Insight into the pathophysiology of cardiometabolic diseases using multiple omics approaches

Wang, W.

Citation

Wang, W. (2024, November 27). *Insight into the pathophysiology of cardiometabolic diseases using multiple omics approaches*. Retrieved from https://hdl.handle.net/1887/4169999

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/4169999

Note: To cite this publication please use the final published version (if applicable).

Insight into the Pathophysiology of Cardiometabolic Diseases using Multiple Omics Approaches

Wenyi Wang

Insight into the Pathophysiology of Cardiometabolic Diseases using Multiple Omics Approaches

Copyright © W,Wang, 2024

Cover design: OurFridayGoldfish Design Studio and Wenyi Wang
Layout: Wenyi Wang
Printed by: ProefschriftMaken
ISBN: 978-94-6510-201-6

The work described in this PhD thesis was conducted at the Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands

W,Wang was supported by the Chinese Scholarship Council. Part of the research described in this thesis was supported by Leiden University Fund (LUF).
Financial support from the Dutch Heart Foundation is gratefully acknowledged.

All right reserved. No part of this thesis may be transformed, reproduced or transmitted in any form by any means without prior permission of the author.

Insight into the Pathophysiology of Cardiometabolic Diseases using Multiple Omics Approaches

Proefschrift

ter verkrijging van

de graad van doctor aan de Universiteit Leiden,

op gezag van rector magnificus prof.dr.ir. H. Bijl,

volgens besluit van het college voor promoties

te verdedigen op Woensdag 27 November 2024

klokke 13:00 uur

door

Wenyi Wang

geboren te Guizhou, China

in 1995

Promotor

Prof. dr. ir. J.A.P. Willems van Dijk

Co-promotor

Dr. ir. D. van Heemst

Dr. R. Noordam

Leden promotiecommissie

Prof. dr. A.M.J.M. van den Maagdenberg

Prof. dr. S.C. Cannegieter

Prof. dr. P. Parini (Karolinska Institutet, Stockholm, Sweden)

Prof. dr. P.M. Munroe (Queen Mary University of London, London, UK)

To my parents
献给我的父母

Table of Contents

Chapter 1 **9**

General Introduction *9*

Part I Insulin Sensitivity **25**

Chapter 2 **27**

Differential insulin sensitivity of NMR-based metabolomics measures in a two-step hyperinsulinemic euglycemic clamp study *27*

Part II Pathophysiology of Type 2 Diabetes Mellitus **51**

Chapter 3 **53**

Assessment of the bi-directional relationship between blood mitochondrial DNA copy number and type 2 diabetes mellitus: a multivariable-adjusted regression and Mendelian Randomization study *53*

Chapter 4 **77**

Clustered Mendelian randomization analyses identify distinct and opposing pathways in the association between genetically influenced insulin-like growth factor-1 and type 2 diabetes mellitus *77*

Part III Pathophysiology of Sleep-Associated Dyslipidemia **103**

Chapter 5 **105**

A Large-Scale Genome-Wide Gene-Sleep Interaction Study in 732,564 Participants Identifies Lipid Loci Explaining Sleep-Associated Lipid Disturbances *105*

Part IV Integration of Pharmacometrics and Epidemiology **133**

Chapter 6 **135**

A novel approach for pharmacological substantiation of safety signals using plasma concentrations of medication and administrative/healthcare databases: A case study using Danish registries for an FDA warning on lamotrigine *135*

Part V Discussion, Future perspective and Appendices **155**

Chapter 7 **157**

Discussion and Future Perspective *157*

Chapter 8 **171**

Appendices *171*

- Summary 172
- Samenvatting 176
- PhD portfolio 181
- Publications and Manuscript 182
- Curriculum Vitae 183
- Acknowledgements 184

Chapter 1

General Introduction

General Introduction

Cardiometabolic diseases cover a wide spectrum of conditions, including insulin resistance, metabolic syndrome and prediabetes, and clinical conditions such as type 2 diabetes mellitus and cardiovascular diseases [1]. Cardiometabolic diseases are a major threat to society, contributing to morbidity and mortality globally [2, 3]. According to the International Diabetes Federation, diabetes mellitus affected approximately 537 million people worldwide in 2021, and this figure is expected to rise to 784 million by 2045 [4]. Similarly, cardiovascular diseases are a leading cause of death worldwide [5]. In 2019, 523 million people developed cardiovascular diseases, which led to 18.6 million deaths worldwide [6].

To reduce the burden of cardiometabolic diseases, many studies have aimed to identify biomarkers associated with increased risk of developing cardiometabolic disease. Identified biomarkers for increased risk of cardiometabolic diseases include, but are not limited to, dyslipidemia, hyperglycemia, hypertension and obesity [7]. However, the direction of the effect in some of the observed associations and the question whether or not some of associations between biomarkers and diseases are causal are uncertain. In addition, the mechanisms underlying cardiometabolic diseases are complex and are affected by multiple genetic and environmental factors. Recently, the emergence of -omics technologies (e.g., genomics and metabolomics) has provided opportunities to explore the causality of biomarkers of cardiometabolic diseases in observational studies, and to gain insight into the underlying pathways leading to disease onset [8, 9].

The overall aims of this thesis are to examine the directions of effect and to investigate causality of associations between (patho)physiological biomarkers and type 2 diabetes mellitus, and between behavioral risk factors and cardiometabolic traits (i.e., dyslipidemia traits). In addition, a novel framework for further study of treatment-response relationship for (cardiometabolic) diseases is introduced.

Insulin sensitivity

Decreased insulin sensitivity is a hallmark of type 2 diabetes. Insulin is a peptide hormone containing 51 amino acids, which is synthesized and secreted by β cells of pancreatic islets [10]. Insulin secretion is biphasic after

stimulation by glucose, with a rapid peak in the early phase followed by a continuous and gentle increase during the second phase [11]. Insulin maintains blood glucose homeostasis by regulating carbohydrate, lipid and protein metabolism. It promotes glucose uptake mainly by the liver, skeletal muscle and adipose tissues [12]. In the liver, insulin inhibits gluconeogenesis and glycogenolysis, and activates synthesis of glycogen [13]. In skeletal muscle, insulin increases the rate of glucose transport across the cell membrane, glycogen synthesis and glycolysis [14]. In adipose tissue, insulin stimulates glucose transport, synthesis of fatty acids, and inhibits lipolysis [14]. Thus, under normal physiological conditions, insulin lowers glucose levels. However, in case of insulin resistance, when biochemical processes (such as glucose uptake) do not respond to insulin properly, hyperglycemia will occur, leading to more insulin secretion. The increased insulin secretion can only compensate to a certain extent for the poor response to insulin. At some point, with increasing insulin resistance, insulin production will have reached its maximum and insulin producing cells may start failing [15]. This vicious cycle will eventually result in type 2 diabetes mellitus [16]. Therefore, understanding insulin sensitivity is crucial for better understanding the pathogenesis of type 2 diabetes mellitus.

Hyperinsulinemic-euglycemic clamp study

Clinically, the hyperinsulinemic-euglycemic clamp study is the golden standard for direct assessment of insulin sensitivity in humans [17]. The hyperinsulinemic-euglycemic clamp study is usually performed in the morning when individuals are in fasting condition. During the procedure, individuals are intravenously infused with insulin at a constant rate to raise the plasma insulin concentration to a steady level above the normal fasting insulin level. Meanwhile, to maintain the plasma glucose level within the normal range, glucose is infused at a rate dependent on the insulin sensitivity of the individual [18]. When a steady state is achieved, the glucose infusion rates are the estimates of the amount of glucose uptake by all the tissues under the defined plasma insulin concentration, which is a reflection of the tissues’ insulin sensitivity [19].

Metabolomics

"Omics" technologies, collectively known as the comprehensive investigation of an entire spectrum of biological molecules in a given biological system or process, have revolutionized medical and biological research [20]. These include genomics, epigenomics, transcriptomics, metabolomics and proteomics analyses. Metabolomics has become a powerful tool in biological and epidemiological studies, as it is thought to reflect the biological changes resulting from interaction between genetic variation, gene expression and the environment [21]. Metabolomics can simultaneously quantify hundreds of small-molecule metabolites including amino acid, lipids, and fatty acids by using different analytical platforms [22].

The most frequently used detection platforms are nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) coupled with different separation techniques [23]. In this thesis, Nightingale Health, a commercial NMR-based metabolomics platform, was used for metabolite measurements. Metabolomic patterns of insulin resistance and type 2 diabetes mellitus have been reported by numerous studies, and changes in metabolomic profiles have also long been associated with insulin resistance and type 2 diabetes mellitus [24, 25].

Mitochondrial DNA copy number

Mitochondria play a critical role in generating metabolic energy in the form of adenosine triphosphate (ATP) through the process of oxidative phosphorylation [26]. Mitochondrial dysfunction, the inability of mitochondria to provide sufficient ATP to cellular demands, has long been associated with insulin resistance and type 2 diabetes mellitus [27]. Different mechanisms of mitochondrial dysfunction leading to insulin resistance and type 2 diabetes mellitus have been proposed including decreased mitochondrial content, dysfunctional mitochondrial biogenesis, and impaired mitochondrial function resulting in increased cellular stress and reactive oxygen species (ROS) [28, 29].

In the opposite direction, it has been observed that the size or shape of mitochondria was changed in diabetic patients, and that hyperglycemia could induce mitochondrial fragmentation, which has led to the hypothesis that type 2 diabetes mellitus may result in mitochondrial dysfunction [30]. Therefore, the causal direction of the association between mitochondrial dysfunction and type 2 diabetes is uncertain. Due to the difficulty of direct measurement of mitochondrial function, mitochondrial dysfunction has been

approximated by mitochondrial DNA copy number (mtDNA-CN), which is a measure of the amount of mitochondrial genomes per cell [31]. Type 2 diabetes mellitus has been clearly associated with mtDNA-CN in several studies[32-34], but it is unknown whether this association is causal or bi-directional.

Insulin-like growth factor-1 (IGF-1)

IGF-1 is a hormone encoded by the IGF-1 gene, which plays an important role in growth, cell differentiation and metabolism [35, 36]. IGF-1 is mainly synthesized and secreted by hepatocytes in the liver after stimulation by growth hormone [37]. With a molecular structure similar to that of insulin, IGF-1 is able to bind to the insulin-like growth factor-1 receptor (IGF-1R) with high affinity and to insulin receptor with lower affinity to regulate carbohydrate, lipid and protein metabolism [38]. IGF-1 functions similarly to insulin in glucose metabolism, but differently from growth hormone. For example, IGF-1 reduces blood glucose levels, whereas excessive growth hormone causes insulin insensitivity and hyperglycemia [39]. Studies have shown that IGF-1 is linked to the development of type 2 diabetes mellitus, however it is unclear whether higher or lower levels of IGF-1 levels are related to the risk of developing type 2 diabetes [40, 41].

Lifestyle risk factors for dyslipidemia

Lifestyle plays an important role in cardiometabolic risk traits such as dyslipidemia. Dyslipidemia associated with cardiometabolic risk can be characterized by elevated triglycerides, decreased HDL and/or increased LDL in the plasma [42]. It is a common feature in patients with type 2 diabetes mellitus and is a major risk factor of developing (atherogenic) cardiovascular diseases in these individuals [43]. Genetic variants associated with dyslipidemia have been widely studied, and over 750 single nucleotide polymorphisms (SNPs) have been associated with blood lipid levels to date [44]. In addition to genetic factors, lifestyle factors such as sleep duration and physical activity have been associated with lipid profiles in epidemiological studies [45]. Some studies have associated long or short sleep duration with adverse lipid profiles [46-49]. However, the interaction effect between genetic variants and sleep duration on dyslipidemia is poorly understood.

A previous sleep-by-SNP interaction analysis conducted in 126,926 participants identified 59 novel genetic variants associated with lipids levels mapping to genes that regulate adiposity, neuropsychiatric and inflammatory traits [50]. To further disentangle the underlying biological pathways driving sleep-associated lipid disturbances, higher statistical power for performing interaction analyses is needed that can be achieved by further increasing the sample size for genome-lifestyle interaction analyses.

Genomics and Genome-Wide Association Studies (GWAS)

Genomics is the study of DNA sequences across the entire genome of a cell or organism, and is the most mature type of omics in the omics field [51]. Frequently used genotyping technologies include microarray analysis for common genetic variants, whole exome sequencing (WES), and whole-genome sequencing (WGS) which can also identify rare variants [52]. Genome wide data on genetic variation can be associated with traits to identify disease-related or trait-related genetic susceptibility loci across the human genome at the population level [52, 53]. The beforementioned types of genomics analyses provide opportunities to improve insight in underlying biological mechanisms influencing traits or diseases [54, 55].

Mendelian randomization (MR) and causal inference

Genomics and GWAS have also provided tools to investigate the causality of the association between traits and diseases using MR analysis. MR analysis is a statistical method using genetic variants as instrumental variables to provide evidence for possible causal effects in the observational association between a risk factor and an outcome [56]. MR analyses make use of the assumption that genotypes are generally protected from confounding and reverse causation given their fixed nature and Mendel's First and Second Laws of Inheritance [57]. Analogous to a randomized placebo-controlled clinical trial that randomly allocates participants to either an intervention or a placebo group (treatment and control groups), the participants in MR are naturally randomized at conception based on their genotype carrier status of risk factors into carrier and non-carrier groups (**Figure 1**) [58, 59]. A valid MR analysis needs to satisfy three core assumptions: (1) Instrumental variables are associated with the exposure; (2) Instrumental variables are not associated with confounders that are related to the exposure and the outcome;

(3) Instrumental variables exclusively affect the outcome through the exposure [60].

Clustered MR

In original MR analyses, large heterogeneity among the causal estimates of individual genetic instruments were frequently seen and interpreted as possible bias or invalidity of the results. However, genetic variants may also affect risk factors or disease outcomes through distinct pathways with different effect sizes and/or with different effect directions, resulting in heterogeneity of causal estimates in the overall MR analysis.

In order to identify potentially distinct pathways of the effect of risk factors on disease outcomes, *Christopher NF et al* developed a method to cluster variants with similar causal estimates [61]. With this method, genetic variants used as genetic instrumental variables were clustered based on their individual causal effect in relation to the disease outcome. These clusters or subgroups of genetic instruments had relatively small heterogeneity in their individual causal estimates with the underlying idea that this might reflect different biology. The authors describing this method successfully applied it to demonstrate the existence of genetic variant clusters of blood pressure on coronary artery disease risk [61].

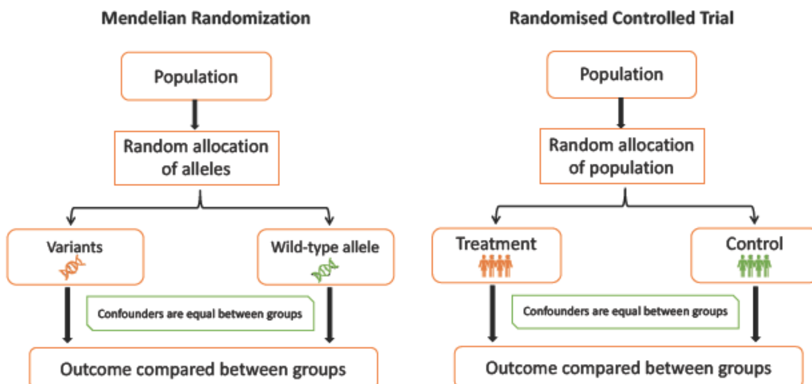


Figure 1 Comparison of the designs of a MR study and a randomized controlled trial

Incorporating pharmacometrics into (pharmaco-)epidemiology

The ultimate goal of applying up-to-date techniques and analytical methods to study disease pathophysiology is to treat patients. After identifying candidate drugs, the response to treatment with this drug is greatly impacted by drug concentrations which are highly individual-specific. Drug concentrations are generally only available in few types of investigations such as randomized clinical trials (RCTs) where blood samples were collected and analyzed. Given the facts that RCTs have ethical limitations, are time consuming and very costly, observational studies are more frequently performed in which a variety of health-related data is collected. However, to the best of our knowledge, most of these observational studies lack information on drug concentrations [62]. In order to make best use of observational data and to understand how the drug concentration can associate with the outcome, a framework, PHARMACOM-EPI, was applied, which can provide guidance for predicting drug concentrations based on dosage, and associate these with the clinical outcome of interest. [63].

This framework integrates pharmacometrics and (pharmaco-)epidemiology disciplines with the aim of providing pharmacological substantiation for pharmaco-epidemiological studies [63]. Pharmacometrics is defined as "the science of developing and applying mathematical and statistical methods to characterize, understand and predict a drug's pharmacokinetics, pharmacodynamics, and biomarker-outcomes behavior" [64]. Pharmacometrics-based approaches are frequently applied in quantifying drug concentrations, disease progression and trial dose design, and are very powerful tools for drug development and regulatory decisions [65]. Distinct pharmacometric models should be employed for different purposes. Pharmacokinetic models can be used to describe dose and concentration relationships, and pharmacodynamic models can be applied to describe the relationship between drug concentration and response [66]. By using pharmacokinetics models, individual and population concentrations of drugs can be predicted [67], which can help make optimal use of observational data to address medical problems. By implementing pharmaco-epidemiological methods, the predicted concentrations using pharmacometrics can be linked to the clinical outcomes.

Main study populations

Leiden Longevity Study

The Leiden Longevity Study [68] recruited 461 families including at least two living long-lived siblings with age higher than 89 years or 91 years for men or women, respectively. The offspring of the long-lived individuals and their partners were also included in this study. The data of the hyperinsulinemic-euglycemic clamp study from 24 participants from the Leiden Longevity Study was used in **Chapter 2**.

UK Biobank

UK Biobank [69] is a prospective cohort study with approximately 500,000 individuals aged between 40 to 69 years at recruitment (2006-2010) from across the United Kingdom. A variety of health-related information including biological measurements, lifestyle indicators, biomarkers in blood and urine, and imaging of the body and brain were collected through interviews, questionnaires, physical measures as well as electronic medical records. Genetic data was also available for 488,377 participants from UK Biobank. The phenotypic and genomic data collected in the UK Biobank was used in **Chapter 3**, **Chapter 4** and **Chapter 5** to allow us to perform solid and well-powered statistical analyses.

Nationwide Danish registry–based cohort study

Danish nationwide administrative and healthcare registers are among the oldest national hospital registries in the world [70]. Based on a unique personal identification number for each Danish citizen, data from different sources and register-based information can be linked. It provides a wide range of health and society related data, from disease identification to medical treatments. All the disease are recorded according to International Classification of Diseases. The Danish nationwide registry-based cohort data is applied in **Chapter 6**.

Thesis Outline

Part I: Insulin Sensitivity

The aim of **Part I** of this thesis is to identify metabolomic measures related to differential effects of insulin. Insulin resistance is a major risk factor for cardiometabolic diseases, especially type 2 diabetes. The changes of metabolomic measures in individuals with insulin resistance have been widely studied [24, 25]. However, the direct effect of insulin on the metabolomic profile in healthy individuals (e.g., irrespective of insulin resistance) has not been described yet. The aim of **Chapter 2** is to disentangle the effects of insulin on metabolomic measures from those induced by glucose in non-diabetic individuals by using data derived from a two-step hyperinsulinemic euglycemic clamp study.

Part II: Pathophysiology of Type 2 Diabetes Mellitus

The aim of **Part II** of this thesis is to study the pathophysiology of type 2 diabetes by identifying (causal) risk factors for type 2 diabetes. Insulin-like growth factor-1 (IGF-1) is a hormone regulating cellular growth, proliferation and survival. It has been associated with various diseases including type 2 diabetes, but the evidence for a possible causal association between IGF-1 and type 2 diabetes is inconsistent, although such relationship is biologically plausible. In **Chapter 3**, we performed prospective analyses to associate IGF-1 and type 2 diabetes using data from the large UK Biobank. In addition, we examined the potential causal association between IGF-1 concentration and development of type 2 diabetes by employing Mendelian Randomization and clustered Mendelian Randomization. In **Chapter 4**, blood mitochondrial DNA copy number (mtDNA-CN), an approximation of mitochondrial function, was associated with type 2 diabetes. The potential causal association between blood mtDNA-CN and type 2 diabetes was also assessed by using bidirectional Mendelian Randomization.

Part III: Pathophysiology of Sleep-Associated Dyslipidemia

The aim of **Part III** of this thesis is to study the role of sleep duration in regulating lipid levels. Short sleep duration is an important environmental factor contributing to higher risk of cardiometabolic diseases [71]. Short sleep

and long sleep have both been associated with abnormal lipids profiles, likely through distinct pathways. In **Chapter 5,** in a large sample of up to 730,000 individuals, we aimed to elucidate the biology of sleep duration associated abnormal lipids levels by using multi-ancestry genome-wide gene-sleep interaction analysis to identify novel and established genomic loci for cardiometabolic traits including high-density lipoprotein cholesterol, low-density lipoprotein cholesterol and triglycerides.

Part IV: Integration of Pharmacometrics and Epidemiology

The aim of **Part IV** of this thesis is to introduce an unique methodological framework (PHARMACO-EPI) which can be used in future studies. PHARMACO-EPI is a method integrating pharmacometrics and pharmacoepidemiology to predict drug concentrations based on drug dosages. In **Chapter 6**, this methodological framework was applied to explore the association between anti-seizure drug concentrations and mortality in older patients with epilepsy by using Danish nationwide administrative and healthcare databases.

Part V: Discussion, Future perspective and Appendices

Lastly, an overview of the main findings and future perspectives is provided in **Chapter 7**. In this chapter, the strengths and limitations of the studies in this thesis are also discussed.

Reference

[1]Guo F, Moellering DR, Garvey WT (2014) The progression of cardiometabolic disease: validation of a new cardiometabolic disease staging system applicable to obesity. Obesity (Silver Spring) 22(1): 110-118. 10.1002/oby.20585
[2]de Waard AM, Hollander M, Korevaar JC, et al. (2019) Selective prevention of cardiometabolic diseases: activities and attitudes of general practitioners across Europe. Eur J Public Health 29(1): 88-93. 10.1093/eurpub/cky112
[3]Vincent GE, Jay SM, Sargent C, Vandelanotte C, Ridgers ND, Ferguson SA (2017) Improving Cardiometabolic Health with Diet, Physical Activity, and Breaking Up Sitting: What about Sleep? Front Physiol 8: 865. 10.3389/fphys.2017.00865
[4]Federation ID (2021) IDF Diabetes Atlas. Available from https://diabetesatlas.org/data/en/world/
[5]Glovaci D, Fan W, Wong ND (2019) Epidemiology of Diabetes Mellitus and Cardiovascular Disease. Curr Cardiol Rep 21(4): 21. 10.1007/s11886-019-1107-y
[6]Roth GA, Mensah GA, Johnson CO, et al. (2020) Global Burden of Cardiovascular Diseases and Risk Factors, 1990-2019: Update From the GBD 2019 Study. J Am Coll Cardiol 76(25): 2982-3021. 10.1016/j.jacc.2020.11.010
[7]Chatterjee A, Harris SB, Leiter LA, Fitchett DH, Teoh H, Bhattacharyya OK (2012) Managing cardiometabolic risk in primary care: summary of the 2011 consensus statement. Can Fam Physician 58(4): 389-393, e196-201
[8]deGoma EM, Knowles JW, Angeli F, Budoff MJ, Rader DJ (2012) The evolution and refinement of traditional risk factors for cardiovascular disease. Cardiol Rev 20(3): 118-129. 10.1097/CRD.0b013e318239b924
[9]Tahir UA, Gerszten RE (2020) Omics and Cardiometabolic Disease Risk Prediction. Annu Rev Med 71: 163-175. 10.1146/annurev-med-042418-010924
[10]Mayer JP, Zhang F, DiMarchi RD (2007) Insulin structure and function. Peptide Science 88(5): 687-713. https://doi.org/10.1002/bip.20734
[11]Pfeifer MA, Halter JB, Porte D (1981) Insulin secretion in diabetes mellitus. The American Journal of Medicine 70(3): 579-588. https://doi.org/10.1016/0002-9343(81)90579-9
[12]Wilcox G (2005) Insulin and insulin resistance. Clin Biochem Rev 26(2): 19-39
[13]Park SY, Gautier JF, Chon S (2021) Assessment of Insulin Secretion and Insulin Resistance in Human. Diabetes Metab J 45(5): 641-654. 10.4093/dmj.2021.0220
[14]Dimitriadis G, Mitrou P, Lambadiari V, Maratou E, Raptis SA (2011) Insulin effects in muscle and adipose tissue. Diabetes Res Clin Pract 93 Suppl 1: S52-59. 10.1016/s0168-8227(11)70014-6
[15]Schinner S, Scherbaum WA, Bornstein SR, Barthel A (2005) Molecular mechanisms of insulin resistance. Diabetic Medicine 22(6): 674-682. https://doi.org/10.1111/j.1464-5491.2005.01566.x
[16]Galicia-Garcia U, Benito-Vicente A, Jebari S, et al. (2020) Pathophysiology of Type 2 Diabetes Mellitus. Int J Mol Sci 21(17). 10.3390/ijms21176275
[17]Gutch M, Kumar S, Razi SM, Gupta KK, Gupta A (2015) Assessment of insulin sensitivity/resistance. Indian J Endocrinol Metab 19(1): 160-164. 10.4103/2230-8210.146874
[18]Robert JJ (1995) [Methods for the measurement of insulin resistance. Hyperinsulinemic euglycemic clamp]. Presse medicale (Paris, France : 1983) 24(15): 730-734
[19]Muniyappa R, Madan R, Varghese RT (2000) Assessing Insulin Sensitivity and Resistance in Humans. In: Feingold KR, Anawalt B, Blackman MR, et al. (eds) Endotext. MDText.com, Inc.
Copyright © 2000-2023, MDText.com, Inc., South Dartmouth (MA)

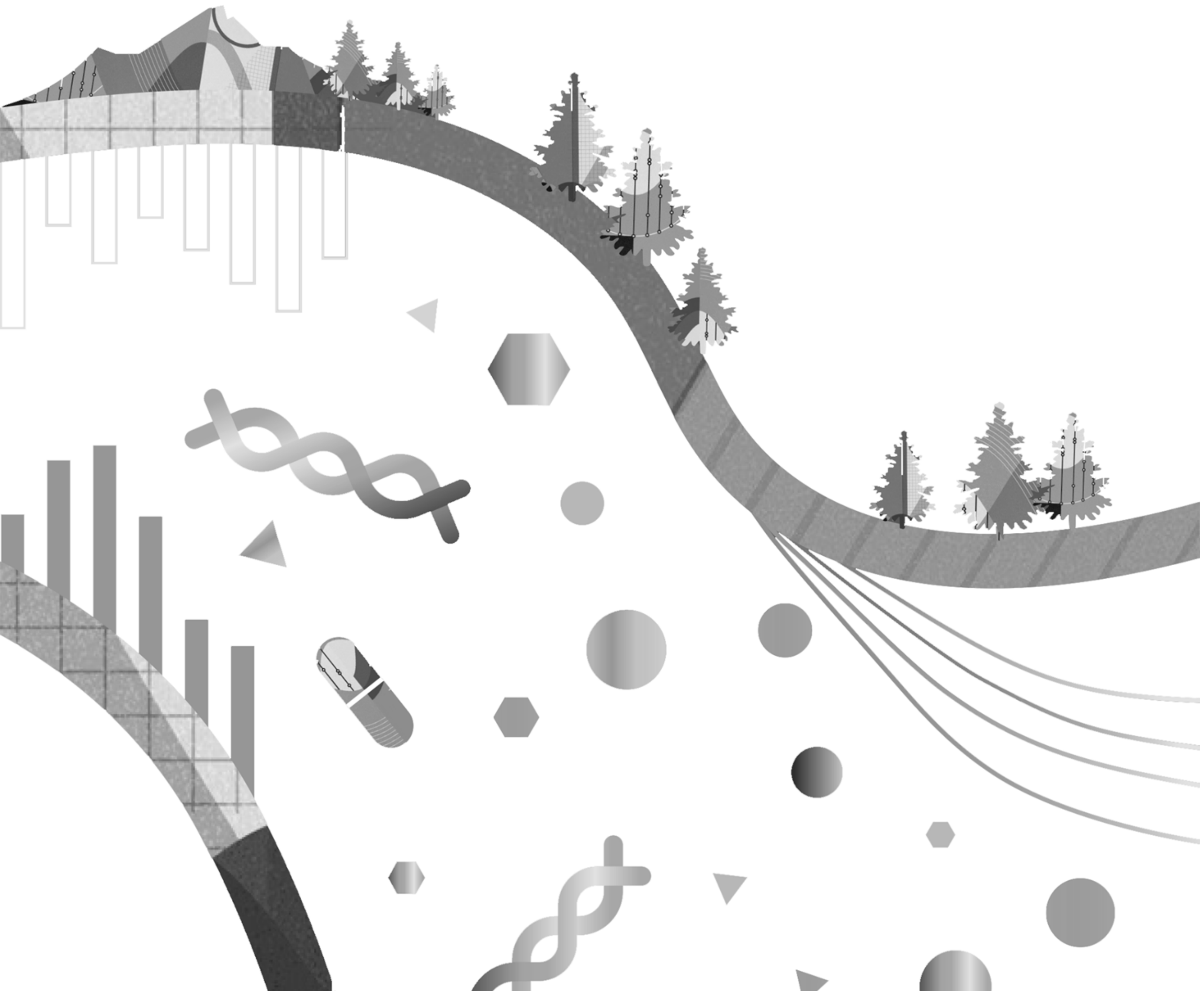
[20]Dai X, Shen L (2022) Advances and Trends in Omics Technology Development. Front Med (Lausanne) 9: 911861. 10.3389/fmed.2022.911861
[21]Rochfort S (2005) Metabolomics reviewed: a new "omics" platform technology for systems biology and implications for natural products research. J Nat Prod 68(12): 1813-1820. 10.1021/np050255w
[22]Dunn WB, Ellis DI (2005) Metabolomics: Current analytical platforms and methodologies. TrAC Trends in Analytical Chemistry 24(4): 285-294. https://doi.org/10.1016/j.trac.2004.11.021
[23]Bujak R, Struck-Lewicka W, Markuszewski MJ, Kaliszan R (2015) Metabolomics for laboratory diagnostics. Journal of Pharmaceutical and Biomedical Analysis 113: 108-120. https://doi.org/10.1016/j.jpba.2014.12.017
[24]Yu D, Moore SC, Matthews CE, et al. (2016) Plasma metabolomic profiles in association with type 2 diabetes risk and prevalence in Chinese adults. Metabolomics 12. 10.1007/s11306-015-0890-8
[25]Palmer ND, Stevens RD, Antinozzi PA, et al. (2015) Metabolomic Profile Associated With Insulin Resistance and Conversion to Diabetes in the Insulin Resistance Atherosclerosis Study. The Journal of Clinical Endocrinology & Metabolism 100(3): E463-E468. 10.1210/jc.2014-2357
[26]McBride HM, Neuspiel M, Wasiak S (2006) Mitochondria: more than just a powerhouse. Curr Biol 16(14): R551-560. 10.1016/j.cub.2006.06.054
[27]Kim JA, Wei Y, Sowers JR (2008) Role of mitochondrial dysfunction in insulin resistance. Circ Res 102(4): 401-414. 10.1161/circresaha.107.165472
[28]Petersen KF, Shulman GI (2006) Etiology of Insulin Resistance. The American Journal of Medicine 119(5, Supplement 1): S10-S16. https://doi.org/10.1016/j.amjmed.2006.01.009
[29]Krako Jakovljevic N, Pavlovic K, Jotic A, et al. (2021) Targeting Mitochondria in Diabetes. Int J Mol Sci 22(12). 10.3390/ijms22126642
[30]Rovira-Llopis S, Bañuls C, Diaz-Morales N, Hernandez-Mijares A, Rocha M, Victor VM (2017) Mitochondrial dynamics in type 2 diabetes: Pathophysiological implications. Redox Biol 11: 637-645. 10.1016/j.redox.2017.01.013
[31]Longchamps RJ, Castellani CA, Yang SY, et al. (2020) Evaluation of mitochondrial DNA copy number estimation techniques. PLoS One 15(1): e0228166. 10.1371/journal.pone.0228166
[32]Memon AA, Sundquist J, Hedelius A, Palmér K, Wang X, Sundquist K (2021) Association of mitochondrial DNA copy number with prevalent and incident type 2 diabetes in women: A population-based follow-up study. Sci Rep 11(1): 4608. 10.1038/s41598-021-84132-w
[33]Reiling E, Ling C, Uitterlinden AG, et al. (2010) The association of mitochondrial content with prevalent and incident type 2 diabetes. J Clin Endocrinol Metab 95(4): 1909-1915. 10.1210/jc.2009-1775
[34]Lee HK, Song JH, Shin CS, et al. (1998) Decreased mitochondrial DNA content in peripheral blood precedes the development of non-insulin-dependent diabetes mellitus. Diabetes Res Clin Pract 42(3): 161-167. 10.1016/s0168-8227(98)00110-7
[35]LeRoith D, Yakar S (2007) Mechanisms of disease: metabolic effects of growth hormone and insulin-like growth factor 1. Nat Clin Pract Endocrinol Metab 3(3): 302-310. 10.1038/ncpendmet0427
[36]Laron Z (2001) Insulin-like growth factor 1 (IGF-1): a growth hormone. Mol Pathol 54(5): 311-316. 10.1136/mp.54.5.311
[37]Blum WF, Alherbish A, Alsagheir A, et al. (2018) The growth hormone-insulin-like growth factor-I axis in the diagnosis and treatment of growth disorders. Endocr Connect 7(6): R212-r222. 10.1530/ec-18-0099

[38]Kineman RD, Del Rio-Moreno M, Sarmento-Cabral A (2018) 40 YEARS of IGF1: Understanding the tissue-specific roles of IGF1/IGF1R in regulating metabolism using the Cre/loxP system. J Mol Endocrinol 61(1): T187-t198. 10.1530/jme-18-0076
[39]Clemmons DR (2004) The relative roles of growth hormone and IGF-1 in controlling insulin sensitivity. J Clin Invest 113(1): 25-27. 10.1172/jci20660
[40]Larsson SC, Michaëlsson K, Burgess S (2020) IGF-1 and cardiometabolic diseases: a Mendelian randomisation study. Diabetologia 63(9): 1775-1782. 10.1007/s00125-020-05190-9
[41]Teppala S, Shankar A (2010) Association between serum IGF-1 and diabetes among U.S. adults. Diabetes Care 33(10): 2257-2259. 10.2337/dc10-0770
[42]Gujral J, Gupta J (2023) Pediatric Dyslipidemia. In: StatPearls. StatPearls Publishing
Copyright © 2023, StatPearls Publishing LLC., Treasure Island (FL)
[43]Mooradian AD (2009) Dyslipidemia in type 2 diabetes mellitus. Nat Clin Pract Endocrinol Metab 5(3): 150-159. 10.1038/ncpendmet1066
[44]Graham SE, Clarke SL, Wu KH, et al. (2021) The power of genetic diversity in genome-wide association studies of lipids. Nature 600(7890): 675-679. 10.1038/s41586-021-04064-3
[45]Enkhmaa B, Surampudi P, Anuurad E, Berglund L (2000) Lifestyle Changes: Effect of Diet, Exercise, Functional Food, and Obesity Treatment on Lipids and Lipoproteins. MDText.com, Inc., South Dartmouth (MA)
[46]Song Q, Liu X, Zhou W, Wu S, Wang X (2020) Night sleep duration and risk of each lipid profile abnormality in a Chinese population: a prospective cohort study. Lipids Health Dis 19(1): 185. 10.1186/s12944-020-01363-y
[47]Abdurahman AA, Bule M, Shab-Bidar S, Rezaei S, Djafarian K (2020) The association between sleep duration and risk of abnormal lipid profile: A systematic review and meta-analysis. Obesity Medicine 18: 100236. https://doi.org/10.1016/j.obmed.2020.100236
[48]Smiley A, King D, Harezlak J, Dinh P, Bidulescu A (2019) The association between sleep duration and lipid profiles: the NHANES 2013-2014. J Diabetes Metab Disord 18(2): 315-322. 10.1007/s40200-019-00415-0
[49]Bos MM, Noordam R, van den Berg R, et al. (2019) Associations of sleep duration and quality with serum and hepatic lipids: The Netherlands Epidemiology of Obesity Study. J Sleep Res 28(4): e12776. 10.1111/jsr.12776
[50]Noordam R, Bos MM, Wang H, et al. (2019) Multi-ancestry sleep-by-SNP interaction analysis in 126,926 individuals reveals lipid loci stratified by sleep duration. Nat Commun 10(1): 5121. 10.1038/s41467-019-12958-0
[51]Hasin Y, Seldin M, Lusis A (2017) Multi-omics approaches to disease. Genome Biology 18(1): 83. 10.1186/s13059-017-1215-1
[52]Uffelmann E, Huang QQ, Munung NS, et al. (2021) Genome-wide association studies. Nature Reviews Methods Primers 1(1): 59. 10.1038/s43586-021-00056-9
[53]Khoury MJ, Millikan R, Little J, Gwinn M (2004) The emergence of epidemiology in the genomics age. International Journal of Epidemiology 33(5): 936-944. 10.1093/ije/dyh278
[54]Rattray NJW, Deziel NC, Wallach JD, et al. (2018) Beyond genomics: understanding exposotypes through metabolomics. Human Genomics 12(1): 4. 10.1186/s40246-018-0134-x
[55]Cano-Gamez E, Trynka G (2020) From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. Front Genet 11: 424. 10.3389/fgene.2020.00424
[56]Emdin CA, Khera AV, Kathiresan S (2017) Mendelian Randomization. JAMA 318(19): 1925-1926. 10.1001/jama.2017.17219

[57]Zheng J, Baird D, Borges M-C, et al. (2017) Recent Developments in Mendelian Randomization Studies. Current Epidemiology Reports 4(4): 330-345. 10.1007/s40471-017-0128-6
[58]Howell AE, Zheng J, Haycock PC, et al. (2018) Use of Mendelian Randomization for Identifying Risk Factors for Brain Tumors. Front Genet 9: 525. 10.3389/fgene.2018.00525
[59]Sanderson E, Glymour MM, Holmes MV, et al. (2022) Mendelian randomization. Nature Reviews Methods Primers 2(1): 6. 10.1038/s43586-021-00092-5
[60]Lyu L, Cai Y, Xiao M, et al. (2022) Causal Relationships of General and Abdominal Adiposity on Osteoarthritis: A Two-Sample Mendelian Randomization Study. J Clin Med 12(1). 10.3390/jcm12010320
[61]Foley CN, Mason AM, Kirk PDW, Burgess S (2020) MR-Clust: Clustering of genetic variants in Mendelian randomization with similar causal estimates. Bioinformatics. 10.1093/bioinformatics/btaa778
[62]Boyko EJ (2013) Observational research--opportunities and limitations. J Diabetes Complications 27(6): 642-648. 10.1016/j.jdiacomp.2013.07.007
[63]Soeorg H, Sverrisdóttir E, Andersen M, Lund TM, Sessa M (2022) The PHARMACOM-EPI Framework for Integrating Pharmacometric Modelling Into Pharmacoepidemiological Research Using Real-World Data: Application to Assess Death Associated With Valproate. Clin Pharmacol Ther 111(4): 840-856. 10.1002/cpt.2502
[64]Ette E.I., Williams P.J. (2007) Pharmacometrics: The Science of Quantitative Pharmacology. John Wiley & Sons, Inc
[65]van der Graaf PH (2012) CPT: Pharmacometrics and Systems Pharmacology. CPT Pharmacometrics Syst Pharmacol 1(9): e8. 10.1038/psp.2012.8
[66]Standing JF (2017) Understanding and applying pharmacometric modelling and simulation in clinical practice and research. Br J Clin Pharmacol 83(2): 247-254. 10.1111/bcp.13119
[67]Hanafin PO, Nation RL, Scheetz MH, et al. (2021) Assessing the predictive performance of population pharmacokinetic models for intravenous polymyxin B in critically ill patients. CPT Pharmacometrics Syst Pharmacol 10(12): 1525-1537. 10.1002/psp4.12720
[68]Willems JM, Trompet S, Eline Slagboom P, de Craen AJ, Westendorp RG (2008) Hematopoietic capacity and exceptional survival: the Leiden Longevity Study. J Am Geriatr Soc 56(11): 2009-2013. 10.1111/j.1532-5415.2008.01933.x
[69]Bycroft C, Freeman C, Petkova D, et al. (2018) The UK Biobank resource with deep phenotyping and genomic data. Nature 562(7726): 203-209. 10.1038/s41586-018-0579-z
[70]Schmidt M, Schmidt SA, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT (2015) The Danish National Patient Registry: a review of content, data quality, and research potential. Clin Epidemiol 7: 449-490. 10.2147/clep.S91125
[71]Knutson KL (2010) Sleep duration and cardiometabolic risk: a review of the epidemiologic evidence. Best Pract Res Clin Endocrinol Metab 24(5): 731-743. 10.1016/j.beem.2010.07.001

Part I

Insulin Sensitivity



Chapter 2

Differential insulin sensitivity of NMR-based metabolomics measures in a two-step hyperinsulinemic euglycemic clamp study

Wenyi Wang, Ko Willems van Dijk, Carolien A Wijsman, Maarten P Rozing, Simon P Mooijaart, Marian Beekman, P Eline Slagboom, J. Wouter Jukema, Raymond Noordam, Diana van Heemst

Metabolomics (2021) 10.1007

The supplemental information for this paper is available online at
https://drive.google.com/drive/folders/1UoSstJZtu2ylarbtkhn6Kmr3rz68hhSu?usp=sharing

Abstract

Background: Insulin is the key regulator of glucose metabolism, but it is difficult to dissect direct insulin from glucose-induced effects. We aimed to investigate the effects of hyperinsulemia on metabolomic measures under euglycemic conditions in nondiabetic participants.

