rat by essential fatty acid preparation: time-course analysis. Neurochemical research. 1998;23(5):627-34.
54. Bourre JM, Francois M, Youyou A, Dumont O, Piciotti M, Pascal G, et al. The effects of dietary alpha-linolenic acid on the composition of nerve membranes, enzymatic activity, amplitude of electrophysiological parameters, resistance to poisons and performance of learning tasks in rats. The Journal of nutrition. 1989;119(12):1880-92.
55. Wang C, Wilson WA, Moore SD, Mace BE, Maeda N, Schmechel DE, et al. Human apoE4-targeted replacement mice display synaptic deficits in the absence of neuropathology. Neurobiology of disease. 2005;18(2):390-8.
56. Covington MB. Omega-3 fatty acids. American family physician. 2004;70(1):133-40.
57. Mahley RW, Weisgraber KH, Huang Y. Apolipoprotein E: structure determines function, from atherosclerosis to Alzheimer's disease to AIDS. Journal of lipid research. 2009;50 Suppl:S183-8.
58. Jofre-Monseny L, Minihane AM, Rimbach G. Impact of apoE genotype on oxidative stress, inflammation and disease risk. Molecular nutrition & food research. 2008;52(1):131-45.
59. Yang ZH, Bando M, Sakurai T, Chen Y, Emma-Okon B, Wilhite B, et al. Long-chain monounsaturated fatty acid-rich fish oil attenuates the development of atherosclerosis in mouse models. Molecular nutrition & food research. 2016.
60. Bensinger SJ, Tontonoz P. Integration of metabolism and inflammation by lipid-activated nuclear receptors. Nature. 2008;454(7203):470-7.
61. Peng DQ, Zhao SP, Nie S, Li J. Gene-gene interaction of PPARgamma and ApoE affects coronary heart disease risk. International journal of cardiology. 2003;92(2-3):257-63.
62. Micallef MA, Garg ML. Anti-inflammatory and cardioprotective effects of n-3 polyunsaturated fatty acids and plant sterols in hyperlipidemic individuals. Atherosclerosis. 2009;204(2):476-82.
63. Raichlen DA, Alexander GE. Exercise, APOE genotype, and the evolution of the human lifespan. Trends in neurosciences. 2014;37(5):247-55.
64. Etnier JL, Nowell PM, Landers DM, Sibley BA. A meta-regression to examine the relationship between aerobic fitness and cognitive performance. Brain research reviews. 2006;52(1):119-30.
65. Angevaren M, Aufdemkampe G, Verhaar HJ, Aleman A, Vanhees L. Physical activity and enhanced fitness to improve cognitive function in older people without known cognitive impairment. The Cochrane database of systematic reviews. 2008(3):CD005381.
66. Szabo AN, McAuley E, Erickson KI, Voss M, Prakash RS, Mailey EL, et al. Cardiorespiratory fitness, hippocampal volume, and frequency of forgetting in older adults. Neuropsychology. 2011;25(5):545-53.
67. Heyn P, Abreu BC, Ottenbacher KJ. The effects of exercise training on elderly persons with cognitive impairment and dementia: a meta-analysis. Archives of physical medicine and rehabilitation. 2004;85(10):1694-704.

68. Head D, Bugg JM, Goate AM, Fagan AM, Mintun MA, Benzinger T, et al. Exercise Engagement as a Moderator of the Effects of APOE Genotype on Amyloid Deposition. Archives of neurology. 2012;69(5):636-43.
69. Woodard JL, Sugarman MA, Nielson KA, Smith JC, Seidenberg M, Durgerian S, et al. Lifestyle and genetic contributions to cognitive decline and hippocampal structure and function in healthy aging. Current Alzheimer research. 2012;9(4):436-46.
70. Rovio S, Kareholt I, Helkala EL, Viitanen M, Winblad B, Tuomilehto J, et al. Leisure-time physical activity at midlife and the risk of dementia and Alzheimer's disease. The Lancet Neurology. 2005;4(11):705-11.
71. Schuit AJ, Feskens EJ, Launer LJ, Kromhout D. Physical activity and cognitive decline, the role of the apolipoprotein e4 allele. Medicine and science in sports and exercise. 2001;33(5):772-7.
72. Smith JC, Nielson KA, Woodard JL, Seidenberg M, Rao SM. Physical activity and brain function in older adults at increased risk for Alzheimer's disease. Brain sciences. 2013;3(1):54-83.
73. Mikkelsen UR, Couppe C, Karlsen A, Grosset JF, Schjerling P, Mackey AL, et al. Life-long endurance exercise in humans: circulating levels of inflammatory markers and leg muscle size. Mechanisms of ageing and development. 2013;134(11-12):531-40.
74. Wei M, Macera CA, Hornung CA, Blair SN. Changes in lipids associated with change in regular exercise in free-living men. Journal of clinical epidemiology. 1997;50(10):1137-42.
75. Thompson PD, Yurgalevitch SM, Flynn MM, Zmuda JM, Spannaus-Martin D, Saritelli A, et al. Effect of prolonged exercise training without weight loss on high-density lipoprotein metabolism in overweight men. Metabolism: clinical and experimental. 1997;46(2):217-23.
76. Bernstein MS, Costanza MC, James RW, Morris MA, Cambien F, Raoux S, et al. Physical activity may modulate effects of ApoE genotype on lipid profile. Arterioscler Thromb Vasc Biol. 2002;22(1):133-40.
77. Hagberg JM, Wilund KR, Ferrell RE. APO E gene and gene-environment effects on plasma lipoprotein-lipid levels. Physiol Genomics. 2000;4(2):101-8.
78. Thompson PD, Tsongalis GJ, Seip RL, Bilbie C, Miles M, Zoeller R, et al. Apolipoprotein E genotype and changes in serum lipids and maximal oxygen uptake with exercise training. Metabolism: clinical and experimental. 2004;53(2):193-202.
79. Hoeke G, Nahon KJ, Bakker LEH, Norkauer SSC, Dinnes DLM, Kockx M, et al. Short-term cooling increases serum triglycerides and small high-density lipoprotein levels in humans. J Clin Lipidol. 2017;11(4):920-8 e2.
80. Seip RL, Otvos J, Bilbie C, Tsongalis GJ, Miles M, Zoeller R, et al. The effect of apolipoprotein E genotype on serum lipoprotein particle response to exercise. Atherosclerosis. 2006;188(1):126-33.
81. Simopoulos AP. An Increase in the Omega-6/Omega-3 Fatty Acid Ratio Increases the Risk for Obesity. Nutrients. 2016;8(3):128.
82. Berlingeri P, Cunningham N, Taylor DM, Knott J, McLean D, Gavan R, et al. Adherence to national exercise guidelines by patients attending emergency departments: A multi-site survey. Emerg Med Australas. 2017;29(3):276-82.
83. Shakersain B, Santoni G, Larsson SC, Faxen-Irving G, Fastbom J, Fratiglioni L, et al. Prudent diet may attenuate the adverse effects of Western diet on cognitive decline. Alzheimer's & dementia : the journal of the Alzheimer's Association. 2016;12(2):100-9.
84. Patterson E, Wall R, Fitzgerald GF, Ross RP, Stanton C. Health implications of high dietary omega-6 polyunsaturated Fatty acids. Journal of nutrition and metabolism. 2012;2012:539426.

85. Kolahdooz F, Ibiebele TI, van der Pols JC, Webb PM. Dietary patterns and ovarian cancer risk. The American journal of clinical nutrition. 2009;89(1):297-304.
86. Koopman JJ, van Bodegom D, Ziem JB, Westendorp RG. An Emerging Epidemic of Noncommunicable Diseases in Developing Populations Due to a Triple Evolutionary Mismatch. The American journal of tropical medicine and hygiene. 2016;94(6):1189-92.
87. van Exel E, Eikelenboom P, Comijs H, Frolich M, Smit JH, Stek ML, et al. Vascular factors and markers of inflammation in offspring with a parental history of late-onset Alzheimer disease. Arch Gen Psychiatry. 2009;66(11):1263-70.
88. van de Rest O, Schutte BA, Deelen J, Stassen SA, van den Akker EB, van Heemst D, et al. Metabolic effects of a 13-weeks lifestyle intervention in older adults: The Growing Old Together Study. Aging. 2016;8(1):111-26.
89. Wijsman CA, Westendorp RG, Verhagen EA, Catt M, Slagboom PE, de Craen AJ, et al. Effects of a web-based intervention on physical activity and metabolism in older adults: randomized controlled trial. Journal of medical Internet research. 2013;15(11):e233.
90. Willett WC, Sacks F, Trichopoulou A, Drescher G, Ferro-Luzzi A, Helsing E, et al. Mediterranean diet pyramid: a cultural model for healthy eating. The American journal of clinical nutrition. 1995;61(6 Suppl):1402S-6S.
91. Scarmeas N, Stern Y, Mayeux R, Manly JJ, Schupf N, Luchsinger JA. Mediterranean diet and mild cognitive impairment. Archives of neurology. 2009;66(2):216-25.
92. Lourida I, Soni M, Thompson-Coon J, Purandare N, Lang IA, Ukoumunne OC, et al. Mediterranean diet, cognitive function, and dementia: a systematic review. Epidemiology. 2013;24(4):479-89.
93. Roman B, Carta L, Martinez-Gonzalez MA, Serra-Majem L. Effectiveness of the Mediterranean diet in the elderly. Clinical interventions in aging. 2008;3(1):97-109.
94. Valls-Pedret C, Sala-Vila A, Serra-Mir M, Corella D, de la Torre R, Martinez-Gonzalez MA, et al. Mediterranean Diet and Age-Related Cognitive Decline: A Randomized Clinical Trial. JAMA internal medicine. 2015;175(7):1094-103.
95. Mitter SS, Oria RB, Kvalsund MP, Pamplona P, Joventino ES, Mota RM, et al. Apolipoprotein E4 influences growth and cognitive responses to micronutrient supplementation in shantytown children from northeast Brazil. Clinics (Sao Paulo). 2012;67(1):11-8.
96. Solomon A, Turunen H, Ngandu T, Peltonen M, Levalahti E, Helisalmi S, et al. Effect of the Apolipoprotein E Genotype on Cognitive Change During a Multidomain Lifestyle Intervention: A Subgroup Analysis of a Randomized Clinical Trial. JAMA neurology. 2018;75(4):462-70.
97. Schmitz J, Gouni-Berthold I. APOC-III Antisense Oligonucleotides: A New Option for the Treatment of Hypertriglyceridemia. Curr Med Chem. 2018;25(13):1567-76.
98. Di Bartolo B, Takata K, Duong M, Nicholls SJ. CETP Inhibition in CVD Prevention: an Actual Appraisal. Curr Cardiol Rep. 2016;18(5):43.



CHAPTER 4.2

Apolipoprotein E genotype, lifestyle and coronary artery disease: gene-environment interaction analyses in the UK Biobank population

Maxime M Bos, Lina de Vries, Patrick CN Rensen, Ko Willems van Dijk, Gerard Jan Blauw, Diana van Heemst, Raymond Noordam

Submitted

ABSTRACT

Carriers of the *APOE ε4* genotype have an increased risk for developing coronary artery disease (CAD), but there is preliminary evidence that lifestyle factors interact with *APOE* genotype on CAD risk. Here, we assessed the interactions of physical activity, oily fish intake and polyunsaturated fatty acid (PUFA) intake with *APOE* genotype on risk of incident cardiovascular disease in a large population of middle-aged individuals. The present study was embedded in the UK Biobank population and comprised 344,092 European participants (mean age: 56.5 years, 45.7% men) without a history of CAD. Information regarding physical activity, oily fish intake and PUFA intake was collected through questionnaires, and information on incident CAD through linkage with hospital admission records. Analyses were performed using Cox proportional hazard models adjusted for age and sex. A higher physical activity level and a higher intake of oily fish were associated with a lower incidence of CAD. These associations were similar across all APOE isoform groups (p-values for interaction > 0.05). A higher PUFA intake was only associated with a lower CAD risk in *APOE ε4* carriers (hazard ratio: 0.76, 95% confidence interval: 0.62 – 0.90), however, no statistically significant interaction was observed ($p\text{-value}_{interaction}$ = 0.137). While higher physical activity, fish intake and PUFA intake all decreased the risk of CAD, no evidence for interaction of these lifestyle factors with *APOE* genotype was observed in UK Biobank participants. Interventions intended to reduce cardiovascular risk might therefore be similarly effective across the APOE isoform carriers.

INTRODUCTION

Despite the introduction of cholesterol-lowering medication and improved revascularization treatments, coronary artery disease (CAD) is still one of the most common causes of morbidity and mortality in the general population[1]. Much of the research in this area is focused on disentangling the pathophysiology of CAD and on the identification of novel targets for disease prevention. Large initiatives have been undertaken to investigate the genetics of CAD pathogenesis[2]. Genetic variation in the *APOE* gene has been widely recognized to increase the risk of CAD, which has also been confirmed by genome-wide association studies[2-4]. There is a standing notion that interactions between lifestyle and genetics may affect the response to cholesterol lowering medication and the susceptibility to CAD [5].

In order to explore the biological mechanisms via which lifestyle may modify disease traits, gene-environment interactions have been investigated[6-10]. We recently reviewed the current evidence on the existence of *APOE*-lifestyle interactions in the development of age-related diseases, including CAD, and argued that the beneficial effect of high physical activity and a high intake of oily fish might be largest for *APOE* ε4 carriers[11]. A population-based cross-sectional survey performed in 1,708 randomly selected participants aged 35 to 74 years, showed that a high intensity level of physical activity was associated with an increase of high-density lipoprotein (HDL) cholesterol and a decrease of triglyceride levels, specifically in *APOE* ε4 carriers[12]. Furthermore, It has been hypothesized that specifically PUFA intake may have a beneficial effect in *APOE* ε4 carriers. Small-scale intervention studies indeed indicate that polyunsaturated fatty acid (PUFA) intake and physical activity may have specific beneficial effects on CAD (risk factors) in *APOE* ε4 carriers[13-15]. However, a study done in 136,701 high CAD risk participants did not indicate that the *APOE*-lipid level association was modified by a high omega-3 intake[16].

Given these contrasting results, there is a need for large scale studies examining effect modification of the *APOE* gene by lifestyle factors such as oily fish intake, PUFA and physical activity on incident CAD. Recently, large studies like the UK Biobank cohort that are well-powered and have data available on large numbers of incident CAD cases have become available as a platform to address such research questions In the present study, we assessed whether the association between *APOE* genotype and incident CAD risk is modified by oily fish and PUFA intake and physical activity levels in middle-aged individuals without a history of CAD from the UK Biobank cohort.

METHODS

Study setting and population

The UK Biobank cohort is a prospective general population cohort. Baseline assessments took place between 2006 and 2010 in 22 different assessment centers across the United Kingdom [17]. A total of 502,628 participants between the age of 40 and 70 years were recruited from the general population. Invitation letters were sent to eligible adults registered to the National Health Services (NHS) and living within a 25 miles distances from one of the study assessment centers. At the study assessment center, participants completed a questionnaire through touchscreen that included topics as sociodemographic characteristics, physical and mental health, lifestyle and habitual food intake. All participants from the UK Biobank cohort provided written informed consent, and the study was approved by the medical ethics committee. The project was completed under project number 32292.

In the present study, genotyped European-ancestry participants without a history of cardiovascular disease were followed till the development of the study outcome, death or the end of the study period (March 31, 2017; N = 363,745). Participants with missing data on questionnaire-based oily fish intake (N = 1,654) and physical activity frequency per week (N = 16,432) were excluded from the study. In total, the present study was conducted in 345,659 participants. Additionally, we performed analyses in a subsample of the study population with data on polyunsaturated fatty acids (PUFA) intake (N = 52,478).

APOE genotyping

UK Biobank genotyping was conducted by Affymetrix using a bespoke BiLEVE Axium array for approximately 50,000 participants; the remaining participants were genotyped using the Affymetrix UK Biobank Axiom array. All genetic data were quality controlled centrally by UK Biobank resources. More information on the genotyping processes can be found online (https://www.ukbiobank.ac.uk). SNPs in *APOE* determining the isoform (notably rs7412 and rs429358) were directly genotyped. Both genetic variants were in Hardy-Weinberg equilibrium (p-value > 0.05). As a reference group, we used participants who are homozygous for the *APOE* ε3 allele (ε3/ε3 genotype). The *APOE* ε4 group consisted of individuals with the genotype ε3/ε4 and ε4/ε4. The *APOE* ε2 group consisted of individuals with the genotype ε2/ε3 and ε2/ε2. We excluded participants with other (rarer) genotypes (e.g. ε2/ε4).

Lifestyle exposures

Via touchscreen questionnaires, information on the frequency of oily fish intake and physical activity per week was collected[18]. Via the same questionnaires, we determined the number of days per week at which participants had more than 10 minutes of vigorous physical activity[19], which was defined as "doing physical activity that made you sweat or breathe hard". Groups for oily fish intake and physical activity were divided based on whether individuals reported oily fish intake of at least once per week or were active for at least one day per week. In order to assess whether there is a dose-response relationship for oily fish intake and physical activity on a lower CAD incidence, we formed groups based on whether individuals had oily fish intake or physical activity only once, twice, or three or more times per week. As a reference group we used those individuals who reported to not have any intake of oily fish or were not active for any day of the week.

In a subset of the population, the frequency of intake of 200 consumed food items and drinks over the previous 24 hours was collected with a 24-hour dietary recall questionnaire (25) based on which the average intake of macro- and micronutrients was calculated. Information regarding polyunsaturated fatty acid (PUFA) intake was obtained via this questionnaire. The groups for PUFA intake were based on the median PUFA intake in which a higher than median PUFA intake was considered high and a lower than the median intake was considered as low and used as a reference group in our analyses.

Cardiovascular disease outcomes

Information on incident cardiovascular disease was collected through information from the data provided by the NHS record systems. Diagnoses were coded according to the International Classification of Diseases (ICD)[17]. Here, the study outcome was CAD which we defined as: angina pectoris (I20), myocardial infarction (I21), acute and chronic ischemic heart disease (I24 and I25) and stroke (I63 and I64).

Covariates

Body Mass Index (BMI) was calculated by dividing the weight in kilograms by the height in meters squared, which were measured objectively at the study center. Participants were asked to remove shoes and heavy outer clothing before weighting. During the visit of the assessment center, participants completed touchscreen questionnaires regarding smoking status (never, previous or current), frequency of alcohol consumption, disease status (e.g. diabetes mellitus) and medication usage (lipid lowering medication and blood pressure lowering medication).

Statistical analyses

Characteristics of the study population were examined at baseline and expressed as means (standard deviations), medians (interquartile ranges; for non-normally distributed variables only), and proportions.

We examined the association between physical activity, oily fish intake and PUFA intake on incident CAD in a population without a history of CAD using cox proportional hazard models adjusted for age and sex in R (version 3.6.1) using the survival package (version 2.44-1.1)[20, 21]. Results were visualized using the R-based packages ggplot2, survminer and the metafor package[22-24]. Participants were followed till the first CAD event, death or the end of the study period (March 31, 2017), whichever came first. Results are presented as the hazard ratios with the accompanying 95% confidence intervals. Since individuals who are at high-risk for development of CAD may alter their lifestyle, in a sensitivity analysis we excluded participants who used lipid lowering medication, blood pressure lowering medication or had a clinical diagnosis of diabetes mellitus. In order to formally test for an interaction, we added an interaction term between the *APOE* genotype and the lifestyle factor to the model.

RESULTS

Characteristics of the study population

A total of 345,659 white Caucasian participants were included in the present study. In **Table 1**, the population characteristics for the study population are presented stratified by APOE isoform. In general, study characteristics are comparable between subgroups. However, statin use, total cholesterol levels and LDL-cholesterol are lower in individuals with the *APOE* ε2 genotype as compared to *APOE* ε3 carriers and higher for *APOE* ε4 carriers. In **Supplementary Table 1**, the population characteristics are presented when stratified based on lifestyle factors. Individuals with a low fish intake are slightly younger than those with a high fish intake. Moreover, individuals with a high physical activity level are more often males, have a lower statin use, a lower use of blood pressure lowering medication and have less diabetes as compared to individuals with a low physical activity level.

Incident cardiovascular disease per APOE isoform

A total of 12,806 participants had an incident event of coronary artery disease (CAD) during a median follow-up period of 8.11 years. **Figure 1** depicts the event-free survival probability per *APOE* isoform. The highest event-free survival probability was observed in ApoE ε2 carriers and the lowest probability in ApoE ε4 carriers with accompanying hazard ratios of 0.93 (95% confidence interval (CI) 0.88 – 0.98) and 1.10 (95%CI: 1.06 – 1.14).

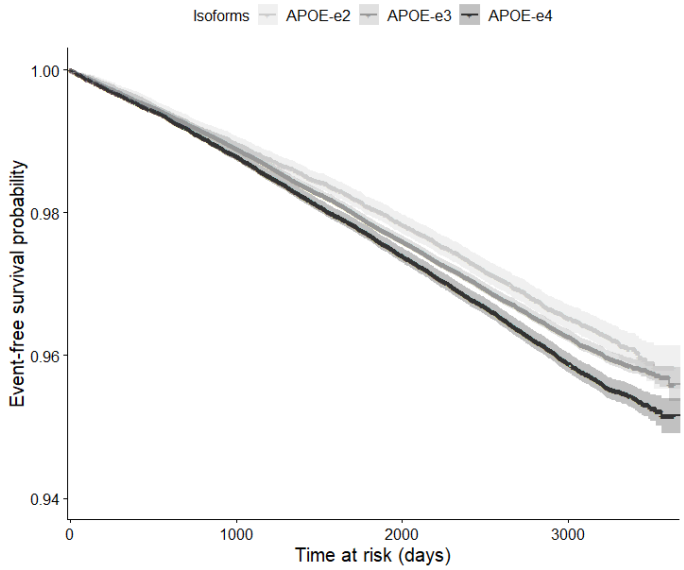


Figure 1. Cardiovascular disease-free survival per *APOE* isoform.

Table 1. *Baseline characteristics of the participants in the UK Biobank, stratified by APOE genotype*

	APOE ε2 *N= 45,570*	**APOE ε3** *N= 207,909*	**APOE ε4** *N= 92,180*
Demographics			
Age in years, mean (SD)	56.5 (8.0)	56.4 (8.0)	56.3 (8.0)
Sex, N (%male)	20,445 (44.9)	92,612 (44.5)	40,930 (44.4)
Lifestyle variables			
BMI, mean (SD)	27.3 (4.7)	27.2 (4.7)	27.1 (4.7)
Smoking status, N (%current)	4,199 (9.2)	19,221 (9.2)	8,236 (8.9)
Oily fish intake in days per week, median (IQR)	2 [1-2]	2 [1-2]	2 [1-2]
10 minutes of vigorous PA in days per week, median (IQR)	1 [0-3]	1 [0-3]	1 [0-3]
Cardiovascular risk factors			
Statin use, N (%yes)	4,613 (10.1)	29,120 (14.0)	15,394 (16.7)
Blood pressure lowering medicines use, N (%yes)	8,039 (17.6)	37,763 (18.2)	16,823 (18.3)
Diabetes, N (%yes)	1,955 (4.3)	8,613 (4.1)	3,492 (3.8)
Cholesterol in mmol/L, mean (SD)	5.4 (1.0)	5.8 (1.1)	5.9 (1.2)
HDL-cholesterol in mmol/L, mean (SD)	1.5 (0.4)	1.5 (0.4)	1.4 (0.4)
LDL-cholesterol in mmol/L, median (IQR)	3.2 [2.7 – 3.7]	3.6 [3.0 – 4.2]	3.7 [3.2 – 4.3]
Triglycerides in mmol/L, median (IQR)	1.5 [1.1 – 2.3]	1.4 [1.0 – 2.1]	1.5 [1.1 – 2.2]

Abbreviations: BMI, body mass index; IQR, interquartile range; N, number; PA, physical activity; SD, standard deviation.