Methods: We assessed concentrations of 151 metabolomic measures throughout a two-step hyperinsulinemic euglycemic clamp procedure. We included 24 participants (50% women, mean age=62 [s.d.=4.2] years) and metabolomic measures were assessed under baseline, low-dose (10 mU/m2/min) and high-dose (40 mU/m2/min) insulin conditions. The effects of low- and high-dose insulin infusion on metabolomic measures were analyzed using linear mixed-effect models for repeated measures.

Results: After low-dose insulin infusion, 90 metabolomic measures changed in concentration (p<1.34e-4), among which glycerol (beta [Confidence Interval] = -1.41 [-1.54, -1.27] s.d., p=1.28e-95) and three-hydroxybutyrate (-1.22 [-1.36, -1.07] s.d., p=1.44e-61) showed largest effect sizes. After high-dose insulin infusion, 121 metabolomic measures changed in concentration, among which branched-chain amino acids showed the largest additional decrease compared with low-dose insulin infusion (e.g., Leucine, -1.78 [-1.88, -1.69] s.d., P=2.7e-295). More specifically, after low- and high-dose insulin infusion, the distribution of the lipoproteins shifted towards more LDL-sized particles with decreased mean diameters.

Conclusion: Metabolomic measures are differentially insulin sensitive and may thus be differentially affected by the development of insulin resistance. Moreover, our data suggests insulin directly affects metabolomic measures previously associated with increased cardiovascular disease risk.

Introduction

Insulin is an anabolic peptide hormone secreted by the pancreas in response to increased blood glucose levels to activate various mechanisms that decrease blood glucose levels [1]. Insulin has broad metabolic effects, which include increasing the rate of glycolysis in fat and muscle, decreasing the rate of lipolysis in fat, decreasing the rate of fatty acid oxidation in muscle and liver, and increasing the rate of protein synthesis in fat, muscle and liver tissue [2, 3]. Insulin resistance is a common pathophysiological consequence of obesity in which body cells are unable to raise a potent physiological response to insulin. Insulin resistance precedes the development of type 2 diabetes and is an independent risk factor of cardiovascular disease [4-6].

Insulin sensitivity is frequently assessed on the basis of the ratio between fasting insulin and glucose levels calculated as the homeostatic model assessment for insulin resistance (HOMA-IR) index [7]. However, an abnormal HOMA-IR index does not provide insight into the tissue-specific origin of the insulin resistance. Insulin not only increases glucose uptake by peripheral tissues such as muscle and fat, but insulin also decreases endogenous glucose production through suppression of gluconeogenesis in the liver and both processes may be affected differentially by insulin resistance [8]. A two-step hyperinsulinemic euglycemic clamp analysis was used to assess whole-body insulin sensitivity and a glucose tracer was included to distinguish hepatic and peripheral insulin resistance [9-11]. During the first step a low dose insulin will predominantly act on the liver, whereas during the second step the higher dose will also have a major effect on peripheral tissues such as muscle and fat tissue [12]. Administration of a low insulin dose has thus been used to assess the insulin sensitivity of endogenous glucose production by the liver, while administration of a higher insulin dose has been used to additionally assess the insulin sensitivity of glucose uptake by peripheral tissues, particularly skeletal muscle and fat [13].

Metabolomic measures are thought to reflect the interaction between proteins encoded by the genome and the environment, such as diet and lifestyle [14]. Numerous platforms have become available which can be exploited to determine the concentrations of a plethora of metabolomic measures in cells and body fluids [15, 16]. Metabolomic measures have been performed to characterize the response to glucose administration in individuals with varying levels of insulin sensitivity [17, 18]. These analyses

have provided insight into the physiological responses and pathophysiological processes underlying disease [19].

Previous studies have shown that multiple blood metabolomic measures are associated with increased insulin resistance and type 2 diabetes [20, 21]. However, the specific effects of hyperinsulinemia, in the absence of major changes in blood glucose levels, on liver and peripheral tissues in determining blood metabolomic measures have not been fully described in healthy individuals. Therefore, the aim of this study was to investigate the responses of metabolomic measures to two different insulin dosages in a two-step hyperinsulinemic euglycemic clamp study in healthy middle-aged individuals without diabetes mellitus.

Methods

Study population and study design

All participants were selected from the Leiden Longevity Study (LLS) [22]. Participants were selected based on the following inclusion criteria: middle-age (50-75 years old), BMI from 22 kg/m^2 to 30 kg/m^2 and living in the proximity of the research center (< 45 minutes by car). Exclusion criteria were: 1) fasting plasma glucose > 6.9 mmol/L [23]; 2) presence of endocrine, renal, hepatic or other significant chronic diseases; 3) use of medication known to influence lipolysis, glucose metabolism or growth hormone secretion; 4) recent weight changes or attempts to lose weight (> 3 kg weight change within last 3 months); 5) smoking; 6) extensive sporting activities (> 10 h per week); 7) inaccessible peripheral veins for intravenous catheter insertion for the assessment by clinical examination and routine laboratory tests. Of the 87 participants that were approached, 17 participants did not fulfill the inclusion criteria (19%), 44 participants refused participation (51%), and 26 participants agreed to participate in the study (30%). Two participants did not finish the study due to medical technical reasons. In total, 24 participants were included in this experiment. Sixteen individuals participated as couples (eight couples) and eight participated as singletons. The Medical Ethical Committee of the Leiden University Medical Center (LUMC) approved the design of the study and all participants gave their written informed consent.

Serum samples were acquired during a two-step hyperinsulinemic euglycemic clamp study (**Figure 1**). All clamp studies started at 8:00 in the morning after an overnight fast. At 08:30 hours (t=0 min), an adjusted primed (17.6 µmol/kg) continuous infusion (0.22 µmol/kg/min) of [6,6-^{2}H2] glucose (enrichment 99.9%; Cambridge Isotopes, Cambridge, MA, USA) was started and lasted for 360 min. At 9:00 hours (t = 30 min), a primed (1.6 µmol/kg), continuous (0.11 µmol/kg/min) infusion of [$^2H_5$]-glycerol (Cambridge Isotopes) was started and continued throughout the study. After two hours of glucose infusion (t = 120 min), low dose human recombinant insulin (10 $mU/m^2/min$, Actrapid, Novo Nordisk Pharma BV, Alphen aan den Rijn, the Netherlands) was infused continuously for 2 hours. After this, high dose insulin (40 $mU/m^2/min$) was infused (t = 240 min) for 2 hours. During the insulin infusion, exogenous glucose 20% enriched with 3% [6,6-$^2H_2$] -glucose was infused at a variable rate to maintain the plasma glucose level at approximately 5.0 mmol/L. Blood samples were taken at the start of the

study, and subsequently every 10 minutes from 90 to 120, from 210 to 240 and from 330 to 360 min. All participants underwent a two-step hyperinsulinemic euglycemic protocol and blood samples were taken for the measurement of 151 metabolomic measures (**Figure 1**). For the three examined conditions, we measured 3 samples as the baseline sample (measured at 95, 105 and 115 minutes after the start), 4 samples as low-dose insulin (measured at 210, 220, 230 and 240 minutes after the start), and another 4 samples as high-dose insulin (measured at 330, 340, 350, 360 minutes after the start). The study population and study design have been described in more detail elsewhere [24].

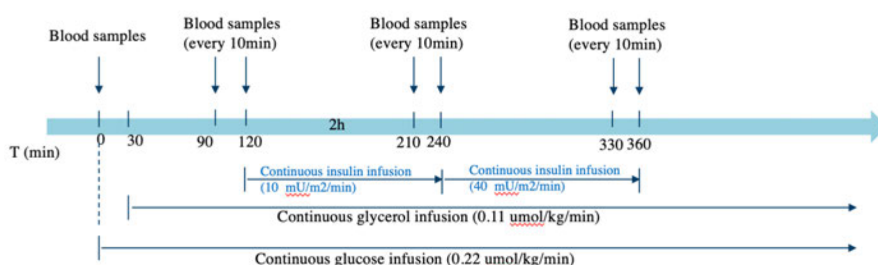


Figure 1 Study design of the two-step hyperinsulinemic euglycemic study

Metabolomics analysis

151 lipid and metabolite concentrations in fasting serum samples (ratios not included) were measured using a high-throughput proton NMR metabolomics platform (Nightingale Health Ltd., Helsinki, Finland) [25]. This method provides quantification of lipoprotein subclass profiling with lipid concentrations within 14 lipoprotein subclasses. The 14 subclass sizes were defined as follows: extremely large VLDL with particle diameters from 75 nm upwards and a possible contribution of chylomicrons, five VLDL subclasses (average particle diameters of 64.0 nm, 53.6 nm, 44.5 nm, 36.8 nm, and 31.3 nm), IDL (28.6 nm), three LDL subclasses (25.5 nm, 23.0 nm, and 18.7 nm), and four HDL subclasses (14.3 nm, 12.1 nm, 10.9 nm, and 8.7 nm). Within the lipoprotein subclasses the following components were quantified: total cholesterol, total lipids, phospholipids, free cholesterol, cholesteryl esters, and triglycerides. The mean size for VLDL, LDL and HDL particles were calculated by weighting the corresponding subclass diameters with their particle concentrations. Furthermore, the majority of the metabolomic

measures that were determined belong to classes of apolipoproteins, cholesterol, fatty acids, glycerides, phospholipids, amino acids, fluid balance, glycolysis-related metabolites, inflammation, and ketone bodies. Detailed experimentation and applications of the NMR metabolomics platform have been described previously [25], as well as representative coefficients of variations (CVs) for the metabolomic measures [26].

Statistical analyses

Characteristics of the study population were presented as percentages (for dichotomous variables) and mean values (with standard deviation [s.d.]).

Missing metabolomic measurement data, which was most frequently due to levels below the limit of detection, were imputed by the half of the minimum of the measured value in the dataset for a specific metabolomic measure. During visual inspection of the data, we observed a decrease in the concentration of albumin during the course of the experiment (**Supplementary Figure 1**), which indicates that blood concentrations were increasingly diluted over time of the study period, as found previously [27]. Therefore, with the exception of the VLDL diameter, LDL diameter, HDL diameter, estimated description of fatty acid chain length, and estimated degree of unsaturation (being all not expressed as mmol/L), correction for dilution of metabolomic measure concentrations was done by normalization to the concentration of albumin. This was done by dividing the concentrations of metabolomic measures by the concentration of albumin (consequently concentrations are expressed per mmol/L/mmol/L albumin). After this correction step, data was log-transformed and subsequently standardized (mean=0, s.d.=1) to approximate a normal distribution and to make all metabolomic measures comparable in unit and in magnitude of effect. Outliers were defined as a value with > 4 s.d. from the mean, and were excluded from the dataset for the analyses prior to any further analyses. Taking into account time-dependent within-person variation in concentrations of the metabolomic measures, a linear mixed-effect model for repeated measures was applied to explore the changes in metabolomic measures' concentrations dependent on different insulin infusion doses within individuals (including the difference between two dose groups compared with the baseline measurement and differences between low dose group and high dose group). In order to further explore the insulin sensitivity of branched chain amino acids (BCAA), we calculated the

percentage changes of the BCAAs after high dose insulin infusion compared with baseline for all individuals. Subsequently, we assessed the correlation of between the percentage change in BCAAs with glucose infusion rate (GIR), which is a measure of whole-body insulin sensitivity, and with glucose disposal rate (GDR), which is a measure of peripheral insulin sensitivity.

The statistical analyses were conducted in the R software (Version 3.6.2), and subsequent data visualization was performed in either Python (2.7) or using the ggplot2 package in R [28].

We corrected the results for multiple testing using Bonferroni. As conventional Bonferroni correction is too stringent given the high correlations between multiple of the included metabolic measures, we corrected for the number of independent metabolic measures instead, using methodology that has been described before by Li et al [29]. Based on this method, we corrected for 37 independent metabolomic measures. Hence, we considered a P-value of 0.00135 (notably 0.05/37) the threshold for statistical significance.

Results

Characteristics of the study population and metabolomic measures

The characteristics of all participants are shown in **Table 1**. In total, 24 participants comprising 12 women and 12 men were included in this study. These participants were clinically healthy with a mean age of 62 (s.d.=4.2) years, mean body mass index of 25.8 kg/m^2 (s.d.=1.8), mean fasting plasma glucose of 5.0 mmol/L (s.d.=0.5) and mean fasting plasma insulin of 6.2 mU/L (s.d.=2.8). Average values of each metabolomic measure within different dose groups were provided in **Supplementary Table 1**.

Table 1 Characteristics of study population

Characteristics	Total
N	24
Men, N (%)	12 (50)
Age in years, mean (s.d.)	62.0 (4.2)
Body mass index in kg/m^2, mean (s.d.)	25.8 (1.8)
Plasma glucose after fasting in mmol/L, mean (s.d.)	5.0 (0.5)
Plasma insulin after fasting in mU/L, mean (s.d.)	6.2 (2.8)

*Plasma glucose and plasma insulin after fasting were calculated based on 17 participants due to the missing values of 7 participants

Changes in metabolomic measures at low-dose insulin infusion

The standardized mean differences in metabolomic measures between baseline and low-dose insulin are summarized in **Figure 2**, and presented in more detail in **Supplementary Table 2**. A total of 90 out of the 151 analyzed metabolomic measures significantly changed in concentration after 10 mM insulin infusion. In particular, after infusing low dose insulin for two hours, the concentrations of glycerol and three-hydroxybutyrate were materially decreased with betas of, respectively, -1.41 [-1.54, -1.27] s.d. ($P=1.28e^{-95}$), and -1.22 [-1.36, -1.07] s.d. ($P=1.44e^{-61}$). Other metabolomic measures that majorly

decreased in concentration during this phase of the experiment included acetate (beta= -0.76 [-0.88, -0.63]; P=1.52e^{-32}), citrate (beta= -0.64 [-0.82, -0.46] s.d.; P=6.95e^{-12}), acetoacetate (beta= -0.57 [-0.71, -0.43] s.d.; P=4.68e^{-15}), LDL diameter (beta= -0.43 [-0.56, -0.30] s.d.; P=2.71e^{-11}) and medium-sized HDL. In contrast, the concentration of pyruvate, and the degree of fatty acid unsaturation increased with betas of 0.59 [0.40, 0.78] s.d. (P=7.46e^{-10}) and 0.48 [0.38, 0.57] s.d (P=1.53e^{-22}) respectively. In addition, the majority of LDL-sized particles and the concentration of apo-lipoprotein B (ApoB) increased in concentration after low-dose insulin infusion.

Figure 2 Circular plot of metabolomic measures after the low dose insulin infusion compared with baseline. Red bars stand for positive betas and blue bars stand for negative betas. The floating dots represents the significance of betas with standard of P-value <0.00135.

Changes in metabolomic measures at high-dose insulin infusion

Figure 3 summarizes the mean changes of metabolomic measures after high dose insulin infusion, and **Supplementary Table 3** presents the results in more detail. 121 out of the 151 metabolomic measures changed significantly in concentration with 40 mM insulin infusion compared with the baseline. The concentrations of glycerol, leucine, isoleucine and valine largely decreased with betas of, respectively, -1.72 [-1.85, -1.59] s.d. ($P=6e^{-142}$), -1.78 [-1.88, -1.69] s.d. ($P=2.7e^{-295}$), -1.65 [-1.77, -1.54] s.d. ($P=3.8e^{-174}$) and -1.53 [-1.63, -1.44] s.d. ($P=6.6e^{-243}$) compared with baseline. In addition, the concentrations of acetate, three-hydroxybutyrate, acetoacetate, tyrosine, glutamine and citrate also decreased. The concentrations of medium, large, extra-large and super extra-large VLDL particles and medium HDL and large HDL decreased after the high-dose insulin infusion. The concentration of pyruvate, lactate, total cholesterol in HDL3, and the degree of fatty acid unsaturation increased significantly with betas of, respectively, 1.29[1.10, 1.48] s.d. ($P=2.15e^{-41}$), 1.29 [1.09, 1.49] s.d. ($P=1.91e^{-37}$), 1.04 [0.91, 1.17] s.d. ($P=1.40e^{-54}$) and 0.74 [0.65, 0.84] s.d. ($P=2.40e^{-51}$). The concentrations of almost all LDL-sized particles, small and extra small VLDL and ApoB also increased significantly after the high dose insulin infusion. In addition, LDL diameter decreased.

Figure 4 showed the percentage changes of isoleucine between high dosage insulin and baseline, which indicated that the magnitude of the changes in isoleucine was correlated with glucose infusion rate. A stronger decrease of isoleucine concentration was found in individuals with higher glucose infusion rates. Similar patterns of change in leucine and valine dependent on the glucose infusion rates were observed (**Supplementary Figures 2 and 3**). Changes in BCAAs at high dose insulin infusion were also positively correlated with glucose disposal rate (GDR) with r=0.68, p=000022, r=0.61, p=0.0017, and r=0.5, p=0.012, for isoleucine, leucine, and valine respectively.

Figure 3 Circular plot of metabolomic measures after high dose insulin infusion compared with baseline. Red bars stand for positive betas and blue bars stand for negative betas. The floating dots represents the significance of betas with standard of P-value < 0.00135

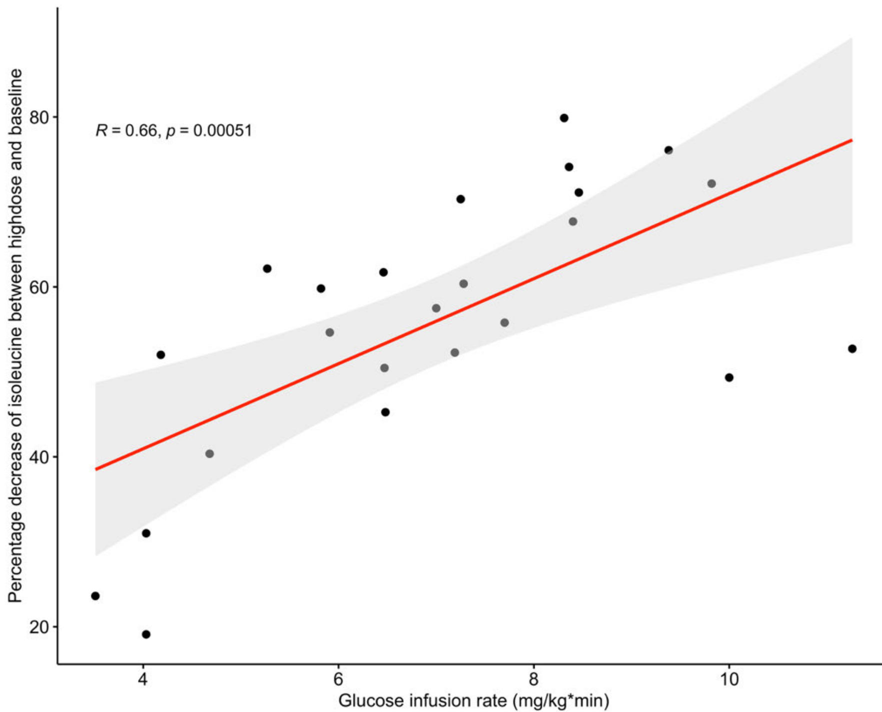


Figure 4 Percentage changes of isoleucine in high dose insulin infusion compared with baseline. Black points represent individuals. Red line is regression line and light grey area represent 95% confidence interval.

Differential changes in metabolomic measures between high and low dose insulin infusion

Figure 5 shows that 99 metabolomic measures changed significantly after high dose insulin infusion compared with low dose insulin infusion and **Supplementary Table 4** provides the results in detail. Apparent additional decreases were specifically seen in the concentrations of branched-chain amino acids, acetate, tyrosine, acetoacetate, glutamine and LDL diameter. Among these significantly changed metabolomic measures, the largest additional changes of concentrations were in branched-chain amino acids, which decreased with betas of -1.39 [-1.47, -1.30] s.d. ($P=2.4e^{-244}$) for leucine, -1.24 [-1.35, -1.14] s.d. ($P=1.6e^{-123}$) for isoleucine and -1.20 [-1.27, -1.12] s.d. ($P=5.8e^{-211}$) for valine compared with low-dose insulin infusion. Furthermore, the concentrations of acetate, tyrosine, acetoacetate, glutamine, extra-large

and super extra-large VLDL particles also sharply decreased. In contrast, the concentrations of lactate (beta= 1.16 [0.97, 1.35] s.d.; P=2.38e^{-33}), pyruvate (beta= 0.98 [0.80, 1.16] s.d.; P=1.95e^{-27}) and HDL3C (beta= 0.67[0.54, 0.79] s.d.; P=3.72e^{-25}) greatly increased. In addition, LDL diameter further decreased.

Based on the results from the low- and high-dose analyses, a beta-beta plot comparing low-dose insulin infusion and high-dose insulin infusion was generated (**Figure 6**). Metabolomic measures on the diagonal line (Y = X) have reached their maximal response already at low-dose insulin infusion, whereas those that deviate from this line show a dose-dependent response. Most notably, leucine, isoleucine, valine and lactate, but also LDL diameter showed a clear additional effect at high-dose insulin infusion beyond that of low-dose insulin infusion.

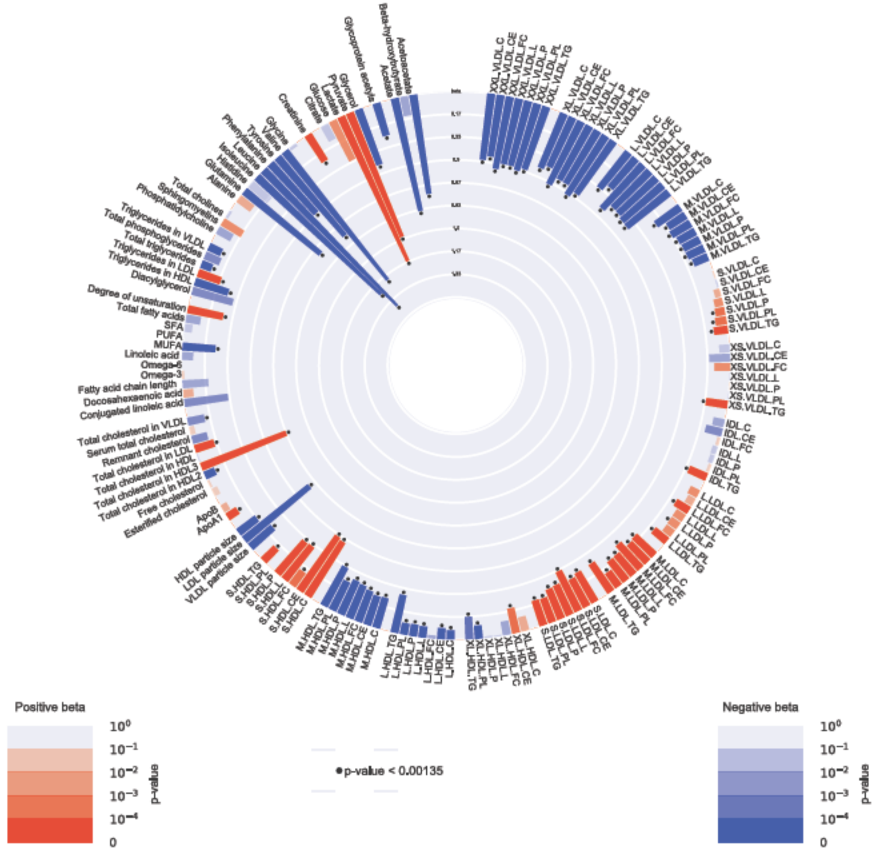


Figure 5 Circular plot of changes in metabolomic measures after high dose insulin infusion compared with low dose insulin infusion. Red bars stand for positive betas and blue bars stand for negative betas. The floating dots represents the significance of betas with standard of P-value < 0.00135.

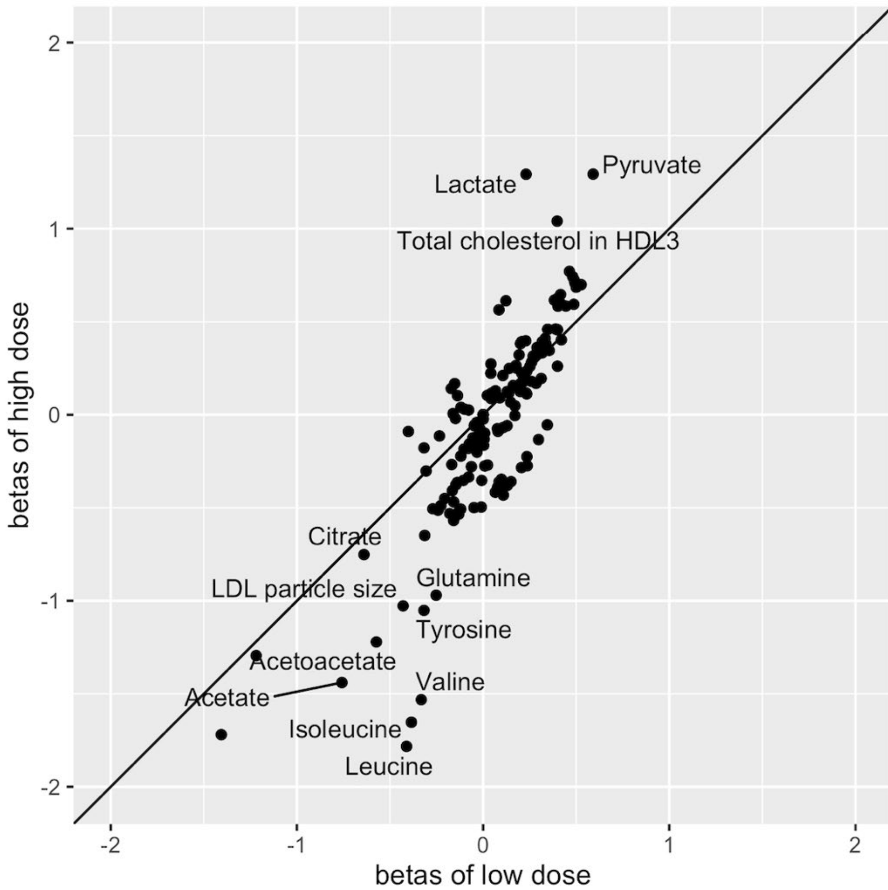


Figure 6 Beta-beta plot of high dose insulin infusion versus low dose insulin infusion. Metabolomic measures on the diagonal line have reached their maximal response already at low dose insulin infusion, whereas those that deviate from this line show a dose-dependent response.

Conclusions and discussion

This study explored the changes of blood metabolomic measures in 24 clinically healthy individuals during a hyperinsulinemic euglycemic clamp study. We found that a large number of metabolomic measures changed significantly in concentration in response to low- and high-dose insulin infusion under euglycemic conditions. A total of 90 out of the 151 analyzed metabolomic measures changed significantly at low-dose insulin infusion, while 121 metabolomic measures changed significantly at high-dose insulin infusion. Some metabolomic measures seemed to have reached their maximum response already at low-dose (10 mU/m^2/min) insulin infusion, whereas other metabolomic measures showed an additional response at high-dose (40 mU/m^2/min)) insulin infusion. This shows that ^{1}H-NMR-based metabolomic measures are differentially insulin sensitive.

2

At low-dose insulin infusion, the largest changes in metabolomic measures comprised glycerol, pyruvate, citrate, three-hydroxybutyrate, acetate, acetoacetate and LDL diameter. These changes are thought to mainly occur via the liver. Low-dose hyperinsulinemia-euglycemia stimulates glycolysis in the liver. This results in the increased production and turnover of pyruvate, which could leak into the circulation [30]. Glycolysis also increases the demand on the mitochondrial citric acid cycle, which requires citrate and this could explain the decrease in citrate. A potential alternative explanation for the observation of the decrease in glycerol after low-dose (10 mU/m^2/min) insulin infusion is an effect of insulin on inhibiting hormone sensitive lipase which would cause a decrease in adipose tissue lipolysis. It had been shown previously that suppression of lipolysis in adipose tissue is very insulin sensitive, and differences in adipose tissue lipolysis between individuals with type 1 diabetes and healthy controls could be detected at insulin doses as low as 4 mU/m^2/min [31]. It is a limitation of the current study that, although (tracer) data on the rate of glycerol appearance were available, the insulin dosages applied might not be low enough to accurately assess potential differences in suppression of lipolysis in adipose tissue (and relate these to some of the observed changes in metabolites). After an overnight fast, ketogenesis is activated in the liver to meet the energy demand of the body (in particular the brain) and ketone bodies such as three-hydroxybutyrate, acetate and acetoacetate are formed in this process [32]. After the infusion of glucose and insulin during the clamp procedure, which is performed in the fasted state, the secretion of ketone bodies is acutely inhibited which explains their decreased concentrations [33].

At high-dose intravenous insulin infusion, the largest changes in metabolomic measures included not only glycerol, pyruvate, lactate, citrate, three-hydroxybutyrate, acetate, acetoacetate, but also leucine, isoleucine, valine, tyrosine, and glutamine. In addition, all sizes of VLDL particles decreased, all sizes of LDL particles increased and mean LDL diameter decreased. High-dose insulin infusion is thought to affect processes in peripheral tissues such as muscle and fat in addition to processes in the liver. In muscle, insulin promotes the synthesis of proteins and suppresses proteolysis [34], which could explain the observed large decrease in concentrations of amino acids including leucine, isoleucine, valine, tyrosine and glutamine. The observed correlation between percentage changes of BCAAs and glucose infusion rate provides additional evidence that the observed decrease of BCAAs at high dose-insulin infusion are positively related to whole-body insulin sensitivity. In addition to these effects, the decrease of medium, large, extra-large and super extra-large VLDL particles, and the increase of the majority of LDL particles, small VLDL, extra small VLDL and cholesterol concentration are indicative of an increased clearance of triglycerides. Triacylglycerol in the large VLDL particles is hydrolyzed by lipoprotein lipase resulting in the formation of smaller VLDL and LDL [35].

The increased concentration of ApoB at low-dose insulin infusion could be explained by stimulation of de-novo lipogenesis in the liver and an increased VLDL production [36]. Interestingly, high-dose insulin infusion had a more pronounced effect on the lipoprotein profile without additionally affecting the concentration of ApoB. This could be explained by direct or indirect high-dose insulin induced changes in the activity of plasma proteins that affect the interchange of components between lipoproteins in the plasma, such as phospholipid transfer protein (PLTP) and cholesteryl ester transfer protein (CETP) [37, 38].

Multiple of the metabolomic measures that we identify as being insulin dependent in the present study have been described before in relation to cardiovascular disease risk. For example, triglycerides within all lipoproteins have been associated with increased risk of incident myocardial infraction and ischemic stroke [39]. In addition, high cholesterol levels in HDL particles have been associated with a lower risk of developing coronary heart disease, myocardial infraction and ischemic stroke [40]. A main driver of atherosclerotic cardiovascular disease is LDL cholesterol which is increased by low-dose insulin and further increased by high-dose insulin infusion [41]. Additionally, the increase of LDL particle number and the decrease of LDL

particle size have also been associated with increased risk of cardiovascular disease [42]. Both parameters are found to be insulin dose dependent in our study. Moreover, the increase of beta-hydroxybutyrate in circulating blood has been associated with an increased intracranial carotid artery atherosclerosis [43]. Branched-chain amino acid were also identified to be associated with incident cardiovascular disease [44]. The insulin dose dependently increased levels of these cardiovascular-disease risk associated metabolomic measures suggest that increased insulin has atherogenic properties independent of glucose concentrations.

Some metabolomic measures identified to be insulin sensitive in our study have also been identified as potential biomarkers for the risk to develop type 2 diabetes. Specifically, higher levels of the BCAAs (leucine, isoleucine and valine) and the aromatic amino acids (tyrosine and phenylalanine) have been associated with increased risk of type 2 diabetes and have the potential to predict the future development of diabetes [45]. These amino acids were also among the metabolites that showed the largest changes in response to the high dose insulin infusion in our study. These data could be interpreted as indicating that decreased insulin sensitivity of amino acids leucine, isoleucine, valine, tyrosine and phenylalanine are predictive for the increased risk of developing type 2 diabetes. However, it has also been demonstrated by Mendelian randomization analysis that higher levels of the branched chain amino acids themselves are causally associated with the risk of type 2 diabetes [46]. Whether increased levels of branched chain amino acids are both consequence and cause of insulin resistance/type 2 diabetes remains to be established.

Our study has provided insight into the direct effects of insulin on changes of metabolomic measures in apparently healthy people under euglycemic conditions. A limitation of this study is the limited sample size, which does not allow specific subgroup analyses. In addition, the age of participants ranged from 50 to 75 years old, which means the results might not apply to younger ages. Moreover, it is important to note that the present study population was selected based on their health and partly on their propensity to become long-lived. This might have introduced bias in our study.

In conclusion, the majority of the plasma metabolomic measures determined by an 1H-NMR metabolomics platform are sensitive to insulin and a large fraction of these responses are insulin dose-dependent. It thus seems likely that some of these metabolomics measures will be differentially affected by

the development of insulin resistance. Since low- and high-dose insulin levels are assumed to target, respectively, the liver and the liver plus peripheral organs (i.e. muscle and fat), our data provide insight into the direct role of insulin on specific processes in the liver and the peripheral tissues. Moreover, our data showed insulin-specific effects on metabolomic measures such as LDL particle number and size, which have previously associated with an increased risk of cardiovascular disease. The implications of this study are to not only avoid the chronic hyperinsulinemia that is associated with insulin resistance, but also to avoid frequent hyperinsulinemia that is caused by frequent snacking as a means to reduce exposure to an atherogenic lipoprotein profile.

Reference

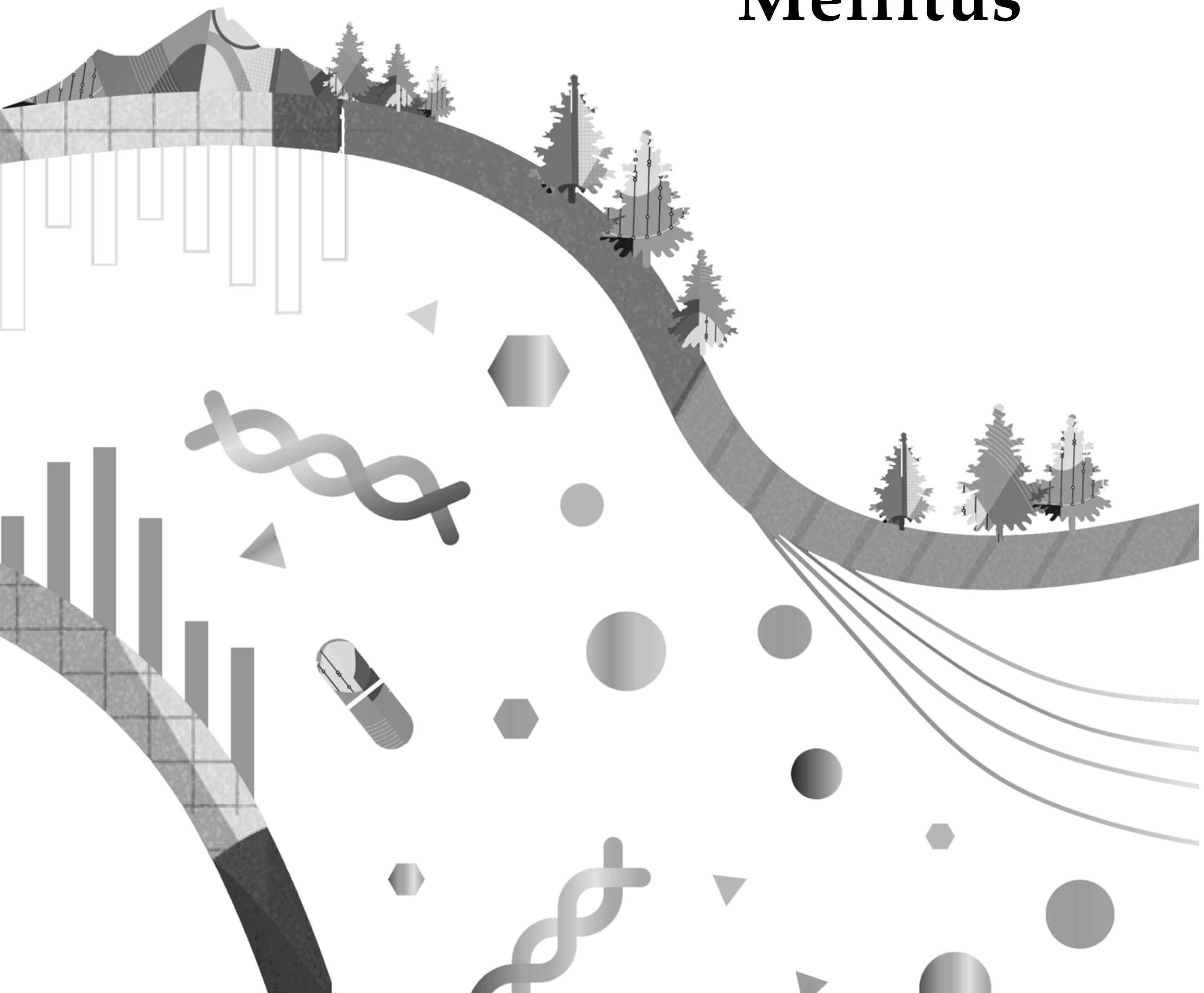
[1]Tokarz VL, MacDonald PE, Klip A (2018) The cell biology of systemic insulin function. J Cell Biol 217(7): 2273-2289. 10.1083/jcb.201802095
[2]Dimitriadis G, Mitrou P, Lambadiari V, Maratou E, Raptis SA (2011) Insulin effects in muscle and adipose tissue. Diabetes Res Clin Pract 93 Suppl 1: S52-59. 10.1016/s0168-8227(11)70014-6
[3]Phillips SM (2008) Insulin and muscle protein turnover in humans: stimulatory, permissive, inhibitory, or all of the above? Am J Physiol Endocrinol Metab 295(4): E731. 10.1152/ajpendo.90569.2008
[4]Ormazabal V, Nair S, Elfeky O, Aguayo C, Salomon C, Zuñiga FA (2018) Association between insulin resistance and the development of cardiovascular disease. Cardiovascular Diabetology 17(1): 122. 10.1186/s12933-018-0762-4
[5]Roberts CK, Hevener AL, Barnard RJ (2013) Metabolic syndrome and insulin resistance: underlying causes and modification by exercise training. Compr Physiol 3(1): 1-58. 10.1002/cphy.c110062
[6]Taylor R (2012) Insulin Resistance and Type 2 Diabetes. Diabetes 61(4): 778. 10.2337/db12-0073
[7]Gutch M, Kumar S, Razi SM, Gupta KK, Gupta A (2015) Assessment of insulin sensitivity/resistance. Indian J Endocrinol Metab 19(1): 160-164. 10.4103/2230-8210.146874
[8]Wallace TM, Levy JC, Matthews DR (2004) Use and Abuse of HOMA Modeling. Diabetes Care 27(6): 1487-1495. 10.2337/diacare.27.6.1487
[9]Muniyappa R, Lee S, Chen H, Quon MJ (2008) Current approaches for assessing insulin sensitivity and resistance in vivo: advantages, limitations, and appropriate usage. Am J Physiol-Endocrinol Metab 294(1): E15-E26. 10.1152/ajpendo.00645.2007
[10]Steele R (1959) Influences of glucose loading and of injected insulin on hepatic glucose output. Ann N Y Acad Sci 82: 420-430. 10.1111/j.1749-6632.1959.tb44923.x
[11]Finegood DT, Bergman RN, Vranic M (1987) Estimation of endogenous glucose production during hyperinsulinemic-euglycemic glucose clamps. Comparison of unlabeled and labeled exogenous glucose infusates. Diabetes 36(8): 914-924. 10.2337/diab.36.8.914
[12]Saccà L, Cicala M, Trimarco B, Ungaro B, Vigorito C (1982) Differential effects of insulin on splanchnic and peripheral glucose disposal after an intravenous glucose load in man. J Clin Invest 70(1): 117-126. 10.1172/jci110583
[13]Bazotte RB, Silva LG, Schiavon FP (2014) Insulin resistance in the liver: deficiency or excess of insulin? Cell Cycle 13(16): 2494-2500. 10.4161/15384101.2014.947750
[14]Beger RD, Dunn W, Schmidt MA, et al. (2016) Metabolomics enables precision medicine: "A White Paper, Community Perspective". Metabolomics 12(10): 149-149. 10.1007/s11306-016-1094-6
[15]Bukowiecka-Matusiak M, Chmielewska-Kassassir M, Szczesna D, Wozniak LA (2016) Metabolomic Insight into Lipid and Protein Profile in Diabetes Using Mass Spectrometry. Mini Rev Med Chem 16(14): 1167-1174. 10.2174/1389557516666160722133534
[16]Liu X, Locasale JW (2017) Metabolomics: A Primer. Trends Biochem Sci 42(4): 274-284. 10.1016/j.tibs.2017.01.004
[17]Shaham O, Wei R, Wang TJ, et al. (2008) Metabolic profiling of the human response to a glucose challenge reveals distinct axes of insulin sensitivity. Mol Syst Biol 4: 214-214. 10.1038/msb.2008.50
[18]Wang Q, Jokelainen J, Auvinen J, et al. (2019) Insulin resistance and systemic metabolic changes in oral glucose tolerance test in 5340 individuals: an interventional study. BMC Med 17(1): 217. 10.1186/s12916-019-1440-4

[19]Wishart DS (2019) Metabolomics for Investigating Physiological and Pathophysiological Processes. Physiological Reviews 99(4): 1819-1875. 10.1152/physrev.00035.2018
[20]Knebel B, Strassburger K, Szendroedi J, et al. (2016) Specific Metabolic Profiles and Their Relationship to Insulin Resistance in Recent-Onset Type 1 and Type 2 Diabetes. The Journal of Clinical Endocrinology & Metabolism 101(5): 2130-2140. 10.1210/jc.2015-4133
[21]Yang Q, Vijayakumar A, Kahn BB (2018) Metabolites as regulators of insulin sensitivity and metabolism. Nat Rev Mol Cell Biol 19(10): 654-672. 10.1038/s41580-018-0044-8
[22]Schoenmaker M, de Craen AJ, de Meijer PH, et al. (2006) Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. Eur J Hum Genet 14(1): 79-84. 10.1038/sj.ejhg.5201508
[23]American Diabetes A (2010) Diagnosis and classification of diabetes mellitus. Diabetes care 33 Suppl 1(Suppl 1): S62-S69. 10.2337/dc10-S062
[24]Wijsman CA, Rozing MP, Streefland TC, et al. (2011) Familial longevity is marked by enhanced insulin sensitivity. Aging Cell 10(1): 114-121. 10.1111/j.1474-9726.2010.00650.x
[25]Soininen P, Kangas AJ, Würtz P, Suna T, Ala-Korpela M (2015) Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Cardiovascular Epidemiology and Genetics. Circulation: Cardiovascular Genetics 8(1): 192-206. 10.1161/circgenetics.114.000216
[26]Kettunen J, Demirkan A, Wurtz P, et al. (2016) Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. Nat Commun 7: 11122. 10.1038/ncomms11122
[27]Li J, Ji L (2005) Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity (Edinb) 95(3): 221-227. 10.1038/sj.hdy.6800717
[28]R Development Core Team (2019). In: computing RAlaefs (ed). R Foundation for Statistical Computing, Vienna, Austria
[29]Li J, Ji L (2005) Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity 95(3): 221-227. 10.1038/sj.hdy.6800717
[30]Guo X, Li H, Xu H, et al. (2012) Glycolysis in the control of blood glucose homeostasis. Acta Pharmaceutica Sinica B 2(4): 358-367. https://doi.org/10.1016/j.apsb.2012.06.002
[31]Schauer IE, Snell-Bergeon JK, Bergman BC, et al. (2011) Insulin resistance, defective insulin-mediated fatty acid suppression, and coronary artery calcification in subjects with and without type 1 diabetes: The CACTI study. Diabetes 60(1): 306-314. 10.2337/db10-0328
[32]Barnett CR, Barnett YA (2003) KETONE BODIES. In: Caballero B (ed) Encyclopedia of Food Sciences and Nutrition (Second Edition). Academic Press, Oxford, pp 3421-3425
[33]Ciaraldi TPaH, R.R. (2004) Insulin Regulation of Ketone Body Metabolism. In: International Textbook of Diabetes Mellitus
[34]Lukens FDW (1964) Insulin and Protein Metabolism. Diabetes 13(5): 451. 10.2337/diab.13.5.451
[35]Bhagavan NV, Ha C-E (2015) Chapter 18 - Lipids III: Plasma Lipoproteins. In: Bhagavan NV, Ha C-E (eds) Essentials of Medical Biochemistry (Second Edition). Academic Press, San Diego, pp 321-337
[36]Julius U (2003) Influence of plasma free fatty acids on lipoprotein synthesis and diabetic dyslipidemia. Exp Clin Endocrinol Diabetes 111(5): 246-250. 10.1055/s-2003-41284
[37]Feingold KR, Grunfeld C (2000) Introduction to Lipids and Lipoproteins. MDText.com, Inc., South Dartmouth (MA)

[38]Van Tol A, Ligtenberg JJ, Riemens SC, van Haeften TW, Reitsma WD, Dullaart RP (1997) Lowering of plasma phospholipid transfer protein activity by acute hyperglycaemia-induced hyperinsulinaemia in healthy men. Scand J Clin Lab Invest 57(2): 147-157. 10.1080/00365519709056383
[39]Holmes MV, Millwood IY, Kartsonaki C, et al. (2018) Lipids, Lipoproteins, and Metabolites and Risk of Myocardial Infarction and Stroke. Journal of the American College of Cardiology 71(6): 620. 10.1016/j.jacc.2017.12.006
[40]Joshi R, Wannamethee SG, Engmann J, et al. (2020) Triglyceride-containing lipoprotein sub-fractions and risk of coronary heart disease and stroke: A prospective analysis in 11,560 adults. Eur J Prev Cardiol 27(15): 1617-1626. 10.1177/2047487319899621
[41]Ference BA, Ginsberg HN, Graham I, et al. (2017) Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel. European Heart Journal 38(32): 2459-2472. 10.1093/eurheartj/ehx144
[42]Campos H, Genest JJ, Blijlevens E, et al. (1992) Low density lipoprotein particle size and coronary artery disease. Arteriosclerosis and Thrombosis: A Journal of Vascular Biology 12(2): 187-195. 10.1161/01.ATV.12.2.187
[43]Vojinovic D, van der Lee SJ, van Duijn CM, et al. (2018) Metabolic profiling of intra- and extracranial carotid artery atherosclerosis. Atherosclerosis 272: 60-65. 10.1016/j.atherosclerosis.2018.03.015
[44]Tobias Deirdre K, Lawler Patrick R, Harada Paulo H, et al. (2018) Circulating Branched-Chain Amino Acids and Incident Cardiovascular Disease in a Prospective Cohort of US Women. Circulation: Genomic and Precision Medicine 11(4): e002157. 10.1161/CIRCGEN.118.002157
[45]Wang TJ, Larson MG, Vasan RS, et al. (2011) Metabolite profiles and the risk of developing diabetes. Nature Medicine 17(4): 448-453. 10.1038/nm.2307
[46]Lotta LA, Scott RA, Sharp SJ, et al. (2016) Genetic Predisposition to an Impaired Metabolism of the Branched-Chain Amino Acids and Risk of Type 2 Diabetes: A Mendelian Randomisation Analysis. PLOS Medicine 13(11): e1002179. 10.1371/journal.pmed.1002179

Part II

Pathophysiology of Type 2 Diabetes Mellitus



Chapter 3

Assessment of the bi-directional relationship between blood mitochondrial DNA copy number and type 2 diabetes mellitus: a multivariable-adjusted regression and Mendelian Randomization study

Wenyi Wang, Jiao Luo, Ko Willems van Dijk, Sara Hägg, Felix Grassmann, Leen M `t Hart, Diana van Heemst, Raymond Noordam

Diabetologia (2022) 10.1007

The supplemental information for this paper is available online at https://drive.google.com/drive/folders/1UoSstJZtu2ylarbtkhn6Kmr3rz68hhSu?usp=sharing

Abstract

Aims/hypothesis: Mitochondrial dysfunction, which can be approximated by blood mtDNA copy number (mtDNA-CN), has been implicated in the pathogenesis of type 2 diabetes mellitus. Thus far, however, insights from prospective cohort studies and Mendelian Randomization (MR) analyses on this relationship are limited. We assessed the association between blood mtDNA-CN and incident type 2 diabetes using multivariable-adjusted regression analyses, and the associations between blood mtDNA-CN and type 2 diabetes and body mass index (BMI) using bi-directional MR.

Methods: Multivariable-adjusted Cox proportional hazard models were used to estimate the association between blood mtDNA-CN and incident type 2 diabetes in 285,967 unrelated European individuals from UK Biobank (UKB) free of type 2 diabetes at baseline. Additionally, a cross-sectional analysis was performed to investigate the association between blood mtDNA-CN and BMI. Furthermore, we assessed the potentially causal relationship between blood mtDNA-CN and type 2 diabetes (N=898,130 from DIAGRAM, N=215,654 from FinnGen) and BMI (N=681,275 from GIANT) using bi-directional two-sample MR.

Results: During a median follow-up of 11.87 years, 15,111 participants developed type 2 diabetes. Participants with higher level of blood mtDNA-CN have lower risk to develop type 2 diabetes (hazard ratio: 0.90; 95% Confidence Interval (CI): 0.89-0.92). After additional adjustment for BMI and other confounders, these results attenuated moderately and remained present. The multivariable-adjusted cross-sectional analyses show higher blood mtDNA-CN was associated with lower BMI (-0.12 (95% CI: -0.14 ~ -0.10) kg/m2. In the bi-directional MR analyses, we found no evidence for causal associations between blood mtDNA-CN and type 2 diabetes, and blood mtDNA-CN and BMI in either direction.

Conclusions/interpretation: The results from the present study indicate that the observed association between low blood mtDNA-CN and higher risk of type 2 diabetes is likely not causal.

Introduction

Type 2 diabetes mellitus, which is characterized by a combination of insulin resistance and insufficient insulin secretion, is a major contributor to the burden of morbidity and mortality worldwide [1]. Multiple risk factors have been associated with an increased risk of developing type 2 diabetes, including obesity, adverse lifestyle and genetic factors [2].

Mitochondria generate most of the cell's need for chemical energy in the form of adenosine triphosphate (ATP), and mitochondrial dysfunction has been implicated in various aspects of the development and complications of type 2 diabetes including insulin resistance, obesity and beta-cell dysfunction [3]. Mitochondrial DNA (mtDNA) is a circular and double-stranded DNA molecule comprising 37 genes, of which thirteen genes are involved in the electron transport chains and generation of ATP to provide energy for cells, while the remaining genes encode proteins involved in the assembly of amino acids into functional proteins [4]. Although mitochondria have been implicated in the onset of type 2 diabetes, only a few of the rare mtDNA genetic variants are associated with diabetes mellitus [5].

MtDNA-copy number (mtDNA-CN) is considered as a proxy for mitochondrial function [6]. Low peripheral blood mtDNA-CN has been associated with various age-related diseases, including type 2 diabetes, cardiovascular disease and cancer [7-9]. A lower mtDNA-CN has also been observed in skeletal muscle and adipose tissue of individuals with obesity or type 2 diabetes [10, 11]. In line, a lower mtDNA-CN in beta cells has also been associated with decreased beta cell function [12].

Thus far, epidemiological studies investigating the relationship between mtDNA-CN and type 2 diabetes are limited. To the best of our knowledge, only four cohort studies have investigated the association between blood mtDNA-CN and type 2 diabetes, among which three were conducted using a prospective study design [13-16]. Two out of the four studies showed that low blood mtDNA-CN was associated with an increased risk of developing type 2 diabetes. However, these studies were performed in a relatively small population with a low number of (incident) type 2 diabetes cases, which warrants replication of the findings in larger study samples. In addition, none of these studies have examined the question whether the relationship between blood mtDNA and type 2 diabetes is causal. Therefore, the aim of our study was to investigate the association between blood mtDNA-CN and

incident type 2 diabetes using data from the UK Biobank (UKB). Additionally, the bi-directional relationship between blood mtDNA-CN and type 2 diabetes and body mass index (BMI) was assessed by bi-directional two-sample Mendelian Randomization (MR) analyses [17].

Method

Prospective analyses

Study population

The data used in the present study was derived from the UKB. The UKB is a large-scale prospective cohort study containing in-depth genetic and health information from over 500,000 participants aged 40-70 years at recruitment from the United Kingdom (UK). Participants were recruited between 2006 and 2010 at 22 assessment centers across the UK. Baseline examinations in all participants included physical measurements (e.g., blood pressure), lifestyle and environmental information (e.g., diet, exercise, smoking, alcohol), biological sampling (e.g., blood and urine), medical history, etc. The UKB study was approved by the North-West Multi-center Research Ethics Committee. Access to information to invite participants was approved by the Patient Information Advisory Group from England and Wales, and National Information Governance Board for Health and Social Care. All participants in the UKB study provided written informed consent and detailed research ethics approval can be found on official website (https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us/ethics). The present study was conducted using accepted proposal 22474.

The genotype data released from UKB contains genotype of 488,377 participants from diverse ancestries. A total of 195,203 participants were removed based on the following exclusion criteria: i) samples did not meet the quality control criteria (N=81,552); ii) participants are related (N=38,642); iii) participants are from non-European ancestry (N=65,498); iv) informed consents were withdrawn (N=71); v) standard deviation (SD) of autosomal probes (last step of mtDNA-CN calculation) > 0.37 (N=9,440). More details on exclusion criteria can be found elsewhere [18]. After filtering, a total of 293,174 unrelated European participants were included in the present study.

Computation of mtDNA copy number in blood

Blood mtDNA-CN, the exposure of our interest, was calculated based on the intensities of genotyping probes on the mitochondrial chromosome on the Affymetrix array in the UKB participants. The pipeline for calculating mtDNA-CN from intensities of probes mapping to the mitochondrial

genome has been described in detail previously [18, 19]. Briefly, the relative amount of mtDNA hybridized to the array at each probe was the log2 transformed ratio (L2R) of the observed genotyping probe intensity divided by the reference intensity. Initial values of mtDNA-CN were computed as the median L2R values across all 265 variants passing quality control on the mitochondrial DNA. The single mtDNA-CN estimated for each individual was derived by computation of weighted mtDNA-CN by using initial values and, normalization of mtDNA-CN to a mean of 0 and SD of 1 over 96 genotyping plates. Finally, quality control was performed by eliminating individuals with SD of autosomal probe intensities higher than 0.37. More details on calculations and used analysis pipeline of mtDNA-CN can be found online (https://github.com/GrassmannLab/MT_UKB).

Outcome definition

Prevalent and incident diagnosis of type 2 diabetes was identified in UKB as the date of first appearance of non-insulin-dependent diabetes mellitus (data-field 130708 in the UKB Database). This variable has been composed through a standard algorithm, performed by the UKB data management team, combining the data derived from hospital admissions (through linkage with the medical records from the National Health Service), general practitioners, death records, and self-report. Detailed information of the algorithm and process could be found online: https://biobank.ndph.ox.ac.uk/showcase/ukb/docs/first_occurrences_outcomes.pdf. Based on the data of first appearance of type 2 diabetes and the data of enrollment information, we defined whether a case was prevalent (before enrolment) or incident (after enrolment).

Covariates

Covariates used in the study included genotyping batch, genetic principal component 1, genetic principal component 2, white blood cell counts, platelet counts, demographic measures (age, sex), anthropometric measures (BMI in kg/m^2, waist circumference in centimeter, height in centimeter), self-reported lifestyles (use of cholesterol-lowering medication [Yes/No], smoking status [Never/Past/Current], physical activities [hours of being physically active per week]) and family (maternal + paternal) history of type 2 diabetes [Yes/No]. Information on the covariates was collected at the moment of study recruitment.

Power calculation

The statistical power in the MR analysis on type 2 diabetes was calculated using an online tool for binary outcome (https://shiny.cnsgenomics.com/mRnd/). Using the data of Diagram consortium, we had sufficient power to detect a causal association with an odds ratio of at least 1.07 per SD increase in blood mtDNA-CN using either of summary statistics from UKB and Longchamp et. al paper with significance level of 0.05 (**Electronic supplementary material (ESM) Fig. 1**)

Statistical analyses

Study characteristics for the population at the baseline were expressed as mean with SD stratified by quartiles of blood mtDNA-CN. Missing values of parameters were imputed using the package MICE which uses Bayesian polytomous regression for prediction of categorical values, predictive mean matching models for continuous missing values, and logistic regression model for prediction of binary missing values [20].

Participants without type 2 diabetes at baseline were followed until the occurrence of type 2 diabetes, mortality or loss of follow-up, whichever occurred first. Kaplan-Meier survival curves of blood mtDNA-CN in quartiles was provided. For the primary analysis, the association between blood mtDNA-CN as continuous variable on type 2 diabetes was assessed by using Cox proportional hazard model. As secondary analyses, the analyses were repeated with blood mtDNA-CN stratified into quartiles. The 1st and 4th quartiles represent the lowest 25% and the highest 25% of blood mtDNA-CN, respectively. Three multivariable-adjusted Cox proportional hazard models were used to analyze the association between blood mtDNA-CN and type 2 diabetes: model 1 was adjusted for genotyping batch, the first 2 genetic principal components to correct for possible population stratification, white blood cell counts, platelet counts, age and sex; model 2 was additionally adjusted for BMI (in kg/m^2); model 3 was additionally adjusted for use of cholesterol lowering medication (Yes/No), smoking status (Never/Past/Current), physical activities (hours of being physical active per week), waist circumference (Centimeter) and family history of type 2 diabetes (Yes/No).

The potential interactions between blood mtDNA-CN and sex and age were tested by adding an (multiplicative) interaction term to the fully adjusted model. In addition, we also performed prespecified subgroup analyses stratified by sex and age (<=50 years, 50~60 years, >=60 years).

In addition, multivariable-adjusted cross-sectional analyses were performed to investigate the association between blood mtDNA-CN and BMI by linear regression adjusted for genotyping batch, first 2 genetic principal components, white blood cell counts, platelet counts, age, sex, use of cholesterol lowering medication, smoking status, physical activities and family history of type 2 diabetes. Additionally, similar models were applied to investigate the association between blood mtDNA-CN and whole-body lean mass, whole-body fat mass, liver enzymes including alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase (ALP) and gamma glutamyl transferase (GGT). All parameters except blood mtDNA-CN were log-transformed to approximate a normal distribution. The linear model of blood mtDNA-CN on whole-body lean mass was adjusted for genotyping batch, the first 2 genetic principal components, white blood cell counts, platelet counts, age, sex, height and fat mass. A similar model was used to investigate the association between blood mtDNA-CN and whole-body fat mass by replacing the fat mass with lean mass. The linear models of blood mtDNA-CN on liver enzymes were adjusted for genotyping batch, the first 2 genetic principal components, white blood cell counts, platelet counts, age, sex and glomerular filtration rate. The estimates derived from all linear models were exponentiated to assess the original effect of each parameter per 1 SD increase in blood mtDNA-CN.

Bi-directional Mendelian Randomization analyses

Exposure data source

Genome-wide association study (GWAS) for blood mtDNA-CN in the unrelated European population of the UKB was performed to provide a list of independent lead single nucleotide polymorphisms (SNPs) for the following use as instrumental variables for the exposure as well as to use the same data as outcome set in the bi-directional MR. UKB genotyping and imputation data released in March 2018 was used to perform GWAS (https://www.ukbiobank.ac.uk/media/cffi4mx5/ukb-genotyping-and-imputation-data-release-faq-v3-2-1.pdf). The analyses were done on the autosomal chromosomes adjusting for age, sex, white blood cell counts,

platelet counts and the first 10 genetic principal components. We excluded SNPs with a minor allele frequency <0.01 as well as SNPs with an imputation quality <0.3. SNPs with p-values smaller than 5e-8 were extracted and stored for the MR analyses. Instrumental variables were identified after clumping process (window size = 10,000 kb, $R^2 < 0.001$) which estimates linkage disequilibrium between SNPs by using the European population from the 1000 genomes project. GWAS in the present study were performed using linear mixed models implemented in the program BOLT_LMM (version 2.3.2) [21]. Visualization of the results was performed using the R-based packages ggplot2 [22] and EasyStrata [23]. Variance in blood mtDNA-CN explained by genetic variants was calculated by using equations that have been described previously [24].

In addition, 129 significant SNPs (p<5e-8) associated with blood mtDNA-CN from the paper of *Longchamps. et. al* [25] were used as exposure as validation of our initial analyses only based on UKB data, which were provided in **ESM Table 1**. This GWAS study was performed in 465,809 European individuals from Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) and the UKB, adjusting for age, sex, principal components, DNA collection site, family structure, and cell composition.

Outcome data source

Summary statistics of type 2 diabetes were extracted from two large databases, notably the Diabetes Genetics Replication and Meta-analysis (Diagram) consortium (available at: https://www.diagram-consortium.org/downloads.html) and the FinnGen study (available at: https://finngen.gitbook.io/documentation/). The summary statistics of BMI was retrieved from The Genetic Investigation of Anthropometric Traits (GIANT) consortium, which is available in the IEU GWAS database (available at: https://gwas.mrcieu.ac.uk/datasets/ieu-b-40/). More information of these data sources can be found in **ESM Table 2**.

MR analyses

Two-sample MR analyses were performed using the R-based statistical package TwoSampleMR (available at: http://github.com/MRCIEU/TwoSampleMR) [26]. This package is able to

harmonize the effect sizes of SNPs for exposure and outcome, and connects to database with thousands of complete GWAS summary data.

The study design for MR analyses is shown in **Figure 1**. In the first set of two-sample MR analyses, we considered blood mtDNA-CN (UKB) (GWAS was performed in the present study) as exposure, and type 2 diabetes (DIAGRAM and FinnGen) and BMI (GIANT) as outcome. Analyses were repeated using data from the *Longchamps et al* paper as exposure [25]. The second set of MR analyses handled reverse causality taking type 2 diabetes (DIAGRAM) and BMI (GIANT) as exposure, and blood mtDNA-CN (UKB) from GWAS as outcome.

The primary MR analysis was conducted using the inverse variance weighted (IVW) method. The magnitudes of the causal effect were estimated by weighted regression of the average SNP-outcome effect on the average SNP-exposure effect with the intercept constrained to zero [27]. Given its assumption that all genetic instruments included are valid, this method could yield biased estimates in the presence of unbalanced horizontal pleiotropy, especially in case the analysis contained many genetic instruments. Therefore, we also performed two additional sensitivity analyses by using other MR models, the weighted median estimator approach and the MR Egger. The weighted median estimator approach implemented the weighted median of the Wald ratios for all instrumental variables, which is able to tolerate up to (not including) 50% of the weight coming from invalid instrumental variables [28]. The MR Egger method is sensitive to horizontal pleiotropy and the estimated value of intercept can be interpreted as an estimate of average pleiotropic effect across the genetic variants [29]. Results for MR analyses were reported in either odds ratio for binary outcomes or estimates for continuous outcomes with accompanied 95% confidence interval.

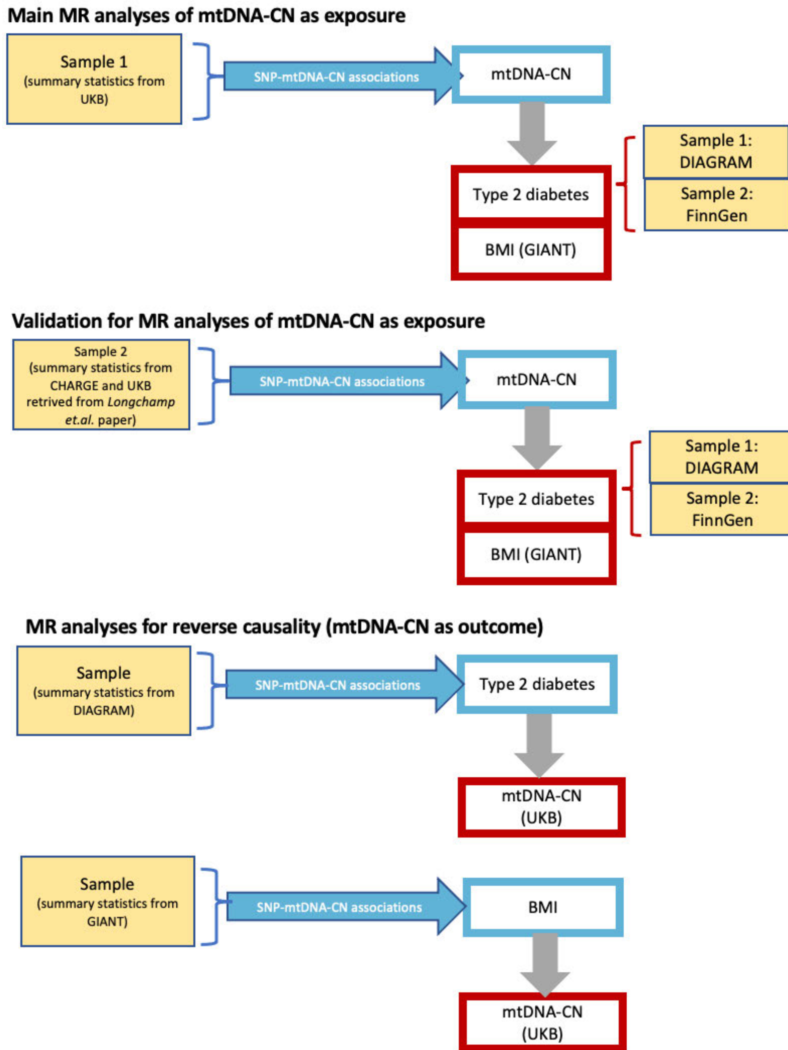


Figure 1 Study design diagram for bi-directional MR analyses. Gray arrows present MR effect estimates, traits in blue rectangles present exposures, traits in red rectangles present outcomes and GWAS data used for traits presented in yellow rectangles. mtDNA-CN: mitochondrial DNA copy number; BMI: body mass index; DIAGRAM: Diabetes Genetics Replication and Meta-analysis consortium; GIANT: The Genetic Investigation of Anthropometric Traits consortium.

Results

Study population

The multivariable-adjusted analyses were conducted in a final sample of 285,967 individuals after excluding 7,207 individuals with type 2 diabetes at baseline. During the follow-up (median follow-up time: 11.9 years; interquartile range: 11.2-12.6 years), 15,111 individuals developed type 2 diabetes. The characteristics of participants stratified by quartile of blood mtDNA-CN after imputation are shown in **Table 1** and the original baseline characteristics before imputation are shown in **ESM Table 3**. The mean age for the whole population is 56.8 (SD = 8.00) years and 54.1% participants are women.

Multivariable-adjusted analyses on incident type 2 diabetes

The Kaplan-Meier curve (**Figure 2A**) illustrated that participants with a higher level of blood mtDNA-CN had a proportionally lower risk of incident type 2 diabetes.

The results from the multivariable-adjusted prospective analyses are presented in **Figure 2B** and **ESM Table 4**. The multivariable-adjusted Cox proportional models show the higher level of blood mtDNA-CN is associated with lower risk of developing type 2 diabetes with hazard ratio (HR) of 0.90 (95% Confidence Interval (CI): 0.89-0.92) for model 1, 0.95 (95% CI: 0.93-0.97) for model 2 and 0.95 (95% CI: 0.94-0.97) for model 3 per 1 SD increase in blood mtDNA-CN. Similarly, taking the participants in the 4th quartile as reference group, model 1 shows participants in quartile 1 for blood mtDNA-CN had the highest risk to develop type 2 diabetes (HR: 1.30; 95% CI: 1.24-1.36); in addition, both quartile 2 and quartile 3 had higher risk of incident type 2 diabetes compared with the reference group (respectively, HR: 1.17; 95% CI: 1.12-1.23, and HR: 1.08; 95% CI: 1.03-1.13). Results did somewhat attenuate after adjustment for BMI (Model 2), and remained similar after further adjustment for other considered confounders (Model 3).

We did not observe any interaction between blood mtDNA_CN with sex or age with p-values for interaction of 0.06 and 0.61, respectively, as was also observed in subgroup analyses (**ESM Table 5**, **ESM Table 6**).

Higher blood mtDNA-CN was associated with lower BMI (-0.12 (95% CI: -0.14 ~ -0.10) kg/m2) in multivariable-adjusted cross-sectional analyses. The results from linear models of blood mtDNA-CN with whole-body lean mass, whole-body fat mass and liver enzymes are shown in **ESM Table 7**. Blood mtDNA-CN was not associated with whole-body lean mass, ALP, AST and ALT. A 1% decrease in whole-body fat mass and GGT were observed per SD increase in blood mtDNA-CN.

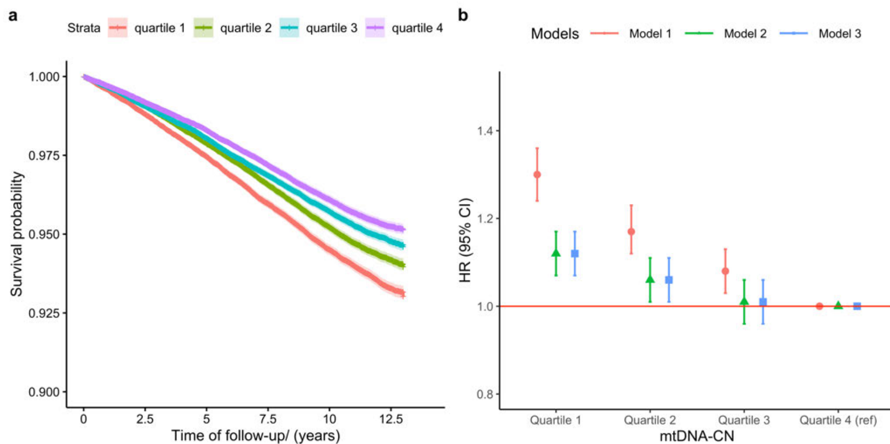


Figure 2 (**A**) Kaplan-Meier curve presenting time to develop incident type 2 diabetes by quartile of mtDNA copy number. X-axis represents follow-up time in years. Y-axis represents the overall survival probability of developing type 2 diabetes. (**B**) Hazard ratio for incident type 2 diabetes by quartiles of mtDNA copy number. Model 1 was adjusted for genotyping batch, principal component 1, principal component 2, white blood cell counts, platelet counts, age and sex; model 2 was additionally adjusted for BMI based on model 1; model 3 was additionally adjusted for BMI, medication use for cholesterol, smoking status, physical activities, waist circumference and family history of type 2 diabetes based on model 1.

Bi-directional Mendelian Randomization analyses

In our newly-conducted GWAS of blood mtDNA-CN in UKB, 3075 significant SNPs ($P < 5e-8$) were identified, of which 55 were independent lead SNPs ($R^2<0.001$), which were used in the MR analyses as exposure (**ESM Table 8**). These independent lead SNPs explained 1.3% of variance in blood mtDNA-CN levels. The -log10(p-value) plot and QQ plot for GWAS of blood mtDNA-CN are presented in **ESM Fig. 2**.

The results from the MR analyses taking 55 SNPs of blood mtDNA-CN from the UKB as exposure are shown in **Figure 3** and **ESM Table 9**. We observed that the odds ratio per SD increase of blood mtDNA-CN on type 2 diabetes using the IVW method was 1.07 (95% CI: 0.89-1.29) for DIAGRAM as outcome data source and 1.11 (95% CI: 0.90-1.38) for FinnGen as outcome data source. Similar results were obtained when we used the previously identified SNPs from the larger *Longchamps et al* study [25] as exposure (**ESM Fig. 3** and **ESM Table 9**) which explained 2.2% of variance in blood mtDNA-CN levels. The results from the MR analyses of blood mtDNA-CN on BMI using either UKB data or data from *Longchamps et al.* as exposure gave similar results indicating no evidence for a causal association between blood mtDNA-CN and BMI (**Figure 4** and **ESM Table 10**). The results from weighted median and MR Egger were consistent with the IVW method in the above analyses.

The reverse MR was applied to investigate the potential causal association between genetically-influenced BMI and type 2 diabetes on blood mtDNA-CN, respectively. **Figure 5** illustrated the estimates of type 2 diabetes on blood mtDNA-CN and BMI on blood mtDNA-CN. For both exposures, we did not find evidence for a causal association, and consistent results were derived in the weighted median and MR Egger analyses (**ESM Table 11**).

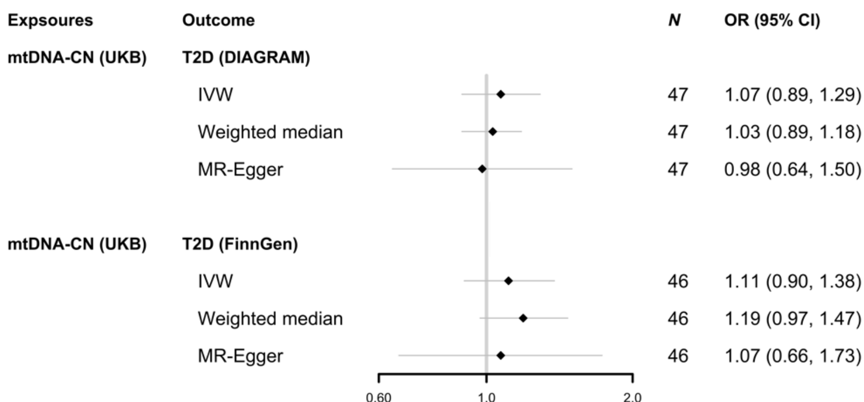


Figure 3 Forest plot of MR results for mtDNA-CN (UKB) and type 2 diabetes (DIAGRAM and FinnGen). N: number of SNPs analyzed in the MR analyses; OR: odds ratio per 1 SD increase in mtDNA-CN; 95% CI: 95% confidence interval; IVW: inverse variance weighted.

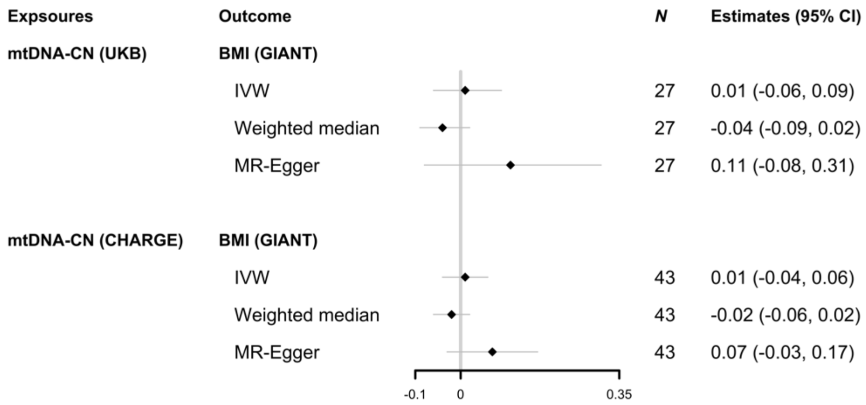


Figure 4 Forest plot of MR results for mtDNA-CN (UKB and CHARGE) and BMI (GIANT). N: number of SNPs analyzed in the MR analyses; Estimates are reported per 1 SD increase in mtDNA-CN; 95% CI: 95% confidence interval; IVW: inverse variance weighted.