Incident cardiovascular disease per lifestyle group

As shown in **Figure 2**, oily fish intake was associated with a lower incidence of CAD in carriers of the ε2 isoform (hazard ratio (HR): 0.78 [95% confidence interval (CI): 0.66 – 0.91]), in carriers of the ε3 isoform (HR: 0.76 [95%CI: 0.71 – 0.82]), and in carriers of the ε4 isoform (HR: 0.83 [95%CI: 0.74 – 0.92]). Physical activity was associated with a lower CAD incidence in *APOE* ε2 carriers (HR: 0.81 [95%CI: 0.73 – 0.89]), in *APOE* ε3 carriers (HR: 0.77 [95%CI: 0.73 – 0.80]) as well as in ApoE ε4 carriers (HR: 0.76 [95%CI: 0.71 – 0.81]). Here, we did not find evidence for an interaction with fish intake or physical activity with *APOE* genotype on the incidence of CAD. A high PUFA intake was associated with a lower incidence of CAD (HR: 0.76 [95%CI: 0.62 – 0.90], only in ε4 carriers, however no interaction was observed (**Figure 2**). Moreover, no clear dose-response relationship was observed for either fish intake or physical activity and CAD risk (**Supplementary**

Figure 1). However, when the frequency of fish intake or physical activity is taken into account, an interaction for once a week of physical activity in the group with ApoE ε2 and ApoE ε3 and an interaction for once a week of fish intake in the group with ApoE ε3 and ApoE ε4 was observed (**Supplementary Figure 1**).

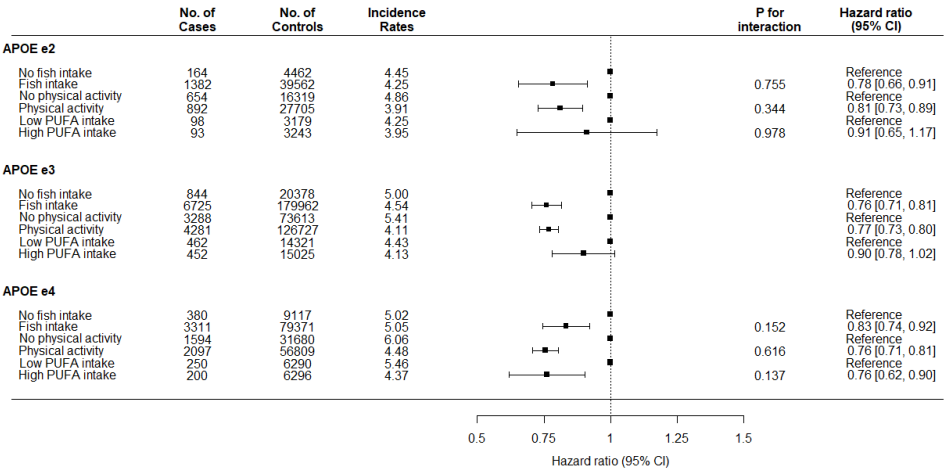


Figure 2. *Hazard ratios for CAD incidence for fish intake, physical activity and polyunsaturated fatty acid (PUFA) intake, stratified by APOE isoform.*

As a sensitivity analysis, we excluded participants who used lipid-lowering medication, blood pressure-lowering medication or with self-reported diabetes mellitus. Results were comparable to those obtained in our previous analyses (**Supplementary Figures 2 and 3**).

DISCUSSION

In the present study, we assessed whether there is evidence for lifestyle-*APOE* interactions on incident CAD in middle-aged individuals of the UK Biobank. Here, we reported no evidence for interaction between fish intake, physical activity and polyunsaturated fatty acid (PUFA) intake with *APOE* genotype on incident CAD in the large European UK Biobank population. However, we showed that, independent of *APOE* genotype, a higher intake of fish and a higher physical activity level both associated with a lower CAD risk. A higher PUFA intake was only associated with a lower CAD risk in *APOE* ε4 carriers, however, no formal statistical interaction was observed.

ApoE ε4 has a different binding affinity for lipoprotein particles or the LDL-receptor than the other isoforms[25]. ApoE ε4 has a higher binding ability for triglyceride-rich lipoproteins, such as chylomicron remnants and very-low-density-lipoprotein (VLDL) particles, than for LDL particles[26]. This results in a diminished clearance of LDL-cholesterol, resulting in higher LDL-cholesterol levels as compared to carriers of APOE ε2 and ε3[27]. We hypothesized that carriers of the ε4 isoform may benefit differently from a higher physical activity and a higher intake of oily fish, however, we did not find evidence supporting this hypothesis. In the present study, for individuals with a high physical activity level, the incidence rate for CAD was lower than for those with a low physical activity level. These lower incidence rates were similar across the different *APOE* carrier groups. Therefore, a higher physical activity is likely to be beneficial irrespective of *APOE* genotype. Moreover, in our study the incidence rate of CAD decreased with a higher oily fish intake, in carriers of all ApoE isoforms and was not higher in carriers of the ε4 isoform. Interestingly, a higher intake of PUFA associated with a lower CAD incidence only in *APOE* ε4 carriers points towards a possible lifestyle-*APOE* interaction on incident CAD. One possible explanation is that intake of PUFA may result in a lower concentration of VLDL and thereby may increase the uptake of LDL by the liver, resulting in lower LDL levels[11]. However, these observations likely have no clear benefit at a clinical level. Since the group with information regarding PUFA intake was relatively small this may have resulted in limited statistical power and, therefore, further research to investigate the effect on a larger population level is warranted.

Although *APOE*-lifestyle interactions on cognitive function have been hypothesized as well[11], no significant interaction with cognitive function was observed in the UK Biobank population previously[28]. The lack of a *APOE*-lifestyle interactions in this population may be a reflection of some sort of selection bias where the current study population includes in general more healthy *APOE* ε4 carriers. Indeed, there is evidence of a 'healthy volunteer' bias in the UK Biobank sample[29]. Moreover, a genome-wide association

study on habitual physical activity in the UK Biobank identified *APOE* to be one of the strongest associations with physical activity[19]. The association was markedly stronger among older participants, therefore it may be that the older *APOE* risk allele carriers are particularly enriched for healthy lifestyle. One explanation could be that individuals with a known familial history of dementia and cardiovascular disease purposefully increase their physical activity levels and intake of oily fish. This bias may have resulted in an underestimation of the effect of exercise and fish intake. Alternatively, participants of the UK Biobank are still in good health or are slightly too young to show significant effects of the *APOE* genotype on CAD incidence. Moreover, additional lifestyle interactions, not covered in the present study, with *APOE* isoform may be possible. However, it should be noted that recent large-scale genome-wide interaction efforts were not able to identify interactions between physical activity, sleep duration, alcohol intake or smoking on blood lipid levels, limiting the potential of APOE-lifestyle interactions in the pathogenesis of CAD[6-10].

A strength of this study is that we used data from the UK Biobank and therefore, to the best of our knowledge, this is the largest study to test for an interaction of lifestyle factors and *APOE* on incident CAD. However, the current analyses have only been performed in participants of Caucasian descent, thereby hampering the translation to individuals with a different ancestry. This is of particular importance since variability exists in *APOE*-related disease prevalence in different ancestry groups.

Based on our results, it is unlikely that there is a significant interaction between genetic variation in *APOE* and various lifestyle factors on the development of incident CAD in a population of middle-aged and older individuals free of CAD history. Therefore, it seems likely that interventions intended to reduce cardiovascular risk via increased physical activity and increased intake of oily fish will be equally effective in carriers of the APOE ε4 genotype as in the other *APOE* genotype carriers and may thus have considerable health benefits irrespective of *APOE* isoform.

Acknowledgements
The present UK Biobank project was conducted under project number 32292.

Funding
Statistical analyses were performed using the facilities of the Dutch Super computer (Surfsara, Inc; access granted by the Dutch Research Council to RN). DvH was supported by the European Commission funded project HUMAN (Health-2013-INNOVATION-1-602757).

REFERENCES

1. Andersson C, Vasan RS. Epidemiology of cardiovascular disease in young individuals. Nat Rev Cardiol. 2018;15(4):230-40.
2. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet. 2015;47(10):1121-30.
3. CARDIoGRAMplusC4D Consortium, Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. Nature genetics. 2013;45(1):25-33.
4. Song Y, Stampfer MJ, Liu S. Meta-analysis: apolipoprotein E genotypes and risk for coronary heart disease. Ann Intern Med. 2004;141(2):137-47.
5. Hagberg JM, Wilund KR, Ferrell RE. APO E gene and gene-environment effects on plasma lipoprotein-lipid levels. Physiol Genomics. 2000;4(2):101-8.
6. Bentley AR, Sung YJ, Brown MR, Winkler TW, Kraja AT, Ntalla I, et al. Multi-ancestry genome-wide gene-smoking interaction study of 387,272 individuals identifies new loci associated with serum lipids. Nat Genet. 2019;51(4):636-48.
7. de Vries PS, Brown MR, Bentley AR, Sung YJ, Winkler TW, Ntalla I, et al. Multiancestry Genome-Wide Association Study of Lipid Levels Incorporating Gene-Alcohol Interactions. Am J Epidemiol. 2019;188(6):1033-54.
8. Kilpelainen TO, Bentley AR, Noordam R, Sung YJ, Schwander K, Winkler TW, et al. Multi-ancestry study of blood lipid levels identifies four loci interacting with physical activity. Nat Commun. 2019;10(1):376.
9. Rao DC, Sung YJ, Winkler TW, Schwander K, Borecki I, Cupples LA, et al. Multiancestry Study of Gene-Lifestyle Interactions for Cardiovascular Traits in 610 475 Individuals From 124 Cohorts: Design and Rationale. Circ Cardiovasc Genet. 2017;10(3).
10. Noordam R, Bos MM, Wang H, Winkler TW, Bentley AR, Kilpelainen TO, et al. Multi-ancestry sleep-by-SNP interaction analysis in 126,926 individuals reveals lipid loci stratified by sleep duration. Nat Commun. 2019;10(1):5121.
11. Bos MM, Noordam R, Blauw GJ, Slagboom PE, Rensen PCN, van Heemst D. The ApoE epsilon4 Isoform: Can the Risk of Diseases be Reduced by Environmental Factors? J Gerontol A Biol Sci Med Sci. 2019;74(1):99-107.
12. Bernstein MS, Costanza MC, James RW, Morris MA, Cambien F, Raoux S, et al. Physical activity may modulate effects of ApoE genotype on lipid profile. Arterioscler Thromb Vasc Biol. 2002;22(1):133-40.
13. Minihane AM, Khan S, Leigh-Firbank EC, Talmud P, Wright JW, Murphy MC, et al. ApoE polymorphism and fish oil supplementation in subjects with an atherogenic lipoprotein phenotype. Arterioscler Thromb Vasc Biol. 2000;20(8):1990-7.
14. Olano-Martin E, Anil E, Caslake MJ, Packard CJ, Bedford D, Stewart G, et al. Contribution of apolipoprotein E genotype and docosahexaenoic acid to the LDL-cholesterol response to fish oil. Atherosclerosis. 2010;209(1):104-10.
15. Lee JH, Hong SM, Shin YA. Effects of exercise training on stroke risk factors, homocysteine concentration, and cognitive function according the APOE genotype in stroke patients. J Exerc Rehabil. 2018;14(2):267-74.

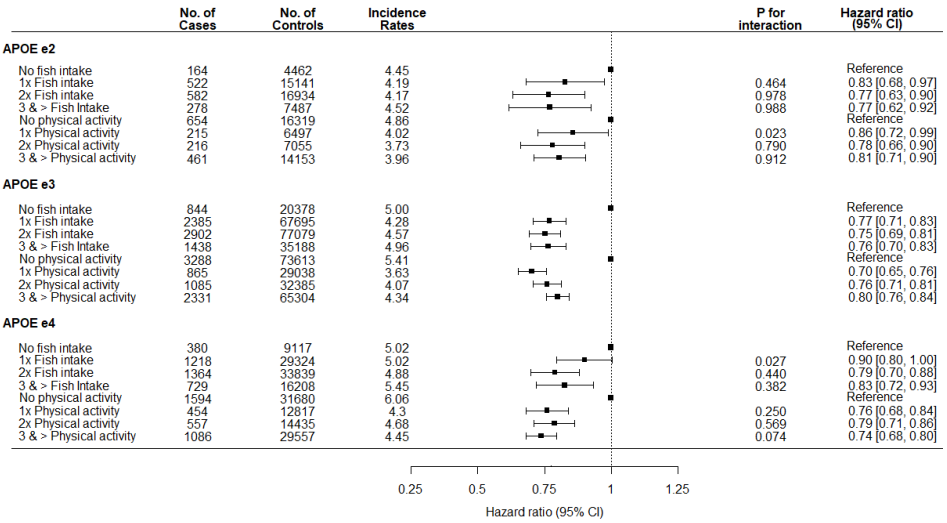
16. Harris WS, Pottala JV, Thiselton DL, S AV, Baedke AM, Dayspring TD, et al. Does APOE genotype modify the relations between serum lipid and erythrocyte omega-3 fatty acid levels? J Cardiovasc Transl Res. 2014;7(5):526-32.
17. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12(3):e1001779.
18. Bradbury KE, Young HJ, Guo W, Key TJ. Dietary assessment in UK Biobank: an evaluation of the performance of the touchscreen dietary questionnaire. J Nutr Sci. 2018;7:e6.
19. Klimentidis YC, Raichlen DA, Bea J, Garcia DO, Wineinger NE, Mandarino LJ, et al. Genome-wide association study of habitual physical activity in over 377,000 UK Biobank participants identifies multiple variants including CADM2 and APOE. Int J Obes (Lond). 2018;42(6):1161-76.
20. Team RC. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. 2019;Vienna, Austria.
21. T T. A Package for Survival Analysis in S. 2015.
22. H. W. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag. 2009.
23. Kosinski AKaM. survminer: Drawing Survival Curves using 'ggplot2'. 2019.
24. Viechtbauer W. Conducting meta-analyses in (R) with the (metafor) package. Journal of Statistical Software. 2010;36(3):1-48.
25. Hui DY, Innerarity TL, Mahley RW. Defective hepatic lipoprotein receptor binding of beta-very low density lipoproteins from type III hyperlipoproteinemic patients. Importance of apolipoprotein E. J Biol Chem. 1984;259(2):860-9.
26. Weisgraber KH. Apolipoprotein E distribution among human plasma lipoproteins: role of the cysteine-arginine interchange at residue 112. J Lipid Res. 1990;31(8):1503-11.
27. Mahley RW, Rall SC, Jr. Apolipoprotein E: far more than a lipid transport protein. Annu Rev Genomics Hum Genet. 2000;1:507-37.
28. Lyall DM, Celis-Morales C, Lyall LM, Graham C, Graham N, Mackay DF, et al. Assessing for interaction between APOE epsilon4, sex, and lifestyle on cognitive abilities. Neurology. 2019;92(23):e2691-e8.
29. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. Am J Epidemiol. 2017;186(9):1026-34.

Supplementary Table 1. *Baseline characteristics of the participants in the UK Biobank, stratified by lifestyle factors.*

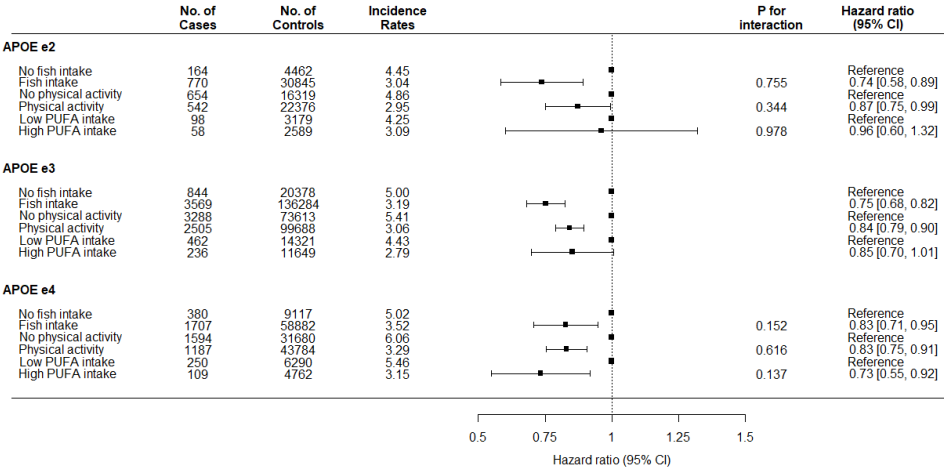
	Low fish intake
	N = 35,344
Demographics	
Age in years, mean (SD)	53.9 (8.1)
Sex, N (%male)	16,581 (46.9)
Lifestyle variables	
BMI, mean (SD)	27.6 (5.1)
Smoking status, N (%current)	
Oily fish intake in days per week, median (IQR)	0 [0 – 0]
10 minutes of vigorous PA in days per week, median (IQR)	1 [0 – 3]
Cardiovascular risk factors	
Statin use, N (%yes)	4,185 (11.8)
Blood pressure lowering medicines use, N (%yes)	5,709 (16.2)
Diabetes, N (%yes)	1,579 (4.5)
Cholesterol in mmol/L, mean (SD)	5.6 (1.1)
HDL-cholesterol in mmol/L, median (IQR)	1.3 [1.1 – 1.6]
LDL-cholesterol in mmol/L, median (IQR)	3.5 [2.9 – 4.1]
Triglycerides in mmol/L, median (IQR)	1.6 [1.1 – 2.3]
Genotype	
APOE ε2, N (%yes)	4,626 (13.1)
APOE ε3, N (%yes)	21,221 (60.0)
APOE ε4, N (%yes)	9,497 (26.9)

Abbreviations: BMI, body mass index; IQR, interquartile range; N, number; PA, physical activity; PUFA, polyunsaturated fatty acids; SD, standard deviation.

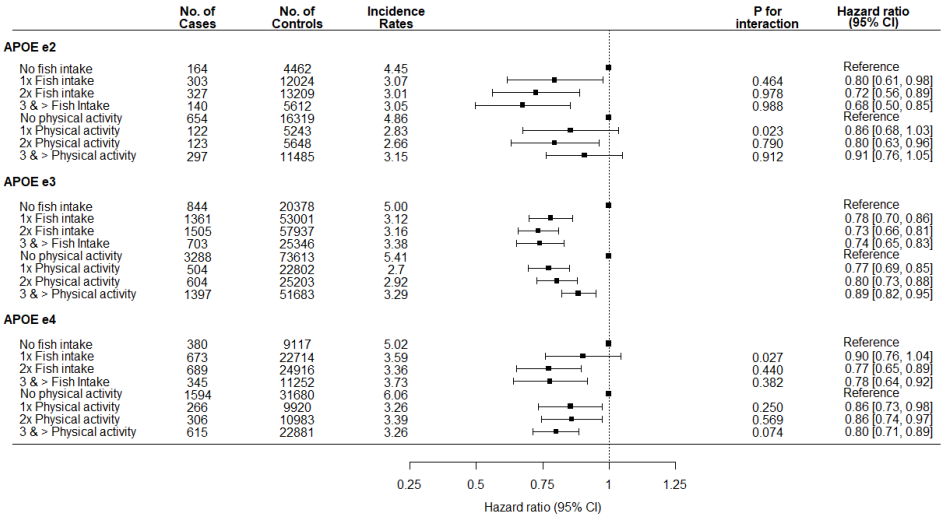
High fish intake *N = 310,313*	Low physical activity *N = 127,146*	High physical activity *N = 218,511*	Low PUFA intake *N = 24,600*	High PUFA intake *N = 25,309*
56.7 (8.0)	57.2 (7.8)	56.0 (8.2)	56.6 (8.0)	55.8 (8.2)
137,406 (44.3)	50,825 (40.0)	103,162 (47.2)	9,680 (39.3)	12,219 (48.3)
27.2 (4.6)	27.9 (5.2)	26.8 (4.3)	27.0 (4.6)	27.0 (4.7)
2 [1 – 2]	2 [1 – 2]	2 [1 – 2]	2 [1 – 2]	1 [1 – 2]
1 [0 – 3]	0 [0 – 0]	3 [2 – 4]	2 [0 – 3]	2 [0 – 3]
44,942 (14.5)	21,861 (17.2)	27,266 (12.5)	3,502 (14.2)	3,425 (13.5)
56,916 (18.3)	27,966 (22.0)	34,659 (15.9)	4,219 (17.2)	4,199 (16.6)
12,481 (4.0)	6,928 (5.4)	7,132 (3.3)	943 (3.8)	961 (3.8)
5.8 (1.1)	5.8 (1.1)	5.8 (1.1)	5.8 (1.1)	5.7 (1.1)
1.4 [1.2 – 1.7]	1.4 [1.2 – 1.7]	1.4 [1.2 – 1.7]	1.5 [1.2 – 1.7]	1.4 [1.2 – 1.7]
3.6 [3.0 – 4.2]	3.6 [3.0 – 4.2]	3.6 [3.0 – 4.1]	3.6 [3.0 – 4.2]	3.6 [3.0 – 4.1]
1.5 [1.0 – 2.1]	1.5 [1.1 – 2.2]	1.4 [1.0 – 2.1]	1.4 [1.0 – 2.0]	1.4 [1.0 – 2.1]
40,944 (13.2)	16,973 (13.3)	28,597 (13.1)	3,277 (13.3)	3,336 (13.2)
186,687 (60.2)	76,900 (60.5)	131,008 (60.0)	14,783 (60.1)	15,477 (61.2)
82,682 (26.6)	33,273 (26.2)	58,906 (27.0)	6,540 (26.6)	6,496 (25.7)



Supplementary Figure 1. *Hazard ratios for CAD incidence for categorized fish intake and physical activity, stratified by APOE isoform.*



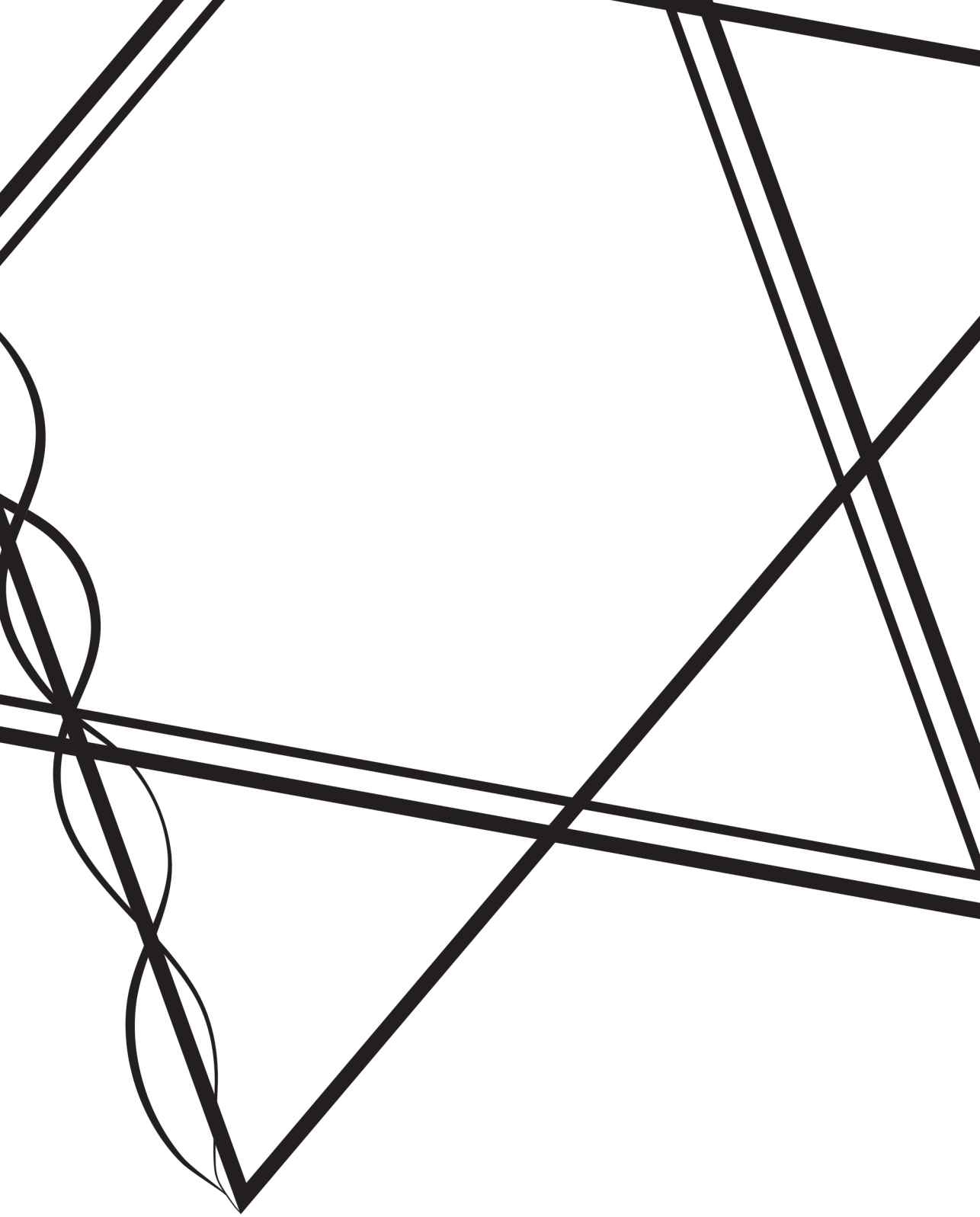
Supplementary Figure 2. *Hazard ratios for CAD incidence for fish intake, physical activity and PUFA intake, stratified by APOE isoform with exclusion of individuals that used lipid-lowering medication, blood pressure-lowering medication or had a clinical diagnosis of diabetes mellitus.*