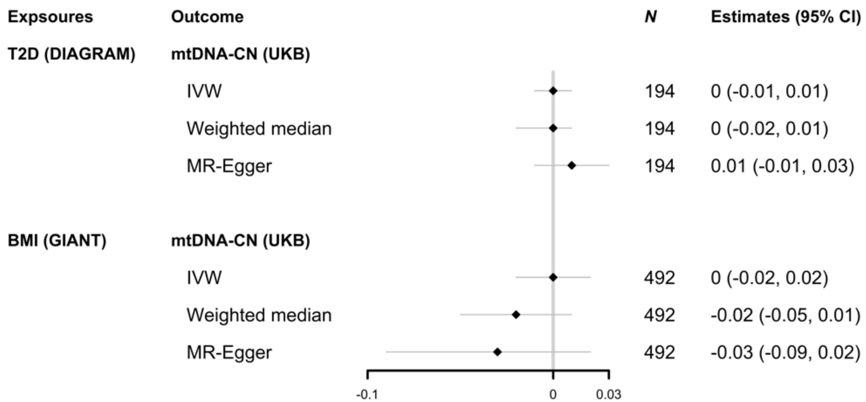


Figure 5 Forest plot of MR results for BMI (GIANT) and type 2 diabetes (DIAGRAM) associating with mtDNA-CN (UKB). N: number of SNPs analyzed in the MR analyses; Estimates were reported as having type 2 diabetes versus not having type 2 diabetes for MR analyses between type 2 diabetes and mtDNA-CN; Estimates were reported as per 1 kg/m2 increase in BMI for MR analyses between BMI and mtDNA-CN; 95% CI: 95% confidence interval; IVW: inverse variance weighted.

Table 2 Baseline characteristics of the population by quartile of mtDNA copy number

	Quartile 1 (N=71492)	Quartile 2 (N=71492)	Quartile 3 (N=71491)	Quartile 4 (N=71492)	Overall (N=285967)
mtDNA-CN (SD)	-1.12 [-5.09, -0.658]	-0.313 [-0.658, -0.00423]	0.303 [-0.00420, 0.656]	1.14 [0.656, 4.68]	-0.00423 [-5.09, 4.68]
Sex (Female)	37587 (52.6%)	38340 (53.6%)	38911 (54.4%)	39801 (55.7%)	154639 (54.1%)
Age (years)	57.4 (8.01)	56.9 (8.00)	56.6 (7.97)	56.1 (7.97)	56.8 (8.00)
White blood cell count	7.36 (1.83)	6.96 (1.72)	6.72 (1.70)	6.40 (2.50)	6.86 (1.99)
Platelet count	247 (57.8)	252 (57.8)	256 (59.0)	259 (62.9)	253 (59.6)
BMI (kg/m^2)	27.6 (4.88)	27.4 (4.66)	27.2 (4.57)	26.9 (4.47)	27.3 (4.65)
Waist circumference (cm)	91.0 (13.6)	90.3 (13.3)	89.7 (13.1)	88.8 (12.9)	90.0 (13.2)
Whole-body fat mass (kg)	25.2 (9.80)	24.8 (9.41)	24.5 (9.26)	24.0 (9.05)	24.6 (9.39)
Whole-body lean mass (kg)	53.7 (11.5)	53.5 (11.5)	53.2 (11.4)	53.0 (11.4)	53.3 (11.5)
Physical activity (hours/week)	26.4 (33.9)	27.0 (34.4)	26.9 (33.9)	27.1 (34.0)	26.8 (34.1)
Cholesterol-lowering medicine	11941 (16.7%)	11468 (16.0%)	10974 (15.4%)	10430 (14.6%)	44813 (15.7%)

	Quartile 1 (N=71492)	Quartile 2 (N=71492)	Quartile 3 (N=71491)	Quartile 4 (N=71492)	Overall (N=285967)
Smoking status					
Never	38362 (53.7%)	39170 (54.8%)	39715 (55.6%)	40525 (56.7%)	157772 (55.2%)
Previous	25027 (35.0%)	24996 (35.0%)	24806 (34.7%)	24904 (34.8%)	99733 (34.9%)
Current	8103 (11.3%)	7326 (10.2%)	6970 (9.7%)	6063 (8.5%)	28462 (10.0%)
Family history of T2D	11262 (15.8%)	10959 (15.3%)	11125 (15.6%)	11156 (15.6%)	44502 (15.6%)
Follow-up time (years)	11.2 (2.30)	11.3 (2.12)	11.4 (2.06)	11.5 (1.99)	11.4 (2.13)
Height (cm)	169 (9.21)	169 (9.24)	169 (9.26)	169 (9.25)	169 (9.24)
ALT (U/L)	20.3 [3.10, 495]	20.2 [3.18, 490]	20.1 [3.25, 425]	19.7 [3.10, 491]	20.1 [3.10, 495]
ALP (U/L)	81.4 [8.00, 1360]	80.4 [14.8, 1420]	79.7 [14.2, 1360]	78.9 [14.1, 1420]	80.1 [8.00, 1420]
AST (U/L)	24.5 [5.10, 947]	24.4 [3.30, 711]	24.4 [5.10, 572]	24.3 [4.40, 584]	24.4 [3.30, 947]
GGT (U/L)	26.9 [5.50, 1170]	26.1 [5.10, 1120]	25.8 [5.00, 1160]	25.1 [5.20, 1160]	26.0 [5.00, 1170]

Continuous variables were presented as mean with standard deviation (SD) or median with minimum and maximum values, and categorial variables were presented as number with percentages. BMI: body mass index; T2D: type 2 diabetes; ALT: alanine aminotransferase; AST: aspartate aminotransferase; ALP: alkaline phosphatase; GGT: gamma glutamyl transfer

Discussion and Conclusion

We found a lower level of blood mtDNA-CN to be associated with higher risk of incident type 2 diabetes. A weak association between blood mtDNA-CN and lower BMI was also found. However, no associations were observed between lower genetically-influenced blood mtDNA-CN and higher type 2 diabetes risk nor higher BMI, suggesting that the observed associations are likely not causal. Also in the reverse direction, no evidence was found for a causal association of type 2 diabetes and of BMI with blood mtDNA-CN. Further regression analyses showed that blood mtDNA-CN was not associated with whole-body lean mass, AST, ALT and ALP, but was associated with whole-body fat mass and GGT, however, with very small observed effect sizes.

Three previous follow-up studies have investigated the association between blood mtDNA-CN and incident type 2 diabetes [13, 15, 16]. Our findings are in line with the results of *Memon et.al.* who found a lower level of blood mtDNA-CN to be associated with increased risk of incident type 2 diabetes in a follow-up study of 2387 middle-aged Swedish women [13]. However, research conducted in the Atherosclerosis Risk in Communities (ARIC) study found that mtDNA-CN in circulating leukocytes was not associated with risk of developing type 2 diabetes among 7713 white participants [15]. Same results as ARIC study were reported by *Erwin Reiling* using European participants from Botnia and Rotterdam studies [16]. Our finding of an association of higher blood mtDNA-CN with lower BMI is similar to the findings of *Liu Xue et.al* [30] and *Skuratovskaia et. al.* [31]. The divergent results on type 2 diabetes may be due to the different characteristics of the populations, different covariates used in the Cox models, the data source of type 2 diabetes and the measurement methods for mtDNA-CN. However, the extremely large UK biobank cohort make it possible to observe the small effect sizes of the associations and the discrepancy may be explained by the differences in power.

We found no evidence for the hypothesis that low blood mtDNA-CN drives the risk of type 2 diabetes and obesity from the results of the MR analyses. This hypothesis is based on the assumption that blood mtDNA-CN reflects overall mitochondrial function including the mitochondrial function of tissues that play a role in type 2 diabetes like muscle and liver. We further elaborated on this assumption by performing linear regression analyses between blood mtDNA-CN and whole-body lean and fat mass, and liver

enzymes. We found one SD increase in blood mtDNA-CN led to no change in lean mass, AST, ALT and ALP, and small decreases in GGT. This indicates that blood mtDNA-CN may not be a marker for mitochondrial function in muscle and a weak marker for mitochondrial function in liver. Although one study in monkeys reported that mitochondrial function in blood was associated with mitochondrial function in muscle [32], human trials have failed to show the correlation between mitochondrial respiration in peripheral blood and muscle in both women and men. This is in agreement with our finding that blood mtDNA-CN may not be a good marker for muscle mitochondrial function [33, 34]. The associations that have been observed between mtDNA-CN and liver disease are not consistent. Increased liver mtDNA-CN was found in patients with non-alcoholic fatty liver disease [35, 36], while decreased blood and liver mtDNA-CN was found in patients with biliary atresia [37, 38]. Although blood mtDNA-CN was associated with liver enzymes GGT in our study, given the small effect sizes, the blood mtDNA-CN is a weak marker for mitochondrial functional in the liver.

Our data showed a clear observational association between blood mtDNA-CN and incident type 2 diabetes. Since the MR analyses did not find evidence for causality, it was expected that adjusting the observational association for BMI and other considered confounders would significantly attenuate the association. However, this was not the case and only adjusting for BMI moderately attenuated the association. The most likely explanation is that the remaining association is due to residual confounding. In addition to unmeasured residual confounding, some measured variables may not be very precise. For example, the physical activities were collected based on the recall of participants, which may not be accurate.

Our study has some strengths and limitations. One of the strengths of our study is that we included 285,967 unrelated European individuals without type 2 diabetes, which is a large sample size and large number (N=15,111) of incident type 2 diabetes cases. To our knowledge, this is the largest investigation on blood mtDNA-CN and type 2 diabetes. One of the limitations was that blood mtDNA-CN may not be a reflection of mitochondrial function in muscle and liver, which are relevant in type 2 diabetes pathogenesis, and therefore blood mtDNA-CN might not be a suitable marker for the involvement of these tissues in type 2 diabetes. Another limitation is that we only included European populations and the results cannot be extrapolated to other populations with different ancestries. In addition, there was some overlap in the data sources for exposure and

outcome in the two-sample MR analyses. However, Minelli *et.al* provide evidence that two-sample MR methods perform well in large one-sample MR studies even when the assumption of independence does not hold. Only results derived with the MR-Egger regression analysis should be interpreted with caution [39].

In conclusion, our results showed that blood mtDNA-CN was prospectively associated with incident type 2 diabetes, but MR analyses did not provide evidence for causality.

Reference

[1]Zheng Y, Ley SH, Hu FB (2018) Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nature Reviews Endocrinology 14(2): 88-98. 10.1038/nrendo.2017.151
[2]Wu Y, Ding Y, Tanaka Y, Zhang W (2014) Risk factors contributing to type 2 diabetes and recent advances in the treatment and prevention. Int J Med Sci 11(11): 1185-1200. 10.7150/ijms.10001
[3]Montgomery MK, Turner N (2015) Mitochondrial dysfunction and insulin resistance: an update. Endocrine Connections 4(1): R1-R15. 10.1530/EC-14-0092
[4]Dimauro S, Davidzon G (2005) Mitochondrial DNA and disease. Ann Med 37(3): 222-232. 10.1080/07853890510007368
[5]Kraja AT, Liu C, Fetterman JL, et al. (2019) Associations of Mitochondrial and Nuclear Mitochondrial Variants and Genes with Seven Metabolic Traits. Am J Hum Genet 104(1): 112-138. 10.1016/j.ajhg.2018.12.001
[6]Castellani CA, Longchamps RJ, Sun J, Guallar E, Arking DE (2020) Thinking outside the nucleus: Mitochondrial DNA copy number in health and disease. Mitochondrion 53: 214-223. https://doi.org/10.1016/j.mito.2020.06.004
[7]Ashar FN, Zhang Y, Longchamps RJ, et al. (2017) Association of Mitochondrial DNA Copy Number With Cardiovascular Disease. JAMA Cardiology 2(11): 1247-1255. 10.1001/jamacardio.2017.3683
[8]Reznik E, Miller ML, Şenbabaoğlu Y, et al. (2016) Mitochondrial DNA copy number variation across human cancers. Elife 5. 10.7554/eLife.10769
[9]Lee HK, Song JH, Shin CS, et al. (1998) Decreased mitochondrial DNA content in peripheral blood precedes the development of non-insulin-dependent diabetes mellitus. Diabetes Res Clin Pract 42(3): 161-167. 10.1016/s0168-8227(98)00110-7
[10]Kaaman M, Sparks LM, van Harmelen V, et al. (2007) Strong association between mitochondrial DNA copy number and lipogenesis in human white adipose tissue. Diabetologia 50(12): 2526-2533. 10.1007/s00125-007-0818-6
[11]Ritov VB, Menshikova EV, He J, Ferrell RE, Goodpaster BH, Kelley DE (2005) Deficiency of subsarcolemmal mitochondria in obesity and type 2 diabetes. Diabetes 54(1): 8-14. 10.2337/diabetes.54.1.8
[12]Nile DL, Brown AE, Kumaheri MA, et al. (2014) Age-related mitochondrial DNA depletion and the impact on pancreatic Beta cell function. PLoS One 9(12): e115433. 10.1371/journal.pone.0115433
[13]Memon AA, Sundquist J, Hedelius A, Palmér K, Wang X, Sundquist K (2021) Association of mitochondrial DNA copy number with prevalent and incident type 2 diabetes in women: A population-based follow-up study. Sci Rep 11(1): 4608. 10.1038/s41598-021-84132-w
[14]Fazzini F, Lamina C, Raftopoulou A, et al. Association of mitochondrial DNA copy number with metabolic syndrome and type 2 diabetes in 14 176 individuals. Journal of Internal Medicine n/a(n/a). https://doi.org/10.1111/joim.13242
[15]DeBarmore B, Longchamps RJ, Zhang Y, et al. (2020) Mitochondrial DNA copy number and diabetes: the Atherosclerosis Risk in Communities (ARIC) study. BMJ Open Diabetes Research & Care 8(1): e001204. 10.1136/bmjdrc-2020-001204
[16]Reiling E, Ling C, Uitterlinden AG, et al. (2010) The association of mitochondrial content with prevalent and incident type 2 diabetes. J Clin Endocrinol Metab 95(4): 1909-1915. 10.1210/jc.2009-1775
[17]Emdin CA, Khera AV, Kathiresan S (2017) Mendelian randomization. Jama 318(19): 1925-1926
[18]Luo J, Noordam R, Jukema JW, et al. (2021) Low mitochondrial copy number drives atherogenic cardiovascular disease: evidence from prospective cohort analyses

in the UK Biobank combined with Mendelian Randomization. medRxiv: 2021.2007.2001.21259854. 10.1101/2021.07.01.21259854
[19]Hägg S, Jylhävä J, Wang Y, Czene K, Grassmann F (2021) Deciphering the genetic and epidemiological landscape of mitochondrial DNA abundance. Human Genetics 140(6): 849-861. 10.1007/s00439-020-02249-w
[20]van Buuren S, Groothuis-Oudshoorn K (2011) mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software 45(3): 1 - 67. 10.18637/jss.v045.i03
[21]Loh PR, Tucker G, Bulik-Sullivan BK, et al. (2015) Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat Genet 47(3): 284-290. 10.1038/ng.3190
[22]Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York
[23]Winkler TW, Kutalik Z, Gorski M, Lottaz C, Kronenberg F, Heid IM (2015) EasyStrata: evaluation and visualization of stratified genome-wide association meta-analysis data. Bioinformatics (Oxford, England) 31(2): 259-261. 10.1093/bioinformatics/btu621
[24]Teslovich TM, Musunuru K, Smith AV, et al. (2010) Biological, clinical and population relevance of 95 loci for blood lipids. Nature 466(7307): 707-713. 10.1038/nature09270
[25]Longchamps R, Yang S, Castellani C, et al. (2021) Genetic analysis of mitochondrial DNA copy number and associated traits identifies loci implicated in nucleotide metabolism, platelet activation, and megakaryocyte proliferation, and reveals a causal association of mitochondrial function with mortality. bioRxiv: 2021.2001.2025.428086. 10.1101/2021.01.25.428086
[26]Hemani G, Tilling K, Davey Smith G (2017) Orienting the causal relationship between imprecisely measured traits using GWAS summary data. PLoS Genet 13(11): e1007081. 10.1371/journal.pgen.1007081
[27]Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan N, Thompson J (2017) A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. Statistics in Medicine 36(11): 1783-1802. https://doi.org/10.1002/sim.7221
[28]Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG (2017) Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants. Epidemiology 28(1): 30-42. 10.1097/EDE.0000000000000559
[29]Burgess S, Thompson SG (2017) Interpreting findings from Mendelian randomization using the MR-Egger method. European journal of epidemiology 32(5): 377-389. 10.1007/s10654-017-0255-x
[30]Liu X, Longchamps RJ, Wiggins KL, et al. (2021) Association of mitochondrial DNA copy number with cardiometabolic diseases. Cell Genomics 1(1): 100006. https://doi.org/10.1016/j.xgen.2021.100006
[31]Skuratovskaia DA, Sofronova JK, Zatolokin PA, et al. (2018) Additional evidence of the link between mtDNA copy number and the body mass index. Mitochondrial DNA Part A 29(8): 1240-1244. 10.1080/24701394.2018.1436170
[32]Tyrrell DJ, Bharadwaj MS, Jorgensen MJ, Register TC, Molina AJA (2016) Blood cell respirometry is associated with skeletal and cardiac muscle bioenergetics: Implications for a minimally invasive biomarker of mitochondrial health. Redox Biology 10: 65-77. https://doi.org/10.1016/j.redox.2016.09.009
[33]Rose S, Carvalho E, Diaz EC, et al. (2019) A comparative study of mitochondrial respiration in circulating blood cells and skeletal muscle fibers in women. Am J Physiol Endocrinol Metab 317(3): E503-e512. 10.1152/ajpendo.00084.2019
[34]Hedges CP, Woodhead JST, Wang HW, et al. (2019) Peripheral blood mononuclear cells do not reflect skeletal muscle mitochondrial function or adaptation

to high-intensity interval training in healthy young men. J Appl Physiol (1985) 126(2): 454-461. 10.1152/japplphysiol.00777.2018

[35]Kamfar S, Alavian SM, Houshmand M, et al. (2016) Liver Mitochondrial DNA Copy Number and Deletion Levels May Contribute to Nonalcoholic Fatty Liver Disease Susceptibility. Hepat Mon 16(12): e40774-e40774. 10.5812/hepatmon.40774

[36]Ma C, Liu Y, He S, et al. (2020) Association Between Leukocyte Mitochondrial DNA Copy Number and Non-alcoholic Fatty Liver Disease in a Chinese Population Is Mediated by 8-Oxo-2'-Deoxyguanosine. Front Med (Lausanne) 7: 536. 10.3389/fmed.2020.00536

[37]Udomsinprasert W, Poovorawan Y, Chongsrisawat V, Vejchapipat P, Jittikoon J, Honsawek S (2019) Leukocyte mitochondrial DNA copy number as a potential biomarker indicating poor outcome in biliary atresia and its association with oxidative DNA damage and telomere length. Mitochondrion 47: 1-9. https://doi.org/10.1016/j.mito.2019.04.006

[38]Tiao M-M, Lin T-K, Kuo F-Y, et al. (2007) Early Stage of Biliary Atresia Is Associated With Significant Changes in 8-Hydroxydeoxyguanosine and Mitochondrial Copy Number. Journal of Pediatric Gastroenterology and Nutrition 45(3)

[39]Minelli C, Del Greco M F, van der Plaat DA, Bowden J, Sheehan NA, Thompson J (2021) The use of two-sample methods for Mendelian randomization analyses on single large datasets. International Journal of Epidemiology. 10.1093/ije/dyab084

Chapter 4

Clustered Mendelian randomization analyses identify distinct and opposing pathways in the association between genetically influenced insulin-like growth factor-1 and type 2 diabetes mellitus

Wenyi Wang, Ephrem Baraki Tesfay, Jan Bert van Klinken, Ko Willems van Dijk, Andrzej Bartke, Diana van Heemst, Raymond Noordam

Int J Epidemiol (2022) 10.1093

The supplemental information for this paper is available online at https://drive.google.com/drive/folders/1UoSstJZtu2ylarbtkhn6Kmr3rz68hhSu?usp=sharing

Abstract

Background: There is inconsistent evidence for the causal role of serum insulin-like growth factor-1 (IGF-1) concentration in the pathogenesis of human age-related diseases as type 2 diabetes (T2D). Here, we investigated the association between IGF-1 and T2D using (clustered) Mendelian Randomization (MR) analyses in the UK Biobank.

Methods: We conducted Cox proportional hazard analyses in 451,232 European-ancestry individuals of the UK Biobank (55.3% women, mean age at recruitment 56.6 years), among which 13,247 individuals developed type 2 diabetes during up to 12 years of follow-up. In addition, we conducted two-sample MR analyses based on independent single nucleotide polymorphisms (SNPs) associated with IGF-1. Given the heterogeneity between the MR effect estimates of individual instruments (P-value for Q statistic=4.03e-145), we also conducted clustered MR analyses. Biological pathway analyses of the identified clusters were performed by overrepresentation analyses.

Results: In the Cox proportional hazard models, with IGF-1 concentrations stratified in quintiles, we observed that participants in the lowest quintile had the highest relative risk of type 2 diabetes (hazard ratio (HR): 1.31; 95% CI: 1.23-1.39). In contrast, in the two-sample MR analyses, higher genetically-influenced IGF-1 was associated with a higher risk of type 2 diabetes. Based on the heterogeneous distribution of MR effect estimates of individual instruments, six clusters of genetically determined IGF-1 associated either with a lower or a higher risk of type 2 diabetes were identified. The main clusters in which a higher IGF-1 was associated with a lower risk of type 2 diabetes consisted of instruments mapping to genes in the growth-hormone signaling pathway, whereas the main clusters in which a higher IGF-1 was associated with a higher risk of type 2 diabetes consisted of instruments mapping to genes in pathways related to amino acid metabolism and genomic integrity.

Conclusion: The IGF-1 associated SNPs used as genetic instruments in MR analyses showed a heterogeneous distribution of MR effect estimates on the risk of type 2 diabetes. This was likely explained by differences in the underlying molecular pathways that increase IGF-1 concentration and differentially mediate the effects of IGF-1 on type 2 diabetes.

Introduction

Insulin-like growth factor-1 (IGF-1) is a pleiotropic hormone that plays a major role in cellular growth, proliferation and survival [1]. The secretion of IGF-1, predominately by the liver, is promoted by growth hormone (GH) while conversely, IGF-1 in the circulation feeds back centrally to hypothalamus to inhibit GH secretion [2]. The availability of free IGF-1 in the blood is regulated by its association with distinct insulin-like growth factor binding proteins (IGFBPs) which can increase IGF-1 half-life or block its binding to IGF-1 receptors [3]. IGF-1 plays a very important role in aging process and it was found that the level of IGF-1 increases during puberty and decreases gradually with increasing age during adulthood [4, 5]. IGF-1 was also found to be involved in the pathophysiology of various diseases, including cancer, neurodegenerative disease, cardiovascular disease, and type 2 diabetes mellitus [6]. Several (prospective) cohort studies have found that lower levels of IGF-1 were associated with an increased risk of impaired glucose tolerance, increased insulin resistance and hence the development of diabetes mellitus [7, 8].

In contrast to these multivariable-adjusted association analyses, a recent study demonstrated that a higher genetically-influenced IGF-1 concentration was associated with a higher risk of developing type 2 diabetes using Mendelian Randomization (MR) analyses [9]. MR is an approach to determine whether the association between risk factors and outcome is causal by using genetic variants as instrumental variables [10], and therefore does not meet the requirements for triangulation of observations done in epidemiology [11]. However, in some circumstances, there is clear heterogeneity in the causal effects of the individual single nucleotide polymorphisms (SNPs) that are used as instrumental variables, which may indicate either pleiotropy or differences in biological pathways contributing to high levels of the exposure [12]. Clustered MR was recently developed to provide a means to address the heterogeneity in causal effects by clustering variants that show similar individual causal estimates on the outcome [12]. Previously, such context-dependent MR analyses have been proposed to provide more biological perspective in causal associations [13, 14].

We hypothesized that heterogeneity in causal effects of individual variants could be a reflection of different biological mechanisms involved in the association between IGF-1 and type 2 diabetes. For example, variation influenced by processes causing insufficient GH signaling may have a

different impact on T2D than variation influenced by processes causing increased GH resistance. Therefore, in this study, we aimed to investigate the association between IGF-1 and incident type 2 diabetes followed by clustered MR analyses [15] in the large UK Biobank population, and explored the possible biological pathways involved in the clustered causal associations.

Methods

Prospective analyses

Study setting and study population

The UK Biobank is a very large prospective cohort study with over 500,000 participants aged 40-69 years at recruitment across the entire United Kingdom (UK) [16]. Participants were recruited between 2006 and 2010 in 22 assessment centers across the UK. Baseline examinations in all participants included physical measures, collection of blood, urine and saliva, a self-completed touch-screen questionnaire, and a brief computer-assisted interview to investigate sociodemographic, family history, environmental factors, lifestyle, psychosocial factors, etc.

The present project was accepted under project number 22474. We restricted the analyses to the UK Biobank participants who reported to be of European ancestry including British, Irish and any other European background, who had information available on serum IGF-1 concentration, and who were in the full release imputed genomics datasets.

Biochemical analyses

Biological samples were collected to measure biochemical markers including IGF-1 at baseline (2006-2010) comprising ~480,000 participant samples. Serum levels of IGF-1 were analyzed using chemiluminescent Immunoassay (DiaSorin Liaison XL) with a one-step sandwich. Coefficients of variation derived from the internal quality control samples of the low, medium, and high IGF-1 concentrations ranged from 6.03-6.18%. More information on assay performance of the UK Biobank Biomarker Project can be found online (https://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=1227).

Outcome definitions

Prevalent and incident diagnosis of type 2 diabetes mellitus was identified in UK Biobank as the date of first appearance of non-insulin-dependent diabetes mellitus (data-field 130708 in the UK Biobank Database). This variable has been composed through a standard algorithm combining the

data derived from hospital admissions (through linkage with the medical records from the National Health Service), general practitioners, death records, and through self-report. Based on the date of first appearance and the data of enrollment, we defined whether a case was prevalent (before enrolment) or incident (after enrolment).

Statistical analyses

The cross-sectional association between IGF-1 and prevalent type 2 diabetes was assessed by means of logistic regression model adjusting for age at recruitment, sex and body mass index (BMI).

Participants without diabetes mellitus at baseline were followed until the occurrence of type 2 diabetes, mortality or loss of follow-up, whichever occurred first. The association between IGF-1 levels (as continuous and stratified variables) and incidence of type 2 diabetes in the UK Biobank cohort was assessed using Cox proportional hazard models. Participants were categorized into five groups based on quintiles of IGF-1 concentration. Quintile 1 (lowest 20%) and quintile 5 (highest 20%) were used separately as reference groups to calculate the hazard ratio (HR), respectively. Potential confounders included sex, age at recruitment and baseline BMI based on height and weight measured at the assessment centers. The analyses were conducted in R using the survival package (version 3.2-7) [17]. The Kaplan Meier curve was plotted to visualize the difference of survival probability between IGF-1 quintiles and whether the proportionality assumption holds.

Mendelian randomization analyses

Exposure data source

Genotyping and genetic imputations

Genome-wide genotype data for all 500,000 UK Biobank participants generated using Affymetrix UK BiLEVE Axiom array (initial 50,000 participants) and the Affymetrix UK Biobank Axiom Array (remaining 450,000 participants), which genotyped around 850,000 variants. All genetic data were quality controlled centrally by UK Biobank resources. In addition, UK Biobank resources performed centralized imputations on approximately 96 million genotypes using the UK10K haplotype [18], 1000 Genomes Phase

3 [19], and Haplotype Reference Consortium (HRC) reference panels [20]. Autosomal SNPs were pre-phased using SHAPEIT3 and imputed using IMPUTE4. Related individuals were identified by estimating kinship coefficients for all pairs of samples using only markers weakly informative of ancestral background. More information on the genotyping processes and genetic imputation can be found online (https://biobank.ctsu.ox.ac.uk/crystal/label.cgi?id=263)

Genome-wide association analyses

Genome-wide association studies (GWAS) on continuous IGF-1 concentrations were performed on all European individuals to provide a list of independent lead SNPs to be used as instrumental variables for exposure (IGF-1) in the MR analyses. Analyses were performed using linear mixed models implemented in the program BOLT_LMM (version 2.3.2) [21]. We adjusted the analyses for age, sex and the first 10 principal components, and corrected for the Kinship matrix to correct for familial relationships in the UK Biobank population. Analyses were done on the autosomal chromosomes only. SNPs with a minor allele frequency <0.01 as well as SNPs with an imputation quality <0.3 were excluded. P-values of SNPs smaller than $5e^{-8}$ were extracted and stored for the MR analyses. Visualization of the results was performed using the R-based packages ggplot2 [22] and EasyStrata [23] (www.genepi-regensburg.de/easystrata).

Outcome data source

For this analysis, the instruments for outcome (type 2 diabetes) were acquired from the DIAGRAM Consortium, which is a publicly-available summary-statistics GWAS meta-analysis of European ancestry comprising 32 studies and 898,130 individuals (74,124 type 2 diabetes cases and 824,006 controls) [24].

The summary statistics of outcome for glycated haemoglobin (HbA1c), fasting glucose, C-reactive protein (CRP) level and BMI were retrieved from an online source (https://gwas.mrcieu.ac.uk/) under the GWAS IDs ieu-b-104 (HbaA1c, N = 46,368), ieu-b-114 (fasting glucose, N = 133,010), ieu-b-35 (CRP, N = 204,402) and ieu-b-40 (BMI, N = 681,275).

Mendelian randomization analyses

Two-sample MR was performed with summary-based statistics of GWAS using the R-based statistical package Two Sample MR (available at: http://github.com/MRCIEU/TwoSampleMR) [25]. This statistical package also connects to a large library of exposures from published GWAS to use as instrumental variables, which is aligned with the online GWAS catalogue.

For the present study, we performed clumping process (window size = 10,000 kb, $R2 < 0.001$) with the European samples from the 1000 genomes project which were used to estimate linkage disequilibrium (LD) between SNPs. Among the pairs of SNPs with R2 above the specified threshold ($R2 = 0.001$), only the SNPs with the lowest P-value were retained to provide a list of independent lead SNPs from the MR analyses. Otherwise, the statistical power of MR analyses would be overestimated (e.g., underestimated standard errors of the summary estimates of the MR analyses). SNPs present in UK Biobank, but absent from the LD reference panel, were removed. On the basis of the significant independent lead SNPs (P-value<5e-8), we assessed their possible association with type 2 diabetes.

Methods for MR analyses of summary-level data based on a two-sample design have been described in detail previously [26, 27]. Using inverse-variance-weighted (IVW) analyses, we combined the effects of the individual genetic instruments to obtain a genetically-determined association between exposure and outcome under the assumption of the absence of horizontal pleiotropy. However, given the large number of genetic instruments included in the present analyses, there is a high probability that at least some SNPs show pleiotropic effect. To test whether possible pleiotropic effects could bias the overall effect estimates (horizontal pleiotropy), we performed the sensitivity analyses MR Egger regression [28] and weighted median estimator (WME) analyses [29]. MR Egger does not force the regression line through the intercept and is, therefore, able to test for the presence of directional pleiotropy, and WME estimator assumes at least 50% of the instruments included in the MR analyses were valid.

Clustered Mendelian randomization analyses

Clustered MR analyses were conducted to identify groups of genetic variants that have similar causal estimates of the exposure on the outcome, which has been described in detail previously[15]. Briefly, this method performs

likelihood-based clustering on a sample of ratio-estimates and standard errors of genetic variants. A MR-Clust mixture model is firstly built for ratio estimates and the mixture proportion for cluster k, after which these two parameters are estimated by using expectation-maximization (EM) algorithm and optimized by using maximum likelihood estimates (MLE). This MR-Clust mixture model automatically accounts for the possibilities that some or all of the ratio-estimates may be drawn under two distribution – a null distribution (the ratio-estimates are centered around zero) and a junk distribution (highly dispersed ratio-estimates considered as outliers in the sample). The presence of null and junk clusters minimizes false-positive findings from the model and no need to interpret. The maximization step to optimize the parameter values resembles the inverse-variance weighted estimates of the causal effect of the exposure on the outcome. Lastly, the number of substantive clusters are estimated based on the values derived from MLEs for each value of k possible substantive clusters by minimizing the Bayesian information criterion (BIC) which helps to avoid overparameterization. Therefore, if the causal estimates of each genetic variant on the outcome were similar (e.g., their ratio-estimates were similar in direction, magnitude and precision), it was divided into different clusters in which the included genetic variants were more homogeneous concerning the causal estimates. The inclusion probability of SNPs in each cluster was higher than 0.8.

In order to identify distinct causal effects of genetic variants derived from GWAS of continuous IGF-1 levels on type 2 diabetes, we made use of the R-based MR-Clust package (https://github.com/cnfoley/mrclust). MR-Clust performs likelihood-based clustering on Wald ratio-estimates and accompanied standard errors. Genetic instruments within a cluster share similar causal estimates (e.g., Wald ratio-estimates are similar in direction, magnitude and precision) of the causal effect of the exposure on the outcome. MR analyses were repeated on all clusters to investigate the causal effects of each cluster on type 2 diabetes. The stability of the clusters was measured by repeating the same analyses using 1000 randomly generated subsets containing 200 SNPs and with the full set of 332 SNPs associated with IGF-1 levels. In addition, cross-validation analyses were performed by randomly sub-dividing the UKB participants into two subsets and repeating the same clustered MR analyses within each subset of the UKB.

Pathway analyses

We performed pathway analysis on all the six IGF-1 clusters to provide insights into biological pathways which could explain for the heterogeneity between causal estimates of genetic variants in different clusters on type 2 diabetes. Gene candidates of each locus were determined with the SNP2Gene function of FUMA (the Functional Mapping and Annotation of Genome-Wide Association Studies platform) [30], where we used a 20kb positional map and included genes whose expression was associated with the locus in GTEx v8. Gene candidates of a locus that were also linked to a gene that had a clear biological relation to IGF-1 were excluded from the analysis. Specifically, we defined the list of prioritized genes as IGF1, IGF2, INS, IGFALS, GHR, GHSR, GH1, GH2, AKT3, FOXO3, which we based on their overlap with the IGF1 gene in the pathway databases that we used in the subsequent analysis.

Pathway analysis was performed using the pathway definitions of the KEGG (a bioinformatics resource for deciphering the genome) and Reactome database, which were downloaded from ConsensusPathDB [31] on 28-10-2021. Gene sets with more than 1000 genes or less than 2 were excluded. Since genes with a shared function sometimes colocalize on the genome, we developed a novel method for performing overrepresentation analysis that corrects for the inflation induced by the colocalization effect. In detail, we defined gene clusters of similar function and close proximity on the genome by merging all genes that were within a 100kb distance on the genome (GENCODE v38) and that co-occurred in at least one pathway. This operation reduced the total number of genes that were covered in the KEGG and Reactome pathways from 11855 to 9123 gene clusters and is completely independent from the set of gene candidates obtained from the GWAS results. Subsequently the pathway gene sets and the gene candidate set were mapped on the gene clusters, and the overrepresentation was done in terms of the gene clusters where loci that mapped multiple times to genes within the same cluster were counted only once. Overrepresentation analysis was done using Fisher's exact test and results were corrected for multiple testing with the Benjamini Hochberg procedure where pathways with the same definition were counted as one. Only pathways with false discovery rate (FDR) > 0.1 were shown in the present study.

In addition, as follow-up analyses, we conducted two-sample MR analyses to associate IGF-1 and several metabolic traits related to type 2 diabetes including HbA1c, fasting glucose, CRP and BMI.

Results

Characteristics of the study population

In total, 451,232 European participants without diabetes mellitus at baseline were included in our study of whom 13,247 developed type 2 diabetes in up to 12 years of follow-up (**Table 1**). Of the participants not developing type 2 diabetes, 55.7% were women, the mean age at recruitment was 56.5 (SD=8.1) years, the average BMI was 27.1 (SD=4.5) kg/m^2 and the mean IGF-1 levels were 21.5 (SD=5.6) nmol/L. Of the participants developing type 2 diabetes during follow-up, women accounted for 41.9%, the mean age at recruitment was 59.3 (SD=7.2) years, the average BMI was 31.7 (SD=5.6) kg/m^2 and the mean IGF-1 level was 19.9 (SD=6.5) nmol/L.

Table 1 Characteristics of the UK Biobank study population for prospective analyses

	Controls	Cases	Total
N	437985	13247	451232
Age at recruitment, in years	56.5 (8.1)	59.3 (7.2)	56.6 (8.0)
Time to diagnosis, in years	-	5.3 (2.5)	-
% of women	55.7	41.9	55.3
BMI, in kg/m^2	27.1 (4.5)	31.7 (5.6)	27.2 (4.6)
IGF-1 levels in nmol/L	21.5 (5.6)	19.9 (6.5)	21.5 (5.6)

Data presented as means with standard deviation (SD) or as stated otherwise. Information on BMI was missing for 7599 controls and 417 cases. Information on age at diagnosis was missing for 1228 cases. Information on IGF-1 levels is missing for 29889 controls and 990 cases. IGF-1: insulin-like growth factor-1; BMI: body mass index.

Association between IGF-1 and prevalent and incident type 2 diabetes

The cross-sectional analyses showed that a higher level of IGF-1 was significantly associated with a lower odds of prevalent type 2 diabetes (Odds ratio: 0.98; 95% CI: 0.97-0.98, per nmol/L increase of IGF-1).

Multivariable-adjusted Cox proportional hazard model analyses were performed to evaluate the association between IGF-1 levels (as quintile and continuous variables) and incident type 2 diabetes. Kaplan Meier curve (**Figure 1**) illustrated that participants in the lower IGF-1 quintiles (quintile 1and 2) had proportionally lower risk of type 2 diabetes compared to

participants in the higher quintiles (quintile 3, 4 and 5), which is consistent with the result of Cox proportional hazard model when IGF-1 is assessed as a continuous variable. The results showed higher level of IGF-1 are associated with lower risk of incident type 2 diabetes with HR of 0.98 per nmol/L IGF-1 and 95% confidence interval (CI) of 0.97-0.98.