Supplementary Figure 3. *Hazard ratios for CAD incidence for categorized fish intake and physical activity, stratified by APOE isoform with exclusion of individuals that used lipid-lowering medication, blood pressure-lowering medication or had a clinical diagnosis of diabetes mellitus.*

PART IV

Thyroid status and diabetes



CHAPTER 5.1

Thyroid Signaling, Insulin Resistance, and 2 Diabetes Mellitus: A Mendelian Randomization Study

Maxime M Bos, Roelof AJ Smit, Stella Trompet, Diana van Heemst, Raymond Noordam

J Clin Endocrinol Metab. 2017;102(6):1960–1970

ABSTRACT

Increasing evidence suggests an association between thyroid stimulating hormone (TSH), free thyroxine (fT4) and deiodinases with insulin resistance and type 2 diabetes mellitus (T2D). We examined whether TSH and fT4 levels and deiodinases are causally associated with insulin resistance and T2D using Mendelian randomization (MR). We selected twenty genetic variants for TSH level and four for fT4 level (identified in a GWAS meta-analysis of European-ancestry cohorts) as instrumental variables for TSH and fT4 level, respectively. We used summary data from genome-wide association studies (GWAS) on the outcomes type 2 diabetes mellitus (T2D) (DIAGRAM; 12,171 cases, 56,862 controls) and glycemic traits in nondiabetics (MAGIC; N=46,186 for fasting glucose and insulin and N=46,368 for HbA1c). To examine whether the associations between TSH/fT4 levels and the glycemic traits are causal, we combined the effects of the genetic instruments. Furthermore, we examined the associations between 16 variants in *DIO1*, *DIO2*, and *DIO3* and T2D and glycemic traits. We found no evidence of an association between the combined genetic instrumental variables for TSH and fT4 and the study outcomes. For example, we did not observe a genetically determined association between high TSH level and T2D (Odds ratio: 0.91 per standard deviation TSH increase; 95% confidence interval: 0.78;1.07). Selected genetic variants in *DIO1* (e.g., rs7527713) were associated with measures of insulin resistance. We found no evidence of a causal association between circulatory levels of TSH and fT4 with insulin resistance and T2D, but found suggestive evidence that *DIO1* affects glucose metabolism.

INTRODUCTION

Due to the increased proportion of obesity[1, 2], and increased life expectancy[3-5], type 2 diabetes mellitus (T2D) is becoming a major public health challenge. Key mechanisms involved in T2D are insulin resistance in muscle, adipose tissue and liver, and impaired insulin secretion due to deterioration of pancreatic β-cell function[3]. As recently described by Peppa *et al.,* there is increasing evidence for the existence of an association between endocrine disorders, like disturbed thyroid function or (subclinical) hypothyroidism, and altered glucose-insulin homeostasis[6]. For example, several studies have reported that subclinical hypothyroidism is related to insulin resistance and T2D[7, 8].

The concentration of thyroid hormones in the circulation is regulated by the hypothalamic pituitary thyroid (HPT) axis. Thyrotropin releasing hormone (TRH) secreted by the hypothalamus regulates synthesis and release of TSH from the pituitary gland, which stimulates the production and secretion of the thyroid hormones by the thyroid gland. Via a classical feedback loop, thyroid hormones inhibit the production of hypothalamic TRH and pituitary TSH. In target tissues, type 1 and 2 deiodinases convert the prohormone thyroxine (T4) into the active hormone triiodothyronine (T3), while type 3 deiodinase converts T4 into inactive reverse triiodothyronine (rT3)[9, 10].

Previously, it has been shown that patients with subclinical hypothyroidism have lower insulin sensitivity [7, 11]. In line, in euthyroid individuals, higher fT4 levels and/or lower TSH levels have been associated with a higher insulin sensitivity[11, 12], and with a lower risk of developing T2D[13]. Of the deiodinases, genetic variation in deiodinase 2 has been associated with risk of T2D[14], but results between studies as these are inconsistent[15-17]. These studies suggest thyroid hormone metabolism might exert an effect on insulin sensitivity and T2D, but the biological mechanisms for these observations are largely unclear. Causality of observational associations (e.g., low TSH in euthyroid individuals and higher insulin sensitivity) cannot be ascertained because of unmeasured confounding and/or reverse causality. In case of reverse causality, insulin resistance might suppress thyroid function, and consequently lead to relatively higher TSH and lower fT4 levels.

One method used to ascertain causality of observational associations, free of confounding and reverse causality, is Mendelian randomization (MR)[18, 19]. This method uses genetic variants as instrumental variables for the exposure of interest[20]. The levels of TSH and fT4 are, in part, genetically determined[21], and several loci have been identified in a large genome-wide association study (GWAS)[22]. To date, however, no studies have investigated the relation between genetic variants associated with TSH and fT4 concentration and measures of insulin resistance and T2D. On the other hand,

genetic variation in *DIO2* has been studied before in relation to T2D, but this needs to be explored in more depth to confirm earlier results. With respect to the other deiodinases, deiodinase 1 is highly expressed in liver tissue, an organ which is pivotal in glucose-insulin homeostasis[23] and IGF-1 production[24]. Deiodinase 3 is an inactivation deiodinase associated with rT3 concentration, and thus inhibits thyroid hormone action. However, to the best of our knowledge, no studies have been performed on deiodinase 1 and 3 in relation to insulin resistance and T2D. In addition to individual participant data, Mendelian randomization studies can also be conducted using the summary statistics data of previously conducted GWAS meta-analyses[25]. Within the present study, we aimed to investigate whether there is evidence for a causal association between circulatory TSH and fT4 levels and glycemic traits and T2D using Mendelian randomization. Furthermore, to provide more insights in the role of thyroid function in target tissues in the pathogenesis of T2D, we additionally tested genetic variants in deiodinase 1, 2, and 3 in relation to measures of insulin resistance and T2D.

MATERIALS AND METHODS

Selection of Single Nucleotide Polymorphisms associated with TSH or fT4

For the present Mendelian randomization study, we selected the lead single nucleotide polymorphisms (SNPs) for all genetic loci that have been shown to independently associate with the levels of TSH or fT4 (p-value < $5e^{-8}$) as genetic instrumental variables for TSH and fT4 levels, respectively. These loci were extracted from a meta-analysis of genome-wide association studies (GWAS) performed by Porcu *et al*. in individuals of European ancestry[22], which is the largest GWAS on TSH and fT4 level to date. This meta-analysis identified twenty independent genetic variants associated with higher serum TSH level in 26,420 individuals and four independent genetic variants associated with higher serum fT4 level in 17,520 individuals. In this meta-analysis, TSH and fT4 were standardized to approximate a normal distribution. The identified genetic variants explained a total of 5.64% and 2.30% of total variation in TSH and fT4 serum concentration, respectively. Within this GWAS analysis, all individuals with a TSH <0.4 and >4.0 mIU/L were excluded.

Selection of Single Nucleotide Polymorphisms associated with deiodinases

Adapted from Panicker *et al*., we selected nine SNPs mapped in *DIO1,* three SNPs in *DIO2* and four SNPs in *DIO3*, which cover 100%, 85% and 71%, respectively, of the common HapMap-based variation in these genes (minor allele frequency >10%) with r^2 greater than 0.8[26].

Data sources and outcome definition

In the present study, we used T2D and measures of insulin resistance as outcomes. For this, we used publically available summary statistics datasets of two large GWAS conducted by the DIAbetes, Genetics Replication and Meta-analysis (DIAGRAM) consortium and the Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC). These datasets contain the summary level meta-analysis data of these GWAS, comprising the per-allele beta estimates of the SNPs on the outcomes, accompanying standard errors and the effect alleles. The data of the DIAGRAM consortium comprised the stage I meta-analysis of 12 different cohort studies of European ancestry[27], including 12,171 cases of T2D and 56,862 controls. T2D was defined based on a fasting glucose >6.9 mmol/L, treatment with glucose-lowering agents and/or diagnosis by a general practitioner or medical specialist. The used data of MAGIC comprised a meta-analysis of 21 different cohort studies of European ancestry investigating genetic variants associated with glycemic traits[28]. In total, 46,186 individuals without diabetes were included in the meta-analysis. From the MAGIC consortium we used all glycemic traits, i.e. fasting glucose and insulin, the homeostatic model assessment for insulin resistance

(HOMA-IR) and pancreatic β-cell function (HOMA-B) as study outcomes. Within the GWAS meta-analysis, fasting insulin, HOMA-IR and HOMA-B data were log transformed to approximate a normal distribution. Additionally, we used data from a second GWAS meta-analysis from MAGIC on HbA1c (N = 46,368)[29]. Furthermore, we used data from MAGIC on fasting glucose and insulin that have been adjusted for body mass index (BMI), as thyroid hormone is known to be associated with BMI[30]. Potential mediation of effects by BMI (e.g., effects of thyroid function on glucose metabolism through BMI) was accounted for in this analysis, and the direct effect could be studied.

Statistical analyses

To limit bias from including weak genetic instrumental variables, i.e. those that explain little of the variance in the exposure of interest, we calculated the F-statistic as a measure of strength for each genetic instrument. In line with the existing literature, we considered an F-statistic of 10 or more as being of sufficient strength[31]. We then explored whether any of the individual genetic instruments or genetic variants in the deiodinase genes were associated with T2D or glycemic traits. For these analyses, we corrected for multiple testing with Bonferroni based on the number of variants included in each analysis (TSH: $\alpha = 0.05/20 = 0.0025$; fT4: $\alpha = 0.05/4 = 0.0125$; DIOs: $\alpha = 0.05/16 = 0.0031$).

By performing summary level Mendelian randomization analyses, we aimed to separately combine the twenty genetic instrumental variables for TSH level and the four genetic instrumental variables for fT4 level to obtain a genetically determined (causal) association between the TSH and fT4 levels on the study outcomes (notably T2D, fasting glucose and insulin, HOMA-IR and HOMA-B, and HbA1c). The resulting estimate can be interpreted as the change in outcome per unit increase in genetically raised exposure of interest (e.g., change in log(odds) for T2D per standard deviation increase in genetically raised TSH levels). Analogous to pooling estimates from different studies in conventional meta-analysis using inverse-variance weighting (IVW), we weighted this combined estimate by the inverse of the variance of the per-allele effect on the outcome (T2D or glycemic trait) for each genetic instrument. However, the effect estimate retrieved from this analysis might be biased as some of the genetic instrumental variables could be invalid because of pleiotropy. Biological pleiotropy is a phenomenon where a genetic variant also influences other traits than the exposure of interest of the study that influence the outcome of interest, which may therefore bias the results of a Mendelian randomization study (e.g., a genetic variant known to influence TSH associates with T2D, but not solely through modifying TSH level). If these pleiotropic effects across the genetic variants do not balance out, this might skew the mean genetically determined estimate of the exposure on the outcome. Such bias is formally called 'directional pleiotropy'. To take into account potential bias in this study

caused by directional pleiotropy, we conducted two sensitivity analyses, knowing MR-Egger regression[32] and weighted median estimator analyses[33]. With MR-Egger analyses, we were able to formally test for the presence of directional pleiotropy [32]. These analyses were repeated using summary-level statistics data of the GWAS analyses on fasting glucose and insulin adjusted for body mass index. The unadjusted source codes for these methods for R, as provided online by the authors[32, 33] were used for the calculations of the combined effect of the individual genetic instrumental variables. For the combined effect of the genetic instruments on the study outcomes, a two-sided p-value below 0.05 was considered statistically significant.

RESULTS

Effect of individual genetic instruments for TSH or fT4 levels on various study outcomes
All individual genetic instruments for both TSH and fT4 levels had an F-statistic above 10 (TSH: median 42.5; range 32.1 – 245.4; fT4: median 52.1; range 34.1 – 132.3), and were therefore considered to be of sufficient strength to be used in the present study. The associations between the individual genetic instruments for TSH levels and T2D and the glycemic traits are presented in **Figure 1.**

Of the individual genetic instruments, we observed an association between rs9472138 in *VEGFA* and T2D (**Online Supplementary Table 1**), and between rs3813582 in *MAF/LOC440389* and HOMA-IR after correction for multiple testing. Within the analysis adjusted for BMI, only rs9472138 in *VEGFA* was associated with fasting glucose after correction for multiple testing (**Figure 2)**. None of the individual genetic instruments for fT4 were associated with any of the study outcomes (**Online Supplementary Figure 1**).

Combined effect of genetic instrumental variables for TSH and fT4 levels on the study outcomes
When using inverse-variance weighted analyses to combine the effects on the outcomes of the individual genetic instruments, we found no evidence that there was an association between TSH level and the risk of T2D (**Table 1 and Figure 3**; odds ratio = 0.91 per 1 standard deviation [SD] higher TSH, 95% confidence interval [CI] = 0.78; 1.07). Furthermore, again using inverse-variance weighted analyses, we found no evidence that there was an association between TSH level and the other studied glycemic traits, such as fasting glucose and insulin. The mean effect estimates remained similar with MR-Egger regression and median weighted estimator analyses. Nevertheless, we found some suggestive evidence using median weighted estimator analysis that there might be an association between a higher TSH level and lower risk of T2D (β = -0.139 [95% CI:-0.301; 0.023]). We found no evidence, using MR-Egger regression analysis, that the intercept deviated from zero, which indicates that we found no evidence that the effect estimates were biased due to directional pleiotropy. In addition, similar results were observed when we used the GWAS datasets on glycemic traits adjusted for BMI (results not shown). Similarly, we did not find evidence for an association between a higher fT4 level and any of the study outcomes (**Table 1 and Online Supplementary Figure 2**).

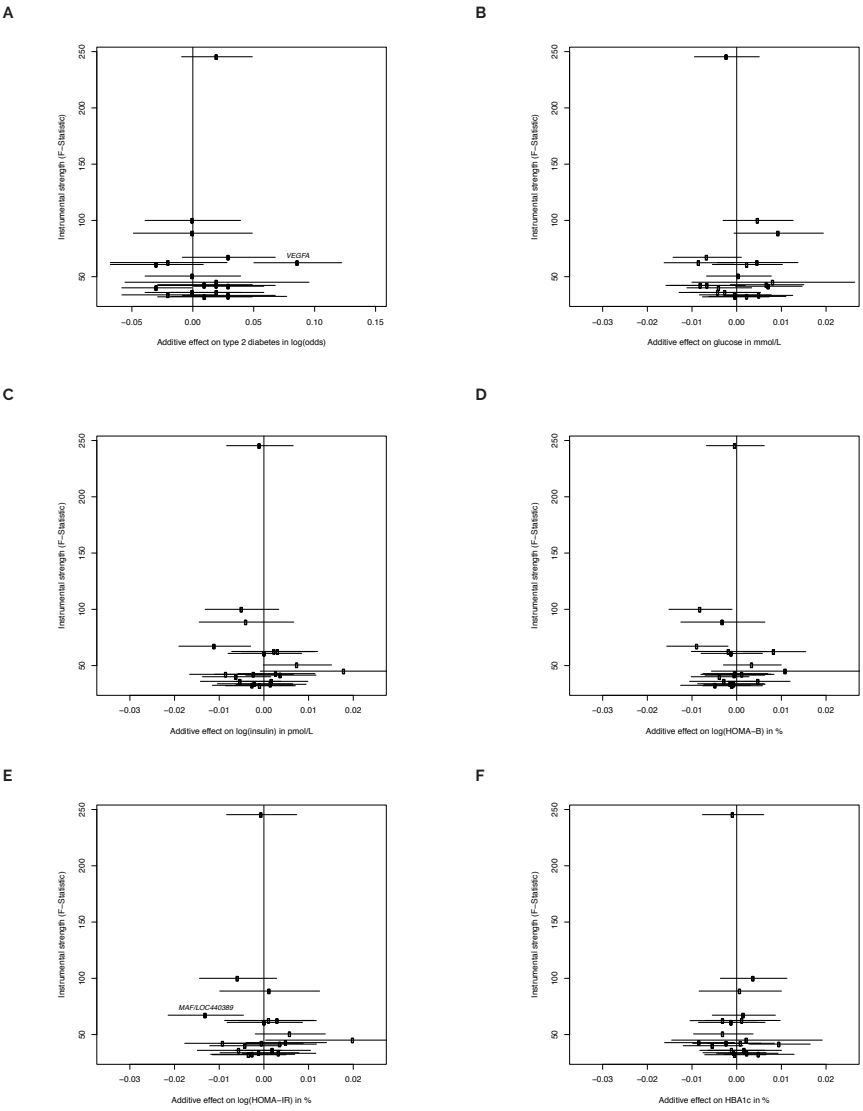


Figure 1. *The effect for individual genetic instrumental variables for TSH levels and study outcomes.* Results are displayed as the additive beta estimates with 95% confidence intervals. The X-axis presents the additive (per allele) effect for each of the individual genetic instrumental variables and the study outcome. Full written name of gene indicates a significant association after Bonferroni correction (p = 0.0025). The association between twenty genetic markers for TSH levels and (**A**) type 2 diabetes mellitus in log(odds), (**B**) fasting glucose in mmol/L per unit, (**C**) fasting log(insulin) in pmol/L, (**D**) log(homeostatic model assessment for beta cell function (HOMA-B)) in %, (**E**) log(homeostatic model assessment for insulin resistance (HOMA-IR)) in %, and (**F**) glycated hemoglobin (HbA1C) in %.

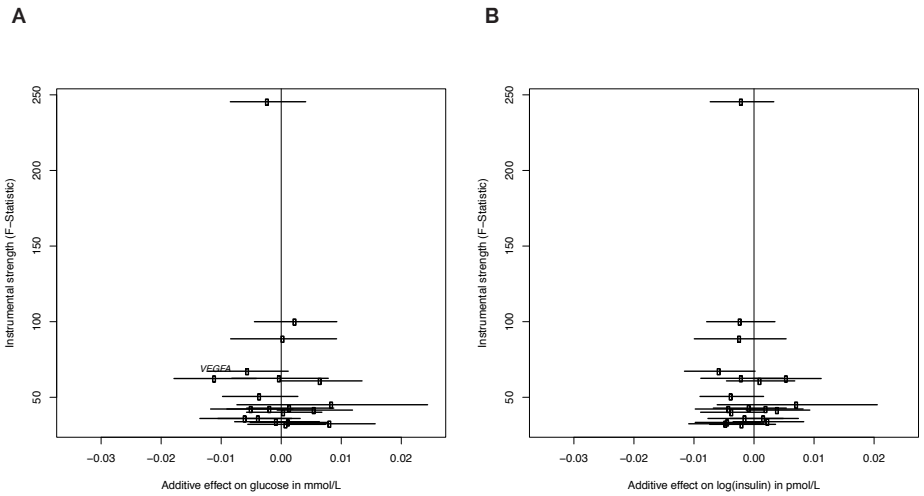


Figure 2. *The effect for individual genetic instrumental variables for TSH levels and BMI adjusted fasting glucose and insulin.* Results are displayed as the additive beta estimates with 95% confidence intervals. The X-axis presents the additive (per allele) effect for each of the individual genetic instrumental variable and the study outcome. Full written name of gene indicates a significant association after Bonferroni correction (p = 0.0025). (**A**) BMI-adjusted fasting glucose in mmol/L and (**B**) BMI-adjusted fasting log(insulin) in pmol/L.

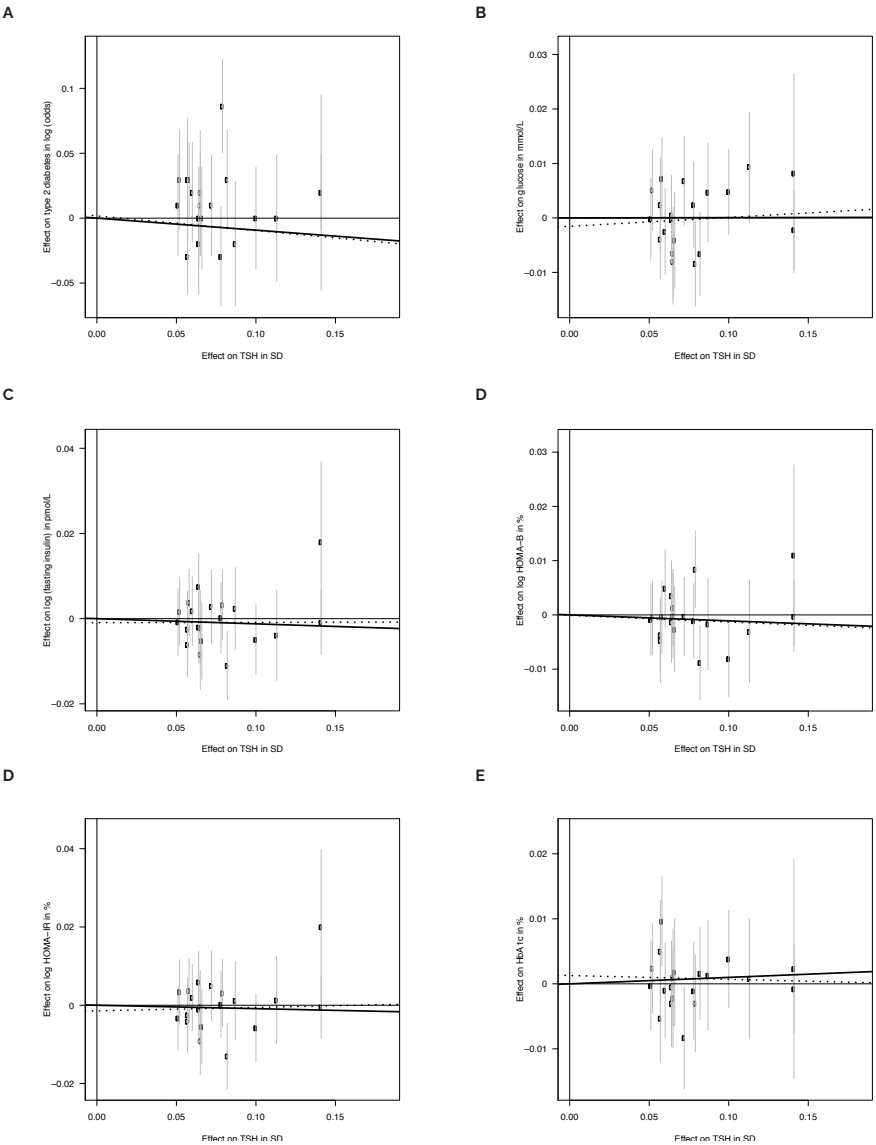


Figure 3. *The combined effect of the individual genetic instrumental variables for TSH levels on the study outcomes.* Results of the individual genetic instruments for TSH levels displayed as the causal effect on study outcome with 95% confidence interval per standard deviation higher serum level of TSH. The solid line represents the regression line of the inverse-variance weighted approach to combine the individual genetic instruments. The dashed line represents the regression line of the MR-Egger regression analysis to combine the individual genetic instruments. Data presented for (**A**) type 2 diabetes mellitus in log(odds), (**B**) fasting glucose in mmol/L, (**C**) fasting log(insulin) in pmol/L, (**D**) log(homeostatic model assessment for beta cell function (HOMA-B)) in %, (**E**) log(homeostatic model assessment for insulin resistance (HOMA-IR)) in %, and (**F**) glycated hemoglobin (HbA1c) in %.