More specifically, individuals in quintile 2 had a lower risk of type 2 diabetes (HR: 0.77; 95% CI: 0.73-0.81) than individuals in quintile 1 (**Table 2**), as did those in quintile 3 (HR: 0.70; 95% CI: 0.66-0.74), quintile 4 (HR: 0.68; 95% CI: 0.64-0.72) and quintile 5 (HR: 0.76; 95% CI: 0.72-0.81). Taking participants in quintile 5 as the reference group, a higher risk was observed in quintile 1 (HR: 1.31; 95% CI: 1.23-1.39), and those in quintile 3 (HR: 0.92; 95% CI: 0.86-0.98) and 4 (HR: 0.89; 95% CI: 0.83-0.95) had a lower risk of developing type 2 diabetes.

Table 2 Hazard ratio of incident type 2 diabetes according to quintiles of insulin-like growth factor-1 concentration

	Quintile 1	**Quintile 2**	**Quintile 3**	**Quintile 4**	**Quintile 5**
Ranges of IGF-1 (nmol/L)	1.44-16.73	16.73-19.93	19.93-22.63	22.63-25.77	25.77-126.77
Hazard ratio (95% CI)	Reference	0.77 (0.73-0.81)	0.70 (0.66-0.74)	0.68 (0.64-0.72)	0.76 (0.72-0.81)
Hazard ratio (95% CI)	1.31 (1.23-1.39)	1.01 (0.94-1.07)	0.92 (0.86-0.98)	0.89 (0.83-0.95)	Reference

*This table presents the hazard ratio (HR) and 95% confidence interval (CI) of incident type 2 diabetes by insulin-like growth factor-1 concentration in quintiles. The second row showed hazard ratios of incident type 2 diabetes using quintile 1 as reference group. The third row showed hazard ratios of incident type 2 diabetes using quintile 5 as reference group. IGF-1: insulin-like growth factor-1

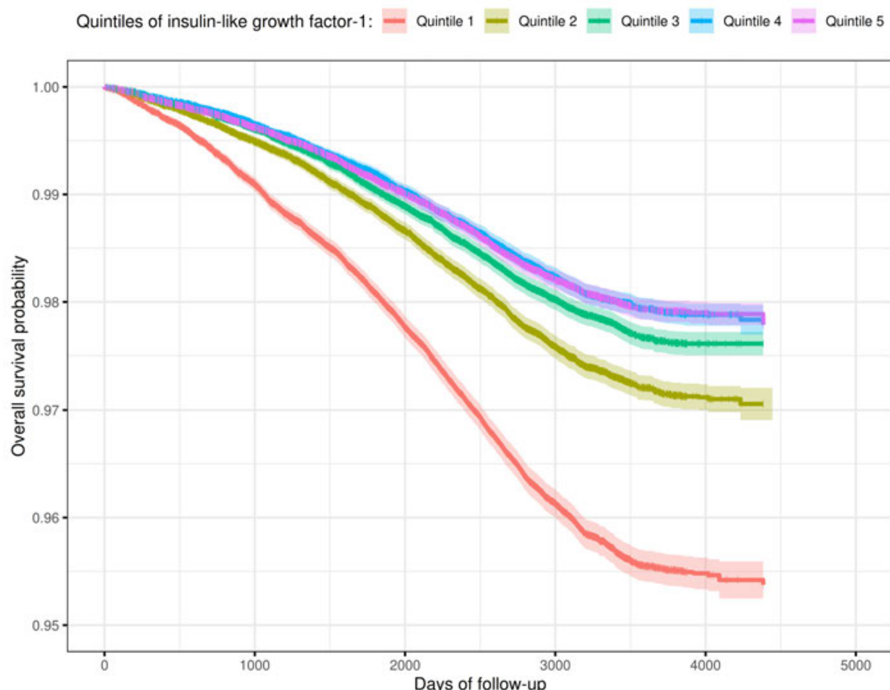


Figure 1 Kaplan Meier survival curves displaying the time to develop incident type 2 diabetes by quintiles of insulin-like growth factor-1 concentrations. The x-axis presents days of follow-up and the y-axis presents overall survival probability. Censoring is indicated by vertical marks.

Mendelian Randomization analyses

There were 95,877 significant SNPs (P<5e-8) identified in the GWAS of continuous IGF-1 (**Supplementary Figure S1**) of which 387 independent lead SNPs were derived. After quality control performed by the TwoSample MR package (e.g. excluding palindromic SNPs and SNPs with intermediate allele frequencies), 323 SNPs were used in the MR analyses (**Supplementary Table S1**). The MR estimates assessing the effect of IGF-1 on type 2 diabetes showed that a 1 nmol/L increase in IGF-1 was associated with a 1% higher risk of type 2 diabetes in the IVW analyses (OR: 1.01; 95% CI: 1.00-1.02). Similar results were obtained from MR Egger (OR: 1.02; 95% CI: 1.00-1.05) and WME analyses (OR: 1.01; 95% CI: 1.00-1.02).

Clustered Mendelian randomization analyses

We observed large heterogeneity in the individual MR effect estimates (**Figure 2**) as was also evidenced by the Q statistic ($P=4.03e^{-145}$). The individual MR effect estimates were clustered into a total of 6 clusters with an inclusion probability of SNPs higher than 0.8 (**Figure 3, Supplementary Table S2**). MR estimates from different methods assessing the effect of 6 clusters on type 2 diabetes are presented in **Figure 4**, **Table 3** and **Supplementary Figure S2**. In cluster 1 (IVW: OR:1.54, 95% CI:1.43-1.65), cluster 2 (IVW: OR:1.03, 95% CI:1.02-1.04) and cluster 5 (IVW: OR:1.20, 95% CI:1.18-1.22), higher levels of IGF-1 level were associated with a higher risk of type 2 diabetes. On the other hand, cluster 3 (IVW: OR:0.92, 95% CI:0.91-0.94), cluster 4 (IVW: OR:0.62, 95% CI:0.58-0.67) and cluster 6 (IVW: OR:0.80, 95% CI:0.76-0.85) showed that higher levels of IGF-1 were associated with lower risk of type 2 diabetes. The results from sensitivity analyses by using weighted median estimator and MR-Egger did not materially differ from the result of the IWV method. The results from the stability test of the MR clust produce demonstrated IGF-1 clusters were consistently constructed across different random subsets of IGF-1 SNPs and across different runs with the same set of 332 IGF-1 SNPs (**Supplementary Figure S3 and S4).** The cross-validation analyses also provide evidence that at least 3 clusters for IGF-1 are reproducible (**Supplementary Figure S5 and S6**)

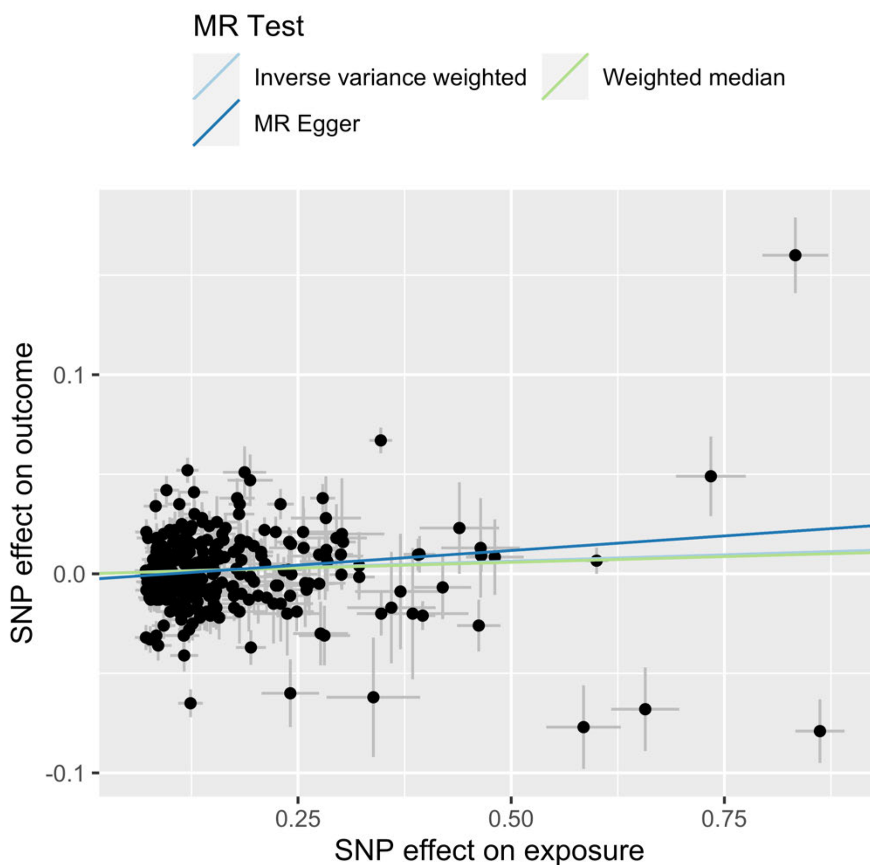


Figure 2 Scatter plots of the MR effect estimates of continuous IGF-1 on type 2 diabetes derived from different MR tests. The x-axis is the genetic association between SNPs and IGF-1 and the y-axis is the genetic association between SNPs and type 2 diabetes. Analyses were conducted using the inverse variance weighted, weighted median and MR Egger methods. The slope of each line presents the estimated MR effect for each method. MR: Mendelian Randomization; IGF-1: insulin-like growth factor-1; SNPs: single nucleotide polymorphisms.

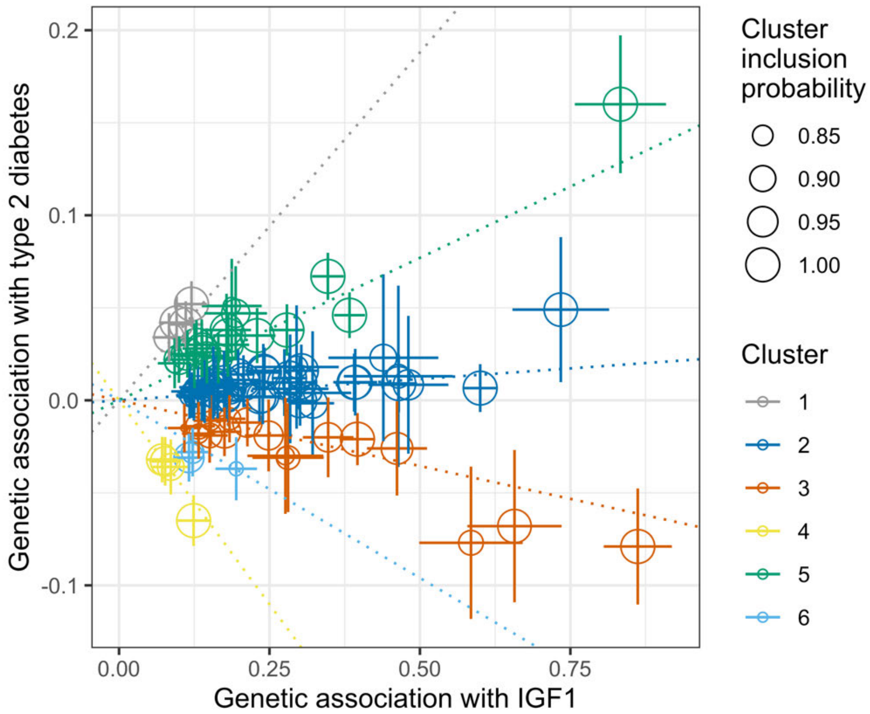


Figure 3 Scatter plots of the MR effect estimates of continuous IGF-1 on type 2 diabetes for clustered MR analyses. The x-axis is the genetic association between SNPs and IGF-1 and the y-axis is the genetic association between SNPs and type 2 diabetes. MR: Mendelian Randomization; IGF-1: insulin-like growth factor-1; SNPs: single nucleotide polymorphisms.

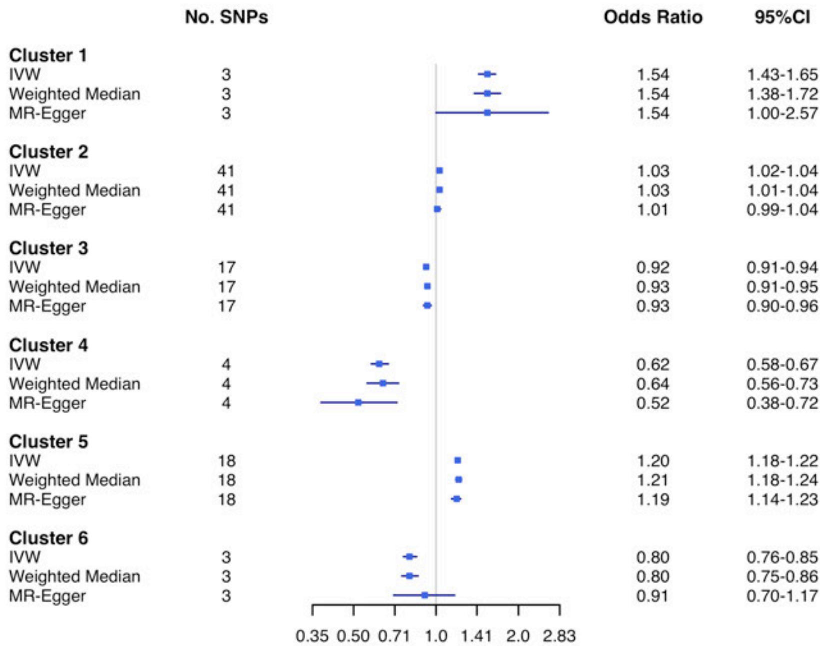


Figure 4 Forest plot of the MR estimates and MR sensitivity analyses for the 6 clusters of genetically-predicted IGF-1 levels on type 2 diabetes. Odds ratios measured the associations between IGF-1 clusters and type 2 diabetes. No. SNPs: number of SNPs used in the separate MR analyses; 95% CI: 95% confidence interval; MR: Mendelian Randomization; IGF-1: insulin-like growth factor-1; SNPs: single nucleotide polymorphisms.

Table 3 Mendelian Randomization effect estimates derived from different methods for 6 clusters on type 2 diabetes

Clusters	Number of SNPs	Inverse variance weighted	Weighted median	MR Egger
		OR (95% CI)	OR (95% CI)	OR (95% CI)
Cluster 1	3	1.54 (1.43-1.65)	1.54 (1.38-1.72)	1.60 (1.00-2.57)
Cluster 2	41	1.03 (1.02-1.04)	1.02 (1.01-1.04)	1.01 (0.99-1.04)
Cluster 3	17	0.92 (0.91-0.94)	0.93 (0.91-0.95)	0.93 (0.90-0.96)
Cluster 4	4	0.62 (0.58-0.67)	0.64 (0.56-0.73)	0.52 (0.38-0.72)
Cluster 5	18	1.20 (1.18-1.22)	1.21 (1.18-1.24)	1.19 (1.14-1.23)
Cluster 6	3	0.80 (0.76-0.85)	0.80 (0.75-0.86)	0.91 (0.70-1.17)

*This table presents the odds ratio (OR) and 95% confidence interval (CI) of associations between 6 clusters and type 2 diabetes. SNPs: single nucleotide polymorphisms.

Pathway analyses

Gene candidates for each SNPs and genes used for the following pathway analyses can be found in **Supplementary Table S3**. For the pathway analyses, we found that cluster 2 and cluster 3 were mapped to specific pathways. Cluster 2 mapped to 11 pathways including metabolism of xenobiotics by cytochrome P450 and chemical carcinogenesis (**Supplementary Figure S7, Supplementary Table S4**). Cluster 3 mapped to 6 pathways including growth hormone synthesis, secretion and action, prolactin receptor signaling and growth hormone receptor signaling (**Supplementary Figure S8, Supplementary Table S5**). Clusters 1, 4, 5 and 6 did not map to any specific pathway.

Follow-up analyses of the clusters in relation to other traits

The results of follow-up analyses using two-sample MR are shown in **Supplementary Table S6**. Using IVW, high genetically-influenced IGF-1 through cluster 2, 3 and 6 were associated with higher HbA1c. In addition, high genetically-influenced IGF-1 through cluster 5 and cluster 6 were associated with lower CRP and high genetically-influenced IGF-1 through cluster 5 was also associated with lower fasting glucose. In general, results were (directionally) consistent in the weighted median and MR-Egger sensitivity analyses.

Discussion and conclusion

In this study, we investigated the association of IGF-1 and type 2 diabetes using prospective multivariable-adjusted survival analyses followed by MR and clustered MR analyses in the UK Biobank. The results from the MR analyses showed that a genetically-influenced higher level of IGF-1 was associated with a higher risk of type 2 diabetes, which was in contrast to the result from the prospective analyses showing that a higher concentration of IGF-1 was associated with a lower risk of type 2 diabetes. Since the underlying individual genetic instruments showed a heterogeneous distribution of MR effect estimates, clustered MR identified 6 clusters of genetic instruments for IGF-1 with different associations with type 2 diabetes, which mapped to distinct molecular pathways. Collectively, our results indicate that the association between IGF-1 and the risk of developing type 2 diabetes is context-dependent.

Findings from other prospective studies regarding the association between IGF-1 and type 2 diabetes have been inconsistent. Our results from the prospective analyses were in line with another cohort study showing high levels of IGF-1 were associated with a lower risk of type 2 diabetes mellitus risk during 4.5 years of follow-up [7]. However, some nested case-cohort studies suggested there was no association between total IGF-1 levels and the risk of type 2 diabetes [32, 33]. In addition, a cohort study found that the association between free IGF-1 and type 2 diabetes was dependent on the level of insulin in women [34]. Interestingly, a recent study that used a data driven approach to metabolically subtype the UK biobank participants found that the subgroup associated with increased prevalence and incidence of type 2 diabetes exhibited a decreased level of IGF-1[35] . The inconsistency of these findings reinforces the notion that the association between IGF-1 and type 2 diabetes is context-dependent which is in line with the distinct biological mechanisms identified by the clustered MR analyses in our study.

Many studies found a J- or U- shaped association between IGF-1 and type 2 diabetes [7, 36] or insulin resistance [37]. For example, one study showed that individuals with both low- or high- IGF-1 levels were at increased risk of developing diabetes in a prospective cohort study [36]. Similarly, a U-shaped association between IGF-1 and measures of insulin resistance was found in a cross-sectional study in Danish adults [37]. Partly in line with these results, we observed a J-shaped relationship between IGF-1 and type 2 diabetes with

particularly low levels of IGF-1 to be associated with an increased risk of developing type 2 diabetes.

The result of MR analyses of continuous IGF-1 and type 2 diabetes in our study was supported by a recent publication showing genetically determined higher levels of IGF-1 were associated with increased risk of developing type 2 diabetes [38]. However, we observed large between-SNP heterogeneity, and the MR effect estimates were not proportional to each other. We identified 6 main clusters of IGF-1 genetic instruments with distinct effects on type 2 diabetes by using clustered MR. After mapping the genetic instruments to genes and overrepresentation analysis using the KEGG and Reactome databases, several clusters were mapped to specific pathways.

To gain insight in the biological basis of the identified IGF-1 genetic clusters in relation to type 2 diabetes, we performed pathway analyses and identified 2 clusters associated with specific pathways. Cluster 2 (in which higher genetically-influenced IGF-1 was associated with a higher risk of type 2 diabetes) mapped to multiple pathways including metabolism of xenobiotics by cytochrome P450 and chemical carcinogenesis. In addition, in MR follow-up analyses using two-sample MR showed cluster 2 was associated with higher HbA1c, which supports our findings on type 2 diabetes. However, the exact interpretation of this cluster is complicated, and requires additional research. It is tempting to speculate that since the cytochrome P450 system is likely involved in the breakdown of HbA1c, it influences the risk of type 2 diabetes by affecting the residence time of HbA1c in blood. In support of this notion, a direct effect of IGF-1 on the cytochrome P450 system has been demonstrated before [39].

Cluster 3 (in which a higher level of genetically-influenced IGF-1 was associated with a lower risk of type 2 diabetes) was mapped to pathways related to GH synthesis, secretion and action, and GH receptor signaling. However, counterintuitively, cluster 3 was associated with higher levels of Hb1Ac, despite being associated with a lower risk of type 2 diabetes. Although this finding seems difficult to interpret, this could mean that as IGF-1 plays an important role in growth and development, higher IGF-1 may have resulted in a larger pancreas, which confers an advantage in a non-diabetic situation (as what was used in the Hb1Ac analyses), by enabling a stronger response to rising glucose levels, resulting in lower HbA1C. However, under conditions of chronically elevated insulin demand, such a stronger pancreatic response may in the end cause an accelerated exhaustion of the insulin production and give a rapid development of a more insulin-

dependent-like phenotype of type 2 diabetes. Previously, IGF-1 has been shown to interact with beta cells[40], and to have direct effects on pancreas development and beta cell compensation to insulin resistance [41]. In support of the biological interpretation of the genes mapped to cluster 3, dysregulation of GH receptor signaling and the GH-IGF-1 axis can lead to multiple diseases such as type 2 diabetes. Mice with liver IGF-1 deficiency had a fourfold increase in GH levels. Upon treatment with a GH antagonist, these mice had decreased blood glucose and insulin levels and increased peripheral insulin sensitivity compared with mice with liver IGF-1 deficiency. These data indicate that the GH/IGF-1 axis plays a balancing role in insulin sensitivity and thus type 2 diabetes [42]. In addition, an epidemiological cross-sectional study showed that IGF-1 was associated with type 2 diabetes risk, but this association varied depending on the insulin levels: in individuals with low levels of insulin IGF-1 decreased type 2 diabetes risk and in individuals with high levels of insulin IGF-1 increased type 2 diabetes risk [34].

The main strength of our study is the extremely large sample size allowing stratification of the genetic instruments with sample statistical power. A limitation of the present study is that the IGF-1 level used was total IGF-1 concentration and not free IGF-1 (e.g., relative to the concentration of IGF binding proteins). Furthermore, the study was performed in a study of European-ancestry participants. Translation of the results to participants of non-European ancestry should be done with caution. Although the SNP-outcome dataset (DIAGRAM) used in two-sample MR analyses had an overlap of 49.3% with the SNP-exposure dataset (UKB only), we expect that the effect of this overlap is limited given the very large sample size, since it has been shown that one sample MR performs well in large samples [43]. In addition, non-fasting blood sampling has likely resulted in somewhat increased variation in IGF-1 concentration and consequently a slight reduction in statistical power[44]. However, given that this variation is most likely independent from genetic make-up, no bias in the results was expected from this. Although sensitivity analyses in a random subset of the UK Biobank and clustering using random subsets of SNP-exposure relationships gave similar results, results warrant replication in an independent sample using SNP-exposure associations derived from a fully independent study sample. Lastly, as our multivariable-adjusted cox-proportional hazard models indicated a possible non-linear association between IGF-1 levels and type 2 diabetes, this was not taken into account in the MR analyses which were done on continuous levels. Whether the different clusters could possibly contribute to a non-linear association is currently not known.

Furthermore, the more recently introduced methods for non-linear MR were not applicable here given the overlap in the SNP-exposure and SNP-outcome dataset (both UK Biobank); the extend of bias introduced by one sample non-linear MR (which was shown to be limited for the more classical two-sample MR methodologies [43]) is currently unknown.

In conclusion, from clustering analyses, we found that genetically determined IGF-1 was associated with both a higher and a lower risk of development of type 2 diabetes, which is likely mediated by distinct biological mechanisms. Therefore, the total concentration of IGF-1 does not provide insight into the risk of developing type 2 diabetes.

References

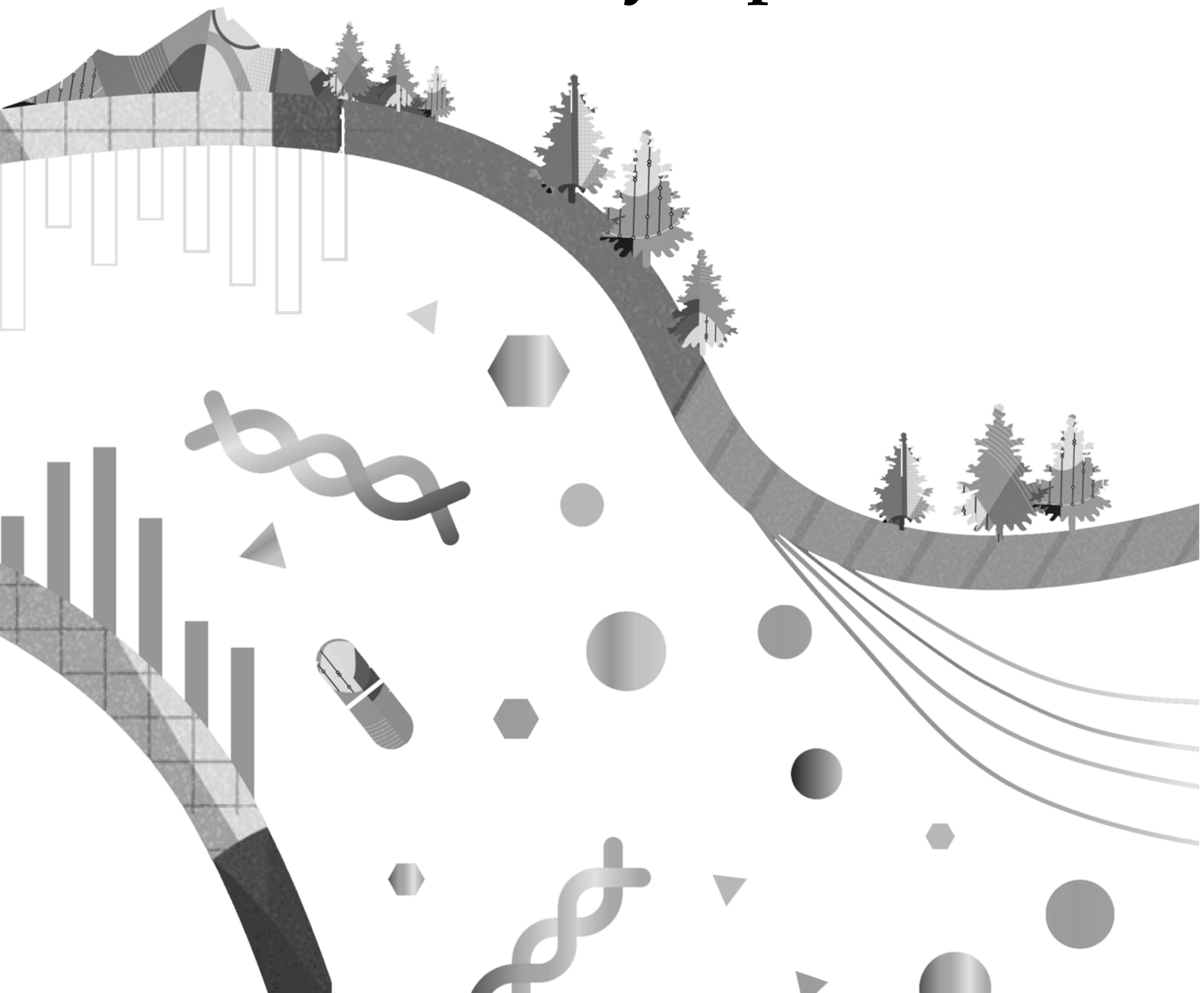
[1]Rajpathak SN, Gunter MJ, Wylie-Rosett J, et al. (2009) The role of insulin-like growth factor-I and its binding proteins in glucose homeostasis and type 2 diabetes. Diabetes/Metabolism Research and Reviews 25(1): 3-12. https://doi.org/10.1002/dmrr.919
[2]Nicholls AR, Holt RI (2016) Growth Hormone and Insulin-Like Growth Factor-1. Front Horm Res 47: 101-114. 10.1159/000445173
[3]Aguirre GA, De Ita JR, de la Garza RG, Castilla-Cortazar I (2016) Insulin-like growth factor-1 deficiency and metabolic syndrome. J Transl Med 14: 3-3. 10.1186/s12967-015-0762-z
[4]Poudel SB, Dixit M, Neginskaya M, et al. (2020) Effects of GH/IGF on the Aging Mitochondria. Cells 9(6). 10.3390/cells9061384
[5]Sasako T, Ueki K (2016) [Insulin/IGF-1 signaling and aging]. Nihon Rinsho 74(9): 1435-1440
[6]Rosenzweig SA (2020) The Continuing Evolution of Insulin-like Growth Factor Signaling. F1000Res 9: F1000 Faculty Rev-1205. 10.12688/f1000research.22198.1
[7]Sandhu MS, Heald AH, Gibson JM, Cruickshank JK, Dunger DB, Wareham NJ (2002) Circulating concentrations of insulin-like growth factor-I and development of glucose intolerance: a prospective observational study. Lancet 359(9319): 1740-1745. 10.1016/s0140-6736(02)08655-5
[8]Rajpathak SN, McGinn AP, Strickler HD, et al. (2008) Insulin-like growth factor-(IGF)-axis, inflammation, and glucose intolerance among older adults. Growth Horm IGF Res 18(2): 166-173. 10.1016/j.ghir.2007.08.004
[9]Larsson SC, Michaëlsson K, Burgess S (2020) IGF-1 and cardiometabolic diseases: a Mendelian randomisation study. Diabetologia 63(9): 1775-1782. 10.1007/s00125-020-05190-9
[10]Emdin CA, Khera AV, Kathiresan S (2017) Mendelian Randomization. JAMA 318(19): 1925-1926. 10.1001/jama.2017.17219
[11]Lawlor DA, Tilling K, Davey Smith G (2016) Triangulation in aetiological epidemiology. Int J Epidemiol 45(6): 1866-1886. 10.1093/ije/dyw314
[12]Foley CN, Mason AM, Kirk PDW, Burgess S (2021) MR-Clust: clustering of genetic variants in Mendelian randomization with similar causal estimates. Bioinformatics 37(4): 531-541. 10.1093/bioinformatics/btaa778
[13]Noordam R, Oudt CH, Bos MM, Smit RAJ, van Heemst D (2018) High-sensitivity C-reactive protein, low-grade systemic inflammation and type 2 diabetes mellitus: A two-sample Mendelian randomization study. Nutr Metab Cardiovasc Dis 28(8): 795-802. 10.1016/j.numecd.2018.03.008
[14]Verkouter I, de Mutsert R, Smit RAJ, et al. (2020) The contribution of tissue-grouped BMI-associated gene sets to cardiometabolic-disease risk: a Mendelian randomization study. International journal of epidemiology 49(4): 1246-1256. 10.1093/ije/dyaa070
[15]Foley CN, Mason AM, Kirk PDW, Burgess S (2020) MR-Clust: Clustering of genetic variants in Mendelian randomization with similar causal estimates. Bioinformatics. 10.1093/bioinformatics/btaa778
[16]Sudlow C, Gallacher J, Allen N, et al. (2015) UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 12(3): e1001779. 10.1371/journal.pmed.1001779
[17]Therneau TM (2020) A Package for Survival Analysis in R. Available from https://CRAN.R-project.org/package=survival

[18]Walter K, Min JL, Huang J, et al. (2015) The UK10K project identifies rare variants in health and disease. Nature 526(7571): 82-90. 10.1038/nature14962
[19]Auton A, Brooks LD, Durbin RM, et al. (2015) A global reference for human genetic variation. Nature 526(7571): 68-74. 10.1038/nature15393
[20]McCarthy S, Das S, Kretzschmar W, et al. (2016) A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet 48(10): 1279-1283. 10.1038/ng.3643
[21]Loh PR, Tucker G, Bulik-Sullivan BK, et al. (2015) Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat Genet 47(3): 284-290. 10.1038/ng.3190
[22]Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York
[23]Winkler TW, Kutalik Z, Gorski M, Lottaz C, Kronenberg F, Heid IM (2015) EasyStrata: evaluation and visualization of stratified genome-wide association meta-analysis data. Bioinformatics (Oxford, England) 31(2): 259-261. 10.1093/bioinformatics/btu621
[24]Mahajan A, Taliun D, Thurner M, et al. (2018) Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet 50(11): 1505-1513. 10.1038/s41588-018-0241-6
[25]Hemani G, Zheng J, Elsworth B, et al. (2018) The MR-Base platform supports systematic causal inference across the human phenome. Elife 7. 10.7554/eLife.34408
[26]Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG (2015) Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. Eur J Epidemiol 30(7): 543-552. 10.1007/s10654-015-0011-z
[27]Noordam R, Smit RA, Postmus I, Trompet S, van Heemst D (2016) Assessment of causality between serum gamma-glutamyltransferase and type 2 diabetes mellitus using publicly available data: a Mendelian randomization study. Int J Epidemiol 45(6): 1953-1960. 10.1093/ije/dyw306
[28]Bowden J, Davey Smith G, Burgess S (2015) Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. International Journal of Epidemiology 44(2): 512-525. 10.1093/ije/dyv080
[29]Bowden J, Davey Smith G, Haycock PC, Burgess S (2016) Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol 40(4): 304-314. 10.1002/gepi.21965
[30]Watanabe K, Taskesen E, van Bochoven A, Posthuma D (2017) Functional mapping and annotation of genetic associations with FUMA. Nature Communications 8(1): 1826. 10.1038/s41467-017-01261-5
[31]Kamburov A, Wierling C, Lehrach H, Herwig R (2009) ConsensusPathDB--a database for integrating human functional interaction networks. Nucleic Acids Res 37(Database issue): D623-D628. 10.1093/nar/gkn698
[32]Drogan D, Schulze MB, Boeing H, Pischon T (2016) Insulin-Like Growth Factor 1 and Insulin-Like Growth Factor–Binding Protein 3 in Relation to the Risk of Type 2 Diabetes Mellitus: Results From the EPIC–Potsdam Study. American Journal of Epidemiology 183(6): 553-560. 10.1093/aje/kwv188
[33]Similä ME, Kontto JP, Virtamo J, et al. (2019) Insulin-like growth factor I, binding proteins -1 and -3, risk of type 2 diabetes and macronutrient intakes in men. British Journal of Nutrition 121(8): 938-944. 10.1017/S0007114519000321
[34]Rajpathak SN, He M, Sun Q, et al. (2012) Insulin-like growth factor axis and risk of type 2 diabetes in women. Diabetes 61(9): 2248-2254. 10.2337/db11-1488
[35]Mulugeta A, Hyppönen E, Ala-Korpela M, Mäkinen V-P (2021.02.01) Metabolic subgroups and cardiometabolic multimorbidity in the UK Biobank. medRxiv: 21250893, preprint: not peer reviewed. 10.1101/2021.02.01.21250893
[36]Schneider HJ, Friedrich N, Klotsche J, et al. (2011) Prediction of incident diabetes mellitus by baseline IGF1 levels. Eur J Endocrinol 164(2): 223-229. 10.1530/eje-10-0963

[37]Friedrich N, Thuesen B, Jørgensen T, et al. (2012) The association between IGF-I and insulin resistance: a general population study in Danish adults. Diabetes care 35(4): 768-773. 10.2337/dc11-1833
[38]Larsson SC, Michaëlsson K, Burgess S (2020) IGF-1 and cardiometabolic diseases: a Mendelian randomisation study. Diabetologia. 10.1007/s00125-020-05190-9
[39]Zhang B, Shozu M, Okada M, et al. (2010) Insulin-like growth factor I enhances the expression of aromatase P450 by inhibiting autophagy. Endocrinology 151(10): 4949-4958. 10.1210/en.2010-0294
[40]van Haeften TW, Twickler TB (2004) Insulin-like growth factors and pancreas beta cells. Eur J Clin Invest 34(4): 249-255. 10.1111/j.1365-2362.2004.01337.x
[41]Kido Y, Nakae J, Hribal ML, Xuan S, Efstratiadis A, Accili D (2002) Effects of mutations in the insulin-like growth factor signaling system on embryonic pancreas development and beta-cell compensation to insulin resistance. J Biol Chem 277(39): 36740-36747. 10.1074/jbc.M206314200
[42]Yakar S, Setser J, Zhao H, et al. (2004) Inhibition of growth hormone action improves insulin sensitivity in liver IGF-1-deficient mice. J Clin Invest 113(1): 96-105. 10.1172/JCI17763
[43]Minelli C, Del Greco MF, van der Plaat DA, Bowden J, Sheehan NA, Thompson J (2021) The use of two-sample methods for Mendelian randomization analyses on single large datasets. Int J Epidemiol 50(5): 1651-1659. 10.1093/ije/dyab084
[44]Levine ME, Suarez JA, Brandhorst S, et al. (2014) Low protein intake is associated with a major reduction in IGF-1, cancer, and overall mortality in the 65 and younger but not older population. Cell metabolism 19(3): 407-417. 10.1016/j.cmet.2014.02.006

Part III

Pathophysiology of Sleep-Associated Dyslipidemia



Chapter 5

A Large-Scale Genome-Wide Gene-Sleep Interaction Study in 732,564 Participants Identifies Lipid Loci Explaining Sleep-Associated Lipid Disturbances

Raymond Noordam, Wenyi Wang, Pavithra Nagarajan, Heming Wang, Michael R Brown, Amy R Bentley, Aldi T Kraja, John L Morrison, Jeffrey R O'Connel, Songmi Lee, Karen Schwander, Yan V Sun, Traci M Bartz, Lisa de les Fuentes, Mary F Feitosa, Xiuqing Guo, Xu Hanfei, Sarah E Harris, Zhijie Huang, Qin Hui, Mart Kals, Christophe Lefevre, Massimo Mangino, Yuri Milaneschi, Peter van der Most, Natasha L Pacheco, Nicholette D Palmer, Varun Rao, Rainer Rauramaa, Quan Sun, Dina Vojinovic, Yujie Wang, Zhao Wei, Stefan Weiss, Qian Yang, Tabara Yasuharu, Wangying Zhu, Md Abu Yusuf Ansari, Huagues Aschard, Pramod Anugu, Themistocles L Assimes, John Attia, Laura D Baker, Christie Ballantyne, Lydia Bazzano, Eric Boerwinkle, Brain Cade, Hung-hsin Chen, Wei Chen, Yii-Der Ida Chen, Zekai Chen, Kelly Cho, Ileana De Anda-Duran, Latchezar Dimitrov, Anh Do, Todd Edwards, Tariq Faquih, Aroon Hingorani, Susan P Fisher-Hoch, J. Michael Gaziano, Sina A Gharib, Ayush Giri, Mohsen Ghanbari, Hans-Jörgen Grabe, Mariaelisa Graff, C Charles Gu, Jiang He, Sami Heikkinen, James Hixson, Yuk-Lam Ho, Michelle M Hood, Serena C Houghton, Carrie A Karvonen-Gutierrez, Takahisa Kawaguchi, Tuomas O Kilpeläinen, Pirjo Komulainen, Henry J Lin, Gregorio V Linchangco Jr., Annemarie I Luik, Jintao Ma, James B Meigs, Joseph B McCormick, Cristina Menni, Ilja M Nolte, Jill M Norris, Lauren E Petty, Hannah G Polikowsky, Laura M Raffield, Stephen S Rich, Renate L Riha, Thomas C Russ, Edward A Ruiz-Narvaez, Colleen M Sitlani, Jennifer A Smith, Harold Snieder, Tamar Sofer, Botong Shen, Jingxian Tang, Kent D Taylor, Maris Teder-Laving, Rima Triatin, Michael Y Tsai, Henry Völzke, Kenneth E. Westerman, Rui Xia, Jie Yao, Kristin L Young, Ruiyuan Zhang, Alan B Zonderman, Xiaofeng Zhu, Jennifer E Below, Simon R Cox, Michelle Evans, Myriam Fornage, Ervin R Fox, Nora Franceschini, Sioban D Harlow, Elizabeth Holliday, M. Arfan Ikram, Tanika Kelly, Timo A Lakka, Deborah A Lawlor, Changwei Li, Ching-Ti Liu, Reedik Mägi, Alisa K Manning, Fumihiko Matsuda, Alanna C Morrison, Matthias Nauck, Kari E North, Brenda WJH Penninx, Michael A Province, Bruce M Psaty, Jerome I Rotter, Tim D Spector, Lynne E Wagenknecht, Ko Willems van Dijk, Lifelines Cohort Study, Cashell E Jaquish, Peter WF Wilso, Patricia A Peyser, Patricia B Munroe, Paul S de Vries, W James Gauderman, Han Chen, Clint L Miller, Thomas W Winkler, Dabeeru C Rao, Susan Redline, Diana van Heemst

Under submission

The supplemental information for this paper is available online at
https://drive.google.com/drive/folders/1UoSstJZtu2ylarbtkhn6Kmr3rz68hhSu?usp=sharing

Abstract

We performed large-scale genome-wide gene-sleep interaction analyses on lipid levels to identify novel genetic variants underpinning the biomolecular pathways of sleep-associated lipid disturbances and to suggest possible druggable targets. We collected data from 55 cohorts with a combined sample size of 732,564 participants (87% European ancestry) with data on lipid traits (high-density lipoprotein [HDL-c] and low-density lipoprotein [LDL-c] cholesterol and triglycerides [TG]). Short (STST) and long (LTST) total sleep time were defined by the extreme 20% of the age- and sex-standardized values within each cohort. Based on cohort-level summary statistics data, we performed meta-analyses for the one-degree of freedom tests of interaction and two-degree of freedom joint tests of the main and interaction effect. In the cross-population meta-analyses, the one-degree of freedom variant-sleep interaction test identified 10 loci (Pint<5.0e-9) not previously observed for lipids. Of interest, the ASPH locus (TG, LTST) is a target for aspartic and succinic acid metabolism previously shown to improve sleep and cardiovascular risk. The two-degree of freedom analyses identified an additional 7 loci that showed evidence for variant-sleep interaction (Pjoint<5.0e-9 in combination with Pint<6.6e-6). Of these, the SLC8A1 locus (TG, STST) has been considered a potential treatment target for reduction of ischemic damage after acute myocardial infarction. Collectively, the 17 (9 with STST; 8 with LTST) loci identified in this large-scale initiative provides evidence into the biomolecular mechanisms underpinning sleep-duration-associated changes in lipid levels. The identified druggable targets may contribute to the development of novel therapies for dyslipidemia in people with sleep disturbances.