Effect of genetic variation in deiodinases and the study outcomes

After correction for multiple testing (**Table 2**), the genetic variants rs11206237, rs2268181 and rs7527713 mapped in *DIO1* were significantly associated with fasting insulin ($\beta_{additive}$ = -0.018 pmol/L [95%CI = -0.029; -0.007]; $\beta_{additive}$ = 0.019 pmol/L [95%CI = 0.008; 0.030]; $\beta_{additive}$ = -0.018 pmol/L [95%CI = -0.028; -0.008], respectively), and with HOMA-IR ($\beta_{additive}$ = -0.018 pmol/L [95%CI = -0.029; -0.007]; $\beta_{additive}$ = 0.020 pmol/L [95%CI = 0.009; 0.031]; $\beta_{additive}$ = -0.019 [95%CI = -0.029; -0.009], respectively). In addition, rs7527713 in *DIO1* was associated with HOMA-B ($\beta_{additive}$ = -0.014 [95%CI = -0.023; -0.005]). None of the three genetic variants in *DIO2* and none of the four SNPs in *DIO3* were significantly associated with T2D or any of the glycemic traits.

Table 1. *Mendelian randomization estimates*

	T2D mellitus in log(odds)	Glucose in mmol/L	Log(insulin) in pmol/L	Log(HOMA-IR) in %	Log(HOMA-B) in %	HbA1c in %
TSH						
IVW	-0.09 (-0.25;0.07)	0.00 (-0.03;0.03)	-0.01 (-0.04;0.02)	-0.01 (-0.04;0.02)	-0.01 (-0.04;0.01)	0.01 (-0.01;0.03)
MR Egger Intercept	0.00 (-0.04;0.04)	-0.00 (-0.01;0.01)	-0.00 (-0.01;0.00)	-0.00 (-0.01;0.01)	$9.3x10^{-5}$ (-0.01;0.01)	0.00 (-0.00;0.01)
Estimate	-0.11 (-0.47;0.24)	0.02 (-0.06;0.10)	0.00 (-0.10;0.10)	0.01 (-0.08;0.10)	-0.01 (-0.08;0.06)	-0.01 (-0.08;0.07)
WME	-0.14 (-0.30;0.02)	-0.01 (-0.04;0.03)	-0.01 (-0.05;0.03)	-0.00 (-0.04;0.04)	-0.01 (-0.04;0.025)	0.00 (-0.03;0.03)
T4						
IVW	0.10 (-0.07;0.27)	0.00 (-0.04;0.05)	0.02 (-0.02;0.05)	0.01 (-0.03;0.05)	0.013 (-0.02;0.04)	-0.01 (-0.04;0.03)
MR Egger Intercept	0.02 (-0.04;0.07)	0.02 (-0.02;0.05)	-0.01 (-0.04;0.02)	-0.01 (-0.03;0.02)	-0.01 (-0.04;0.01)	-0.00 (-0.02;0.02)
Estimate	-0.04 (-1.33;1.25)	-0.15 (-0.47;0.18)	0.10 (-0.24;0.45)	0.06 (-0.27;0.39)	0.12 (-0.16;0.41)	0.01 (-0.26;0.27)
WME	0.09 (-0.10;0.27)	-0.02 (-0.06;0.02)	0.02 (-0.02;0.07)	0.02 (-0.03;0.07)	0.01 (-0.03;0.04)	-0.01 (-0.04;0.03)

Abbreviations: HOMA-IR, homeostatic model assessment for insulin resistance; HOMA-B, homeostatic model assessment for pancreatic beta cell function; IVW, inverse-variance weighted; WME, weighted median estimator. Data presented as beta coefficients with 95% confidence interval per standard deviation higher serum level of thyroid stimulating hormone (TSH) or free prohormone thyroxine (fT4).

Table 2. *The association between genetic variants for deiodinases and the study outcomes*

	T2D mellitus in log(odds)	Glucose in mmol/L	Log (insulin) in pmol/L	Log (HOMA-IR) in %	Log (HOMA-B) in %	HbA1c in %
DIO1						
11206237	0.01 (-0.04;0.06)	-0.00 (-0.01;0.01)	-0.02 (-0.03;-0.01)	-0.02 (-0.03;-0.01)	-0.01 (-0.02;-0.00)	0.00 (-0.01;0.01)
11206244	0.00 (-0.04;0.04)	-0.00 (-0.01;0.01)	0.00 (-0.01;0.01)	0.00 (-0.01;0.01)	0.00 (-0.00;0.01)	0.00 (-0.01;0.01)
2235544	0.01 (-0.02;0.05)	-0.00 (-0.01;0.00)	0.00 (-0.00;0.01)	0.00 (-0.01;0.01)	0.00 (-0.00;0.01)	-0.00 (-0.01;0.01)
2268181	0.01 (-0.04;0.06)	0.01 (-0.01;0.02)	0.02 (0.01;0.03)	0.02 (0.01;0.03)	0.01 (0.00;0.02)	-0.00 (-0.01;0.01)
2294511	0.01 (-0.03;0.05)	0.00 (-0.01;0.01)	-0.01 (-0.02;0.00)	-0.01 (-0.02;0.01)	-0.01 (-0.01;0.00)	0.00 (-0.00;0.01)
2294512	0.03 (-0.01;0.07)	-0.00 (-0.01;0.01)	-0.00 (-0.01;0.01)	-0.00 (-0.01;0.01)	-0.00 (-0.01;0.01)	-0.00 (-0.01;0.01)
4926616	0.02 (-0.03;0.06)	-0.00 (-0.01;0.01)	0.01 (-0.00;0.02)	0.01 (-0.01;0.02)	0.01 (-0.00;0.02)	-0.01 (-0.01;0.00)
731828	0.01 (-0.03;0.04)	-0.00 (-0.01;0.00)	-0.01 (-0.01;0.00)	-0.01 (-0.01;0.00)	-0.00 (-0.01;0.01)	-0.00 (-0.01;0.01)
7527713	0.02 (-0.02;0.07)	-0.00 (-0.01;0.01)	-0.02 (-0.03;-0.01)	-0.02 (-0.03;-0.01)	-0.01 (-0.02;-0.01)	0.01 (-0.00;0.01)
DIO2						
225011	0.00 (-0.03;0.04)	-0.00 (-0.01;0.01)	-0.00 (-0.01;0.01)	-0.00 (-0.01;0.01)	0.00 (-0.01;0.01)	-0.00 (-0.01;0.01)
225014	0.01 (-0.03;0.05)	-0.00 (-0.01;0.01)	0.00 (-0.01;0.01)	0.00 (-0.01;0.01)	0.00 (-0.01;0.01)	-0.00 (-0.01;0.00)
225015	0.01 (-0.03;0.05)	0.00 (-0.01;0.01)	0.00 (-0.01;0.01)	0.00 (-0.01;0.01)	-0.00 (-0.01;0.01)	0.00 (-0.00;0.01)
DIO3						
17716499	0.02 (-0.03;0.07)	0.01 (-0.00;0.01)	0.00 (-0.01;0.01)	0.00 (-0.01;0.01)	-0.00 (-0.01;0.01)	0.00 (-0.01;0.01)
7150269	0.01 (-0.03;0.05)	0.01 (0.00;0.02)	0.00 (-0.01;0.01)	0.00 (-0.01;0.01)	-0.00 (-0.01;0.01)	0.00 (-0.01;0.01)
8011440	0.01 (-0.03;0.06)	-0.01 (-0.02;0.00)	-0.00 (-0.01;0.01)	-0.00 (-0.01;0.01)	0.00 (-0.01;0.01)	-0.00 (-0.01;0.01)
945006	0.05 (-0.04;0.14)	0.00 (-0.01;0.02)	-0.01 (-0.03;0.01)	-0.01 (-0.02;0.01)	-0.00 (-0.02;0.01)	-0.01 (-0.02;0.01)

Abbreviations: HOMA-IR, homeostatic model assessment for insulin resistance; HOMA-B, homeostatic model assessment for pancreatic beta cell function. Data presented as additive beta coefficients with 95% confidence interval.

DISCUSSION

In the present study we examined whether the previously described observational associations between (subclinical) hypothyroidism, characterized by high TSH levels and fT4 levels within the normal range, with measures of insulin resistance and T2D[7, 8] are causal using Mendelian randomization analyses. Within this study, we did not find evidence that the association between TSH and fT4 (as determined in a population with TSH>0.4 mIU/L and TSH<4.0 mIU/L) with T2D and glycemic traits (fasting glucose and insulin, HOMA-IR, HOMA-B, and HbA1c) is causal. However, although effect sizes were relatively small, we found suggestive evidence of an association between selected genetic variants in *DIO1* with glycemic traits, but not with T2D.

The lack of a genetically determined (causal) association between TSH/fT4 level and measures of glucose metabolism suggests that the previously observed association between subclinical hypothyroidism and insulin resistance might be explained by reverse causality or residual confounding. In case of reverse causality, progression of insulin resistance and T2D influences TSH and thyroid hormone levels[34]. In addition, unmeasured or unknown factors not taken into account as confounding variables in the observational studies might have affected the observed association between subclinical hypothyroidism and insulin resistance.

In this study two genetic variants associated with TSH level showed a significant association with the study outcomes after correction for multiple testing. The genetic variant rs3813582 was significantly associated with HOMA-IR. This polymorphism is a noncoding variant in an exonic region of the *LOC440389/MAF* gene, associated with increased thyroid volume[35], a condition more frequently found in patients with insulin resistance[36]. However, the allele that has been associated with higher TSH concentration[22] was associated with lower insulin resistance in the data from the MAGIC consortium. Gene expression of *LOC440389/MAF* was previously found higher in thyroid tissue compared to skeletal muscle, which might suggest a more thyroid-specific function[35]. Further studies are required to gain more information about this SNP and its potential effect on insulin resistance beyond the effects through TSH. Moreover, we observed that rs9472138 in *VEGFA* was associated with T2D and fasting glucose after adjustment for BMI. *VEGFA* encodes a growth factor important in angiogenesis, which is critical for thyroid function due to the high vascularization of this organ for the continuous supply of iodine to synthesize thyroid hormones[37]. Genetic variation in *VEGFA* has not been significantly related to T2D in GWAS studies. Nevertheless, it might still be a novel genetic predictor for T2D. Previous work already

suggested that *VEGFA* gene variants were associated with diabetic nephropathy[38]. Based on our results, we hypothesize that this association is independent of TSH level, and that *VEGFA* is involved in multiple biological pathways.

In the literature, there is evidence of an association between deiodinases and the onset of insulin resistance[14]. The most studied polymorphism in the deiodinases in relation to insulin resistance and T2D is rs225014 in *DIO2*, which has a high allele frequency in various ethnic groups [16]. However, previous studies have shown contradictive results[15, 17]. In this study we assessed the role of deiodinases on insulin resistance and T2D using genetic variants in the deiodinase genes that cover the gene as much as possible[26]. We observed genetic variation in *DIO1* to be significantly associated with fasting insulin, HOMA-IR and HOMA-B. In the data from DIAGRAM and MAGIC, we did not find a significant association between the previously described polymorphism in *DIO2* and increased risk for T2D and any of the glycemic traits in individuals of European ancestry. The discrepancy in the results can be explained by the smaller sample size of the previous studies and by the differences in ethnic background of the different cohorts examining the relationship. However, we showed that there might be a minor effect of deiodinase 1 on insulin resistance. To the best of our knowledge, no data are yet available about this mechanism, which will therefore require further studies. Speculatively, polymorphisms in *DIO1* are highly expressed in liver tissue[23] and have been associated with IGF-1 and IGF-1 related endpoints, such as body height and skeletal muscle mass[24]. Skeletal muscle is an important organ for insulin-stimulated glucose uptake and a lower relative muscle mass was found to associate with insulin resistance[39]. However, this hypothesis should be explored in more detail in future studies.

Strengths of the present study include the large sample size of the used study populations, and the use of summary statistics data which increases efficiency. A limitation of our study was that the twenty independent loci associated with the level of TSH only explained 5.64% of the total variation in TSH concentration, and for fT4 this was only 2.30%[22]. Nevertheless, the F-statistics of all individual genetic instrumental variables was above 10, which should greatly limit the likelihood of weak instrumental variable bias[31]. Additional insights could be provided by the identification of additional loci that are associated with TSH and fT4 concentration. However, the effect sizes of these newly identified loci will probably be lower or they will have a smaller allele frequency. In both cases the strength of the instrumental variables will be lower, and the contribution of these genetic instrumental variables in the combined effect will be lower as well. Another limitation is that the genetic contribution to TSH and fT4 concentration was found to be different between men and women[22], and (subclinical) hypothyroidism is more frequent in women than in men[40]. As the datasets of the study outcomes were

not available stratified by sex, we were not able to address the research questions separately for men and women. Furthermore, as the genetic instrumental variables as well as the summary statistics originated from population of European ancestry, our results may not be generalized to populations of non-European ancestry. A last limitation might be that the GWAS on TSH and fT4 was only performed in euthyroid individuals. However, as the combined score of the twenty identified genetic variants associated with TSH was also associated with a higher risk of having an extreme high TSH level[22], results might be also applicable for patients beyond the normal TSH range.

Taken together, the results in the present study indicate no causal association between genetically determined circulatory TSH and fT4 levels and higher risk of T2D. This conclusion was supported by the lack of significant association between genetically determined circulatory TSH and fT4 with glycemic traits. Moreover, we did observe an association between genetic variants in *DIO1* and insulin resistance, which might suggest that thyroid metabolism affects glucose metabolism in target tissues.

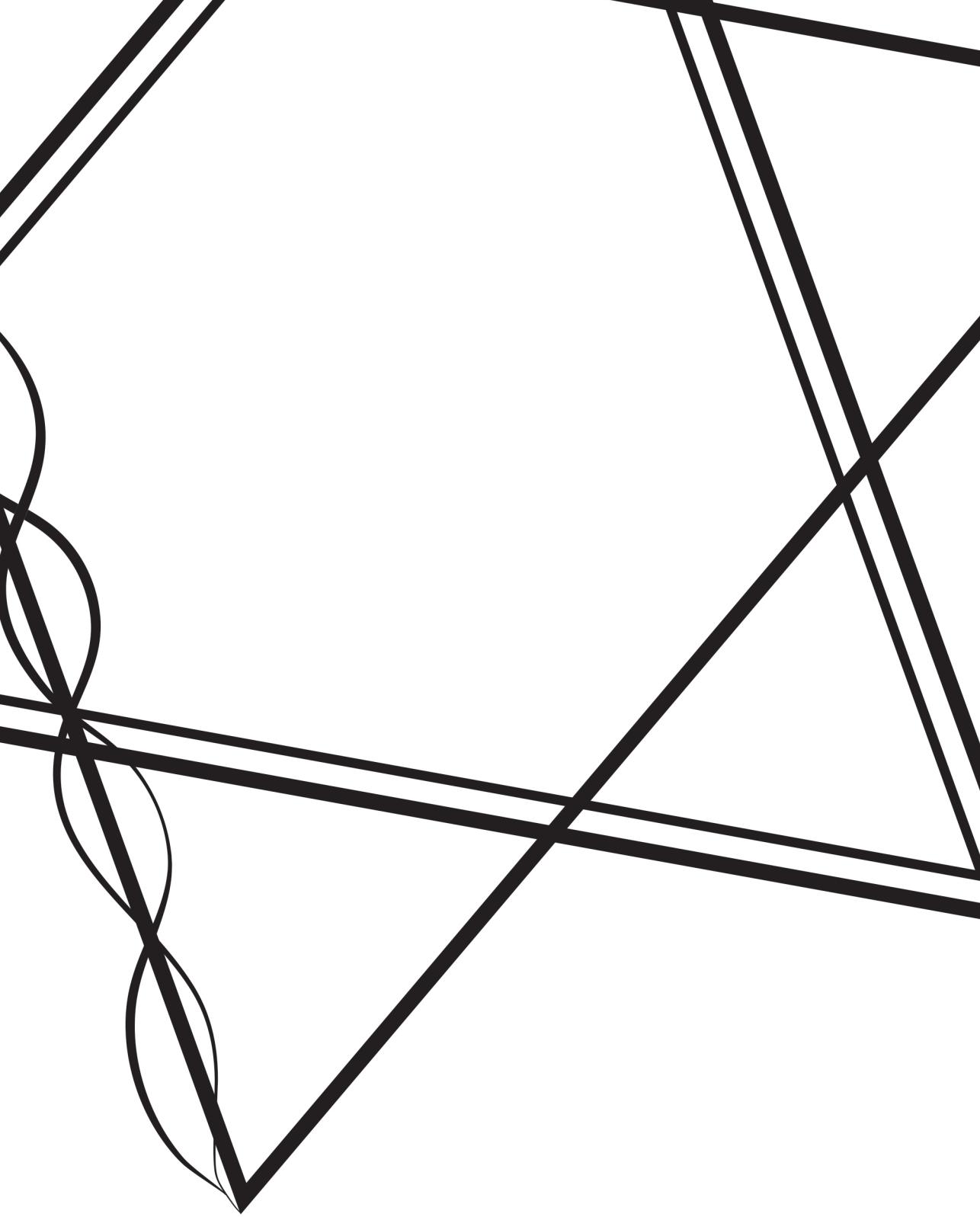
Acknowledgments
We would like to thank the DIAGRAM and MAGIC consortia for making the results files of their meta-analyses available. This study was funded by the European Commission funded projects HUMAN (Health-2013-INNOVATION-1-602757) and THYRAGE (Horizon 2020 research and innovation programme under grant agreement No 666869).

REFERENCES

1. Ogden CL, Carroll MD, Flegal KM. Prevalence of obesity in the United States. Jama. 2014;312(2):189-90.
2. Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of childhood and adult obesity in the United States, 2011-2012. Jama. 2014;311(8):806-14.
3. Chen L, Magliano DJ, Zimmet PZ. The worldwide epidemiology of type 2 diabetes mellitus--present and future perspectives. Nature reviews Endocrinology. 2012;8(4):228-36.
4. van Dieren S, Beulens JW, van der Schouw YT, Grobbee DE, Neal B. The global burden of diabetes and its complications: an emerging pandemic. European journal of cardiovascular prevention and rehabilitation : official journal of the European Society of Cardiology, Working Groups on Epidemiology & Prevention and Cardiac Rehabilitation and Exercise Physiology. 2010;17 Suppl 1:S3-8.
5. Kirkman MS, Briscoe VJ, Clark N, Florez H, Haas LB, Halter JB, et al. Diabetes in older adults. Diabetes care. 2012;35(12):2650-64.
6. Peppa M, Koliaki C, Nikolopoulos P, Raptis SA. Skeletal muscle insulin resistance in endocrine disease. J Biomed Biotechnol. 2010;2010:527850.
7. Maratou E, Hadjidakis DJ, Kollias A, Tsegka K, Peppa M, Alevizaki M, et al. Studies of insulin resistance in patients with clinical and subclinical hypothyroidism. European journal of endocrinology / European Federation of Endocrine Societies. 2009;160(5):785-90.
8. Al Sayed A, Al Ali N, Bo Abbas Y, Alfadhli E. Subclinical hypothyroidism is associated with early insulin resistance in Kuwaiti women. Endocr J. 2006;53(5):653-7.
9. Bianco AC, Salvatore D, Gereben B, Berry MJ, Larsen PR. Biochemistry, cellular and molecular biology, and physiological roles of the iodothyronine selenodeiodinases. Endocrine reviews. 2002;23(1):38-89.
10. Bianco AC, Larsen PR. Cellular and structural biology of the deiodinases. Thyroid : official journal of the American Thyroid Association. 2005;15(8):777-86.
11. Garduno-Garcia Jde J, Alvirde-Garcia U, Lopez-Carrasco G, Padilla Mendoza ME, Mehta R, Arellano-Campos O, et al. TSH and free thyroxine concentrations are associated with differing metabolic markers in euthyroid subjects. European journal of endocrinology / European Federation of Endocrine Societies. 2010;163(2):273-8.
12. Amouzegar A, Kazemian E, Gharibzadeh S, Mehran L, Tohidi M, Azizi F. Association between thyroid hormones, thyroid antibodies and insulin resistance in euthyroid individuals: A population-based cohort. Diabetes & metabolism. 2015;41(6):480-8.
13. Chaker L, Ligthart S, Korevaar TI, Hofman A, Franco OH, Peeters RP, et al. Thyroid function and risk of type 2 diabetes: a population-based prospective cohort study. BMC medicine. 2016;14(1):150.
14. Marsili A, Zavacki AM, Harney JW, Larsen PR. Physiological role and regulation of iodothyronine deiodinases: a 2011 update. Journal of endocrinological investigation. 2011;34(5):395-407.
15. Maia AL, Hwang SJ, Levy D, Larson MG, Larsen PR, Fox CS. Lack of association between the type 2 deiodinase A/G polymorphism and hypertensive traits: the Framingham Heart Study. Hypertension. 2008;51(4):e22-3.
16. Dora JM, Machado WE, Rheinheimer J, Crispim D, Maia AL. Association of the type 2 deiodinase Thr92Ala polymorphism with type 2 diabetes: case-control study and meta-analysis. European journal of endocrinology / European Federation of Endocrine Societies. 2010;163(3):427-34.

17. Mentuccia D, Proietti-Pannunzi L, Tanner K, Bacci V, Pollin TI, Poehlman ET, et al. Association between a novel variant of the human type 2 deiodinase gene Thr92Ala and insulin resistance: evidence of interaction with the Trp64Arg variant of the beta-3-adrenergic receptor. Diabetes. 2002;51(3):880-3.
18. Postmus I, Deelen J, Sedaghat S, Trompet S, de Craen AJ, Heijmans BT, et al. LDL cholesterol still a problem in old age? A Mendelian randomization study. International journal of epidemiology. 2015;44(2):604-12.
19. Voight BF, Peloso GM, Orho-Melander M, Frikke-Schmidt R, Barbalic M, Jensen MK, et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. Lancet. 2012;380(9841):572-80.
20. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Statistics in medicine. 2008;27(8):1133-63.
21. Medici M, Visser WE, Visser TJ, Peeters RP. Genetic determination of the hypothalamic-pituitary-thyroid axis: where do we stand? Endocrine reviews. 2015;36(2):214-44.
22. Porcu E, Medici M, Pistis G, Volpato CB, Wilson SG, Cappola AR, et al. A meta-analysis of thyroid-related traits reveals novel loci and gender-specific differences in the regulation of thyroid function. PLoS genetics. 2013;9(2):e1003266.
23. St Germain DL, Galton VA. The deiodinase family of selenoproteins. Thyroid : official journal of the American Thyroid Association. 1997;7(4):655-68.
24. Verloop H, Dekkers OM, Peeters RP, Schoones JW, Smit JW. Genetics in endocrinology: genetic variation in deiodinases: a systematic review of potential clinical effects in humans. European journal of endocrinology. 2014;171(3):R123-35.
25. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG, Consortium E-I. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. European journal of epidemiology. 2015.
26. Panicker V, Cluett C, Shields B, Murray A, Parnell KS, Perry JR, et al. A common variation in deiodinase 1 gene DIO1 is associated with the relative levels of free thyroxine and triiodothyronine. The Journal of clinical endocrinology and metabolism. 2008;93(8):3075-81.
27. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nature genetics. 2012;44(9):981-90.
28. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nature genetics. 2010;42(2):105-16.
29. Soranzo N, Sanna S, Wheeler E, Gieger C, Radke D, Dupuis J, et al. Common variants at 10 genomic loci influence hemoglobin A(1)(C) levels via glycemic and nonglycemic pathways. Diabetes. 2010;59(12):3229-39.
30. Manning AK, Hivert MF, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nature genetics. 2012;44(6):659-69.
31. Pierce BL, Ahsan H, Vanderweele TJ. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. International journal of epidemiology. 2011;40(3):740-52.

32. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. International journal of epidemiology. 2015;44(2):512-25.
33. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genetic epidemiology. 2016;40(4):304-14.
34. Sandhu MS, Debenham SL, Barroso I, Loos RJ. Mendelian randomisation studies of type 2 diabetes: future prospects. Diabetologia. 2008;51(2):211-3.
35. Teumer A, Rawal R, Homuth G, Ernst F, Heier M, Evert M, et al. Genome-wide association study identifies four genetic loci associated with thyroid volume and goiter risk. American journal of human genetics. 2011;88(5):664-73.
36. Duran AO, Anil C, Gursoy A, Nar A, Inanc M, Bozkurt O, et al. Thyroid volume in patients with glucose metabolism disorders. Arquivos brasileiros de endocrinologia e metabologia. 2014;58(8):824-7.
37. Yamada E, Yamazaki K, Takano K, Obara T, Sato K. Iodide inhibits vascular endothelial growth factor-A expression in cultured human thyroid follicles: a microarray search for effects of thyrotropin and iodide on angiogenesis factors. Thyroid : official journal of the American Thyroid Association. 2006;16(6):545-54.
38. Mooyaart AL, Valk EJ, van Es LA, Bruijn JA, de Heer E, Freedman BI, et al. Genetic associations in diabetic nephropathy: a meta-analysis. Diabetologia. 2011;54(3):544-53.
39. Bijlsma AY, Meskers CG, van Heemst D, Westendorp RG, de Craen AJ, Maier AB. Diagnostic criteria for sarcopenia relate differently to insulin resistance. Age (Dordr). 2013;35(6):2367-75.
40. Fatourechi V. Subclinical hypothyroidism: an update for primary care physicians. Mayo Clinic proceedings. 2009;84(1):65-71.



CHAPTER 5.2

Genetically-determined higher TSH is associated with lower diabetes mellitus risk in individuals with low BMI: results from the UK Biobank

Maxime M Bos, Nicolien A van Vliet, Raymond Noordam, Diana van Heemst

Submitted

ABSTRACT

Thyroid status is hypothesized to be causally related with the risk of diabetes mellitus (DM), but previous results were conflicting possibly due to a complex interaction between TSH, BMI and diabetes. In this study, we aimed to investigate the causal association between thyroid status with DM and glucose homeostasis and to what extent this association is dependent on BMI. This study was performed in British participants from the UK Biobank population. Genetic variants for circulatory TSH and fT4 levels and for BMI were used to calculate weighted genetic risk scores (GRS). We assessed the associations between genetically-determined TSH and fT4 and self-reported DM, and stratified the analyses by BMI. Similar analyses were performed on fasting glucose and Hb1Ac as outcomes among individuals without DM. The present study was performed in 401,773 British participants (mean age of 57.4 years (standard deviation 8.0) of which 45.9% were men), of which 19,773 individuals classified themselves as DM cases. Genetically-determined TSH and fT4 levels were not associated with risk of DM in the total UK Biobank population (odds ratio (OR) 0.96; 95% confidence interval (CI) 0.92,1.01). However, we did observe an interaction between genetically-determined TSH and BMI ($p = 0.06$) in their association with DM. In individuals with a low genetically-determined BMI, genetically-determined higher TSH, and not fT4, was associated with a lower risk for DM (OR 0.91; 95%CI 0.85,0.98). A higher genetically-determined TSH was associated with a lower level of fasting glucose and glycated haemoglobin in the overall population and in the groups stratified based on BMI. Genetically determined fT4 was not associated with fasting glucose or glycated haemoglobin. We found evidence for a potential causal association between higher circulatory TSH, but not fT4, and risk for DM. However, only in those with a lower BMI, suggesting protective effects of TSH only in low-risk populations.

INTRODUCTION

Diabetes mellitus (DM) is a major public health challenge, mainly due to increased prevalence of obesity[1,2]. DM is a heterogeneous disease which is caused by different mechanisms, notably, insulin resistance in muscle, adipose tissue, and liver, and impaired pancreatic insulin secretion[2].

Obesity is a major risk factor of DM development[3]. However, another potential risk factor for DM development is thyroid status, defined by the combined levels of thyroid-stimulating hormone (TSH) and free thyroxine (fT4)[4]. In observational studies, increasing evidence suggests an association between circulating levels of TSH and fT4 with DM[5-8]. However, not all studies support evidence for this association[9].

Moreover, obesity is associated with higher levels of TSH, though the direction of causation is still unclear[10]. Therefore, this complex relation between obesity, thyroid status and DM further complicates studies on causal inference regarding associations between thyroid status and DM. Not all confounder and causal factors are known and/or measured in observational studies and thus it remains unclear whether, and to what extent, thyroid status affects risk of developing DM and how obesity modifies this effect. In a previous study, we used genetic instrument to ascertain causality for associations of circulating levels of TSH and fT4 with DM, but did not find evidence for causality[11]. However, our previous study suffered from some limitations related to the small sample size and lack of (strong) genetic instruments, especially for fT4.

Now, with more than double the number of genetic instruments for TSH and more than quadruple number of genetic variants for fT4 combined with the availability of large sample size individual participant data in UK Biobank, we can readdress this research question in a more rigorous manner. Moreover, in order to study the effect of thyroid status on DM, possible effect modification by body mass index (BMI) should be considered. We hypothesize that obesity has a catalysing effect on the development of DM, especially type 2 diabetes, and could thereby overshadow more subtle causal pathways. By stratification on genetically determined BMI, differential effects of thyroid status on DM can be assessed.

Within the current study, we aimed to investigate the association between thyroid status and glucose homeostasis and the risk of DM in British participants from the UK Biobank. In addition, we stratified our analyses based on genetically-determined BMI in order to test our hypothesis of a possible effect modification by obesity on the association between thyroid status and DM and glucose homeostasis.

METHODS

Study population

For the present study, we included all participants from the UK Biobank with imputed genotype data and self-reported data on DM diagnosis. Between 2006 and 2011, men and women aged 40 till 69 years living within a reasonable travelling distance of one of the 22 assessment centres in the United Kingdom, were invited to participate in the UK Biobank via a population-based register [12]. During their visits to the assessment centres, participants completed questionnaires using a touchscreen device regarding current health and medical history. Body Mass Index (BMI) was established by dividing the weight in kilograms by the height in metres squared. Participants were asked to remove shoes and heavy clothing before weighing [13]. The UK Biobank operates within the terms of an Ethics and Governance Framework and all participants provided signed written informed consent [13,14].

Genotyping and genetic imputations

UK Biobank genotyping was conducted by Affymetrix using a bespoke BiLEVE Axium array for approximately 50,000 participants; the remaining participants were genotyped using the Affymetrix UK Biobank Axiom array. Quality control was centrally executed by UK Biobank. More information on the genotyping processes can be found online (https://www.ukbiobank.ac.uk). Based on the genotyped SNPs, UK Biobank resources performed centralized imputations on the autosomal SNPs using the UK10K haplotype [15], 1000 Genomes Phase 3 [16], and Haplotype Reference Consortium (HRC) reference panels [17]. Autosomal SNPs were pre-phased using SHAPEIT3 and imputed using IMPUTE4. In total, ~96 million SNPs were imputed. Related individuals were identified by estimating kinship coefficients for all pairs of samples using only markers weakly informative of ancestral background.

Selection of single nucleotide polymorphisms associated with TSH, fT4 and BMI

For this study, we selected genetic instruments from published genome-wide association studies, in which the UK Biobank did not contribute. For thyroid status, we selected the lead single nucleotide polymorphisms (SNPs) for all genetic loci that have been shown to independently associate with the circulating levels of TSH (42 loci) or fT4 (21 loci) ($P < 5 \times 10^{-8}$) as genetic instrumental variables for TSH and fT4 levels, respectively[18]. To investigate the combined effect of the thyroid hormone associated risk variants, we calculated a weighted genetic risk score (GRS) for circulating TSH or fT4. For the TSH GRS, we excluded rs13100823 in IGF2BP2 as this locus has been associated with type 2 diabetes mellitus[18]. For the fT4 GRS, we excluded rs11039355 in FNBP4 because of its previous association with body height, body mass index and proinsulin[18]. In addition,

we calculated a weighted genetic risk score (GRS) for BMI, for which we selected the lead SNPs for 97 BMI-associated loci [19]. For the present study, we considered a low and high genetically-determined BMI, which was defined on the basis of the median value in the study population.

Outcome definition

To define cases with diabetes mellitus, the baseline self-reported interview data collected in the full UK Biobank population was used. All participants reporting to have diabetes mellitus. Moreover, the individuals were asked about the age of diagnosis and whether they used insulin within the first year after diagnosis. Based on the age of diagnosis and use of insulin, we subdivided the self-reported DM diagnosis to homogenize the outcome population. These subdivisions were based on the median age of diagnosis (low/high) and use of insulin (yes/no).

Statistical analysis

Characteristics of the study population were expressed as mean with standard deviation for normally distributed measures, and proportions for categorical variables.

We performed multivariable logistic regression analyses to assess the association between the GRSs and DM (subtypes) and linear regression analyses were performed for the associations between the GRSs and the continuous variables glucose and glycated haemoglobin (HbA1C) adjusted for age, sex and four principal components. The resulting estimate is a weighted mean estimate and reflects a standard deviation increase of genetically determined TSH or fT4 on an odds ratio or standard deviation increase of our study outcome. All analyses were adjusted for age, sex and population stratification (first four principal components). To investigate possible effect modification by BMI, our analyses were additionally stratified based on the median of the GRS for BMI. We choose to use a genetically-determined BMI instead of the observational data, in order to minimize the possible effect of reverse causation, treatment and to study life-long exposure to high BMI. In order to formally test for interaction of the GRSs with BMI, we added an interaction term between the thyroid GRS and BMI GRS in their association with DM to the regression models.

In addition, we performed exploratory analyses in which we homogenized the self-reported DM phenotype by age of diagnosis and insulin use. For this, similar multivariable logistic regression analyses were performed as before using a subset of the case groups.

All statistical analyses were performed using R statistical software version 3.5.3. Results were reported as odds ratios (for dichotomous outcomes) or beta estimates (for glucose and Hb1Ac) with 95% confidence intervals.

RESULTS

Population characteristics

After excluding individuals lacking genetic information or those who were of non-European ancestry, this study comprised 401,773 participants with a mean age of 57.4 years (standard deviation 8.0) of which 45.9% were men. A total of 19,773 individuals (4.9%) reported a diagnosis of DM. The population characteristics of the study population are shown in **Table 1** for non-DM cases and DM cases. As compared to controls, DM cases had a higher mean age (57.3 [8.0] versus 60.6 [6.9] years), were more often male (54.1% versus 61.8%), had a higher mean BMI (27.2 [4.6] versus 31.5 [5.8] kg/m^2), had a higher fasting glucose (5.0 [0.8] versus 7.6 [3.4] mmol/L) and a higher HbA1c (35.1 [4.5] versus 52.4 [13.7] %).

Table 1. Characteristics of study population.

	No DM (n=389,122)	DM (n=19,773)
Age at study visit in years	57.3 (8.0)	60.6 (6.9)
Age at diagnosis in years	-	50.3 (14.7)
Sex, N (% male)	54.1	61.8
BMI in kg/m^2	27.2 (4.6)	31.5 (5.8)
Fasting glucose in mmol/L	5.0 (0.8)	7.6 (3.4)
HbA1c in %	35.1 (4.5)	52.4 (13.7)

Data presented as mean with standard deviation or as stated otherwise. Abbreviations: BMI, body mass index; DM, diabetes mellitus; N, number.

Genetically determined TSH and fT4 with DM

A genetically-determined higher TSH did not associate with DM in the overall group (odds ratio (OR) 0.96; 95% confidence interval (CI) 0.92,1.01) (**Figure 1**). In the subgroup with a propensity for low BMI (GRS BMI lower than median), a higher GRS for TSH was associated with a lower risk for DM (OR 0.91; 95%CI 0.85,0.98) (**Figure 1**). However, no associations between genetically-determined fT4 and DM were observed in the main group (OR 0.96; 95%CI 0.90,1.03) and in the group with low (OR 0.95; 95%CI 0.84,1.05) and high genetically-determined BMI (OR 0.99; 95%CI 0.90,1.08). We did find suggestive evidence for an interaction between the genetically-determined TSH and genetically-determined BMI on DM (p-value for interaction: 0.06), however, we did not observe a formal interaction for genetically-determined fT4 and genetically-determined BMI on DM (p-value for interaction: 0.19).

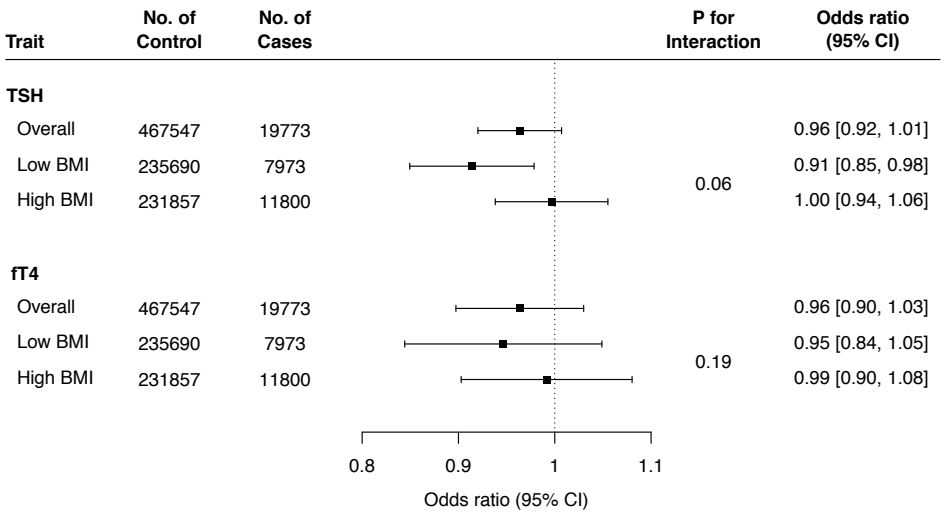


Figure 1. *Associations between the GRS for TSH and fT4 with diabetes mellitus in the overall population and stratified by genetically-determined BMI.*

Genetically determined TSH and fT4 with fasting glucose and glycated haemoglobin

A higher genetically-determined TSH was associated with a lower level of fasting glucose (beta -0.02; 95%CI -0.03,-0.01) in individuals without DM (**Table 2**). Moreover, a SD higher genetically-determined TSH was associated with lower fasting glucose levels (beta -0.02 mmol/L; 95%CI -0.03,-0.01)), in the group of individuals with a low BMI as well as in the group of individuals with a high BMI (beta -0.02 mmol/L; 95%CI -0.00,-0.03)(**Table 2**). The results for the GRS of TSH on glycated haemoglobin showed a similar trend, although with wider confidence intervals. Genetically determined fT4 was not associated with fasting glucose or glycated haemoglobin.

Table 2: *Associations between the GRS for TSH and fT4 with fasting glucose and HbA1c, stratified by genetically-determined BMI.*

			Low BMI		High BMI	
	TSH	T4	TSH	T4	TSH	T4
Fasting glucose	-0.02 (-0.03;-0.01)	-0.00 (-0.01;0.01)	-0.02 (-0.03;-0.01)	0.00 (-0.01;0.02)	-0.02 (-0.03;-0.00)	-0.00 (-0.02;0.01)
HbA1c	-0.03 (-0.07;0.02)	0.02 (-0.05;0.09)	-0.01 (-0.07;0.06)	-0.00 (-0.10;0.10)	-0.05 (-0.11;0.01)	0.05 (-0.04;0.15)

Data presented as SD increase in GRS per SD difference in outcome with accompanying 95% confidence interval. A total of 467,547 individuals are included in the overall analyses, 231,857 in the low BMI group and 235,690 in the high BMI group.

Sensitivity analyses in DM subtypes

In sub-analyses of this study, we also explored the associations in DM subgroups based on initiation of treatment with insulin (analogues) within the first year after diagnosis and on the age at diagnosis. Here we did observe an association between a higher GRS for TSH and a lower risk for DM diagnosed at a younger age for which patients did not require insulin(analogues) within the first year for those with a low BMI (OR 0.87; 95%CI 0.77, 0.98) (**Figure 2**).

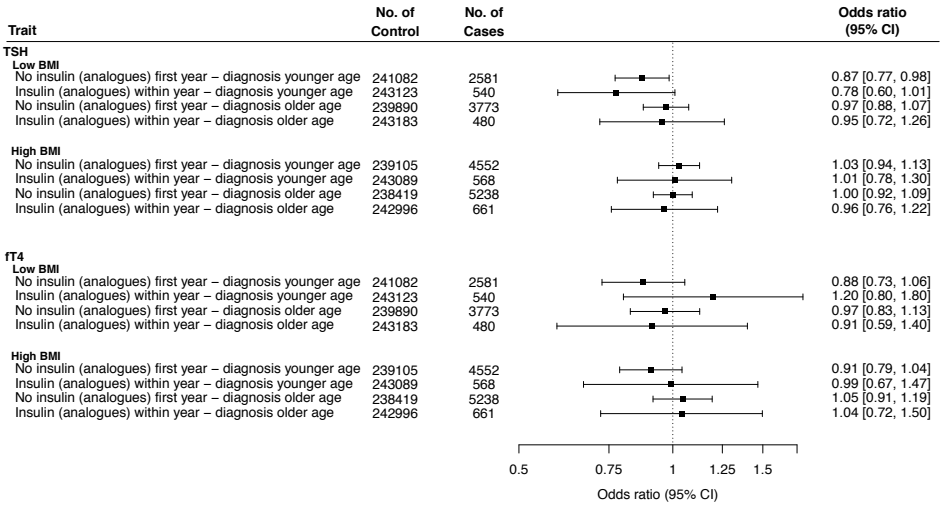


Figure 2. *Associations between the GRS for TSH and fT4 with subgroups of diabetes mellitus stratified by genetically-determined BMI.*

DISCUSSION

In this study, we examined the associations between genetic-risk scores for circulating TSH and fT4 levels with DM and glucose homeostasis. In the total British UK Biobank population, we did not find evidence for an association between genetically-determined TSH or fT4 with DM. However, when stratified based on genetically-determined BMI, higher genetically-determined TSH was associated with a lower risk of DM in the group with a low genetically-determined BMI. Moreover, higher genetically-determined TSH was associated with a lower fasting glucose level, in the overall group and those stratified based on genetically-determined BMI. Though no associations were found between genetically-determined fT4 and any of our study outcomes.

A strength of this study is the use of a large sample size with a large number of DM cases, which allowed for stratified analyses based on genetically-determined BMI and exploratory analyses on DM subgroups. By stratification on genetically determined BMI, the hypothesized catalysing role of obesity was taken out of the equation, revealing the more subtle causal pathway of the HPT-axis on DM. Certain limitations of this study also need to be addressed. The present study made use of self-reported touchscreen based data, which might be prone to measurement error. As measurement error is likely to be unrelated to the genetic factors (e.g., nondifferential misclassification) this likely resulted in a reduced statistical power. Furthermore, because the genetic instrumental variables, as well as the outcome data set, originated from populations of European ancestry, our results may not be generalized to populations of non-European ancestry.

The findings of the current study add to previous research regarding the role of low thyroid status and DM onset. Several observational human studies observed an association of higher TSH level with a higher risk of DM[6-8]. However, not all observational studies showed an association between higher TSH and DM. De Vries et al. (2019) did not observe a relation between plasma TSH levels within the normal range and incident DM in patients at high cardiovascular risk[9]. The lack of a causal association with TSH and fT4 as observed in the overall study population of the current study may suggest that previously observed associations of alterations in thyroid status and DM onset might have resulted from reverse causality and/or residual confounding. One of the potential interfering mechanisms could be central resistance to thyroid hormones commonly seen coinciding with metabolic syndrome[20]. In addition, many other factors such as auto-immune disorders, could cause residual confounding. Furthermore, these findings confirm our previous observations of no association between circulating TSH and fT4 and risk of DM at population level using two-sample Mendelian Randomization analyses with fewer instruments in a smaller study population[11].

The main novel observation of the current study is the association of higher genetically determined circulating TSH levels with a lower risk of DM in individuals with a lower genetically-determined BMI. Two main routes of action can be hypothesized; either a direct effect of TSH or an indirect route via a lower HPT-axis setpoint. For TSH to have a direct effect on tissue function, TSH-receptors (TSHRs) are required. Various extrathyroidal expression of TSH-receptors are described in literature, including in orbital fibroblasts, adipose tissue, bone, skeletal muscle, thymus and kidney[21,22]. TSH could exert its protective effect against DM via adipose tissue. In mice and human adipocytes expression of TSHR was demonstrated previously [23,24]. Adipocytes were also shown to increase lipolysis in response to stimulation with TSH *in vitro* and *in vivo* [23,24]. Furthermore, interaction with the insulin signaling pathway was demonstrated, leading to an inhibition of PI3K resulting in lower rates of adipogenesis [25]. Thus, higher levels of TSH could potentially be protective against accumulation of adipose tissue and thereby reduce the risk of DM. Alternatively, TSH could influence glucose homeostasis through increasing insulin sensitivity and glucose uptake of skeletal muscle. In line, we described a causal association between higher TSH levels and lower fasting glucose levels in this study. Moon *et al*. (2016) have demonstrated a direct stimulatory effect of TSH on insulin receptor substrate (ISR)-1 expression in muscle tissue and improved glucose tolerance[26]. Another potential etiological pathway could be via immunomodulation. TSHR was shown to be present in thymus tissue, and stimulation with TSH increased development and differentiation of T-cells in both rodent and human thymal cell lines[27]. Hence, individuals with higher circulating levels of TSH could be having a more diverse and effective adaptive immune system. Having a diverse arsenal of T-cells prevents auto-immunity and other sources of low-grade inflammation[28]. As low-grade inflammation is a well-established causal risk factor for developing DM, any factor targeting inflammatory pathways could be a potential strategy for prevention of DM [29]. Higher TSH levels could be such an immunomodulating factor protecting against DM. Apart from direct effects of TSH, an indirect effect of higher TSH via a different HPT-axis setpoint could also explain our findings. As expected from the strong inverse relationship between TSH and fT4, virtually all genetic variants for higher TSH are associated with lower circulating levels of fT4 in the original GWAS[18]. Therefore, our observation could be elaborated to an association of higher TSH and lower fT4, i.e. a lower HPT-axis setpoint, with a lower risk of DM. Previously, thyroid status has been linked to multiple components of glucose homeostasis. It has long been known that thyroid hormones induce hepatic gluconeogenesis[30]. Furthermore, thyroid hormones could affect insulin production and secretion in the pancreas[31]. Although thyroid hormones are required for maturation of pancreatic β-cells, senescence is also accelerated by elevated levels of thyroid hormones in these insulin producing cells [32].