Introduction

Serum levels of high-density lipoprotein cholesterol (HDL-c), low-density lipoprotein cholesterol (LDL-c) and triglycerides (TG) are well-characterized risk factors for atherosclerotic cardiovascular disease [1-4]. High LDL-c and TG concentrations have also been shown to causally impact atherosclerotic cardiovascular disease onset [5, 6]. Serum lipid levels are influenced by both environmental and genetic factors [7], and large-scale efforts have identified hundreds of loci associated with increased lipid levels [8-15].

Sleep disturbances are increasingly recognized as important modifiable risk factors for various metabolic diseases including atherosclerotic cardiovascular disease and type 2 diabetes [16, 17]. In 2022, sleep duration was added to the Life's Essentials by the American Heart Association, highlighting the recognition of sleep duration as being important in cardiovascular prevention [18]. Both short and long self-reported habitual sleep duration have been associated with adverse (atherogenic) lipid profiles in epidemiological cohort studies [19-23], and recent Mendelian Randomization studies suggest that both short and long habitual sleep durations as potential causal risk factors for atherogenic cardiovascular disease [24-26]. However, despite these findings, the biomolecular mechanisms underpinning sleep-associated atherogenic cardiovascular disease risk are still poorly understood. Examining gene-lifestyle interactions can be an important tool to identify additional genetic variants associated with the trait of interest as well as provide insights into the biomolecular mechanisms underpinning the trait-outcome association [27, 28]. In previously conducted gene-lifestyle interaction projects performed within the Cohort for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium [29, 30] Gene-Lifestyle Working Group [27], we identified multiple loci interacting with lifestyle exposures to lipid levels [31-34]. In particular, we performed a meta-analysis of 126,926 (predominantly European-ancestry; 20% of the participants defined as having either short or long sleep duration) individuals, which identified multiple loci associated with lipid profiles in the context of short and long sleep duration. Our results particularly suggested that the effect of long sleep duration and short sleep duration may modify lipid profiles through distinct biological pathways [31].

In recent years, more data has become available from large biobank initiatives (i.e., UK Biobank and the Million Veteran Program [35, 36]). These data provide an opportunity to increase the sample size substantially in a

more diverse study population to allow improved statistical power for the detection of additional gene-by-sleep duration interactions on serum lipid levels. Ultimately, such efforts can improve our understanding of the biomolecular mechanisms underpinning sleep-associated lipid disturbances. Here, we conducted a new and updated multi-population sleep duration-by-gene interaction study on lipid profiles in 732,564 participants from five population groups (African [AFR], East Asian [EAS], European [EUR], Hispanic/Latino [HIS] and South Asian [SAS]).

Methods

Overall study design

The study was designed to include cohorts that collected questionnaire-based data on habitual total sleep time and measured blood lipids levels (TG, LDL-c and/or HDL-c). Genome-wide gene×sleep interaction analyses were performed separately by each participating study (and separately for each population group: (AFR, EAS, EUR, HIS, and SAS) following a standardized analysis protocol. Participants 18 years and older were included if they reported a total sleep time between 3 and 14 hours. For studies having habitual total sleep time and lipid levels collected at multiple rounds of visits, the visit with the largest sample size was selected for analysis. Statistical analyses were performed for men and women combined as well as separately for men and women to observe potential effect modification of the variant-sleep interaction effect by sex. Data were subsequently aggregated centrally for quality control and meta-analyses. When applicable, the analysis protocol was reviewed and approved by institutional review boards. Each contributing study was approved by local medical ethics committees and each participant provided written informed consent, in line with the declaration of Helsinki. More information on the individual cohorts is presented in the **Online Supplement**.

Harmonization of Exposure Variables

Data on habitual total sleep time were collected through questionnaires using questions like "On an average day, how long do you sleep?" to calculate short total sleep time (STST) and long total sleep time (LTST). STST and LTST were derived by regressing sleep time on age, sex, and age×sex, or as indicated otherwise (**Supplementary Table 2**). The derived residuals' 20th and 80th percentiles were used as cutoffs: STST=1 if ≤ 20th percentile (otherwise "0"); LTST=1 if ≥ 80th percentile (otherwise "0").

Harmonization of Outcome Variables

We considered 3 lipid traits as outcome variables: LDL-c, HDL-c and TG. For most cohorts, fasting (≥8 hours) LDL-c and TG were used. In UK Biobank (N = 359,962 for the combined sample; 49.1% of the total sample) participants

were not asked to fast prior to blood samples, and therefore the vast majority (>90%) had no ≥8 hours fasting time. For LDL-c and TG, analyses in UK Biobank were done separately for those meeting the fasting criteria and those who did not, and considered as separate cohorts in subsequent meta-analyses. LDL-c was either directly assayed or derived using the Friedewald equation (the latter restricted to those with TG ≤ 400 mg/dL)[37]. LDL-c was corrected 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 clinical practice). If LDL-c was directly assayed, the concentration of LDL-c was corrected by dividing the LDL-c concentration by 0.7. Otherwise (i.e. if LDL-c was derived using the Friedewald equation), then we first divided the concentration of total cholesterol by 0.8 before LDL-c calculationDue to the skewed distribution of HDL-c and TG, we natural log-transformed the concentration prior to the analyses. No transformation for LDL-c was required. All lipid levels were winsorized at 6 standard deviations from the (transformed) mean.

Individual cohort data analyses

Genotype data were restricted to autosomal chromosomes, imputation quality R^2 ≥0.3 and minor allele frequency (MAF) ≥0.05% (**Supplementary Table 1**). After data harmonization, each population-group specific cohort ran 2 regression models for 18 phenotype-exposure-sex combinations (3 phenotypes x 2 exposures x All/Men/Women). Below E denotes the sleep exposure (STST or LTST), Y denotes the lipid trait (LDL-c, HDL-c, TG), and C denotes the vector of covariates mentioned above specific to E. Analyses were preferably conducted by each cohort using either of the three software : LinGxEScanR v1.0 (https://github.com/USCbiostats/LinGxEScanR), GEM v1.4.1 (https://github.com/large-scale-gxe-methods/GEM), or MMAP (https://github.com/MMAP/MMAP.github.io) with robust standard errors (SEs) enforced [38] (**Supplementary Table 1**). In cohorts with related participants, null model residuals (regressing lipid traits on a kinship matrix/genetic covariance matrix) were formulated as the lipid outcome.

The two regression models performed included one-degree of freedom (df) tests for examining the variant-sleep interaction effects, and the two-df-joint test that simultaneously assesses the variant-main and variant-sleep interaction effects [39]. Covariates included population specific principal components of the genotype matrix, cohort-specific confounders (e.g., study center), age, age^2, sex, age×S/LTST, age^2×S/LTST, and sex×S/LTST. Finally,

for a fair comparison of our results with the existent (e.g., standard genome-wide association model for comparison [15]), we also conducted a standard marginal genetic effect model without the consideration of STST or LTST within the same study sample as the variant-interaction analyses.

Centralized cohort-level and meta-level quality control

Cohort-level summary statistics were processed centrally. For quality control (QC), we used the EasyQC2 software (www.genepi-regensburg.de/easyqc2) package in R [40]. Data were filtered for degrees of freedom ≥20 calculated as minor allele count * imputation quality within the unexposed, the exposed, and the total sample. When required, hg37 genomic coordinates were lifted over to hg38 genomic coordinates. Allele frequency discrepancies relative to population-matched TOPMed-imputed 1000G reference panels (Trans-Omics for Precision Medicine imputed 1000Genomes) were assessed, along with genomic control (GC) lambda inflation. Next, meta-level quality control was conducted within population groups (AFR: 13 cohorts, EAS: 5 cohorts, EUR: 30 cohorts, HIS: 7 cohorts, SAS: 1 cohort), with the evaluation of the improper transformation of the outcome variables, unstable numerical computation, or alarming inflation.

Meta-analyses

Meta-analyses were performed for each population group separately and further combined in a cross-population meta-analyses (CPMA). This resulted in a total of 18 meta-analyses per combination of sleep exposure and lipid trait: five population groups (EUR, HIS, EAS, AFR, SAS) and CPMA, and 3 sex groups (all, women, men). Four tests were considered: the marginal genetic effect (B_{M2_G}), the main genetic effect from the interaction model (B_{M1_G}), the interaction effect (B_{M1_GxE}), and the joint main and interaction effects (B_{M1_G}, B_{M1_GxE}) with cohort-level GC correction to correct for possible inflation [41]. METAL software for meta-analysis with inverse-variance weights [28] was used to combine evidence across studies for each of the four tests. CPMA was subsequently executed on the resultant population- specific METAL output results, with population-level GC correction. Due to the low numbers of participants contributing to the HIS, EAS and SAS analyses, only results derived from the CPMA, EUR, and AFR analyses were considered for further interpretation. Furthermore, we only considered variants analyzed in at least 40,000 participants in the main analysis for discovery.

Identification of independent genomic loci

We used EasyStrata2 software in R to prioritize top loci from significant results identified in the one-df interaction and two-df joint tests [42]. We excluded variants within 1 Mb distance of the major histocompatibility complex (MHC) region. Significant variants were identified using the threshold criteria detailed below. (1) Variants with significant one-df interaction effect (P_{int}<5 x 10^{-9}, FDR<0.05) and (2) variants with significant two-df joint effect (P_{joint}<5 x10^{-9} with FDR<0.05) were selected as top variants. To prioritize lead variants from the 2-df joint analysis with evidence for having variant-sleep interaction, we evaluated the 2-df joint lead variants for 1-df interactions and used a Bonferroni correction for the number of 2-df joint variants identified in the respective population-specific group (CPMA, EUR, AFR) [43]. Note that the 2-df joint test and the 1 df interaction effect tests are correlated, so the latter procedure does not offer formal statistical evidence of interaction. Nevertheless, it provides a fast and easy prioritization of variants most likely to be involved in interaction with the sleep variables. All such variants were narrowed down to loci based on a 250 kB distance. Finally, within these regions, independent loci were identified by linkage disequilibrium (LD) r^2 threshold <0.1 using TOPMed-imputed 1000G reference panels. If variants were missing in the LD panels, then the most significant variant within each 500kb region was retained. From the lead variants identified, we additionally extracted the variant information from the sex-stratified analyses to test for heterogeneity of the interaction effects by sex. The heterogeneity of the variant-sleep interaction effect between men and women was tested by performing two-sample Z-tests assuming independence, which were conducted for each interaction loci in the meta-analysis of men and women combined [44].

Gene mapping, functional annotation, and follow-up phenotypic annotations

For the lead variants identified, variant mapping was primarily performed using Functional Mapping and Annotation of Genome-wide Association Studies v1.6.0 (FUMA) [45], and Locuszoom (https://my.locuszoom.org) [46, 47]. At the genomic region level, FUMA's SNP2GENE pipeline was used to annotate a comprehensive list of genes for each top locus, incorporating genomic position, chromatin interaction (FDR <=1 x 10^{-6}, 250bp upstream - 500 bp downstream of the transcription startsite [TSS]), and GTEXv8 eQTL

evidence with the top variant or its variants in LD (r^2>0.1 within 500kb) [45, 48]. At the variant level, PheWeb and Open Target Genetics were queried for significant trait associations ($p<5 \times 10^{-8}$) from past GWAS analyses [49, 50]. At the gene level, we explored the International Mouse Phenotyping Consortium release 19.1 (IMPC), Online Mendelian Inheritance in Man (OMIM; https://omim.org/), PheWeb, Phenotype-Genotype Integrator (PheGenI), Open Target Genetics, and the online drugbank for retrieving information on the genes as potential drug targets (https://go.drugbank.com) [49-52]. All identified mapped protein-coding genes were then queried using FUMA's GENE2FUNC pipeline to identify significant (adjusted p-value<0.05) pathways and traits [45].

Druggability analysis

We investigated the potential druggability of the sleep duration-lipid trait candidate interacting gene targets as previously described [53]. In short, we first used the Drug-Gene Interaction database (DGIdb; v4.2.0) to query high or medium priority sleep-lipid interacting genes to determine the potential druggability of the candidate gene targets. We annotated genes for implicated pathways and functions using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. We annotated the druggability target categories and queried all interacting drugs reported in 43 databases (BaderLabGenes, CarisMolecularIntelligence, dGene, FoundationOneGenes, GO, HingoraniCasas, HopkinsGroom, HumanProteinAtlas, IDG, MskImpact, Oncomine, Pharos, RussLampel, Tempus, CGI, CIViC, COSMIC, CancerCommons, ChemblDrugs, ChemblInteractions, ClearityFoundationBiomarkers, ClearityFoundationClinicalTrial, DTC, DoCM, DrugBank, Ensembl, Entrez, FDA, GuideToPharmacology, JACX-CKB, MyCancerGenome, MyCancerGenomeClinicalTrial, NCI, OncoKB, PharmGKB, TALC, TEND, TTD, TdgClinicalTrial, Wikidata). We queried protein targets for available active ligands in ChEMBL. We queried gene targets in the druggable genome using the most recent druggable genome list established from the NIH Illuminating the Druggable Genome Project (https://github.com/druggablegenome/IDGTargets) available through the Pharos web platform (https://pharos.nih.gov/targets). We also queried FDA-approved drugs, late-stage clinical trials and disease indications in the DrugBank, ChEMBL, and ClinicalTrials.gov databases. We provided results for the top MESH and DrugBank indications and clinical trials.

Results

Study overview

Data from 55 cohorts including five population groups were included: AFR (13 cohorts, N=48,851 [7%]), EAS (4 cohorts, N=8,097 [1%]), EUR (30 cohorts, N=637,166 [87%]), HIS (7 cohorts, N=32,508 [4%]), and SAS (1 cohort, N=7,619 [1%]). The total sample size was 732,564 participants in the CPMA with 149,210 participants as STST and 147,603 participants as LTST. Additional information on the characteristics of each of study sample is presented in **Supplementary Tables 1-3**.

Findings from the one-df variant-sleep interaction analyses

One-df interaction CPMA identified 9 loci displaying evidence for genetic associations with the lipid traits modified by either STST or LTST ($P_{int} < 5 \times 10^{-9}$ in combination with an FDR < 0.05) (**Figure 1**; **Table 1; Supplementary Figures 1-3** for -log(P_{int}) and QQ plots). Of these, we identified 4 variants for TG, 2 variants for LDL-c and 3 variants for HDL-c. These variants have not been observed before in studies on lipid levels (i.e., [15]) nor did we find evience of potential variant main effects in the same study sample (**Supplementary Table 4**). Of the lead variants identified, the 13:50374420:C_T locus (rs14172636; Minor Allele Frequency [MAF] = 0.0087), mapped to the *DLEU1* gene, interacted with STST in its association with both TG ($P_{int} = 2.40 \times 10^{-16}$) and HDL-c ($P_{int} = 4.10 \times 10^{-12}$). To illustrate, in the case of the finding for TG, the effect of the rs14172636-C allele on log-transformed TG level was 0.26 units lower (equivalent to an approximate additive decrease of 22.9%) in those reporting STST, and on HLD-c level was 0.132 units higher (equivalent to an approximate additive increase of 14.1%) in those reporting LTST. We did not find evidence for a 1-df STST-*DLEU1* interaction effect on LDL-c (P_{int} = 0.07). The 8:61617696:C_T locus (rs147261056; MAF: 0.0048), mapped to *ASPH* and *CLVS1*, interacted with LTST in its association with TG ($P_{int} = 2.78 \times 10^{-13}$). The 11:10411707:C_CT locus (rs1847639939; MAF: 0.46), mapped to the *AMPD3* gene, interacted with LTST in its association with LDL-c ($P_{int} = 4.72 \times 10^{-9}$). Other variants identified in the CPMA included 3:162278901A_T (rs162278901; *OTOL1*, LDL-c with LTST; $P_{int} = 2.78 \times 10^{-13}$), 7:72156448:A_G (rs573762901; *CALN1*, HDL with LTST; $P_{int} = 1.43 \times 10^{-10}$), 2:186808058:G_T (rs6760240; *ZSWIM2*, TG with STST; $P_{int} = 1.47 \times 10^{-9}$), 7:102460277:G_T (rs543672875; *ALKBH4*, HDL-c

with STST; P_{int} = 1.51 x 10^{-9}) and 2:184828292:C_T (rs190975828; *ZNF804A*, LDL-c with LTST; P_{int} = 4.72 x 10^{-9}).

One additional variant was identified in the European meta-analysis only. The variant 4:12768773:C_G (rs192018195; P_{int} = 4.81 x 10^{-11}, MAF = 0.0151) mapping to *INTU/SLC25A31/HSPA4L* was identified in the STST analysis on TG, and was just outside the significance boundaries in the CPMA (P_{int} = 5.03 x 10^{-9}). Especially the more rare variants identified in these efforts were unable to be investigated further in the population-specific subgroup analyses as variants did not pass post-meta-analysis QC (**Figure 2**). Of the remaining variants, we only found evidence rs1847639939 was associated with LDL-c in the EUR sample (P_{int} = 1.61 x 10^{-8}), and not in the AFR meta-analysis (P_{int} = 0.74) (**Figure 2**).

An extensive summary of the primary results, including reporting of the results in the sex-specific and population-specific analyses when passing post meta-analysis QC, are presented in **Supplementary Table 4**; additional information on the region of the identified variants is presented in regional plots presented in **Supplementary Figure 4**. With the exception of the lead variants mapped to *ASPH* and *DLEU1*, none were functional variants. No additional variants were identified in the sex-stratified analyses nor did we observe evidence for sex differences ($P_{sex\text{-}Int} > 0.05$) for variants identified with the 1df interaction test.

Loci identified through the two-df variant-sleep interaction meta-analyses

Additional analyses were performed to prioritize potential variant-sleep interactions identified by the 2-degree of freedom of main and interaction effect meta-analyses. In the 2-df CPMA (**Supplementary Table 4 and 5; Supplementary Figure 5**), we identified ($P_{2df} < 5 \times 10^{-9}$ and FDR < 0.05) a total of 1,190 lead variants for the TG-LTST analysis (covering 371 genomic regions), 1,156 lead variants for the TG-STST analysis (covering 312 genomic regions), 1,185 lead variants for the HDL-c-LTST analyses (covering 362 genomic regions), 1,178 lead variants for the HDL-c-STST analyses (covering 358 genomic regions), 1,433 lead variants for the LDL-c-LTST analyses (covering 264 genomic regions), and 1,431 lead variants for the LDL-c-STST analyses.

These lead variants were then tested for 1-df interaction. Here, we used a less stringent *P*-value cut off for 1-df interactions based on the total number of

lead variants identified in the CPMA sample for the three traits and two exposure groups combined (P_{int} < 6.60 x 10^{-6} = 0.05/7,573, Bonferroni-corrected, see **Methods**). Through this process we identified seven additional genetic lead variants showing evidence for 1-df interaction (**Table 2**); of these, five variants were identified for TG (one with LTST, four for STST), one variant for HDL-c (for LTST), and one variant for LDL-c (for LTST) not previously identified for lipid levels nor associated with the lipid trait in the model when not incorporating sleep duration in the same study sample (**Table 2 and Supplementary Table 4 and 5** and for detailed information). Regional plots of the 1-df interaction results of these variants are presented in **Supplementary Figure 8**. In particular, we identified 20:51830403:A_G (rs150607032; *ATP9A*/*NFATC2*/*SALL4*, TG with STST, Pint = 3.59 x 10^{-8}), 11.13058160:C_T (rs59374498; *TEAD1*/*RASSF10*, TG with LTST, P_{int} = 5.71 x 10^{-8}), 10:97769146:A_G (rs191757273; *PYROXD2*, LDL-c with LTST, P_{int} = 7.41 x 10^{-8}), 21:35272725:A_T (rs114083565; *RUNX1*, TG with STST, P_{int} = 8.40 x 10^{-7}), 18:55378517:A_T (rs9949541; *TCF4*, HDL-c with LTST), 2:40094191:A_T (rs34771893; *SLC8A1*, TG with STST, P_{int} = 4.12 x 10^{-6}), and 20:23353740:A_G (rs73319497; *GZF1*/*NPAB*/*CASTL1*/*CAST11*/*NXT1*, TG with STST, P_{int} = 4.47 x 10^{-6}). No evidence was observed that the interaction terms differed for men and women (sex-difference $P_{sex\text{-}Int}$ > 0.05) (**Supplementary Table 4**). We identified no additional variants among the 2-df joint findings showing evidence for 1-df interaction (P_{int} > 1.10 x 10^{-5} and >1.36 x 10^{-4}, respectively; **Supplementary Table 6 and 7 and Supplementary Figures 6 and 7**).

5

Follow-up analyses

Based on the findings identified in the TG-STST analyses (the lipid-sleep combination with most identified variants in the one-df and two-df interaction analyses), and using the GTEx v8 databases, we did not observe evidence for eQTLs enrichement in any particular tissue (P >0.05). Some evidence (p-value = 0.01) was found for enrichment of the Vitamin D receptor pathway (based on the *SLC8A1*, *NFATC2* and *SALL4* genes; based on Wikipathways using the GENE2FUNC in FUMA [45]) (**Supplementary Figure 9**). No evidence for tissue and pathway enrichment was observed for the other loci identified in the exposure-trait combinations.

Druggability analyses

We first queried mapped gene targets from the different analyses using the Drug-Gene Interaction database (DGIdb), which identified seven genes annotated as clinically actionable or members of the druggable genome (Supplemental Table 8a). Several of these gene targets are implicated in calcium signaling (*SLC25A31, SLC8A1, ASPH*), amino acid or purine metabolism (*PYROXD2, AMPD3*), and regulation of gene transcription (*TEAD1, NFATC2, RUNX1*). We identified seven gene targets of FDA-approved drugs evaluated in late-stage clinical trials using DrugBank and ClinicalTrials.gov databases (**Supplemental Table 8b**). *SLC8A1* is a target of the nutraceutical Icosapent (a modified version of omega-3 fatty acid ethyl eicosapentaenoic acid (EPA)), which is used to treat patients with hypertriglyceridemia. *SLC8A1* is also a target of the small molecule inhibitor caldaret, which was investigated for preventing acute myocardial infarction and treating patients with congestive heart failure. *SLC8A1* is also a target of FDA-approved antiarrhythmic Dronedarone (Multaq) to treat patients with atrial fibrillation or atrial flutter. We also identified *SLC25A31*, *ASPH*, and *PYROXD2* as targets of commonly prescribed drugs: beta-blocker Metoprolol, anticoagulant Warfarin, and the attention deficit hyperactivity disorder (ADHD) drug Methylphenidate, respectively, all conditions frequently observed in people with sleep disorders [54-56].

Table 1: Nine variants identified through the 1 degree of freedom interaction analyses in the meta-analyses of men and women combined

Variant	RSid	Effect allele	Exposure	Trait	EAF	Sample Size	Sample	Mapped gene	Interaction Beta	SE	1df p-value
2:184828292:C_T	rs190975828	C	STST	TG	0.9909	557,910	CPMA	*ZNF804A*	-0.102	0.0171	2.99E-09
2:186808058:G_T	rs6760240	G	STST	TG	0.0075	188,049	CPMA	*ZSWIM2*	0.184	0.0304	1.47E-09
3:162278901:A_T	rs162278901	A	LTST	LDL-c	0.0062	41,379	CPMA	*OTOL1*[1]	25.360	3.5968	1.78E-12
4:127678773:C_G	rs192018195	C	LTST	TG	0.9849	208,087	EUR	*INTU, SLC25A31, HSPA4L*	-0.0956	0.0145	4.81E-11
7:72156448:A_G	rs573762901	A	LTST	HDL-c	0.0033	42,445	CPMA	*CALN1*	0.148	0.0230	1.43E-10
7:102460277:G_T	rs543672875	G	STST	HDL-c	0.0024	42,445	CPMA	*ALKBH4*	0.140	0.0231	1.51E-09
8:61617696:C_T	rs147261056	T	LTST	TG	0.0071	212,110	CPMA	*CLVS1, ASPH*	0.185	0.0253	2.78E-13
11:10411707:C_CT	rs1847639939	C	LTST	LDL-c	0.4588	654,182	CPMA	*AMPD3*	0.934	0.1594	4.72E-09
13:50374420:C_T	rs14172636	C	STST	HDL-c	0.9919	196,379	CPMA	*DLEU1*	0.132	0.0190	4.10E-12
13:50374420:C_T	rs14172636	C	STST	TG	0.9913	188,528	CPMA	*DLEU1*	-0.266	0.0325	2.40E-16

Abbreviations: CPMA, Cross Population Meta-Analysis; EAF, Effect Allele Frequency; EUR, European; HDL-c, High-Density Lipoprotein Cholesterol; LDL-c, LowDensity Lipoprotein Cholesterol; SE, standard error; TG, Triglycerides. 1) Identified variant not found in FUMA or in Locuszoom, *OTOL1* gene is closest gene.

Table 2: Additional 7 variants identified through the 2 degree of freedom interaction analyses after prioritization for joint effects in the meta-analyses of men and women combined												
Variant	RSid	Exposure	Effect allele	Trait	EAF	Sample Size	Sample	Mapped gene	2df joint p-value	Interaction Beta	SE	1df interaction p-value
20:51830403:A_G	rs150607032	STST	A	TG	0.0066	532,172	CPMA	*APT9A, NFATC2, SALL4*	1.77E-17	0.098	0.0178	3.59E-08
11:13058160:C_T	rs59374498	LTST	C	TG	0.9752	661,725	CPMA	*TEAD1, RASSF10*	3.34E-38	-0.0358	0.0066	5.71E-08
10:97769146:A_G	rs191757273	LTST	A	LDL-c	0.0023	52,159	CPMA	*PYROXD2*	8.51E-16	-23.9315	4.4475	7.41E-08
21:35272725:A_T	rs114083564	STST	A	TG	0.9855	43,202	CPMA	*RUNX1*	4.45E-29	0.1284	0.0261	8.40E-07
18:55378517:A_T	rs9949541	LTST	A	HDL-c	0.9748	71,290	CPMA	*TCF4*	2.64E-73	0.0478	0.01	1.92E-06
2:40094191:A_T	rs34771893	STST	A	TG	0.0055	557,910	CPMA	*SLC8A1*	1.97E-18	0.1018	0.0221	4.12E-06
20:23353740:A_G	rs73319497	STST	A	TG	0.9781	666,234	CPMA	*GZF1, NAPB, CASTL1, CST11, NXT1*	4.31E-32	-0.036	0.0078	4.47E-06

Abbreviations: CPMA, Cross Population Meta-analysis; EAF, Effect Allele Frequency; HDL-c, High-Density Lipoprotein Cholesterol; LDL-c, Low-Density Lipoprotein Cholesterol; SE, standard error; TG, Triglycerides

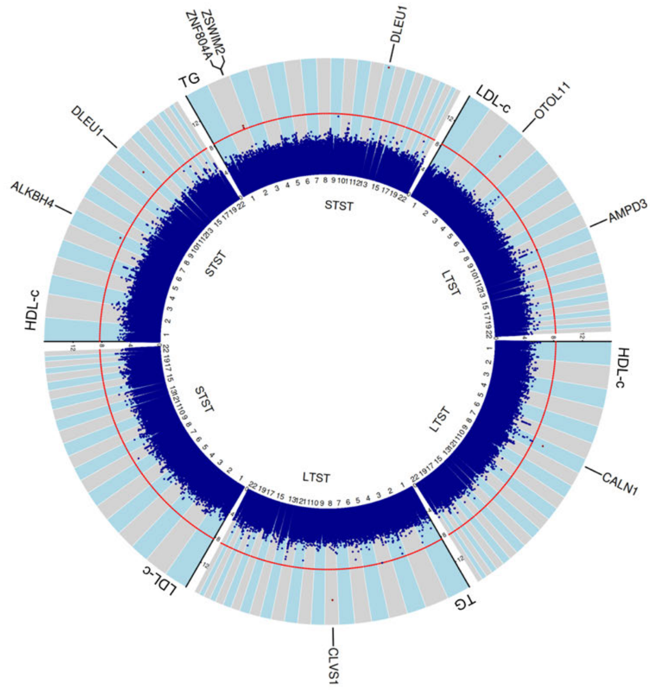
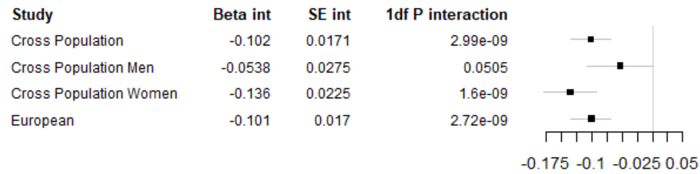
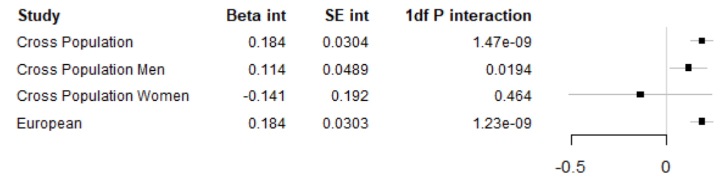
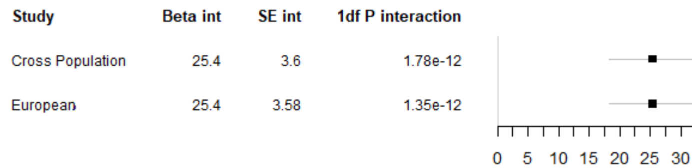
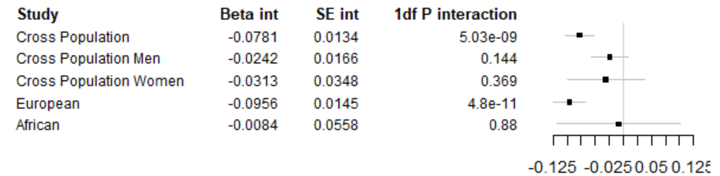


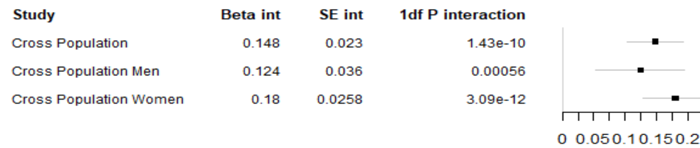
Figure 1: **Circular -log10(Pint) plot of all the 6 main analyses in the cross-population meta-analysis of men and women combined.** ASPH (TG and LTST) maps also at the *CLVS1* locus.

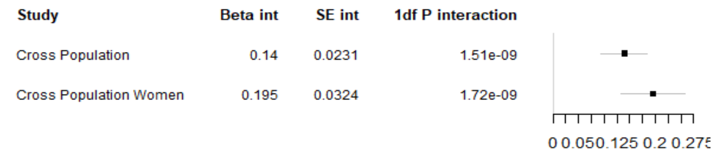

2:184828292:C_T (STST and TG)
Study
Beta int
SE int
1df P interaction
Cross Population -0.102 0.0171 2.99e-09
Cross Population Men -0.0538 0.0275 0.0505
Cross Population Women -0.136 0.0225 1.6e-09
European -0.101 0.017 2.72e-09
-0.175 -0.1 -0.025 0.05


2:186808058:G_T (STST and TG)
Study
Beta int
SE int
1df P interaction
Cross Population 0.184 0.0304 1.47e-09
Cross Population Men 0.114 0.0489 0.0194
Cross Population Women -0.141 0.192 0.464
European 0.184 0.0303 1.23e-09
-0.5
0


3:162278901:A_T (LTST and LDL-c)
Study
Beta int
SE int
1df P interaction
Cross Population 25.4 3.6 1.78e-12
European 25.4 3.58 1.35e-12
0 5 10 15 20 25 30


4:127678773:C_G (LTST and TG)
Study
Beta int
SE int
1df P interaction
Cross Population -0.0781 0.0134 5.03e-09
Cross Population Men -0.0242 0.0166 0.144
Cross Population Women -0.0313 0.0348 0.369
European -0.0956 0.0145 4.8e-11
African -0.0084 0.0558 0.88
-0.125 -0.025 0.05 0.125


7:72156448:A_G (LTST and HDL-c)
Study
Beta int
SE int
1df P interaction
Cross Population 0.148 0.023 1.43e-10
Cross Population Men 0.124 0.036 0.00056
Cross Population Women 0.18 0.0258 3.09e-12
0 0.05 0.1 0.15 0.2


7:102460277:G_T (STST and HDL-c)
Study
Beta int
SE int
1df P interaction
Cross Population 0.14 0.0231 1.51e-09
Cross Population Women 0.195 0.0324 1.72e-09
0 0.05 0.125 0.2 0.275

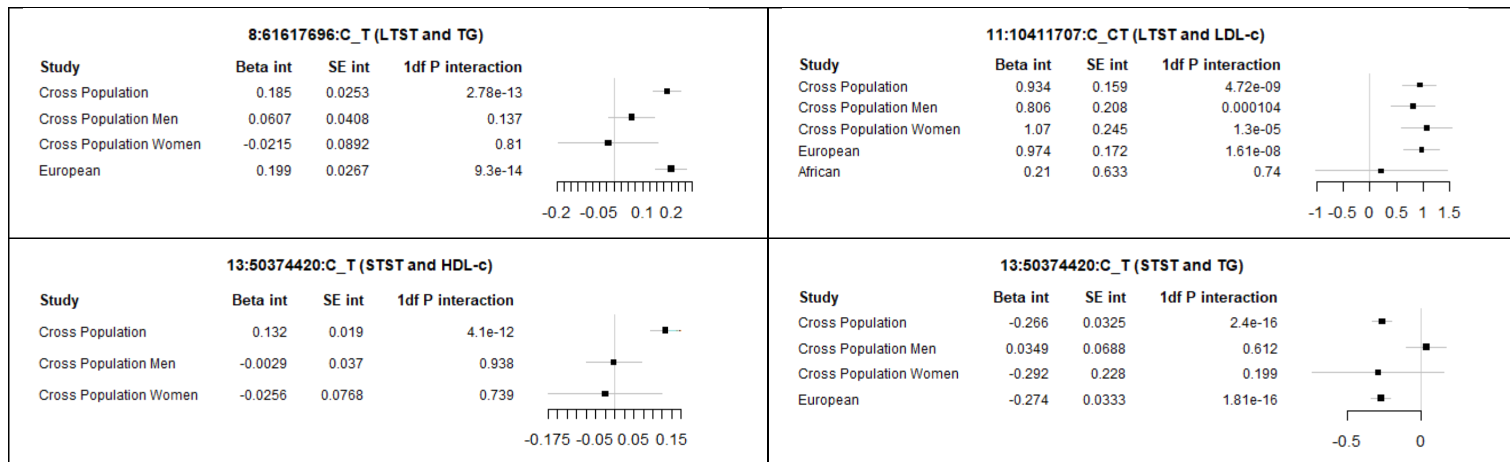


Figure 2: Main results from the 1-degree of freedom interaction analyses in different subgroups. Presented results are the additive variant-interaction effects (log units for TG and HDL-c; mg/dL). Only variants passing all post meta-analysis QC steps were presented. Abbreviations: HDL-c, high-densitivty lipoprotein cholesterol. LDL-c, low-density lipoprotein cholesterol; LTST, long total sleep time; STST, short total sleep time; TG, triglycerides.

Discussion

This large-scale effort identified several variant-lipid trait associations that were modified by either short or long habitual total sleep time, including 10 loci previously-unidentified in relation to lipid levels that interact with either STST or LTST to blood lipid levels without evidence of overlap between STST and LTST results. Using joint meta-analyses, in which the main effect of the variant and the variant-sleep interaction effects are tested jointly, 7 additional genetic lead variants were identified that also showed evidence for interaction with STST or LTST. One of the variants mapped to *DLEU1* and was identified for 2 traits (HDL-c and TG). Moreover, we found distinct variants for STST and LTST interactions– a pattern we also previously reported in a smaller sample for generally higher frequency alleles- suggesting that short and long sleep duration affect the lipid traits through distinct biomolecular mechanisms. Some of the identified genes (most notably *SLC8A1*, *SLC35A31* and *ASPH*) were previously identified as targets for the prevention or treatment of cardiovascular disease and, therefore show promise for additional efforts for further validation and clinical translation.