Here, we specifically studied the effects of circulatory TSH and fT4 on DM onset and glucose homeostasis. However, target tissues customize intracellular thyroid hormone levels to their current needs independently of circulating levels in the blood[33]. Deiodinases are key players in the modulation of the availability of thyroid hormones in target tissues [34]. In previous research of our group, we demonstrated that genetic variation in *DIO1* may affect glucose metabolism [11]. This may be more reflective of target tissue levels of thyroid hormone than the circulating levels. We therefore propose that future studies should focus on the role of deiodinases and availability of thyroid hormones in target tissues on glucose homeostasis and the risk of DM.

CONCLUSION

In this study, genetically-determined circulating TSH was associated with a lower risk of DM in participants with low genetically-determined BMI. Moreover, a higher genetically-determined TSH is associated with lower fasting glucose levels in participants without DM, in the overall population and in individuals with both low and high genetically-determined BMI. We did not find evidence for a causal association between higher circulatory TSH or fT4 levels and any of our study outcomes. These findings may indicate that TSH levels affect glucose homeostasis and that higher TSH levels might protect against DM. Only finding these associations in subgroups at lower risk may indicate a more subtle role of the HPT-axis in the aetiology of DM.

Acknowledgements
We express our gratitude to all individuals who participate in the UK Biobank study. The present UK Biobank project was conducted under project numbers 22474 (to RN) and 32743 (to DvH). Statistical analyses were performed using the facilities of the Dutch Supercomputer (Surfsara, Inc; project 38013 to RN). This study was funded by the European Commission-funded project THYRAGE (to DvH; Horizon 2020 research and innovation program under Grant Agreement Number 666869).

Duality of interest
The authors declare to have no conflict of interest.

Contribution statement
Concept and study design: MMB, NAvV, SPM, RN, DvH. Data analyses: MMB, NAvV. Drafting the initial versions of the manuscript: MMB, NAvV. Data collection: RN. Supervision: RN, DvH. Commenting on draft versions of the manuscript: RN, SPM, DvH. Final approval of the manuscript: MMB, NAvV, SPM, RN, DvH.

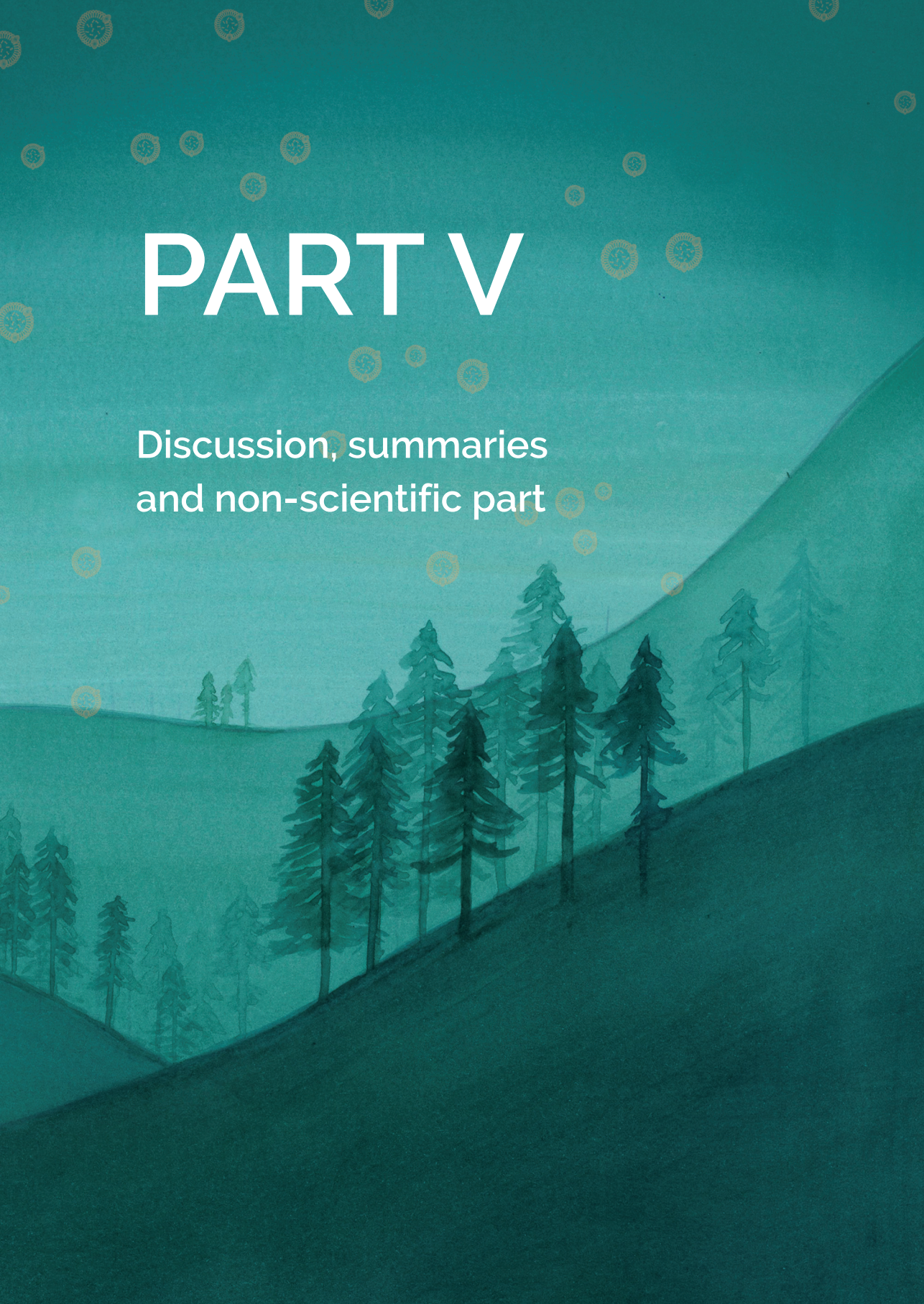
REFERENCES

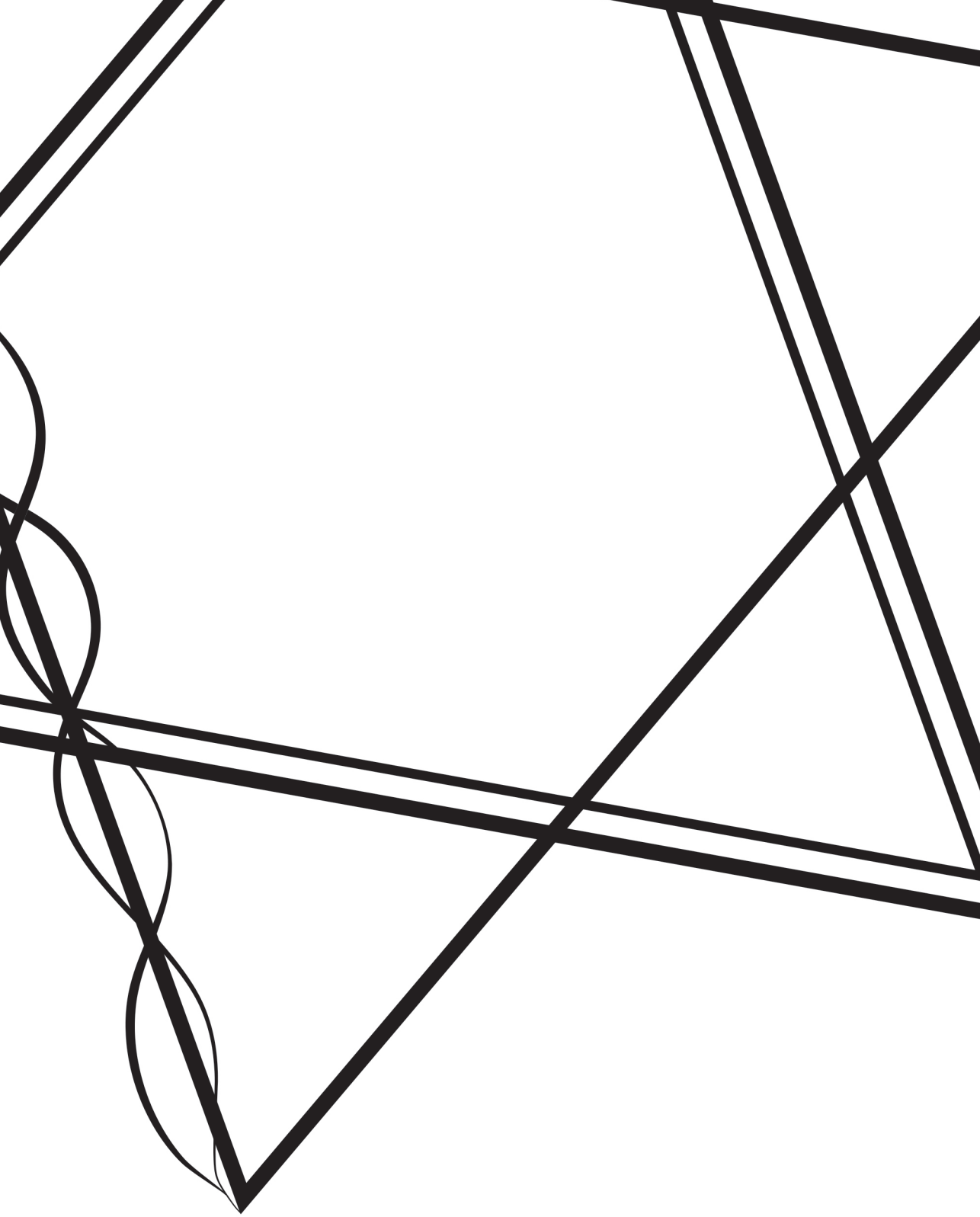
1. Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of obesity among adults: United States, 2011-2012. NCHS Data Brief. 2013(131):1-8.
2. Chen L, Magliano DJ, Zimmet PZ. The worldwide epidemiology of type 2 diabetes mellitus-present and future perspectives. Nat Rev Endocrinol. 2012;8(4):228-236.
3. Kahn SE, Hull RL, Utzschneider KM. Mechanisms linking obesity to insulin resistance and type 2 diabetes. Nature. 2006;444(7121):840-846.
4. Peppa M, Koliaki C, Nikolopoulos P, Raptis SA. Skeletal muscle insulin resistance in endocrine disease. J Biomed Biotechnol. 2010;2010:527850.
5. Chaker L, Ligthart S, Korevaar TIM, et al. Thyroid function and risk of type 2 diabetes: a population-based prospective cohort study. BMC Med. 2016;14(1):150-150.
6. Ittermann T, Schipf S, Dorr M, et al. Hyperthyroxinemia is positively associated with prevalent and incident type 2 diabetes mellitus in two population-based samples from Northeast Germany and Denmark. Nutr Metab Cardiovasc Dis. 2018;28(2):173-179.
7. Wang C. The Relationship between Type 2 Diabetes Mellitus and Related Thyroid Diseases. J Diabetes Res. 2013;2013:390534.
8. Han C, He X, Xia X, et al. Subclinical Hypothyroidism and Type 2 Diabetes: A Systematic Review and Meta-Analysis. PLoS One. 2015;10(8):e0135233.
9. de Vries TI, Kappelle LJ, van der Graaf Y, et al. Thyroid-stimulating hormone levels in the normal range and incident type 2 diabetes mellitus. Acta Diabetol. 2019;56(4):431-440.
10. Kitahara CM, Platz EA, Ladenson PW, Mondul AM, Menke A, Berrington de González A. Body fatness and markers of thyroid function among U.S. men and women. PLoS One. 2012;7(4):e34979.
11. Bos MM, Smit RAJ, Trompet S, van Heemst D, Noordam R. Thyroid Signaling, Insulin Resistance, and 2 Diabetes Mellitus: A Mendelian Randomization Study. J Clin Endocrinol Metab. 2017;102(6):1960-1970.
12. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12(3):e1001779.
13. UKBiobank. UK Biobank: Protocol for a large-scale prospective epidemiological resource 2007.
14. UKBiobank. UK Biobank: Ethics and Governance Framework. 2007.
15. UK10K Consortium, Walter K, Min JL, et al. The UK10K project identifies rare variants in health and disease. Nature. 2015;526(7571):82-90.
16. 1000 Genomes Project Consortium, Auton A, Brooks LD, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68-74.
17. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016;48(10):1279-1283.
18. Teumer A, Chaker L, Groeneweg S, et al. Genome-wide analyses identify a role for SLC17A4 and AADAT in thyroid hormone regulation. Nat Commun. 2018;9(1):4455.
19. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015;518(7538):197-206.
20. Laclaustra M, Moreno-Franco B, Lou-Bonafonte JM, et al. Impaired Sensitivity to Thyroid Hormones Is Associated With Diabetes and Metabolic Syndrome. Diabetes Care. 2019;42(2):303-310.

21. Sorisky A, Bell A, Gagnon A. TSH receptor in adipose cells. Hormone and metabolic research = Hormon- und Stoffwechselforschung = Hormones et metabolisme. 2000;32(11-12):468-474.
22. Williams GR. Extrathyroidal expression of TSH receptor. Annales d'endocrinologie. 2011;72(2):68-73.
23. Janson A, Rawet H, Perbeck L, Marcus C. Presence of thyrotropin receptor in infant adipocytes. Pediatric research. 1998;43(4 Pt 1):555-558.
24. Endo T, Kobayashi T. Expression of functional TSH receptor in white adipose tissues of hyt/hyt mice induces lipolysis in vivo. American journal of physiology Endocrinology and metabolism. 2012;302(12):E1569-1575.
25. Felske D, Gagnon A, Sorisky A. Interacting Effects of TSH and Insulin on Human Differentiated Adipocytes. Hormone and metabolic research = Hormon- und Stoffwechselforschung = Hormones et metabolisme. 2015;47(9):681-685.
26. Moon MK, Kang GH, Kim HH, et al. Thyroid-stimulating hormone improves insulin sensitivity in skeletal muscle cells via cAMP/PKA/CREB pathway-dependent upregulation of insulin receptor substrate-1 expression. Molecular and cellular endocrinology. 2016;436:50-58.
27. van der Weerd K, van Hagen PM, Schrijver B, et al. Thyrotropin acts as a T-cell developmental factor in mice and humans. Thyroid : official journal of the American Thyroid Association. 2014;24(6):1051-1061.
28. King C, Ilic A, Koelsch K, Sarvetnick N. Homeostatic expansion of T cells during immune insufficiency generates autoimmunity. Cell. 2004;117(2):265-277.
29. Lontchi-Yimagou E, Sobngwi E, Matsha TE, Kengne AP. Diabetes mellitus and inflammation. Current diabetes reports. 2013;13(3):435-444.
30. Mullur R, Liu YY, Brent GA. Thyroid hormone regulation of metabolism. Physiological reviews. 2014;94(2):355-382.
31. Medina MC, Molina J, Gadea Y, et al. The thyroid hormone-inactivating type III deiodinase is expressed in mouse and human beta-cells and its targeted inactivation impairs insulin secretion. Endocrinology. 2011;152(10):3717-3727.
32. Aguayo-Mazzucato C, Lee TB, Jr., Matzko M, et al. T3 Induces Both Markers of Maturation and Aging in Pancreatic beta-Cells. Diabetes. 2018;67(7):1322-1331.
33. Luongo C, Dentice M, Salvatore D. Deiodinases and their intricate role in thyroid hormone homeostasis. Nature reviews Endocrinology. 2019;15(8):479-488.
34. Gereben B, Zavacki AM, Ribich S, et al. Cellular and molecular basis of deiodinase-regulated thyroid hormone signaling. Endocr Rev. 2008;29(7):898-938.

PART V

Discussion, summaries and non-scientific part





CHAPTER 6

General discussion and future perspectives

Maxime M Bos

GENERAL DISCUSSION

The aim of this thesis was to study the role and interplay of genetic and environmental factors on cardiometabolic health. Hereby, we aimed to identify potential (causal) biological mechanisms that are of interest as future targets for primary or secondary preventive strategies for cardiometabolic diseases. The first part of this thesis focused on the use of untargeted metabolites in relation to early disturbances in glucose metabolism and insulin sensitivity in middle-aged individuals. The second part was aimed at studying, and potentially identifying, targets that are important in the relations between measures of habitual sleep with cardiometabolic health in the general population. In the third part, we focused on the *APOE* gene and the effect of oily fish intake, physical activity and polyunsaturated fatty acid intake on the onset of age-related diseases, especially coronary artery disease. In the last part of this thesis, we investigated the potential causal association of thyroid status with glucose homeostasis and type 2 diabetes mellitus. In this section of this thesis, I will discuss and further interpret the results from these different chapters. Moreover, the implications of the findings for future preventive and treatment strategies to decrease the burden of cardiometabolic diseases are discussed.

Main findings

The increase in overall life expectancy has resulted in a larger proportion of older individuals worldwide, which is projected to continue to increase[1]. Simultaneously, a rapid rise of individuals with multiple age-related cardiometabolic diseases such as type 2 diabetes mellitus is observed[2]. Therefore, there is an urgent need for the identification of risk factors in an early stage of this condition. In **Chapter 2**, we studied, through untargeted metabolomics, metabolites in relation to measures of insulin resistance in non-diabetic individuals and we replicated the identified metabolites in an independent cohort. We validated the metabolites by assessing their relationship with diabetes mellitus. The findings from **Chapter 2** of this thesis point towards biomarkers of early insulin resistance that may predict type 2 diabetes onset. In particular, several amino acids and saturated fatty acids may contribute to alterations in insulin sensitivity and increased insulin resistance[3]. These findings are in line with previous studies that show that several amino acids, sugar metabolites, and lipids have been associated with a higher risk for diabetes mellitus[4-7]. Especially, several (branched-chain) amino acids are consistently associated with diabetes mellitus[3,5]. Amino acids are essential for protein metabolism and muscle proteins in particular serve as an energy store. When the body is in the anabolic state, amino acids are added to the body's protein pool, while in the catabolic state energy is provided by the breakdown of endogenous proteins to provide the body with amino acids that are used for gluconeogenesis. This occurs when the body is energy deprived as well as when there is an excess of dietary protein. Both in

prediabetic and diabetic individuals, gluconeogenesis has mostly been observed to be higher as compared to controls[8]. Alanine and glutamine are the most important gluconeogenic precursors in the liver. Interestingly, genetic predisposition to type 2 diabetes is associated with increased levels of alanine and genetically-determined higher levels of alanine are associated with an increased risk for type 2 diabetes, thereby providing novel insights into promising causal paths to and from type 2 diabetes[9].

Habitual sleep is increasingly considered as an important factor contributing to cardiometabolic disease onset[10-13]. Previously, in epidemiological cohort studies, both a short and long habitual sleep duration have been associated with risk factors for an adverse cardiometabolic health, but likely via different biological mechanisms[10-13]. The findings as presented in **Chapter 3.1** and **3.2** indicate an observational relationship between habitual sleep duration and sleep quality with glucose metabolism and lipid metabolism. We additionally described that specifically a shorter sleep duration and a poorer sleep quality were associated with higher insulin resistance, higher fasting triglyceride levels and a higher hepatic triglyceride content. However, we demonstrated that it is not sleep duration or sleep quality per se, but rather BMI and associated sleep apnea that drive these associations, since these associations disappeared after adjustment for these factors. Moreover, an innovative research approach that is increasingly used to help disentangle questions of causality – Mendelian Randomization (MR) – was used to study these relations. We showed that using MR, no evidence for a causal association between total, short or long sleep duration, glycemic traits and type 2 diabetes was observed. Therefore, we conclude that previously observed associations of shorter sleep duration and poorer sleep quality with an adverse glucose metabolism and lipid profile, may be explained by BMI and sleep apnea, with no direct effect of sleep duration and quality on glucose metabolism. In a previous study, that assessed the causal association of total sleep duration with 22 prevalent diseases from the Electronic Medical Records in the Partners Biobank (n=16,033), associations were observed with congestive heart failure, obesity, hypertension, restless legs syndrome, and insomnia[14]. While after adjustment for obesity, associations with total sleep duration disappeared for hypertension and insomnia, these associations were maintained for congestive heart failure and restless legs syndrome. Taken together, findings suggest that sleep has no direct effect on energy metabolism and measures of glucose metabolism (and type 2 diabetes), however, sleep may have a direct effect on cardiovascular diseases. In line, another recent study showed that there is a causal effect of short sleep duration on myocardial infarction[15]. In **Chapter 3.3**, we aimed to elucidate the biology of sleep duration-associated cardiovascular risk through gene-lifestyle interactions. We performed short- and long-sleep-SNP interaction analyses in over 125,000 individuals in a large collaborative setting (notably within the Cohorts for

Heart and Ageing Research in Genomics Epidemiology[16,17]) to obtain novel insights in the biological background of sleep duration-associated adverse lipid profiles. A total of 59 novel loci were identified in relation to lipid traits. These loci were previously described in relation to adiposity, inflammation and neuropsychiatric traits. Importantly, the novel lipid loci that we identified with short sleep duration were not identified with long sleep duration, and *vice versa*. Our findings suggest that the biological mechanisms that underlie the relation between short sleep and an adverse lipid profile are distinct from those of long sleep and an adverse lipid profile.

Large initiatives have been performed to investigate the genetic contribution to CVD pathogenesis [18]. Genetic variation in the *APOE* gene has been widely recognized to increase the risk of CVD, which has also been confirmed by genome-wide association studies[18-20]. Genetic variation in *APOE* is a well-established CVD risk factor; however, in some individuals from a number of non-European ancestries (e.g., sub-Saharan), this relationship is not observed[21,22], possibly due to cultural, environmental and lifestyle differences. In **Chapter 4.1**, a hypothesis is proposed regarding potential effect modification by oily fish intake and physical activity on the risk of cardiometabolic and brain diseases associated with genetic variation in *APOE*. We hypothesized a higher level of physical activity and a higher intake of oily fish to decrease the adverse health effects associated with genetic variation in *APOE*, especially in carriers of the ApoE ε4 isoform. This hypothesis was further tested in the largest study to date, embedded in the population-based UK Biobank[23], in **Chapter 4.2**. With a higher level of physical activity and a higher intake of oily fish, we observed a lower incidence of coronary artery disease (CAD) in all ApoE isoform groups. Our results indicate that not only carriers of the ApoE ε4 isoform may benefit from a higher intake of oily fish and a higher physical activity, but that all individuals may benefit from a higher level of physical activity and a higher oily fish intake in regard to CAD incidence. However, this study did not find support for the previous hypothesis that lifestyle interactions are specifically present in individuals carrying the ApoE ε4 isoform. In line, another study that assessed potential interaction with lifestyle factors and *APOE* in relation to cognition failed to prove the existence of a gene-environment interaction[24].

Several studies have observed associations between thyroid status and diabetes mellitus, however causality remained unknown[25]. In **Chapter 5.1**, we focussed on genetically-determined TSH and fT4 in relation to measures of glucose metabolism and type 2 diabetes using MR. We showed that there is no causal association between genetically-determined higher TSH and fT4 with measures of glucose metabolism and type 2 diabetes. However, we found some evidence for an association between genetic variation in the *DIO1* gene and measures of glucose metabolism. In **Chapter 5.2**,

we examined whether the findings of **Chapter 5.1** could be replicated using stronger instruments in a large cohort study of the UK Biobank. Since diabetes mellitus is a heterogeneous disease, which is often not taken into consideration, we took into account different subtypes of diabetes mellitus. Moreover, we assessed the associations in relation to genetically-determined BMI, since BMI is the main risk factor for diabetes mellitus onset. We observed that genetically-determined higher fT4 was associated with diabetes mellitus only in those individuals who had a younger age at diagnosis and did not use insulin or insulin analogues within the first year after diagnosis. When we stratified based on genetically-determined BMI, a higher genetically-determined TSH was associated with diabetes mellitus, however, only in the subgroup with a low genetically-determined BMI. Taken together, these findings may indicate that a higher fT4 level may protect from diabetes mellitus only at a younger age and that a higher TSH level may protect from diabetes mellitus onset, however, only if individuals have a lower BMI.