The variants identified in the present study have not been associated previously with sleep duration [57], other sleep phenotypes (i.e., chronotype, insomnia symptoms or daytime napping) [58-60], or the blood lipid levels that were considered in the present study [15]. The majority of the previously unreported findings in this study are low-frequency variants, that were unlikely to be found in previous studies because they were either not included in the used imputation panels or there was insufficient power. Of the variants identified in the one-df interaction analyses, only the lead variants identified mapped to *ASPH* and *DLEU1* were upstream/downstream transcript variants; all other variants were intronic variants. These findings support the importance of gene-phenotype interaction testing in large studies to explore mechanisms and potential health preventive targets.

A number of the variants identified in the present effort are supported by biological follow-up analyses. Interestingly, we identified *DLEU1* (Deleted In Lymphocytic Leukemia 1), a gene originally identified as a possible tumour suppressor gene and often deleted in patients with B-cell chronic lymphocytic leukemia [61], in both the variant-STST analyses on HDL-c and TG (and not LDL-c). Previously, genome-wide association studies have also increasingly identified this gene with, amongst others, lipid levels [62], fatty

acids [63], anthropometrics [64, 65], immune markers [66], and blood pressure [67]. Furthermore, epigenetic changes in peripheral blood in this gene have been identified in acute myocardial infarction [68]. Although *DLEU1* has not been identified with the habitual sleep variables [57-59, 69], *DLEU1* has been identified to sleep apnea, which is often associated with poor sleep quality and altered sleep duration [70]. We found that the identified variant in *DLEU1* was associated with lower TG and higher HDL-c in individuals reporting short sleep duration, indicative of a lower atherogenic profile. Whether short sleep duration is protective of *DLEU1*-related dyslipidemia, or this variant modifies adverse effects of short sleep duration on lipid levels, cannot be sorted out. The *ASPH* gene was found to be a target for the supplemental Aspartic acid and Succinic acid. Succinate metabolism has been hypothesized as a novel target for myocardial reperfusion injury [71], and elevated plasma succinate levels have been associated with higher levels of cardiovascular risk factors [72].

Our druggability analysis results suggest there are potential drug repositioning opportunities to intervene in common signaling and metabolic pathways implicated in sleep behaviour and lipid metabolism, which could help attenuate serious cardiovascular complications in high-risk patients. One of our top plausible gene targets identified in the 2-degree interaction analyses, *SLC8A1*, is targeted by nutraceutical icosapent. Furthermore, *SLC8A1* has previously been described as a target for the investigational drug caldaret. Caldaret, which acts as a cardioprotective drug modulating intracellular calcium levels, has been previously investigated to reduce infarct sizes in patients with acute myocardial infarction, although did not show positive results [73, 74]. Of interest, *SLC8A1* is affected by the renin angiotensin system [75], which is altered by different sleep conditions [76, 77]. These might present an effective strategy to reduce elevated triglycerides in patients with short sleep duration at risk for cardiovascular complications (e.g., acute myocardial infarction or atrial fibrillation). We also identified several FDA-approved compounds with decades of safe use, which could be evaluated in future preclinical or clinical studies. It is also worth noting the limitations of these predicted drug interactions, which could potentially reflect medication side effects on sleep duration and lipid traits and thus should be interpreted with caution.

We found preliminary evidence for the involvement of the Vitamin D receptor pathway in the association between STST and TG. Although vitamin D itself has not been shown to play any significant role in the onset of cardiovascular disease based on data from randomized clinical trials and

Mendelian randomization [78, 79], the vitamin D receptor appears to be involved in lipid metabolism [80]. Furthermore, genetic variation in the vitamin D receptor gene (*VDR*) has been associated with cardiovascular disease [81]. Accelerated atherosclerosis was observed in *VDR* knock-out mice [82], suggesting that vitamin D receptor signaling inhibits atherosclerosis development. Finally, vitamin D levels have been reported to vary with various sleep outcomes [83], and vitamin D supplementation has been hypothesized to improve sleep [84]. Nevertheless, the role of the vitamin D receptor in the association between sleep and lipid disturbances should be explored in greater detail. No other pathways were identified, also not for the other exposure-trait analyses, possibly due to the fairly low number of genes being identified.

The present study used the largest study sample possible, by considering as much cohorts as possible with available data on genomics, self-reported sleep duration, and concurrent lipid levels. Furthermore, we attempted to standardize the self-reported dichoutomous sleep-exposure variables as much as possible by first taking the age- and sex-adjusted residuals of total sleep time. Despite our efforts to increase sample size in combiniation with increased ancestry diversity compared with our previous effort [31], the vast majority of our study still consisted of cohorts with mainly European-population participants. It is very likely that particular population-specific variant-sleep interactions were missed in the meta-analyses of the non-European populations due to a lack of sufficient statistical power; furthermore, because of low statistical power, we did not present the results from the Hispanic and Asian specific meta-analyses. Future efforts in non-European cohorts, when more data become available, should be further expanded. Although some of the identified loci, despite having low allele frequencies, had some evidence of biological plausibility, they should be further explored in independent samples as we did not have the power to separate cohorts into discovery and independent replication analyses. The present study used information on habitual sleep duration collected through self-report, in which may have measurement error, possibly resulting in lower statistical power. Note that phenotypic and genetic correlations between sleep duration assessed through questionnaire and accelerometry are low to modest as most [57, 69, 85], which suggests phenotypes derived by these methodologies reflect different sleep aspects. Finally, the present study considered sleep as a single dimension, whereas sleep is largely acknowledged to be highly dimensional and complex [86]. Indeed, joint associations between sleep duration and sleep quality have been observed in relation to atherosclerotic cardiovascular disease [26, 87, 88]. However,

detailed data on sleep quality measures were not available in many cohorts, nor was it possible to harmonize these measures when available. Identified variants should, therefore, also be explored in independent samples, as they become available with other sleep variables.

In summary, the present study identified several novel genetic loci associated with lipid traits that were modified by self-reported short- and long total sleep time. The findings yield new insights into the biology underpinning the observed (causal) association between sleep duration and atherosclerotic cardiovascular disease. The observed targets for treatments yield insights into possible prevention of atherosclerotic cardiovascular disease in relation to sleep duration. Additional functional follow-up is required to further characterize the identified genetic variants and to translate the findings to more biological and clinical context.

Reference

[1]Barter P, Gotto AM, LaRosa JC, et al. (2007) HDL cholesterol, very low levels of LDL cholesterol, and cardiovascular events. N Engl J Med 357(13): 1301-1310. 10.1056/NEJMoa064278
[2]Nordestgaard BG, Varbo A (2014) Triglycerides and cardiovascular disease. Lancet 384(9943): 626-635. 10.1016/s0140-6736(14)61177-6
[3]Rader DJ, Hovingh GK (2014) HDL and cardiovascular disease. Lancet 384(9943): 618-625. 10.1016/s0140-6736(14)61217-4
[4]Ridker PM (2014) LDL cholesterol: controversies and future therapeutic directions. Lancet 384(9943): 607-617. 10.1016/s0140-6736(14)61009-6
[5]Holmes MV, Asselbergs FW, Palmer TM, et al. (2015) Mendelian randomization of blood lipids for coronary heart disease. Eur Heart J 36(9): 539-550. 10.1093/eurheartj/eht571
[6]Voight BF, Peloso GM, Orho-Melander M, et al. (2012) Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. Lancet 380(9841): 572-580. 10.1016/S0140-6736(12)60312-2
[7]Zheng Y, Qi L (2014) Diet and lifestyle interventions on lipids: combination with genomics and metabolomics. Clinical Lipidology 9(4): 417-427. 10.2217/clp.14.30
[8]Klarin D, Damrauer SM, Cho K, et al. (2018) Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. Nat Genet 50(11): 1514-1523. 10.1038/s41588-018-0222-9
[9]Liu DJ, Peloso GM, Yu H, et al. (2017) Exome-wide association study of plasma lipids in >300,000 individuals. Nat Genet 49(12): 1758-1766. 10.1038/ng.3977
[10]Willer CJ, Schmidt EM, Sengupta S, et al. (2013) Discovery and refinement of loci associated with lipid levels. Nat Genet 45(11): 1274-1283. 10.1038/ng.2797
[11]Asselbergs FW, Guo Y, van Iperen EP, et al. (2012) Large-scale gene-centric meta-analysis across 32 studies identifies multiple lipid loci. Am J Hum Genet 91(5): 823-838. 10.1016/j.ajhg.2012.08.032
[12]Teslovich TM, Musunuru K, Smith AV, et al. (2010) Biological, clinical and population relevance of 95 loci for blood lipids. Nature 466(7307): 707-713. 10.1038/nature09270
[13]Aulchenko YS, Ripatti S, Lindqvist I, et al. (2009) Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. Nat Genet 41(1): 47-55. 10.1038/ng.269
[14]Kathiresan S, Manning AK, Demissie S, et al. (2007) A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study. BMC Med Genet 8 Suppl 1(Suppl 1): S17. 10.1186/1471-2350-8-s1-s17
[15]Graham SE, Clarke SL, Wu KH, et al. (2021) The power of genetic diversity in genome-wide association studies of lipids. Nature 600(7890): 675-679. 10.1038/s41586-021-04064-3
[16]Van Cauter E, Spiegel K, Tasali E, Leproult R (2008) Metabolic consequences of sleep and sleep loss. Sleep Med 9 Suppl 1(0 1): S23-28. 10.1016/s1389-9457(08)70013-3
[17]Wolk R, Gami AS, Garcia-Touchard A, Somers VK (2005) Sleep and cardiovascular disease. Curr Probl Cardiol 30(12): 625-662. 10.1016/j.cpcardiol.2005.07.002
[18]Lloyd-Jones DM, Allen NB, Anderson CAM, et al. (2022) Life's Essential 8: Updating and Enhancing the American Heart Association's Construct of Cardiovascular Health: A Presidential Advisory From the American Heart Association. Circulation 146(5): e18-e43. 10.1161/cir.0000000000001078

[19]Song Q, Liu X, Zhou W, Wu S, Wang X (2020) Night sleep duration and risk of each lipid profile abnormality in a Chinese population: a prospective cohort study. Lipids Health Dis 19(1): 185. 10.1186/s12944-020-01363-y
[20]Abdurahman AA, Bule M, Shab-Bidar S, Rezaei S, Djafarian K (2020) The association between sleep duration and risk of abnormal lipid profile: A systematic review and meta-analysis. Obesity Medicine 18: 100236. https://doi.org/10.1016/j.obmed.2020.100236
[21]Smiley A, King D, Harezlak J, Dinh P, Bidulescu A (2019) The association between sleep duration and lipid profiles: the NHANES 2013-2014. J Diabetes Metab Disord 18(2): 315-322. 10.1007/s40200-019-00415-0
[22]Bos MM, Noordam R, van den Berg R, et al. (2019) Associations of sleep duration and quality with serum and hepatic lipids: The Netherlands Epidemiology of Obesity Study. J Sleep Res 28(4): e12776. 10.1111/jsr.12776
[23]Bos MM, Goulding NJ, Lee MA, et al. (2021) Investigating the relationships between unfavourable habitual sleep and metabolomic traits: evidence from multi-cohort multivariable regression and Mendelian randomization analyses. BMC Med 19(1): 69. 10.1186/s12916-021-01939-0
[24]Daghlas I, Dashti HS, Lane J, et al. (2019) Sleep Duration and Myocardial Infarction. J Am Coll Cardiol 74(10): 1304-1314. 10.1016/j.jacc.2019.07.022
[25]Ai S, Zhang J, Zhao G, et al. (2021) Causal associations of short and long sleep durations with 12 cardiovascular diseases: linear and nonlinear Mendelian randomization analyses in UK Biobank. Eur Heart J 42(34): 3349-3357. 10.1093/eurheartj/ehab170
[26]Goodman MO, Dashti HS, Lane JM, et al. (2023) Causal Association Between Subtypes of Excessive Daytime Sleepiness and Risk of Cardiovascular Diseases. J Am Heart Assoc 12(24): e030568. 10.1161/jaha.122.030568
[27]Rao DC, Sung YJ, Winkler TW, et al. (2017) Multiancestry Study of Gene-Lifestyle Interactions for Cardiovascular Traits in 610 475 Individuals From 124 Cohorts: Design and Rationale. Circ Cardiovasc Genet 10(3). 10.1161/CIRCGENETICS.116.001649
[28]Manning AK, LaValley M, Liu CT, et al. (2011) Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP x environment regression coefficients. Genet Epidemiol 35(1): 11-18. 10.1002/gepi.20546
[29]Psaty BM, O'Donnell CJ, Gudnason V, et al. (2009) Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. Circ Cardiovasc Genet 2(1): 73-80. 10.1161/CIRCGENETICS.108.829747
[30]Psaty BM, Sitlani C (2013) The Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium as a model of collaborative science. Epidemiology 24(3): 346-348. 10.1097/EDE.0b013e31828b2cbb
[31]Noordam R, Bos MM, Wang H, et al. (2019) Multi-ancestry sleep-by-SNP interaction analysis in 126,926 individuals reveals lipid loci stratified by sleep duration. Nat Commun 10(1): 5121. 10.1038/s41467-019-12958-0
[32]de Vries PS, Brown MR, Bentley AR, et al. (2019) Multiancestry Genome-Wide Association Study of Lipid Levels Incorporating Gene-Alcohol Interactions. Am J Epidemiol 188(6): 1033-1054. 10.1093/aje/kwz005
[33]Bentley AR, Sung YJ, Brown MR, et al. (2019) Multi-ancestry genome-wide gene-smoking interaction study of 387,272 individuals identifies new loci associated with serum lipids. Nat Genet 51(4): 636-648. 10.1038/s41588-019-0378-y
[34]Kilpelainen TO, Bentley AR, Noordam R, et al. (2019) Multi-ancestry study of blood lipid levels identifies four loci interacting with physical activity. Nat Commun 10(1): 376. 10.1038/s41467-018-08008-w

[35]Bycroft C, Freeman C, Petkova D, et al. (2018) The UK Biobank resource with deep phenotyping and genomic data. Nature 562(7726): 203-209. 10.1038/s41586-018-0579-z
[36]Gaziano JM, Concato J, Brophy M, et al. (2016) Million Veteran Program: A mega-biobank to study genetic influences on health and disease. J Clin Epidemiol 70: 214-223. 10.1016/j.jclinepi.2015.09.016
[37]Friedewald WT, Levy RI, Fredrickson DS (1972) Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin Chem 18(6): 499-502
[38]Westerman KE, Pham DT, Hong L, et al. (2021) GEM: scalable and flexible gene-environment interaction analysis in millions of samples. Bioinformatics 37(20): 3514-3520. 10.1093/bioinformatics/btab223
[39]Kraft P, Yen YC, Stram DO, Morrison J, Gauderman WJ (2007) Exploiting gene-environment interaction to detect genetic associations. Hum Hered 63(2): 111-119. 10.1159/000099183
[40]Winkler TW, Day FR, Croteau-Chonka DC, et al. (2014) Quality control and conduct of genome-wide association meta-analyses. Nat Protoc 9(5): 1192-1212. 10.1038/nprot.2014.071
[41]Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26(17): 2190-2191. 10.1093/bioinformatics/btq340
[42]Winkler TW, Kutalik Z, Gorski M, Lottaz C, Kronenberg F, Heid IM (2015) EasyStrata: evaluation and visualization of stratified genome-wide association meta-analysis data. Bioinformatics 31(2): 259-261. 10.1093/bioinformatics/btu621
[43]Thomas D (2010) Methods for investigating gene-environment interactions in candidate pathway and genome-wide association studies. Annu Rev Public Health 31: 21-36. 10.1146/annurev.publhealth.012809.103619
[44]Altman DG, Bland JM (2003) Interaction revisited: the difference between two estimates. BMJ 326(7382): 219. 10.1136/bmj.326.7382.219
[45]Watanabe K, Taskesen E, van Bochoven A, Posthuma D (2017) Functional mapping and annotation of genetic associations with FUMA. Nat Commun 8(1): 1826. 10.1038/s41467-017-01261-5
[46]Boughton AP, Welch RP, Flickinger M, et al. (2021) LocusZoom.js: interactive and embeddable visualization of genetic association study results. Bioinformatics 37(18): 3017-3018. 10.1093/bioinformatics/btab186
[47]Pruim RJ, Welch RP, Sanna S, et al. (2010) LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26(18): 2336-2337. 10.1093/bioinformatics/btq419
[48]Boyle AP, Hong EL, Hariharan M, et al. (2012) Annotation of functional variation in personal genomes using RegulomeDB. Genome Res 22(9): 1790-1797. 10.1101/gr.137323.112
[49]Ghoussaini M, Mountjoy E, Carmona M, et al. (2021) Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. Nucleic Acids Res 49(D1): D1311-D1320. 10.1093/nar/gkaa840
[50]Gagliano Taliun SA, VandeHaar P, Boughton AP, et al. (2020) Exploring and visualizing large-scale genetic associations by using PheWeb. Nat Genet 52(6): 550-552. 10.1038/s41588-020-0622-5
[51]Groza T, Gomez FL, Mashhadi HH, et al. (2023) The International Mouse Phenotyping Consortium: comprehensive knockout phenotyping underpinning the study of human disease. Nucleic Acids Res 51(D1): D1038-D1045. 10.1093/nar/gkac972
[52]Ramos EM, Hoffman D, Junkins HA, et al. (2014) Phenotype-Genotype Integrator (PheGenI): synthesizing genome-wide association study (GWAS) data with existing genomic resources. Eur J Hum Genet 22(1): 144-147. 10.1038/ejhg.2013.96

[53]Kavousi M, Bos MM, Barnes HJ, et al. (2023) Multi-ancestry genome-wide study identifies effector genes and druggable pathways for coronary artery calcification. Nat Genet 55(10): 1651-1664. 10.1038/s41588-023-01518-4

[54]Wang Y, Mei H, Jiang YR, et al. (2015) Relationship between Duration of Sleep and Hypertension in Adults: A Meta-Analysis. J Clin Sleep Med 11(9): 1047-1056. 10.5664/jcsm.5024

[55]Qiu Y, Li X, Zhang X, et al. (2022) Prothrombotic Factors in Obstructive Sleep Apnea: A Systematic Review With Meta-Analysis. Ear Nose Throat J 101(9): NP412-NP421. 10.1177/0145561320965208

[56]Diaz-Roman A, Mitchell R, Cortese S (2018) Sleep in adults with ADHD: Systematic review and meta-analysis of subjective and objective studies. Neurosci Biobehav Rev 89: 61-71. 10.1016/j.neubiorev.2018.02.014

[57]Dashti HS, Jones SE, Wood AR, et al. (2019) Genome-wide association study identifies genetic loci for self-reported habitual sleep duration supported by accelerometer-derived estimates. Nat Commun 10(1): 1100. 10.1038/s41467-019-08917-4

[58]Jones SE, Lane JM, Wood AR, et al. (2019) Genome-wide association analyses of chronotype in 697,828 individuals provides insights into circadian rhythms. Nat Commun 10(1): 343. 10.1038/s41467-018-08259-7

[59]Dashti HS, Daghlas I, Lane JM, et al. (2021) Genetic determinants of daytime napping and effects on cardiometabolic health. Nat Commun 12(1): 900. 10.1038/s41467-020-20585-3

[60]Jansen PR, Watanabe K, Stringer S, et al. (2019) Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. Nat Genet 51(3): 394-403. 10.1038/s41588-018-0333-3

[61]Liu Y, Corcoran M, Rasool O, et al. (1997) Cloning of two candidate tumor suppressor genes within a 10 kb region on chromosome 13q14, frequently deleted in chronic lymphocytic leukemia. Oncogene 15(20): 2463-2473. 10.1038/sj.onc.1201643

[62]Sakaue S, Kanai M, Tanigawa Y, et al. (2021) A cross-population atlas of genetic associations for 220 human phenotypes. Nat Genet 53(10): 1415-1424. 10.1038/s41588-021-00931-x

[63]de Oliveira Otto MC, Lemaitre RN, Sun Q, et al. (2018) Genome-wide association meta-analysis of circulating odd-numbered chain saturated fatty acids: Results from the CHARGE Consortium. PLoS One 13(5): e0196951. 10.1371/journal.pone.0196951

[64]Tachmazidou I, Suveges D, Min JL, et al. (2017) Whole-Genome Sequencing Coupled to Imputation Discovers Genetic Signals for Anthropometric Traits. Am J Hum Genet 100(6): 865-884. 10.1016/j.ajhg.2017.04.014

[65]Pulit SL, Stoneman C, Morris AP, et al. (2019) Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. Hum Mol Genet 28(1): 166-174. 10.1093/hmg/ddy327

[66]Vuckovic D, Bao EL, Akbari P, et al. (2020) The Polygenic and Monogenic Basis of Blood Traits and Diseases. Cell 182(5): 1214-1231 e1211. 10.1016/j.cell.2020.08.008

[67]Plotnikov D, Huang Y, Khawaja AP, et al. (2022) High Blood Pressure and Intraocular Pressure: A Mendelian Randomization Study. Invest Ophthalmol Vis Sci 63(6): 29. 10.1167/iovs.63.6.29

[68]Fernandez-Sanles A, Sayols-Baixeras S, Subirana I, et al. (2021) DNA methylation biomarkers of myocardial infarction and cardiovascular disease. Clin Epigenetics 13(1): 86. 10.1186/s13148-021-01078-6

[69]Jones SE, van Hees VT, Mazzotti DR, et al. (2019) Genetic studies of accelerometer-based sleep measures yield new insights into human sleep behaviour. Nat Commun 10(1): 1585. 10.1038/s41467-019-09576-1

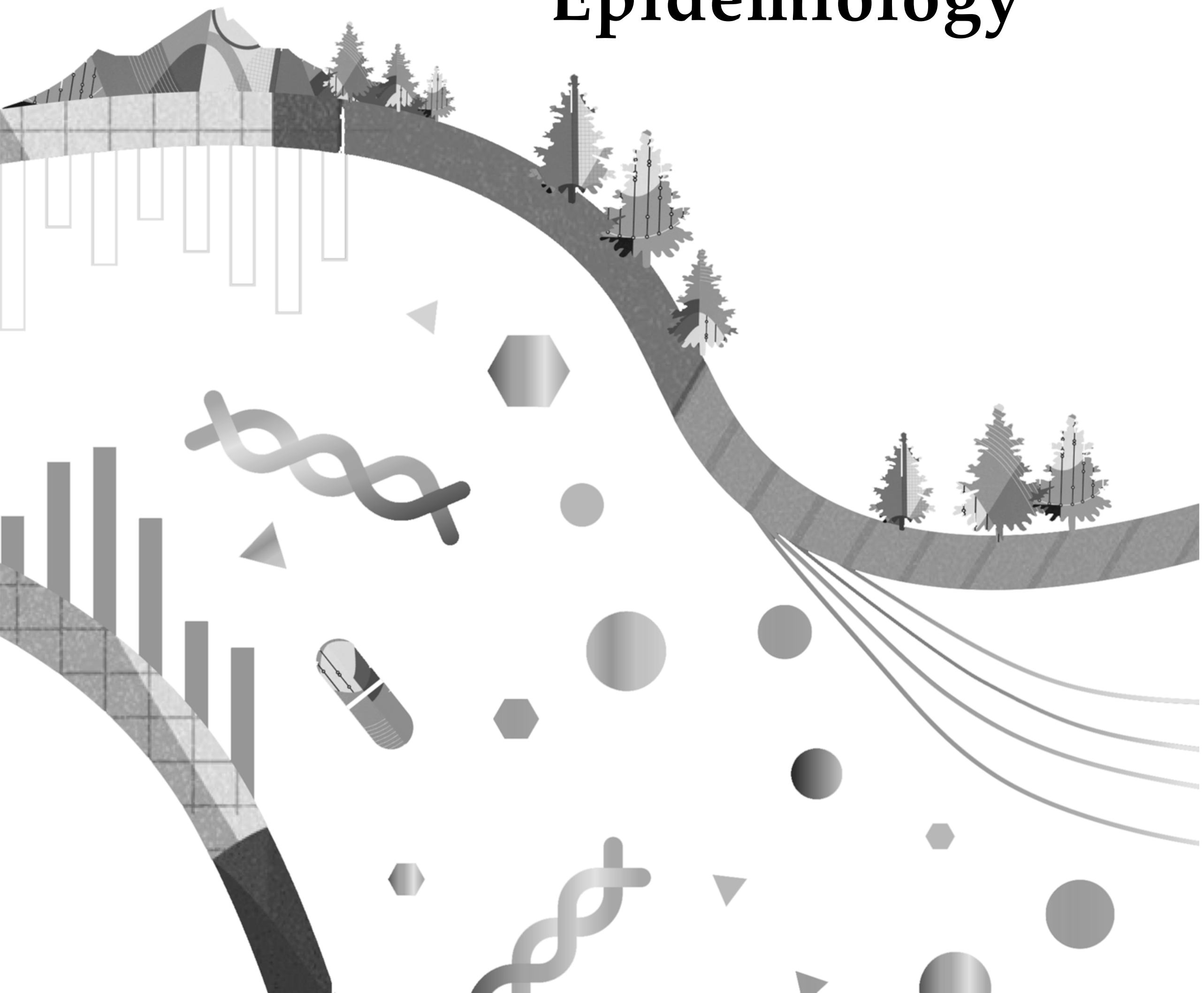
[70]Campos AI, Ingold N, Huang Y, et al. (2023) Discovery of genomic loci associated with sleep apnea risk through multi-trait GWAS analysis with snoring. Sleep 46(3). 10.1093/sleep/zsac308

[71]Pell VR, Chouchani ET, Frezza C, Murphy MP, Krieg T (2016) Succinate metabolism: a new therapeutic target for myocardial reperfusion injury. Cardiovasc Res 111(2): 134-141. 10.1093/cvr/cvw100
[72]Osuna-Prieto FJ, Martinez-Tellez B, Ortiz-Alvarez L, et al. (2021) Elevated plasma succinate levels are linked to higher cardiovascular disease risk factors in young adults. Cardiovasc Diabetol 20(1): 151. 10.1186/s12933-021-01333-3
[73]Jang IK, Weissman NJ, Picard MH, et al. (2008) A randomized, double-blind, placebo-controlled study of the safety and efficacy of intravenous MCC-135 as an adjunct to primary percutaneous coronary intervention in patients with acute myocardial infarction: Evaluation of MCC-135 for left ventricular salvage in acute myocardial infarction (EVOLVE). Am Heart J 155(1): 113 e111-118. 10.1016/j.ahj.2007.08.020
[74]Kawasumi H, Satoh N, Kitada Y (2007) Caldaret, an intracellular Ca2+ handling modulator, limits infarct size of reperfused canine heart. J Pharmacol Sci 103(2): 222-233. 10.1254/jphs.fp0060765
[75]Takeuchi F, Liang YQ, Isono M, Ang MY, Mori K, Kato N (2022) Transcriptomic Response in the Heart and Kidney to Different Types of Antihypertensive Drug Administration. Hypertension 79(2): 413-423. 10.1161/HYPERTENSIONAHA.121.18026
[76]Wang Y, Li CX, Lin YN, et al. (2021) The Role of Aldosterone in OSA and OSA-Related Hypertension. Front Endocrinol (Lausanne) 12: 801689. 10.3389/fendo.2021.801689
[77]Murck H, Uhr M, Ziegenbein M, et al. (2006) Renin-angiotensin-aldosterone system, HPA-axis and sleep-EEG changes in unmedicated patients with depression after total sleep deprivation. Pharmacopsychiatry 39(1): 23-29. 10.1055/s-2006-931476
[78]Thompson B, Waterhouse M, English DR, et al. (2023) Vitamin D supplementation and major cardiovascular events: D-Health randomised controlled trial. BMJ 381: e075230. 10.1136/bmj-2023-075230
[79]Emerging Risk Factors Collaboration and Epic- C. V. D. Vitamin D. Studies Collaboration (2024) Estimating dose-response relationships for vitamin D with coronary heart disease, stroke, and all-cause mortality: observational and Mendelian randomisation analyses. Lancet Diabetes Endocrinol 12(1): e2-e11. 10.1016/S2213-8587(23)00287-5
[80]Martinez-Sena T, Soluyanova P, Guzman C, Valdivielso JM, Castell JV, Jover R (2020) The Vitamin D Receptor Regulates Glycerolipid and Phospholipid Metabolism in Human Hepatocytes. Biomolecules 10(3). 10.3390/biom10030493
[81]Kichaev G, Bhatia G, Loh PR, et al. (2019) Leveraging Polygenic Functional Enrichment to Improve GWAS Power. Am J Hum Genet 104(1): 65-75. 10.1016/j.ajhg.2018.11.008
[82]Szeto FL, Reardon CA, Yoon D, et al. (2012) Vitamin D receptor signaling inhibits atherosclerosis in mice. Mol Endocrinol 26(7): 1091-1101. 10.1210/me.2011-1329
[83]Bertisch SM, Sillau S, de Boer IH, Szklo M, Redline S (2015) 25-Hydroxyvitamin D Concentration and Sleep Duration and Continuity: Multi-Ethnic Study of Atherosclerosis. Sleep 38(8): 1305-1311. 10.5665/sleep.4914
[84]Abboud M (2022) Vitamin D Supplementation and Sleep: A Systematic Review and Meta-Analysis of Intervention Studies. Nutrients 14(5). 10.3390/nu14051076
[85]Jackson CL, Patel SR, Jackson WB, 2nd, Lutsey PL, Redline S (2018) 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 41(6). 10.1093/sleep/zsy057
[86]Chung J, Goodman M, Huang T, Bertisch S, Redline S (2021) Multidimensional sleep health in a diverse, aging adult cohort: Concepts, advances, and implications for research and intervention. Sleep Health 7(6): 699-707. 10.1016/j.sleh.2021.08.005

[87]Arora N, Richmond RC, Brumpton BM, et al. (2023) Self-reported insomnia symptoms, sleep duration, chronotype and the risk of acute myocardial infarction (AMI): a prospective study in the UK Biobank and the HUNT Study. Eur J Epidemiol 38(6): 643-656. 10.1007/s10654-023-00981-x
[88]Arora N, Bhatta L, Skarpsno ES, et al. (2023) Investigating the causal interplay between sleep traits and risk of acute myocardial infarction: a Mendelian randomization study. BMC Med 21(1): 385. 10.1186/s12916-023-03078-0

Part IV

Integration of Pharmacometrics and Epidemiology



Chapter 6

A novel approach for pharmacological substantiation of safety signals using plasma concentrations of medication and administrative/healthcare databases: A case study using Danish registries for an FDA warning on lamotrigine

Wenyi Wang, Vera Battini, Carla Carnovale, Raymond Noordam, Ko Willems van Dijk, Kristian Hay Kragholm, Diana van Heemst, Hiie Soeorg & Maurizio Sessa

Pharmacol Res (2023) 10.1016

The supplemental information for this paper is available online at https://drive.google.com/drive/folders/1UoSstJZtu2ylarbtkhn6Kmr3rz68hhSu?usp=sharing

Abstract

PHARMACOM-EPI is a novel framework to predict plasma concentrations of drugs at the time of occurrence of clinical outcomes. In early 2021, the U.S. Food and Drug Administration (FDA) issued a warning on the antiseizure drug lamotrigine claiming that it has the potential to increase the risk of arrhythmias and related sudden cardiac death due to a pharmacological sodium channel-blocking effect. We hypothesized that the risk of arrhythmias and related death is due to toxicity. We used the PHARMACOM-EPI framework to investigate the relationship between lamotrigine's plasma concentrations and the risk of death in older patients using real-world data. Danish nationwide administrative and healthcare registers were used as data sources and individuals aged 65 years or older during the period 1996 – 2018 were included in the study. According to the PHARMACOM-EPI framework, plasma concentrations of lamotrigine were predicted at the time of death and patients were categorized into non-toxic and toxic groups based on the therapeutic range of lamotrigine (3-15 mg/L). Over 1 year of treatment, the incidence rate ratio (IRR) of all-cause mortality was calculated between the propensities score matched toxic and non-toxic groups. In total, 7286 individuals were diagnosed with epilepsy and were exposed to lamotrigine, 432 of which had at least one plasma concentration measurement The pharmacometric model by *Chavez et al* was used to predict lamotrigine's plasma concentrations considering the lowest absolute percentage error among identified models (14.25%, 95% CI: 11.68-16.23). The majority of lamotrigine associated deaths were cardiovascular-related and occurred among individuals with plasma concentrations in the toxic range. The IRR of mortality between the toxic group and non-toxic group was 3.37 [95% CI: 1.44-8.32] and the cumulative incidence of all-cause mortality exponentially increased in the toxic range. Application of our novel framework PHARMACOM-EPI provided strong evidence to support our hypothesis that the increased risk of all-cause and cardiovascular death was associated with a toxic plasma concentration level of lamotrigine among older lamotrigine users.

Introduction

Recent research has emphasized that the incorporation of information on plasma concentrations is important for effective pharmacoepidemiological research when performing pharmacological substantiation of clinical outcomes. [1–3] Thus far, the integration of data on plasma concentrations in pharmacoepidemiological research has been limited due to the lack of available approaches for performing this task. We have recently developed a novel framework, PHARMACO-EPI, to integrate pharmacometrics into pharmacoepidemiological research with the final goal of predicting plasma concentrations at the time of occurrence of clinical outcomes for their pharmacological substantiation, especially for newly discovered associations from signal detection activities. However, this framework has not been tested yet in real-life cases where regulatory agencies have detected new safety signals.

In early 2021, the U.S. Food and Drug Administration (FDA) issued a warning indicating that lamotrigine has the potential to increase the risk of arrhythmias and related sudden cardiac death. [4] These adverse events may be related to lamotrigine's sodium channel blocking effect and related alteration of the QRS complex of the electrocardiogram, and delay in ventricular conduction is associated with enhanced risk of ventricular arrhythmias and sudden cardiac death in susceptible patients. [4–6] As lamotrigine's sodium channel-blocking effects are dose-dependent [7], this risk may be higher in older epileptic patients aged 65+ exposed to lamotrigine as they have an increased risk of reaching toxic plasma concentration. [8]

We used PHARMACO-EPI [1,2,9,10], to predict the plasma concentration of lamotrigine at death and investigate the association between toxic plasma concentrations and death in older individuals exposed to lamotrigine by using Danish nationwide real-world data.

Method

Data source

Danish nationwide administrative and healthcare registers were used as data sources. Based on a unique personal identification number for each Danish citizen that linked Danish registers, we collected information on residents' date of birth, sex [11], hospital admission/ambulatory visit [12], medication redemptions in community pharmacies [13], and deaths. [14] Additionally, by using the Danish Register for Laboratory Results for Research, we collected information about plasma concentrations of antiseizure medications and results of immunological and biochemical tests performed in Danish patients. [15]

In Denmark, plasma concentration measurements are commonly used in Therapeutic Drug Monitoring (TDM) to optimize the dosage of antiepileptic medication. TDM helps ensure that the medication is within the therapeutic range and that the patient is receiving an effective and safe dose.

Study population

All residents in Denmark aged 65 years or older during the period 1996 – 2018, that were diagnosed with epilepsy, and that have redeemed their first prescription of lamotrigine were included in the study population. The first prescription of lamotrigine following epilepsy diagnosis was used as the index date from which we followed up individuals in the registers. Individuals that redeemed any antiseizure medications twelve months before the index date were excluded. Individuals were classified as having epilepsy if they had been hospitalized or had been in outpatient care with an epilepsy diagnosis (International Classification of Disease code, ICD-8 code 345, ICD-10 codes G40). The diagnosis codes for epilepsy in the Danish registers have been previously validated with a positive predictive value of 81% (95% Confidence Interval (CI): 75%–87%). [16]

Follow-up period & study outcome

The study population was followed from the index date for 365 days until the occurrence of the study outcome (i.e., death) or censoring due to

permanent emigration or end of the data coverage. The study outcome was all-cause of mortality within the follow-up period. We have selected 1-year follow-up based on the consideration that it was expected that the majority of dose-dependent events occurred during the early stages of the treatment.[17]

Identification of accessible pharmacometric models

We used PHARMACOM-EPI [1] to determine the fraction of deaths that are pharmacologically substantiated by toxic plasma concentrations at the time of death or the end of follow-up. According to PHARMACOM-EPI [1], pharmacometric models of lamotrigine developed on the adult population from the most recent systematic reviews [18] of population pharmacokinetic models of lamotrigine were identified. The same systematic review query was used to update the review and identify models that were not previously included in the systematic review. Detailed processes concerning systematic review are shown in Appendix 1.

Application of pharmacometric models to real-world data

Before applying the pharmacometrics model, we characterized them by extracting relevant socio-demographic characteristics of the populations on which the models were developed and parameters from the models including fixed parameters, equations for estimated parameters, inter-individual variability, residual variability, covariates tested, and software and number of compartments. The analytical dataset for pharmacometric models was structured according to the format requirements of NONMEM software [19]. By applying pharmacometric models to real-world data, we obtained population predictions and individual predictions of lamotrigine plasma concentrations for each model. Possible missing covariates required in the pharmacometric models were imputed using published data from similar populations. Missing value in height and weight were imputed based on published Danish reference data for men and women, respectively. [20] Blood urea nitrogen/serum creatinine ratio was imputed based on a normal distribution having a mean 11 and SD 2. Body surface area was calculated using following equation (Body Surface Area= $007184 * Height(m)^{0.725} * Weight(kg)^{0.425}$). Glomerular filtration rate was estimated by equation published by *Levey AS et al*. [21] Similarly, genetic polymorphisms were imputed based on the allele frequencies of genotypes satisfying Hardy-

Weinberg equilibrium obtained from the NCBI database of genetic variation. [22] Daily doses of antiseizure medications were imputed using the Sessa Empirical Estimator [23–26] and missing dosing times were replaced with fixed times, assuming that the latest dose was taken at 8:00 a.m. on the same day of the measurement of plasma concentration. The models' performance was assessed by plotting observed versus predicted population/individual concentrations as two-dimensional density plots with contour lines where each contour shows aggregated data for more than 5 individuals. Trend lines were drawn using local polynomial regression fitting. In addition, the absolute percentage error of population concentration prediction and individual concentration prediction for each model was calculated to identify the pharmacometrics model with the lowest absolute percentage error. The absolute percentage error for population and individual prediction of each pharmacometric model was plotted using a boxplot.

Statistical analysis

Characteristics for the study population at the baseline were reported as mean with standard deviation (SD) for continuous variables and percentage for categorical variables separately between individuals with and without available plasma concentrations of lamotrigine.