Conclusions and Future perspectives

In this thesis, we identified markers of early disturbances of glucose metabolism which may reflect diabetes mellitus onset. The findings of **Chapter 2** point to the possibility of using biomarkers that are indicative of risk of future disease. Early identification of individuals at high risk for diabetes is of importance for prevention, mainly because routine screening misses many cases of prediabetes and early type 2 diabetes[26]. These developments point towards the usage of metabolomics in clinical practice. Since high-throughput metabolomics is an easily assessable and relatively cheap method, it is relevant to study the potential usage of this method in clinical practice since an increase in usage of this method is expected. For example, several metabolites may be used in strategies involved in risk stratification of subjects at increased risk of type 2 diabetes. Moreover, the usage of metabolomics for insulin sensitivity and diabetes onset may be implemented in risk stratification of patients with difficult glucose control. Whether such markers provide accurate prognostic information for individuals without diabetes mellitus is subject to further study.

Our findings from **Chapter 3.1** and **3.2** suggest that a shorter sleep duration and a poorer sleep quality are not associated with glucose metabolism and the lipid profile once BMI and sleep apnea are taken into consideration and no causal association between sleep duration, type 2 diabetes and glycemic traits is observed. In **Chapter 3.3**, we demonstrated that we were able to identify additional lipid loci once we take into account interaction with total sleep duration. The new identified loci mainly associated with adiposity and psychological measures. These findings suggest that shifting the focus to anthropometric and psychiatric factors that are related to adverse habitual

sleep may improve risk management and treatment strategies in cardiometabolic disease prevention. That being said, an adverse sleep profile may still be used as a risk factor in risk stratification for cardiometabolic disease. Collectively, the findings of our study provide insights in sleep-associated lipid biology that are of very high interest in follow-up studies. In order to increase insights into sleep, further studies should also consider the complexity of sleep. Habitual sleep is a reflection not only of sleep duration, but also of factors such as chronotype and insomnia. In order to gain better insights in the biology of sleep in relation to cardiometabolic health, studies should incorporate this multidimensional nature of sleep in their researches. Moreover, metabolomics studies are of interest in relation to sleep, since these methods are reflective of the metabolic state of an individual. Additionally, in the studies as described in this thesis, we only used self-reported measures of sleep. These measures may be vulnerable to biases and therefore studies using more objective measures are warranted. One could think of studies that use accelerometer-derived variables. As a final suggestion, future studies could focus on several lifestyle factors in a jointly manner. Sleep is affected by many factors, including food intake and physical activity, which should all be considered when assessing sleep in relation to cardiometabolic health.

Despite the introduction of cholesterol lowering medication, cardiovascular disease (CVD) is still one of the most common causes of morbidity and mortality in the general population[27]. Much of the current research has been focused on disentangling the biology of CVD pathogenesis and the identification of novel targets for disease prevention. The findings from **Chapter 4.1** and **4.2** of this thesis point toward a possible direct link of physical activity, oily fish intake and polyunsaturated fatty acid intake with incident CVD and CAD. While a higher physical activity, fish intake and PUFA intake both decreased the risk of CAD, no evidence for a statistical environment-*APOE* interaction was observed. Therefore, it seems unlikely that interventions intended to reduce cardiovascular risk show different effects depending on *APOE* genotype. Importantly, we demonstrated that individuals that are at a high genetic risk for CVD and CAD can still benefit from a healthier lifestyle. Thus, the addition of lifestyle advice to existing treatment options for cardiovascular disease and to prevention strategies should be further investigated.

In the last part of this thesis, we focussed on the association of thyroid status with diabetes mellitus and insulin resistance. In **Chapter 5.1**, we demonstrated that there is no evidence for a causal association between circulating TSH and fT4 with type 2 diabetes mellitus and glycemic traits. In **Chapter 5.2**, we provide evidence that in a larger population a higher genetically-determined TSH is causally associated with diabetes mellitus, however, only in participants with a genetically-determined low BMI. This may imply that there still may be a causal association between thyroid status and diabetes,

however, BMI may be a stronger risk factor for diabetes mellitus, thereby overruling the potential protective effect of a higher TSH. Future studies should investigate this hypothesis. Moreover, in the current study, we only assessed the effects of circulatory TSH and fT4 on DM onset. Variations in circulating TSH and fT4 are thought to predominantly reflect the sensitivity of the thyroid gland to feedforward stimulation and the sensitivity of the pituitary gland to feedback inhibition. In contrast, deiodinases have a critical role in the activation of thyroid hormones in target tissues[28]. In **Chapter 5.1**, we demonstrated that genetic variation in *DIO1* may affect glucose metabolism[29]. This result may indicate a role of reduced bioavailability of thyroid hormone in target tissues (such as liver) in glucose metabolism. We therefore propose that future studies should focus on the role of deiodinases and availability of thyroid hormones in target tissues on the risk of DM.

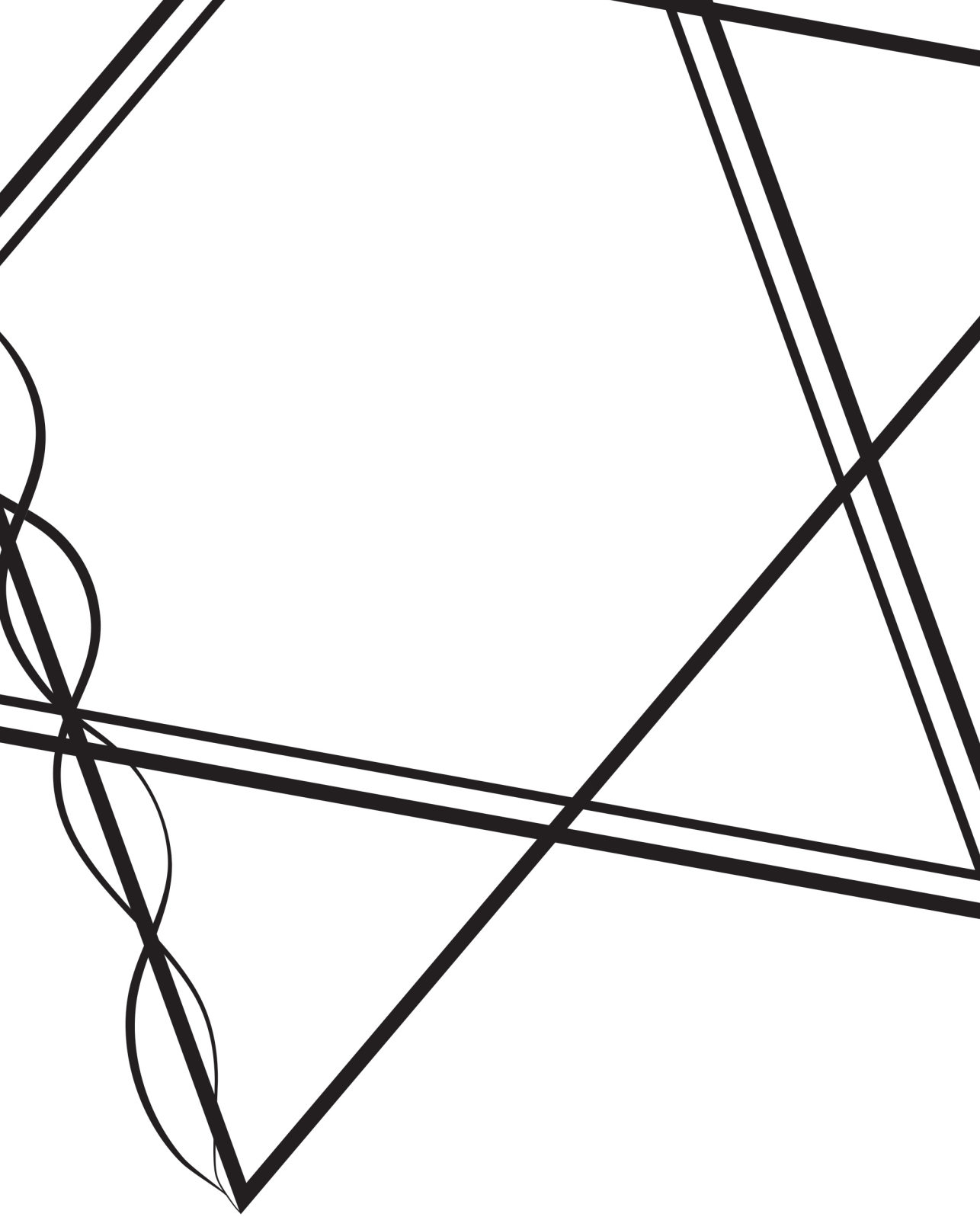
This thesis provides novel insights in several associations that are potentially of importance in cardiometabolic health. However, we should keep in mind that cardiometabolic diseases are very complex and multiple pathways act together in order to maintain cardiometabolic health. For example, another study demonstrated that sleep duration tended to be positively associated with free thyroxine levels and negatively associated with HbA1c and CRP[30]. Moreover, these findings showed that short-sleeping UK adults are more likely to have obesity, a disease with many comorbidities. This extraordinary level of complexity calls for more sophisticated approaches in order to deepen our understanding of the mechanisms that underlie cardiometabolic health and disease. For example, national mega-biobanks such as the UK Biobank, the Million Veteran Program and the China Kadoorie Biobank offer a variety of possibilities to perform research with genomic data in large well-phenotyped populations. In combination with other omics approaches (e.g. metabolomics, proteomics), these studies may provide valuable insights in cardiometabolic health. Identification of genetic determinants of cardiometabolic health, assisted by reliable and cost-effective biomarkers, can help in the further understanding of the individual risk differences in developing cardiometabolic disease. Moreover, the tremendous amount of data that will be generated by these studies asks for more complex modelling of data. Therefore, more advanced machine-learning methods such as methods to perform advanced clustering and classification may be of special interest in this era of big data. Because of the exponential increase in data (complexity), the application of the more traditional bioinformatic tools may become soon outdated. By using more advanced data modelling techniques, several types of data (e.g. clinical and omics) may be analysed simultaneously. When these methods are then used to predict individual risks for disease onset, this may be of value in personalized medicine. In the future, these efforts may lead to a healthcare system in which the risk for an individual (or group of individuals) for certain diseases can be estimated more precisely. Then, future and preventive strategies may be implemented

specifically tailored to an individual's genetic profile to more effectively prevent and treat several cardiometabolic diseases. Eventually, by considering the multidimensionality of cardiometabolic health and the adaptation of future preventive and curative strategies based upon this complexity may add to a more effective prevention and treatment of cardiometabolic disease.

REFERENCES

1. Kontis V, Bennett JE, Mathers CD, Li G, Foreman K, Ezzati M. Future life expectancy in 35 industrialised countries: projections with a Bayesian model ensemble. *Lancet.* 2017;389(10076):1323-1335.
2. Gorelick PB, Furie KL, Iadecola C, et al. Defining Optimal Brain Health in Adults: A Presidential Advisory From the American Heart Association/American Stroke Association. *Stroke.* 2017;48(10):e284-e303.
3. Würtz P, Soininen P, Kangas AJ, et al. Branched-Chain and Aromatic Amino Acids Are Predictors of Insulin Resistance in Young Adults. 2013;36(3):648-655.
4. Del Coco L, Vergara D, De Matteis S, et al. NMR-Based Metabolomic Approach Tracks Potential Serum Biomarkers of Disease Progression in Patients with Type 2 Diabetes Mellitus. *J Clin Med.* 2019;8(5).
5. Gar C, Rottenkolber M, Prehn C, Adamski J, Seissler J, Lechner A. Serum and plasma amino acids as markers of prediabetes, insulin resistance, and incident diabetes. *Crit Rev Clin Lab Sci.* 2018;55(1):21-32.
6. Mook-Kanamori DO, de Mutsert R, Rensen PC, et al. Type 2 diabetes is associated with postprandial amino acid measures. *Arch Biochem Biophys.* 2016;589:138-144.
7. t Hart LM, Vogelzangs N, Mook-Kanamori DO, et al. Blood Metabolomic Measures Associate With Present and Future Glycemic Control in Type 2 Diabetes. *J Clin Endocrinol Metab.* 2018;103(12):4569-4579.
8. Chung ST, Hsia DS, Chacko SK, Rodriguez LM, Haymond MW. Increased gluconeogenesis in youth with newly diagnosed type 2 diabetes. *Diabetologia.* 2015;58(3):596-603.
9. Liu J, van Klinken JB, Semiz S, et al. A Mendelian Randomization Study of Metabolite Profiles, Fasting Glucose, and Type 2 Diabetes. *Diabetes.* 2017;66(11):2915-2926.
10. Wong PM, Manuck SB, DiNardo MM, Korytkowski M, Muldoon MF. Shorter sleep duration is associated with decreased insulin sensitivity in healthy white men. *Sleep.* 2015;38(2):223-231.
11. Reutrakul S, Van Cauter E. Interactions between sleep, circadian function, and glucose metabolism: implications for risk and severity of diabetes. *Annals of the New York Academy of Sciences.* 2014;1311:151-173.
12. Cappuccio FP, Taggart FM, Kandala NB, et al. Meta-analysis of short sleep duration and obesity in children and adults. *Sleep.* 2008;31(5):619-626.
13. Cappuccio FP, D'Elia L, Strazzullo P, Miller MA. Quantity and quality of sleep and incidence of type 2 diabetes: a systematic review and meta-analysis. *Diabetes Care.* 2010;33(2):414-420.
14. Dashti HS, Redline S, Saxena R. Polygenic risk score identifies associations between sleep duration and diseases determined from an electronic medical record biobank. *Sleep.* 2018;42(3).
15. Daghlas I, Dashti HS, Lane J, et al. Sleep Duration and Myocardial Infarction. *J Am Coll Cardiol.* 2019;74(10):1304-1314.
16. Rao DC, Sung YJ, Winkler TW, et al. Multiancestry Study of Gene-Lifestyle Interactions for Cardiovascular Traits in 610 475 Individuals From 124 Cohorts: Design and Rationale. *Circ Cardiovasc Genet.* 2017;10(3):e001649.
17. Psaty BM, O'Donnell CJ, Gudnason V, et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ Cardiovasc Genet.* 2009;2(1):73-80.

18. Nikpay M, Goel A, Won HH, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nature genetics.* 2015;47(10):1121-1130.
19. CARDIoGRAMplusC4D Consortium, Deloukas P, Kanoni S, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nature genetics.* 2013;45(1):25-33.
20. Song Y, Stampfer MJ, Liu S. Meta-analysis: apolipoprotein E genotypes and risk for coronary heart disease. *Annals of internal medicine.* 2004;141(2):137-147.
21. Chen CH, Mizuno T, Elston R, et al. A comparative study to screen dementia and APOE genotypes in an ageing East African population. *Neurobiology of aging.* 2010;31(5):732-740.
22. Gureje O, Ogunniyi A, Baiyewu O, et al. APOE epsilon4 is not associated with Alzheimer's disease in elderly Nigerians. *Annals of neurology.* 2006;59(1):182-185.
23. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine.* 2015;12(3):e1001779.
24. Lyall DM, Celis-Morales C, Lyall LM, et al. Assessing for interaction between APOE ε4, sex, and lifestyle on cognitive abilities. *Neurology.* 2019;92(23):e2691-e2698.
25. Chaker L, Ligthart S, Korevaar TIM, et al. Thyroid function and risk of type 2 diabetes: a population-based prospective cohort study. *BMC Med.* 2016;14(1):150-150.
26. Wilson PWF, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB, Sr. Prediction of Incident Diabetes Mellitus in Middle-aged Adults: The Framingham Offspring Study. *JAMA Internal Medicine.* 2007;167(10):1068-1074.
27. Andersson C, Vasan RS. Epidemiology of cardiovascular disease in young individuals. *Nature reviews Cardiology.* 2018;15(4):230-240.
28. Gereben B, Zavacki AM, Ribich S, et al. Cellular and molecular basis of deiodinase-regulated thyroid hormone signaling. *Endocr Rev.* 2008;29(7):898-938.
29. Bos MM, Smit RAJ, Trompet S, van Heemst D, Noordam R. Thyroid Signaling, Insulin Resistance, and 2 Diabetes Mellitus: A Mendelian Randomization Study. *J Clin Endocrinol Metab.* 2017;102(6):1960-1970.
30. Potter GDM, Cade JE, Hardie LJ. Longer sleep is associated with lower BMI and favorable metabolic profiles in UK adults: Findings from the National Diet and Nutrition Survey. *PloS one.* 2017;12(7):e0182195-e0182195.



CHAPTER 7.1

English summary

ENGLISH SUMMARY

As a consequence of increased ageing of the population, age-related diseases have vastly increased. Cardiovascular disease is the leading cause of death. Moreover, a steep increase in the prevalence of type 2 diabetes mellitus is observed worldwide. In addition to obesity, being one of the main risk factors, there are multiple genetic and nongenetic factors that jointly determine the risk of developing cardiometabolic diseases. Therefore, it is important in research to focus on a better understanding of those factors that are (causally) associated with cardiometabolic diseases. This thesis aims to study the interplay between non-modifiable factors (genetics) and modifiable lifestyle factors (e.g. sleep, nutrition, physical activity) with cardiometabolic and cardiovascular health.

In chapter 2, we aimed to identify plasma metabolites associated with different indices of early disturbances in glucose metabolism and insulin sensitivity. In 233 non-diabetic individuals from the Leiden Longevity Study, we tested whether we could find associations between metabolites and measures of glucose metabolism and insulin resistance. We repeated these analyses in another cohort (the Netherlands Epidemiology of Obesity study) in individuals without diabetes to replicate our findings. Next, we investigated the significant findings in individuals with diabetes mellitus to validate these findings. We identified 12 metabolites to be associated with measures of glucose metabolism. Moreover, five of these metabolites, tyrosine, alanine, valine, tryptophan, and alpha-ketoglutaric acid levels, had a higher mean level in blood in individuals with diabetes mellitus compared to individuals without diabetes mellitus. These results may improve the understanding of the mechanisms involved in disease etiology and thereby may contribute to improved diagnostics of the early metabolic disturbances preceding diabetes mellitus.

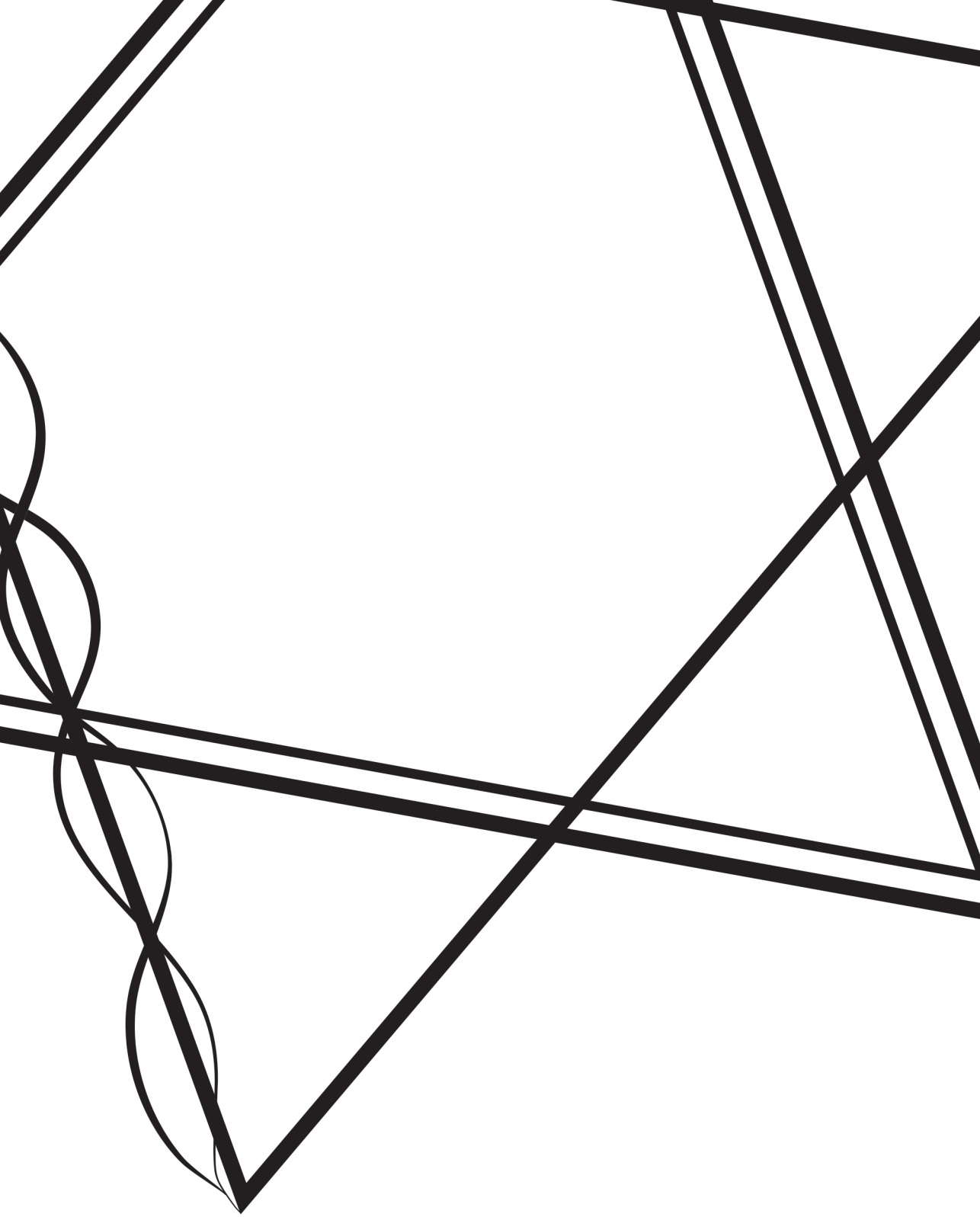
In chapters 3.1, 3.2 and 3.3, we studied the association of sleep traits with cardiometabolic outcomes. It is known that both a short and long sleep duration and a poor sleep quality may affect glucose metabolism as well as serum and hepatic lipid levels. Therefore, we aimed to study the association between sleep duration and quality with serum and hepatic lipid content in chapter 3.1, and with glucose metabolism and insulin sensitivity in chapter 3.2. These studies were performed in the Netherlands Epidemiology of Obesity study. Self-reported sleep duration and quality were assessed using the Pittsburgh Sleep Quality Index questionnaire and outcomes were measures of the lipid profile, hepatic triglyceride content and glycemic traits. Compared with participants with medium sleep (7.0 hours of sleep per night), participants with the shortest sleep (5.0 hours sleep per night) had higher levels of blood lipids and hepatic triglyceride content, and higher insulin resistance. Bad sleep quality as compared with good sleep

quality was associated with higher levels of triglycerides in the blood and higher insulin resistance. However, when we adjusted our analyses for body mass index and the risk of sleep apnea, all these associations disappeared. In addition, we performed a Mendelian Randomization analysis to test for a potential causal relationship of sleep duration with glucose metabolism and diabetes, and we did not find evidence these associations were based on causality. Therefore, we concluded that previously observed cross-sectional associations of shorter sleep duration and poorer sleep quality with an adverse lipid profile and higher insulin resistance, may be explained by body mass index and the risk of sleep apnea, rather than by a direct effect of sleep itself. Therefore, in chapter 3.3, we aimed to elucidate the biological pathways of an adverse lipid profile for both short and long sleepers. We performed short- and long-sleep-SNP interaction analyses in over 125,000 individuals in a large collaborative setting (participating in the gene-lifestyle working group of the Cohorts for Heart and Ageing Research in Genomics Epidemiology) to obtain novel insights in the biological background of sleep duration-associated adverse lipid profiles. A total of 59 novel loci were identified in relation to lipid traits and we identified sleep-interactions for known lipid loci. For short sleep, the loci were previously described in relation to adiposity and inflammation and for long sleep in relation to neuropsychiatric traits. These results contribute to our understanding of the biological mechanisms that underlie sleep-associated adverse health outcomes.

Candidate gene studies and genome-wide association studies found that genetic variation in *APOE* is robustly associated with multiple cardiometabolic diseases and age-related phenotypes. However, large differences in the deleterious effects of this gene occur. In chapter 4.1, we described how a higher intake of fish and polyunsaturated fatty acids (e.g. omega-3 fatty acids) and a higher level of physical activity may be beneficial in preventing cognitive decline and heart disease onset in carriers of the risk *APOE ε4* allele. In chapter 4.2, we tested this hypothesis in the large UK Biobank, and we aimed to investigate the presence of a gene-environment interaction between physical activity, fish intake and polyunsaturated fatty acid intake and *APOE* on incident cardiovascular heart disease. In a population comprised of 344,092 European participants with no history of cardiovascular disease at study inclusion, we observed that a higher level of physical activity and a higher intake of oily fish was associated with a lower risk of new onset cardiovascular disease. This association was similarly observed in both APOE genotype carriers (risk and nonrisk). A higher intake of polyunsaturated fatty acids was only associated with a lower risk of cardiovascular disease in carriers of the risk allele. We did not find evidence for a formal gene-environment interaction on a multiplicative scale on cardiovascular disease onset. These results indicate that a better lifestyle might be similarly effective across all APOE isoform carriers in reducing new cardiovascular disease onset.

Increasing evidence suggests an association between levels of thyroid hormone and insulin resistance, and diabetes mellitus. In chapters 5.1 and 5.2, we aimed to investigate the potential causal association of thyroid hormone status with glucose metabolism and diabetes mellitus. Genetic variants were used in Mendelian Randomization analyses to assess these associations relatively free of residual confounding and/or reverse causation. Genetically-determined thyroid hormone status was not associated with glucose metabolism nor with diabetes mellitus in the study described in chapter 5.1. In chapter 5.2, we used a new study population and a more recent set of genetic instruments. Similar as in chapter 5.1, thyroid hormone status was not associated with diabetes mellitus. However, in a group with a genetically-determined lower body mass index, higher thyroid stimulating hormone (TSH) level was causally associated with a lower risk of diabetes mellitus. This indicates that there might be a causal association between thyroid hormone status and diabetes mellitus, however, only in individuals with a lower body mass index. These findings suggest a potential protective effect of TSH in individuals with a low body mass index.

This thesis has provided novel insights in several associations that are potentially important for cardiometabolic health. However, we should keep in mind that cardiometabolic diseases are very complex and multiple pathways act together in order to maintain cardiometabolic health. This extraordinary level of complexity calls for more sophisticated approaches in order to deepen our understanding of the mechanisms that underlie cardiometabolic health and disease. For example, by using more advanced data modelling techniques, several types of data (e.g. clinical and omics) may be analyzed simultaneously. When these methods are then used to predict individual risks for disease onset, this may be of value in personalized medicine. In the future, these efforts may lead to a healthcare system in which the risk for an individual (or group of individuals) for certain diseases can be estimated more precisely. Eventually, preventive strategies may be implemented, specifically tailored to an individual's genetic and/or metabolomics profile to more effectively prevent and treat cardiometabolic diseases. This may thereby result in a decrease of the burden of cardiometabolic disorders and age-related diseases on the patient and the society as a whole.



CHAPTER 7.2

Nederlandse samenvatting

NEDERLANDSE SAMENVATTING

Als gevolg van de toenemende vergrijzing van de bevolking zijn leeftijdsgebonden ziekten enorm toegenomen. Hart- en vaatziekten zijn momenteel een van de belangrijkste doodsoorzaken. Daarnaast wordt wereldwijd een sterke stijging van de prevalentie van type 2 diabetes mellitus waargenomen. Naast obesitas, als een van de belangrijkste risicofactoren, zijn er meerdere genetische en niet-genetische factoren die samen het risico bepalen op het ontwikkelen van deze cardiometabole ziekten. Het is daarom van belang om te focussen op een beter begrip van de factoren die (causaal) geassocieerd zijn met cardiometabole ziekten. Dit proefschrift heeft tot doel het samenspel tussen genetische factoren en leefstijlfactoren (bv. slaap, voeding, lichamelijke activiteit) met cardiometabole en cardiovasculaire gezondheid te bestuderen.

In hoofdstuk 2 wilden we metabolieten identificeren die geassocieerd zijn met verschillende indices van vroege verstoringen in het glucosemetabolisme en de insulinegevoeligheid. Bij 233 niet-diabetici uit de Leiden Longevity Study hebben we getest of we verbanden konden vinden tussen metabolieten en metingen van het glucosemetabolisme en insulineresistentie. We herhaalden deze analyses in een ander cohort (de Nederlandse epidemiologie van Obesitas – NEO - studie) om de gevonden metabolieten te repliceren. Vervolgens testten we de gevonden associaties in mensen met diabetes mellitus. We hebben 12 metabolieten geïdentificeerd die geassocieerd waren met metingen van het glucosemetabolisme. Bovendien waren de gemiddelde concentraties in het bloed van vijf van deze metabolieten, tyrosine, alanine, valine, tryptofaan en alfaketoglutaarzuur, ook hoger bij personen met diabetes mellitus in vergelijking tot personen zonder diabetes mellitus. Deze resultaten kunnen het inzicht in de mechanismen die betrokken zijn bij de ziekte-etiologie verbeteren en daardoor bijdragen aan een verbeterde diagnostiek van de vroege metabole stoornissen die voorafgaan aan diabetes mellitus.

In de hoofdstukken 3.1, 3.2 en 3.3 hebben we de associatie van slaap met cardiometabole uitkomsten bestudeerd. Het is bekend dat zowel een korte als een lange slaapduur en een slechte slaapkwaliteit het glucosemetabolisme en de bloed- en levervetten niveaus kunnen beïnvloeden. Daarom wilden we het verband bestuderen tussen slaapduur en slaapkwaliteit met bloed- en levervetten in hoofdstuk 3.1 en met glucosemetabolisme en insulinegevoeligheid in hoofdstuk 3.2. Deze onderzoeken zijn uitgevoerd in het Nederlandse epidemiologie van obesitasonderzoek. Zelf gerapporteerde slaapduur en slaapkwaliteit werden verkregen met behulp van de Pittsburgh Sleep Quality Index-vragenlijst en de uitkomsten waren metingen van het lipidenprofiel, hepatisch triglyceridengehalte en glycemische eigenschappen. Vergeleken met gemiddelde slapers (7,0 uur slaap per nacht), hadden de kortste slapers (5,0 uur slaap per nacht) hogere niveaus van bloedvetten,

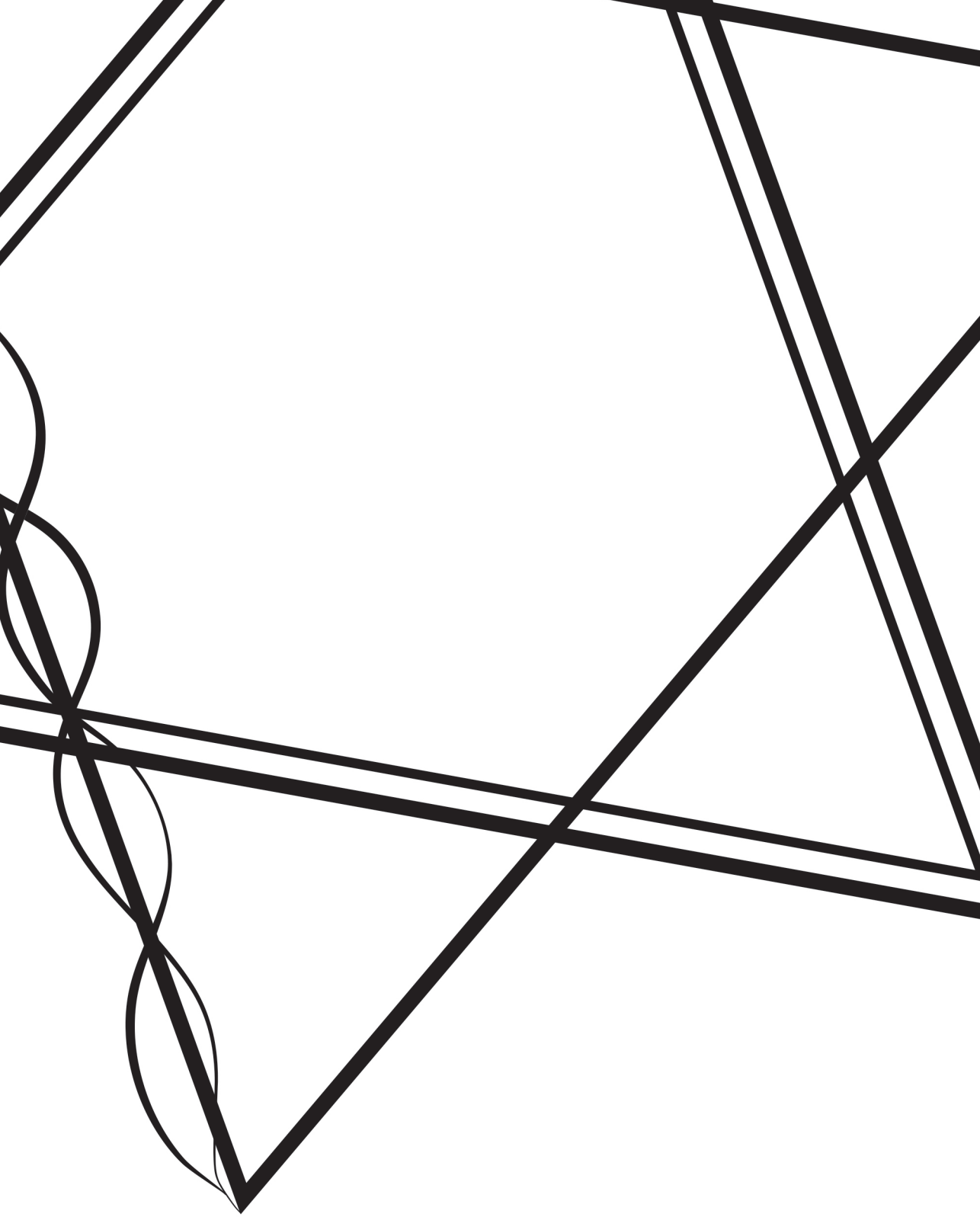
levervetten en insulineresistentie. Slechte slaapkwaliteit in vergelijking met goede slaapkwaliteit was geassocieerd met hogere triglyceridenwaarden in het bloed en een hogere insulineresistentie. Toen we in onze analyses echter rekening hielden met de body mass index en het risico op slaapapneu verdwenen al deze associaties. Daarnaast hebben we een Mendeliaanse randomisatie analyse uitgevoerd om de mogelijke causaliteit in de relatie van slaapduur met glucosemetabolisme en diabetes te testen. We hebben geen bewijs gevonden voor een oorzakelijk verband. Daarom concludeerden we dat eerder waargenomen cross-sectionele associaties tussen een kortere slaapduur en een slechtere slaapkwaliteit met een negatief lipidenprofiel en een hogere insulineresistentie kunnen worden verklaard door de body mass index en het risico op slaapapneu, in plaats van door een direct effect van slaap zelf. Vervolgens hebben we in hoofdstuk 3.3 getracht de biologische routes die leiden tot hogere bloedvetten voor zowel korte als lange slapers op te helderen. We hebben SNP-interactieanalyses uitgevoerd met korte en lange slaap in meer dan 125.000 individuen in een grote samenwerkingsomgeving (met name binnen de Cohorts for Heart and Aging Research in Genomics Epidemiology in de gen-omgevings interactie werkgroep). Er werden in totaal 59 nieuwe loci geïdentificeerd met betrekking tot lipiden in het bloed en we identificeerden slaapinteracties voor bekende lipide loci. Voor korte slaap werden de gevonden loci eerder beschreven in relatie tot adipositas en ontsteking en voor lange slaap in verband gebracht met neuropsychiatrische eigenschappen. Deze resultaten dragen bij aan ons begrip van de biologische mechanismen die ten grondslag liggen aan slaapgerelateerde nadelige gezondheidsresultaten.

Kandidaat-genstudies en genetische associatiestudies hebben aangetoond dat genetische variatie in het *APOE*-gen robuust wordt geassocieerd met meerdere cardiometabole ziekten en leeftijdsgebonden fenotypen. Er zijn echter grote verschillen in de schadelijke effecten van dit gen. In hoofdstuk 4.1 hebben we beschreven hoe een hogere inname van vis en meervoudig onverzadigde vetzuren (bijv. omega-3-vetzuren) en een hoger niveau van lichamelijke activiteit gunstig kunnen zijn bij het voorkomen van cognitieve achteruitgang en het ontstaan van hartaandoeningen bij dragers van het risico *APOE ε4*-allel. In hoofdstuk 4.2 hebben we deze hypothese getest in de grote Britse Biobank de "UK Biobank". Bovendien wilden we de aanwezigheid van een gen-omgevingsinteractie van fysieke activiteit, visinname en meervoudig onverzadigde vetzuurinname met *APOE* op hartaandoeningen onderzoeken. In een populatie van 344.092 Europese deelnemers zonder ziektegeschiedenis van hart- en vaatziekten op het moment van deelname aan de studie, hebben we waargenomen dat een hoger niveau van fysieke activiteit en een hogere inname van vette vis geassocieerd waren met een lager risico op nieuw beginnende hart- en vaatziekten. Een hogere inname van meervoudig onverzadigde vetzuren was alleen geassocieerd met een lager risico op hart- en vaatziekten bij dragers van het risico-allel. We vonden geen bewijs voor

een formele gen-omgevingsinteractie bij het ontstaan van hart- en vaatziekten. Deze resultaten geven aan dat een betere levensstijl voor alle *APOE*-dragers even effectief zou kunnen zijn bij het verminderen van het ontstaan van nieuwe hart- en vaatziekten.

Toenemend bewijs suggereert een verband tussen concentraties van schildklierhormoon in het bloed met insulineresistentie en diabetes mellitus. In de hoofdstukken 5.1 en 5.2 wilden we het mogelijk oorzakelijk verband tussen de schildklierhormoonstatus, het glucosemetabolisme en diabetes mellitus onderzoeken. Genetische varianten werden gebruikt in Mendeliaanse randomisatie-analyses om deze associaties relatief vrij van residuele verstoring en/of omgekeerde oorzakelijkheid te beoordelen. De genetisch bepaalde schildklierhormoonstatus was niet geassocieerd met glucosemetabolisme en evenmin met diabetes mellitus in hoofdstuk 5.1. In hoofdstuk 5.2 hebben we een nieuwe onderzoekspopulatie en een recentere set van genetische instrumenten gebruikt. Ook in deze studie was de schildklierhormoonstatus niet geassocieerd met diabetes mellitus. In een groep met een genetisch lagere body mass index was een hoger schildklierstimulerend hormoon (TSH) -niveau echter wel causaal geassocieerd met een lager risico op diabetes mellitus. Dit geeft aan dat er mogelijk een causaal verband bestaat tussen de schildklierhormoonstatus en diabetes mellitus, maar alleen bij personen met een lagere body mass index. Deze bevindingen suggereren een potentieel beschermend effect van TSH bij personen met een lage body mass index.

Dit proefschrift heeft nieuwe inzichten opgeleverd in verschillende associaties die potentieel belangrijk zijn voor de cardiometabole gezondheid. We moeten er echter rekening mee houden dat cardiometabole ziekten zeer complex zijn en dat meerdere routes samenwerken om de cardiometabole gezondheid te behouden. Dit buitengewone niveau van complexiteit vereist meer verfijnde benaderingen om ons inzicht in de mechanismen die ten grondslag liggen aan cardiometabole gezondheid en ziekte te verdiepen. Door bijvoorbeeld meer geavanceerde gegevensmodelleringstechnieken te gebruiken, kunnen verschillende soorten gegevens (bijv. klinische en omics) tegelijkertijd worden geanalyseerd. Wanneer deze methoden vervolgens worden gebruikt om individuele risico's voor het begin van de ziekte te voorspellen, kan dit van waarde zijn in gepersonaliseerde geneeskunde. Deze inspanningen kunnen in de toekomst leiden tot een zorgstelsel waarin het risico voor een individu (of groep individuen) voor bepaalde ziekten nauwkeuriger kan worden ingeschat. Uiteindelijk kunnen preventieve strategieën worden geïmplementeerd die specifiek zijn afgestemd op het genetische en/of metabolietenprofiel van een individu om zo verschillende cardiometabole ziekten effectiever te voorkomen en te behandelen. Dit kan daardoor resulteren in een vermindering van het effect van cardiometabole aandoeningen en leeftijdsgebonden ziekten bij de patiënt en bij de samenleving als geheel.

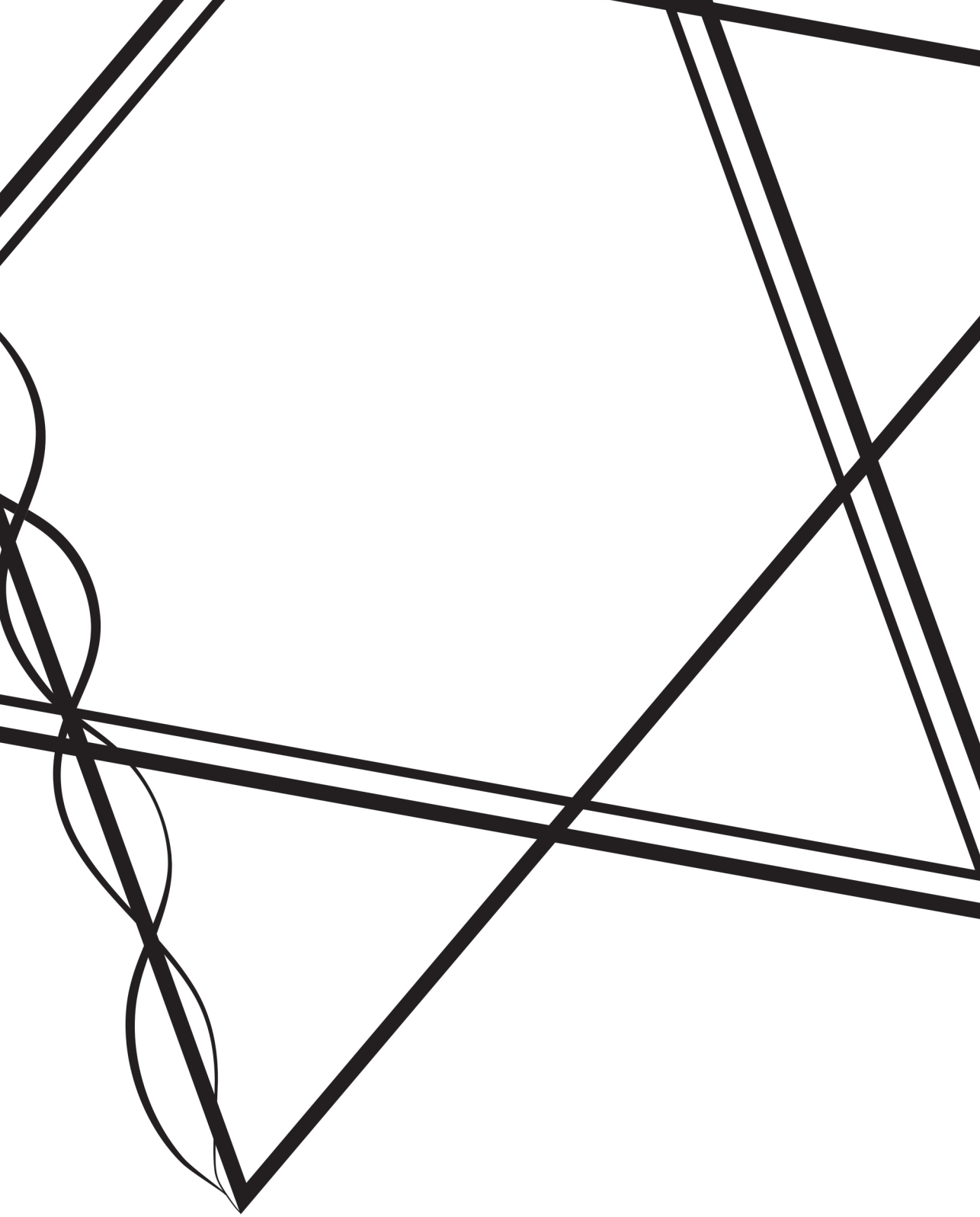


CHAPTER 7.3

Curriculum Vitae

CURRICULUM VITAE

Maxime Milou Bos is geboren op 18 juli 1991 in Purmerend. Ze behaalde haar atheneumdiploma in 2009 aan het Bertrand Russel College in Krommenie. Nadat ze haar diploma behaalde heeft zij de bachelor Biomedische Wetenschappen aan de Universiteit van Amsterdam gedaan. Hierna is zij een master Biomedical Sciences Management gaan doen aan de Universiteit van Leiden. Deze opleiding heeft zij tijdelijk onderbroken om voor anderhalf jaar als salescoach te werken bij Pepperminds in Haarlem. Echter bleek toch haar interesse bij het onderzoek te liggen. Vanaf september 2015 heeft zij haar master Biomedical Sciences Research afgemaakt. Aansluitend heeft zij haar promotieonderzoek gedaan van november 2016 tot oktober 2019 op de afdeling Ouderengeneeskunde in het Leids Universitair Medisch Centrum. Momenteel werkt zij als postdoctoraal onderzoeker op de afdeling Epidemiologie in het Erasmus Medisch Centrum in Rotterdam.



CHAPTER 7.4

Dankwoord

DANKWOORD

Allereerst zou ik alle deelnemers van de verschillende studies waarop de onderzoeken in dit proefschrift zijn gebaseerd willen bedanken.

Daarnaast wil ik graag mijn promotor professor Gerard Jan Blauw en mijn co-promotoren dr. Ir. Diana van Heemst en dr. Raymond Noordam bedanken. Beste Gerard Jan, door jou heb ik de mogelijkheid gehad om mijn promotieonderzoek te doen. Diana, je was een enorme hulp wanneer ik even niet goed wist hoe ik verder moest met mijn onderzoeken. Bedankt voor de verhelderende gesprekken en goede adviezen die ik van je heb mogen ontvangen. Raymond, ondanks dat wij allebei ontzettend koppig waren, heb ik veel van jou kunnen leren. Bedankt dat je altijd bereikbaar was.

De leden van de promotiecommissie wil ik graag bedanken voor het kritisch beoordelen van dit proefschrift.

De medewerkers van de afdeling Ouderengeneeskunde en Epidemiologie. Ik wil jullie bedanken voor de samenwerking, de gezelligheid en de leerzame momenten. In het bijzonder, mijn kamergenoten: Laura, Nicolien, Evie, Marjan, Lisa en Jiao.

Graag wil ik al mijn vrienden bedanken. In het bijzonder gaat mijn dank uit naar Sarah, Melissa, Premika, Angela en Gabriella. Sarah, wij begrijpen elkaar. Bedankt dat jij mijn paranimf wilt zijn.

Mijn grootste dank gaat uit naar mijn geweldige familie.

Beaudine, het betekent heel veel voor mij dat je vandaag naast mij staat. Zoals wij nu naast elkaar staan, zo hoop ik dat wij altijd naast elkaar zullen staan.

Lieve mama, mijn steun en toeverlaat. Wanneer ik zelf even niet meer weet wat ik moet doen, ben jij er altijd om mij te helpen.

Lieve papa, wat zou jij ontzettend trots zijn geweest op dit proefschrift. Ik zou jou in mijn hele leven niet genoeg kunnen bedanken voor alles wat je voor ons over hebt gehad.

Oma, u biedt altijd een luisterend en begrijpend oor; ik kan niet in woorden uitdrukken hoeveel u voor mij betekent. Bedankt dat u er altijd voor mij bent. Opa, u zei het al een aantal jaar geleden: professortje. Ik ben keihard aan het werk om dat ook daadwerkelijk waar te maken.

Diana, Salo, bedankt dat jullie er voor mij zijn.

Mijn vriendin Masuma. De laatste loodjes wegen inderdaad het zwaarst en jij was er om mij bij te staan. Ik kijk uit naar wat de toekomst ons zal brengen.