To predict lamotrigine plasma concentration at death using individual predictions from pharmacometrics models, we used data from individuals with available plasma concentrations among those included in the study population. To infer the results from individuals with lamotrigine plasma concentration to those without plasma concentrations measurements, we performed propensity score matching. The propensity score for receiving or not a plasma concentration measurement was obtained based on risk factors of the outcome including age, sex, comorbidities, and co-medications according to the Elixhauser Comorbidity Index [27] using a logistic regression model. The Elixhauser Comorbidity Index [27] is a validated index that include clinical conditions associated with mortality. For each individual with available plasma concentration measurements, we matched four individuals without lamotrigine concentration using the greedy nearest neighbor method with a maximum caliper distance of 0.2 of the logit of the propensity score. The density plots of the propensity scores before and after matching were plotted. We considered it possible to infer the results from individuals with available plasma concentration to individuals without

plasma concentration if there was at least a 70% overlap of the propensity scores distributions.

To investigate the association between plasma concentration, when available, and the study outcome, we used the pharmacometric model with the lowest absolute percentage error to predict trough concentrations at death or at the end follow-up. Plasma concentrations were categorized into non-toxic and toxic groups based on the reference therapeutic range for lamotrigine (3-15 mg/L). [28] The cumulative incidence rate for all-cause mortality versus lamotrigine plasma concentration was plotted and the incidence rate ratio (IRR) of all-cause mortality between non-toxic and toxic group was calculated. The hazard functions of individuals within the non-toxic range and those with toxic plasma concentration were computed using a Cox regression model and plotted as cumulative hazards functions curves. All the analyses were implemented in R (version 4.0.0, RStudio, Austria).

Ethics

In Denmark, ethical committee approval or individual patient consent is not required according to Danish law, which implies that it is not required to obtain consent for participants in studies based on register data in Denmark.

Results

Study population characteristics

In total, 7286 individuals were diagnosed with epilepsy and were exposed to lamotrigine, 432 of which had at least one plasma concentration measurement. For the whole population, 49.7% of the individuals were men with an average age of 70 (SD: 10.5) years. The population with plasma concentration measurement had a greater mean age [72.7 (SD: 8.1) years] and a higher proportion number of males (53.5%) compared with the population without plasma concentration which had 49.4% of men and a mean age of 69.8 (SD: 10.6) years (**Table 1**). Within those individuals with at least one plasma concentration measurement, frequencies of comorbidities, age and sex did not differ significantly between those in the non-toxic group (n=290) and those who reached toxic ranges (n=142) suggesting comparability between the two groups (**Table 2**).

Table 1 Characteristics of epileptic patients with and without lamotrigine concentration measurements

Variable	Level	Patients without plasma concentration (n= 6849)	Patients with plasma concentration (n=432)	Total (n=7286)
Sex	Male	3385 (49.4)	234 (53.7)	3619 (49.7)
Age	Mean (SD)	69.8 (10.6)	72.7 (8.1)	70 (10.5)
Diabetes	Yes	717 (10.5)	56 (13.0)	773 (10.6)
Cancer	Yes	764 (11.2)	74 (17.1)	838 (11.5)
Arthritis	Yes	264 (3.9)	15 (3.5)	279 (3.8)
Coronary Heart Disease	Yes	1262 (18.4)	82 (19.0)	1344 (18.4)
Dementia	Yes	587 (8.6)	58 (13.4)	646 (8.9)
Respiratory Disorders	Yes	1665 (24.3)	131 (30.3)	1797 (24.7)
Ulcer	Yes	396 (5.8)	33 (7.6)	429 (5.9)
Cerebrovascular Disorders	Yes	3531 (51.6)	233 (53.9)	3765 (51.7)
Dyslipidaemia	Yes	982 (14.3)	95 (22.0)	1077 (14.8)
Depression	Yes	506 (7.4)	29 (6.7)	536 (7.4)
Renal Disorders	Yes	303 (4.4)	33 (7.6)	336 (4.6)
Pacemaker implanted	Yes	2373 (34.6)	99 (22.9)	2474 (34.0)

Data are reported as mean (SD) or frequency (%).

Table 2. Characteristics of epileptic patients within and over lamotrigine therapeutic range

Variable	Level	Non-toxic (n=290)	Toxic (n=142)	P
Sex	Male	159 (54.8)	73 (51.4)	0.57086
Age	Mean (SD)	72.8 (8.4)	72.5 (7.4)	0.68085
Diabetes	Yes	44 (15.2)	12 (8.5)	0.07165
Cancer	Yes	57 (19.7)	17 (12.0)	0.06358
Arthritis	Yes	8 (2.8)	7 (4.9)	0.37992
Coronary Heart Disease	Yes	55 (19.0)	27 (19.0)	1.00000
Dementia	Yes	3.8 (13.1)	20 (14.1)	0.89598
Respiratory Disorders	Yes	84 (29.0)	47 (33.1)	0.44339
Ulcer	Yes	23 (7.9)	10 (7.0)	0.89349
Cerebrovascular Disorders	Yes	153 (52.8)	80 (56.3)	0.54959
Dyslipidaemia	Yes	63 (21.7)	32 (22.5)	0.94615
Depression	Yes	18 (6.2)	11 (7.7)	0.69209
Renal Disorders	Yes	17 (5.9)	16 (11.3)	0.07279
Pacemaker implanted	Yes	70 (24.1)	29 (20.4)	0.45855

Data are reported as mean (SD) or frequency (%).

Pharmacometric models

Seven pharmacometric models satisfying inclusion and exclusion criteria that describe the population pharmacokinetics of lamotrigine were identified, one of which was derived from the systematic search and the rest were from the review paper [18]. Characteristics of the study population for each model were provided in **Supplementary Table 1**.

Application of pharmacometric models to real-world data

Four pharmacometric models successfully predicted individual and population lamotrigine concentrations, while three models failed to perform prediction due to convergence issues. Compared with population prediction, the individual prediction had a stronger correlation with observed plasma concentration as the density closely follows the trend line (**Figure 1**). The individual prediction had a lower absolute percentage error compared with population prediction (**Figure 2**) for all the models. Therefore, the individual prediction was prioritized for the pharmacological substantiation of lamotrigine-related death. Among these four pharmacometric models, the

model by *Chavez et al* [29]had the lowest absolute prediction error for individual prediction with 14.25% [95% CI: 11.68-16.23] and, therefore, it was used to predict plasma concentrations at death/end of the follow-up.

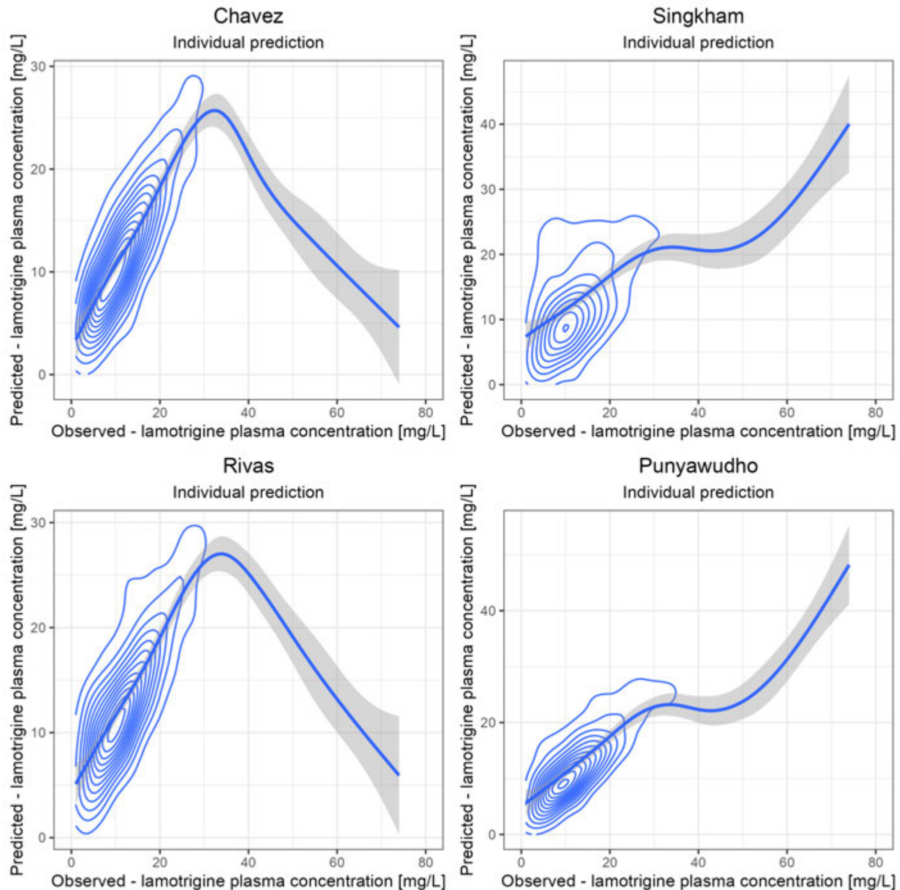


Figure 1. Two-dimensional density plots with contour lines and trend lines of population/individual predictions versus observed plasma concentrations of lamotrigine for four pharmacometric models.

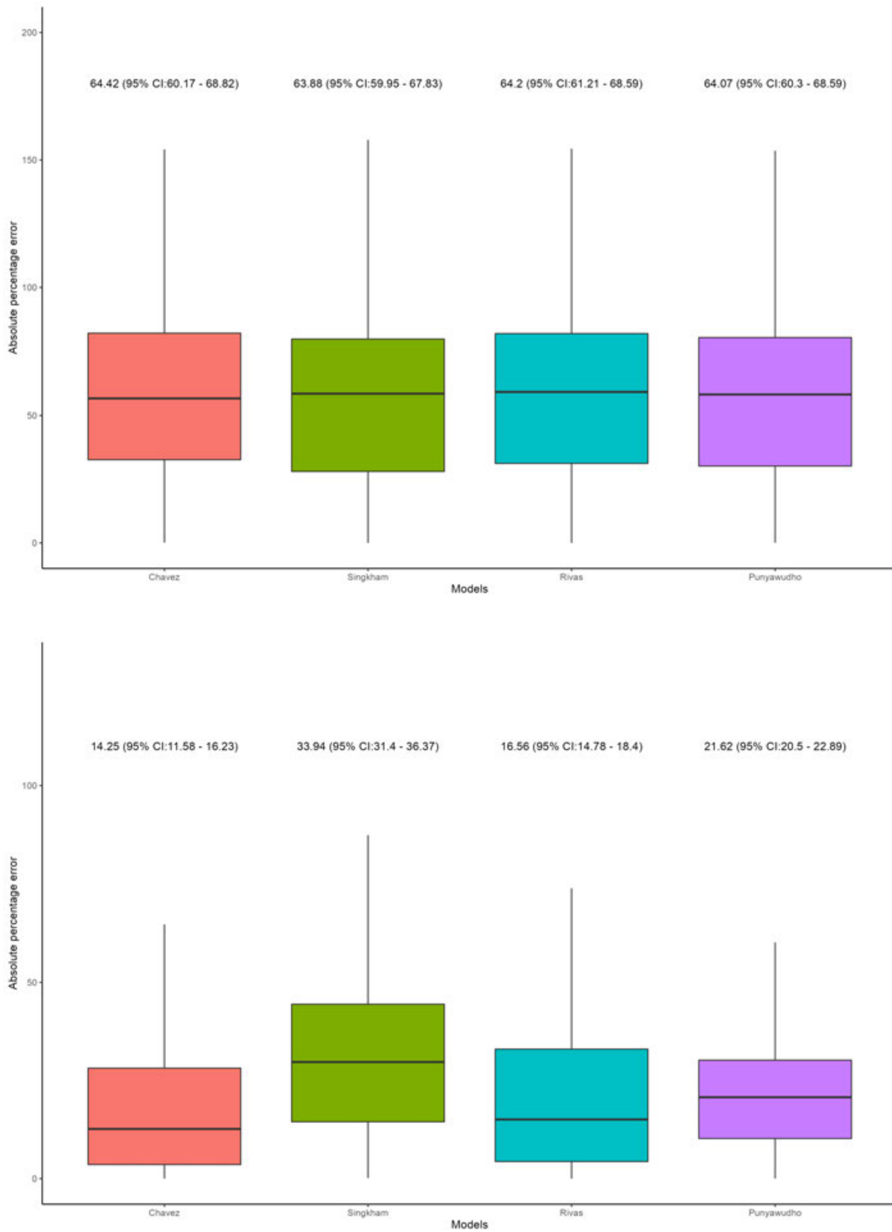


Figure 2. Bar plots of absolute percentage error for population (up) and individual (down) prediction of four pharmacometric models.

Pharmacological substantiation

Based on the individual prediction derived from the model by *Chavez et al*[29], among individuals who died, 10 were in the non-toxic range (n=290, 3.45%) and 16 in the toxic one (n=142, 11.27%). The median plasma concentration for the group within the non-toxic range was 9 mg/L (interquartile range: 6-12 mg/L) and for the group within the toxic range was 21 mg/L (interquartile range: 16-28 mg/L). Cardiovascular diseases were the cause of death in 44% and 60% of people who died in the non-toxic and toxic groups, respectively. The IRR of mortality between the toxic group and non-toxic group was 3.37 [95% CI: 1.44-8.32] and the cumulative incidence of all-cause mortality exponentially increased outside the therapeutic range (**Figure 3**). The cumulative hazard of all-cause mortality was significantly higher in the toxic compared with the non-toxic group (**Figure 4**). After propensity score matching of individuals (i.e., overlapping 99%) with lamotrigine concentration and those without, the densities between the two groups nearly fully overlapped compared with the overlapping before matching of 70% (**Supplementary figures 1 and 2**).

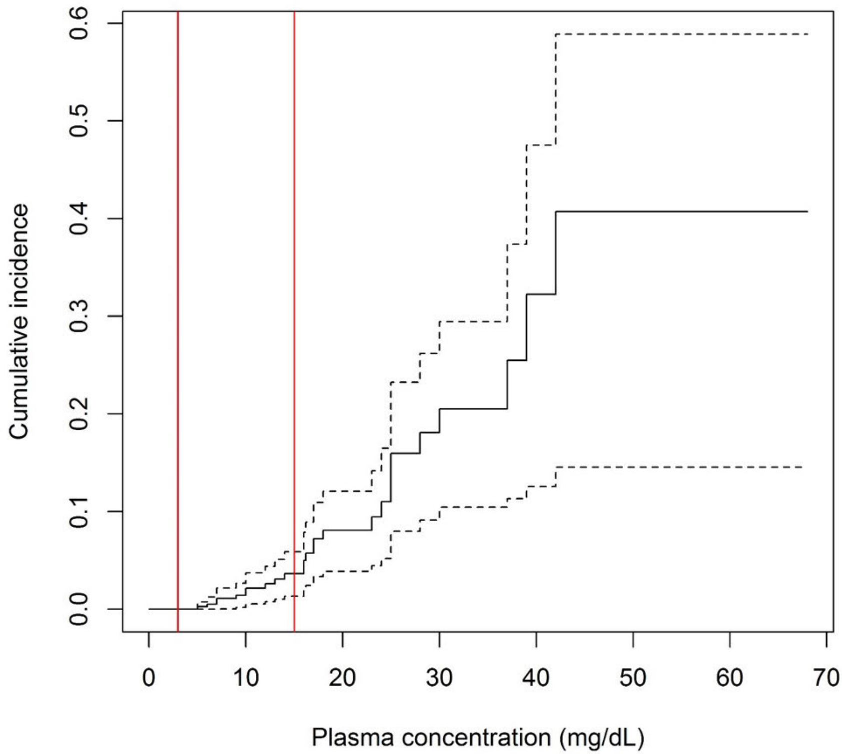


Figure 3. Cumulative incidence curve of all-cause mortality versus plasma concentration. The left horizontal red line indicated lower bound of therapeutic range of lamotrigine (3 mg/L) and the right horizontal red line indicated upper bound of therapeutic range of lamotrigine (15 mg/L). Dotted line = 95% Confidence interval.

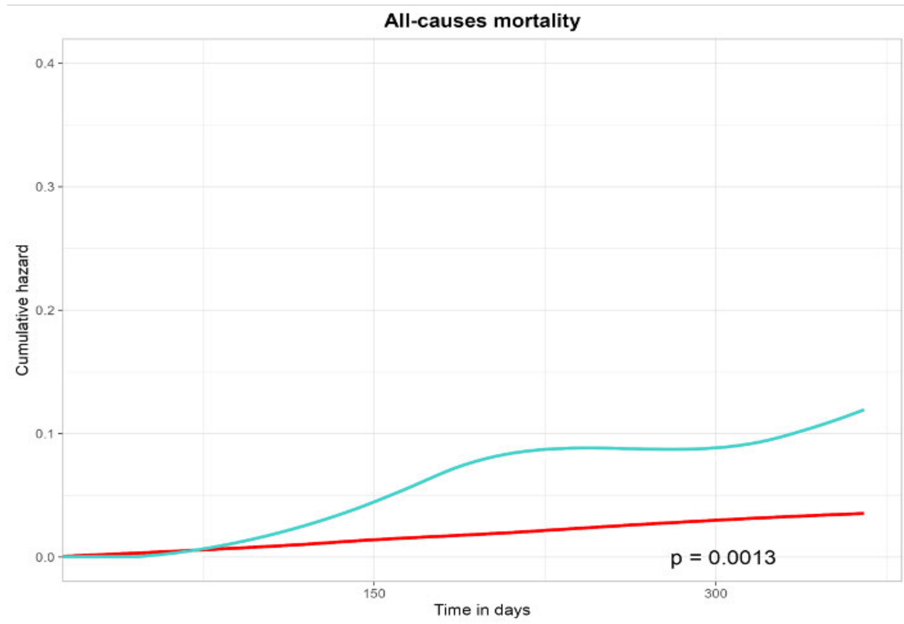


Figure 4. Cumulative hazard of all-cause mortality over time. The red curve is for individuals with therapeutic range and the blue curve is for individuals with toxic range.

Discussion

In this study, we applied a novel framework, PHARMACOM-EPI to real-world data to predict individual plasma concentrations of lamotrigine, which were subsequently categorized into non-toxic and toxic groups for individuals with plasma measurements. We found that individuals within the toxic group had a higher incidence of all-cause and cardiovascular mortality compared with individuals within the non-toxic group. For the first time, it was feasible to use data on plasma concentration levels and a new framework to confirm toxicity hypotheses with administrative and real-world healthcare data. The approach was applied to a case of lamotrigine's triggered cardiovascular death, which had not been previously clarified.
This result suggests that QRS alterations are mainly related to the dose-dependent inhibiting effects of lamotrigine on the sodium channel. [30] In the heart, sodium channels modulate the voltage threshold and the duration of action potentials, thus contributing to conduction and cellular excitability. [31] A key difference between neuronal and cardiac action potentials is their duration: in humans, in the central nervous system, it is between 1 and 10 ms [32], whereas in cardiac ventricles it may vary up to 450 ms. [33] Therefore, any sodium channel isoform sensitive to lamotrigine in the heart will have a much longer inactivation time. In individuals with tachycardia, this mechanism of action enhances an anti-arrhythmic effect. However, in individuals without tachycardia, the increase in the refractory period may cause bradycardia and other related cardiac arrhythmias. Concurrently, late sodium current contributes to longer QT intervals and may thus trigger arrhythmias. [31]

The warning of the FDA on lamotrigine [4] warranted clinical studies: multiple studies on healthcare registries were then performed. However, their results did not clarify whether there was an increased risk of the overall-mortality compared to other antiseizure medications, both in the general adult population [34–36] and in older individuals. [37,38] Our study is the first observational study to use nationwide real-world data for predicting plasma concentration to investigate the association between plasma concentrations of lamotrigine and its related toxicity [39,40]. We found that some patients experienced fatal effects within the therapeutic range, which may be explained by inter-patient variability. [41] Several individual features such as body weight, gender, duration of therapy, and genetic polymorphisms can influence serum levels of lamotrigine, which is why therapeutic drug monitoring of antiepileptic drugs is strongly recommended.

[18] The importance of inter-individual variability among patients might explain the better prediction we found of individual-patient pharmacometric models, which considered individual demographic characteristics reported in our sample.

Strengths and limitations

One of the strengths of our study is the application of a novel framework along with real-world data. Our analysis can therefore be considered appropriate to better explore issues raised by Health Authorities. For instance, we were able to include a wide sample of individuals aged over 65 years (mean age 70 years) and this criterion was based on the FDA's warning concerning the increased risk of arrhythmias and cardiac death mostly in patients with heart diseases. [4] Individuals aged over 65 years that use lamotrigine and have pre-existing cardiovascular disorders are preferable for investigating our aims for the following reasons. 1) The incidence of epilepsy has a bimodal distribution with the highest risk in the older people. [42] 2) In the older population, cerebrovascular diseases (e.g. stroke) are the most common cause of epilepsy and seizures. [43] 3) Age-related changes in body composition and function and the co-existence of epilepsy with other chronic diseases complicates clinical treatment and increases the risk of clinically relevant drug-drug interaction and pharmacokinetics issues which, in turn, increase the risk of reaching toxic plasma concentrations. [44–46].

The major limitation of our study is that we could not assess the relationship between mortality and plasma concentration in all patients using lamotrigine in the Danish registry, since only 432 cases had at least one serum measurement. Therefore, the results indicating the toxic range of lamotrigine associated with a higher incidence of mortality should be inferred cautiously to individuals aged over 65 years without plasma measurements in the Danish register-based cohort. Moreover, in real-world data, there usually are missing covariates such as genetic profiles, but we were able to impute them based on available literature and multiple imputation.

Conclusion

By using our novel framework PHARMACOMEPI, we found strong evidence to support our hypothesis that the increased risk of all-cause and cardiovascular death among older lamotrigine users was associated with a toxic plasma concentration of lamotrigine.

Reference

[1] H. Soeorg, E. Sverrisdóttir, M. Andersen, T.M. Lund, M. Sessa, The PHARMACOM-EPI Framework for Integrating Pharmacometric Modelling Into Pharmacoepidemiological Research Using Real-World Data: Application to Assess Death Associated With Valproate, Clin Pharmacol Ther. 111 (2022) 840–856. https://doi.org/10.1002/CPT.2502.

[2] H. Soeorg, E. Sverrisdóttir, M. Andersen, T.M. Lund, M. Sessa, Artificial Neural Network vs. Pharmacometric Model for Population Prediction of Plasma Concentration in Real-World Data: A Case Study on Valproic Acid, Clin Pharmacol Ther. 111 (2022) 1278–1285. https://doi.org/10.1002/CPT.2577.

[3] J.B. Woillard, M. Labriffe, A. Prémaud, P. Marquet, Estimation of drug exposure by machine learning based on simulations from published pharmacokinetic models: The example of tacrolimus, Pharmacol Res. 167 (2021). https://doi.org/10.1016/J.PHRS.2021.105578.

[4] Studies show increased risk of heart rhythm problems with seizure and mental health medicine lamotrigine (Lamictal) in patients with heart disease | FDA, (n.d.). https://www.fda.gov/drugs/drug-safety-and-availability/studies-show-increased-risk-heart-rhythm-problems-seizure-and-mental-health-medicine-lamotrigine (accessed April 3, 2023).

[5] L. Ingleby-Talecki, S.C. van Dijkman, S.P. Oosterholt, O. Della Pasqua, C. Winter, M. Cunnington, L. Rebar, S. Forero-Schwanhaeuser, V. Patel, J.A. Cooper, A. Bahinski, K.W. Chaudhary, Cardiac sodium channel inhibition by lamotrigine: In vitro characterization and clinical implications, Clin Transl Sci. 15 (2022) 1978–1989. https://doi.org/10.1111/CTS.13311.

[6] G. Morciano, A. Rimessi, S. Patergnani, V.A.M. Vitto, A. Danese, A. Kahsay, L. Palumbo, M. Bonora, M.R. Wieckowski, C. Giorgi, P. Pinton, Calcium dysregulation in heart diseases: Targeting calcium channels to achieve a correct calcium homeostasis, Pharmacol Res. 177 (2022). https://doi.org/10.1016/J.PHRS.2022.106119.

[7] Y. Nakatani, H. Masuko, T. Amano, Effect of lamotrigine on Na(v)1.4 voltage-gated sodium channels, J Pharmacol Sci. 123 (2013) 203–206. https://doi.org/10.1254/JPHS.13116SC.

[8] R.S. Lott, Lamotrigine: treatment of epilepsy in the elderly., Consult Pharm. 18 (2003) 979–92. http://www.ncbi.nlm.nih.gov/pubmed/16563069 (accessed April 3, 2023).

[9] M. Sessa, A.R. Khan, D. Liang, M. Andersen, M. Kulahci, Artificial Intelligence in Pharmacoepidemiology: A Systematic Review. Part 1-Overview of Knowledge Discovery Techniques in Artificial Intelligence, Front Pharmacol. 11 (2020). https://doi.org/10.3389/FPHAR.2020.01028.

[10] M. Sessa, D. Liang, A.R. Khan, M. Kulahci, M. Andersen, Artificial Intelligence in Pharmacoepidemiology: A Systematic Review. Part 2-Comparison of the Performance of Artificial Intelligence and Traditional Pharmacoepidemiological Techniques, Front Pharmacol. 11 (2021). https://doi.org/10.3389/FPHAR.2020.568659.

[11] C.B. Pedersen, The Danish civil registration system, Scand J Public Health. 39 (2011) 22–25. https://doi.org/10.1177/1403494810387965.

[12] D.H.-S. for O. and R. Diseases, undefined 2016, The Danish National Patient Register, Soard.Org. (n.d.). https://www.soard.org/article/S1550-7289(15)01039-4/abstract (accessed April 3, 2023).

[13] D. Gaist, H. Sørensen, J.H.-D. medical bulletin, undefined 1997, The Danish prescription registries., Europepmc.Org. (n.d.). https://europepmc.org/article/med/9377907 (accessed April 3, 2023).
[14] K.H.-L.-S. journal of public health, undefined 2011, The Danish register of causes of death, Journals.Sagepub.Com. 39 (2011) 26–29. https://doi.org/10.1177/1403494811399958.
[15] J. Frederik, H. Arendt, A.T. Hansen, A. Ladefoged, H.T. Sørensen, L. Pedersen, K. Adelborg, Existing data sources in clinical epidemiology: laboratory information system databases in Denmark, Taylor & Francis. 12 (2020) 469–475. https://doi.org/10.2147/CLEP.S245060.
[16] J. Christensen, M. Vestergaard, J. Olsen, P. Sidenius, Validation of epilepsy diagnoses in the Danish National Hospital Register, Epilepsy Res. 75 (2007) 162–170. https://doi.org/10.1016/J.EPLEPSYRES.2007.05.009.
[17] D.W. Kim, H.K. Kim, E.K. Bae, Switching from traditional sodium channel blockers to lacosamide in patients with epilepsy, Seizure. 65 (2019) 172–175. https://doi.org/10.1016/J.SEIZURE.2019.01.027.
[18] J. Methaneethorn, N. Leelakanok, Sources of lamotrigine pharmacokinetic variability: A systematic review of population pharmacokinetic analyses, Seizure. 82 (2020) 133–147. https://doi.org/10.1016/J.SEIZURE.2020.07.014.
[19] R.J. Bauer, NONMEM Tutorial Part I: Description of Commands and Options, With Simple Examples of Population Analysis, CPT Pharmacometrics Syst Pharmacol. 8 (2019) 525–537. https://doi.org/10.1002/PSP4.12404.
[20] J. Tinggaard, L. Aksglaede, K. Sørensen, A. Mouritsen, C. Wohlfahrt-Veje, C.P. Hagen, M.G. Mieritz, N. Jørgensen, O.D. Wolthers, C. Heuck, J.H. Petersen, K.M. Main, A. Juul, The 2014 Danish references from birth to 20 years for height, weight and body mass index, Acta Paediatr. 103 (2014) 214–224. https://doi.org/10.1111/APA.12468.
[21] A.S. Levey, J.P. Bosch, J.B. Lewis, T. Greene, N. Rogers, D. Roth, A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group., Ann Intern Med. 130 (1999) 461–70. https://doi.org/10.7326/0003-4819-130-6-199903160-00002.
[22] S.T. Sherry, M.H. Ward, M. Kholodov, J. Baker, L. Phan, E.M. Smigielski, K. Sirotkin, dbSNP: the NCBI database of genetic variation, Nucleic Acids Res. 29 (2001) 308–311. https://doi.org/10.1093/NAR/29.1.308.
[23] M. Meaidi, H. Støvring, K. Rostgaard, C. Torp-Pedersen, K.H. Kragholm, M. Andersen, M. Sessa, Pharmacoepidemiological methods for computing the duration of pharmacological prescriptions using secondary data sources, Eur J Clin Pharmacol. 77 (2021) 1805–1814. https://doi.org/10.1007/S00228-021-03188-9.
[24] L. Pazzagli, M. Andersen, M. Sessa, Pharmacological and epidemiological considerations while constructing treatment episodes using observational data: A simulation study, Pharmacoepidemiol Drug Saf. 31 (2022) 55–60. https://doi.org/10.1002/PDS.5366.
[25] L. Pazzagli, D. Liang, M. Andersen, M. Linder, A.R. Khan, M. Sessa, Rationale and performances of a data-driven method for computing the duration of pharmacological prescriptions using secondary data sources, Sci Rep. 12 (2022). https://doi.org/10.1038/S41598-022-10144-9.
[26] P.S. Breitenstein, I. Mahmoud, F. Al-Azzawi, S. Shakibfar, M. Sessa, A machine-learning guided method for predicting add-on and switch in secondary data sources: A case study on anti-seizure medications in Danish registries, Front Pharmacol. 13 (2022). https://doi.org/10.3389/FPHAR.2022.954393.

[27] A. Elixhauser, C. Steiner, D.R. Harris, R.M. Coffey, Comorbidity measures for use with administrative data, Med Care. 36 (1998) 8–27. https://doi.org/10.1097/00005650-199801000-00004.
[28] C. Johannessen Landmark, S.I. Johannessen, P.N. Patsalos, Therapeutic drug monitoring of antiepileptic drugs: current status and future prospects, Expert Opin Drug Metab Toxicol. 16 (2020) 227–238. https://doi.org/10.1080/17425255.2020.1724956.
[29] C.E. Chávez-Castillo, S.E. Medellín-Garibay, R. del C. Milán-Segovia, I. Rodríguez-Leyva, S. Romano-Moreno, Dosing Recommendations Based on Population Pharmacokinetics of Lamotrigine in Mexican Adult Patients With Epilepsy, J Pharm Sci. 109 (2020) 2902–2908. https://doi.org/10.1016/J.XPHS.2020.05.030.
[30] J.N. Nogar, A.B. Minns, D.J. Savaser, B.T. Ly, Severe sodium channel blockade and cardiovascular collapse due to a massive lamotrigine overdose, Clin Toxicol (Phila). 49 (2011) 854–857. https://doi.org/10.3109/15563650.2011.617307.
[31] J.A. Contreras Vite, C. Vega Valle, H. Biekeu Mbem, S.M. Boivin, R. Dumaine, Epilepsy-Induced High Affinity Blockade of the Cardiac Sodium Current INa by Lamotrigine; A Potential for Acquired Arrythmias, Pharmaceuticals (Basel). 15 (2022). https://doi.org/10.3390/PH15101208.
[32] B.P. Bean, The action potential in mammalian central neurons, Nat Rev Neurosci. 8 (2007) 451–465. https://doi.org/10.1038/NRN2148.
[33] C. Antzelevitch, R. Dumaine, Electrical Heterogeneity in the Heart: Physiological, Pharmacological and Clinical Implications, Compr Physiol. (2002) 654–692. https://doi.org/10.1002/CPHY.CP020117.
[34] C.Y. Hsu, C.Y. Cheng, J. Der Lee, M. Lee, B. Ovbiagele, Effects of long-term anti-seizure medication monotherapy on all-cause death in patients with post-stroke epilepsy: a nationwide population-based study in Taiwan, BMC Neurol. 21 (2021). https://doi.org/10.1186/S12883-021-02241-5.
[35] D. Larsson, A. Baftiu, C. Johannessen Landmark, M. Von Euler, E. Kumlien, S. Åsberg, J. Zelano, Association Between Antiseizure Drug Monotherapy and Mortality for Patients With Poststroke Epilepsy, JAMA Neurol. 79 (2022) 169–175. https://doi.org/10.1001/JAMANEUROL.2021.4584.
[36] J.B. Olesen, S.Z. Abildstrøm, J. Erdal, G.H. Gislason, P. Weeke, C. Andersson, C. Torp-Pedersen, P.R. Hansen, Effects of epilepsy and selected antiepileptic drugs on risk of myocardial infarction, stroke, and death in patients with or without previous stroke: a nationwide cohort study, Pharmacoepidemiol Drug Saf. 20 (2011) 964–971. https://doi.org/10.1002/PDS.2186.
[37] D. Liang, E. Gardella, K. Kragholm, C. Polcwiartek, M. Sessa, The Relationship Between Valproate and Lamotrigine/Levetiracetam Use and Prognosis in Patients With Epilepsy and Heart Failure: A Danish Register-Based Study, J Card Fail. 28 (2022) 630–638. https://doi.org/10.1016/J.CARDFAIL.2021.07.020.
[38] A.S. Hauser, A.J. Kooistra, E. Sverrisdóttir, M. Sessa, Utilizing drug-target-event relationships to unveil safety patterns in pharmacovigilance, Expert Opin Drug Saf. 19 (2020) 961–968. https://doi.org/10.1080/14740338.2020.1780208.
[39] J. Christensen, B.B. Trabjerg, J.W. Dreier, Cardiac morbidity and mortality associated with the use of lamotrigine, Epilepsia. 63 (2022) 2371–2380. https://doi.org/10.1111/EPI.17339.
[40] S. Diederich, U. Hemmeter, M. Paulmann, M. Mockenhaupt, Effects of dosage in new users of lamotrigine inducing epidermal necrolysis: Results of the German Registry of Severe Skin Reactions, Epilepsia. (2023). https://doi.org/10.1111/EPI.17563.

[41] D. Battino, D. Croci, T. Granata, M. Estienne, F. Pisani, G. Avanzini, Lamotrigine plasma concentrations in children and adults: influence of age and associated therapy, Ther Drug Monit. 19 (1997) 620–627. https://doi.org/10.1097/00007691-199712000-00003.

[42] A. Sen, N. Jette, M. Husain, J.S.-T. Lancet, undefined 2020, Epilepsy in older people, Elsevier. (n.d.). https://www.sciencedirect.com/science/article/pii/S0140673619330648 (accessed April 3, 2023).

[43] M. Brodie, A. Elder, P.K.-T.L. Neurology, undefined 2009, Epilepsy in later life, Elsevier. (n.d.). https://www.sciencedirect.com/science/article/pii/S1474442209702406 (accessed April 3, 2023).

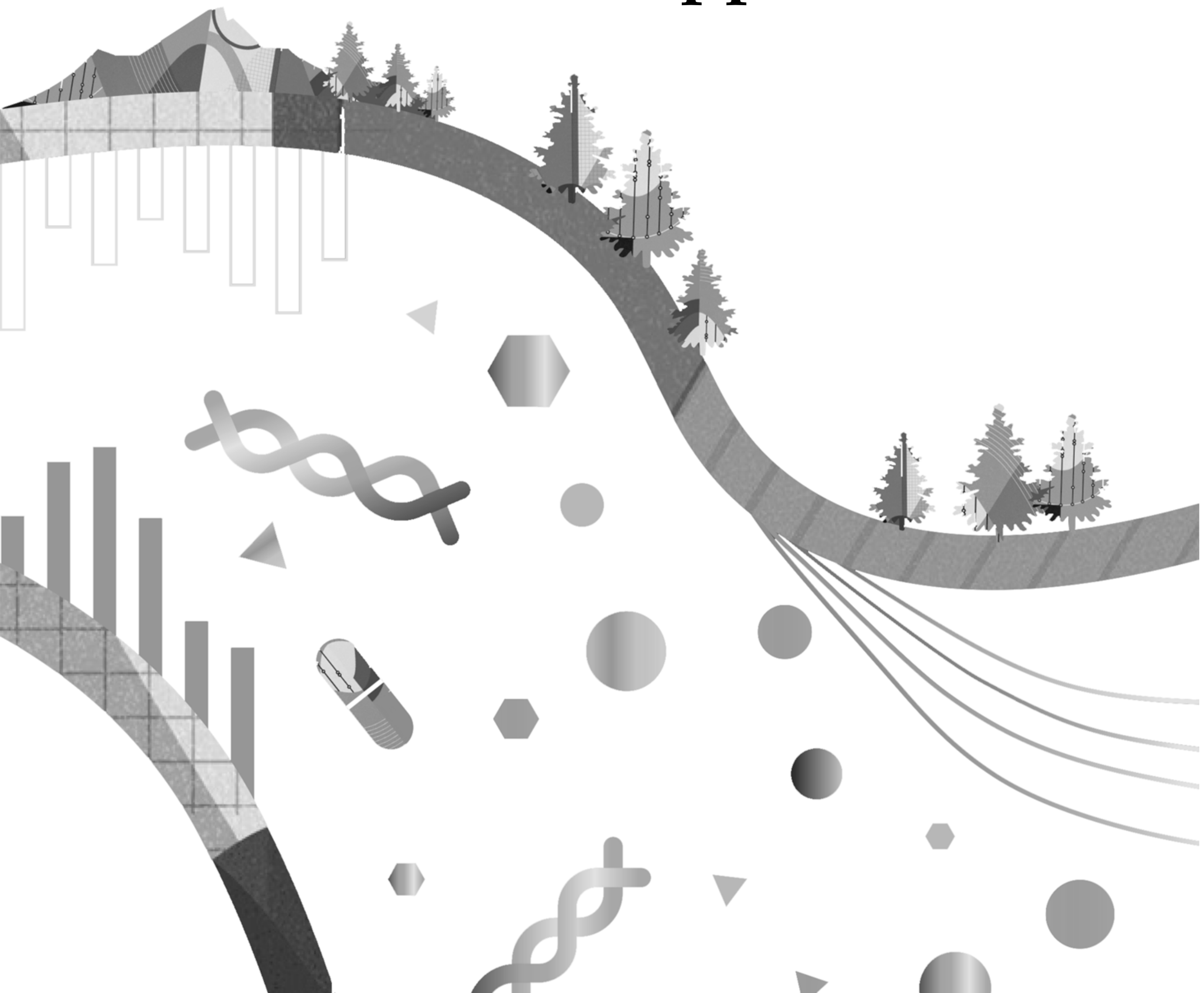
[44] Valproate Sodium Interactions - Drugs.com, (n.d.). https://www.drugs.com/drug-interactions/valproic-acid,valproate-sodium.html (accessed April 3, 2023).

[45] F. Cucchiara, S. Ferraro, G. Luci, G. Bocci, Relevant pharmacological interactions between alkylating agents and antiepileptic drugs: Preclinical and clinical data, Pharmacol Res. 175 (2022). https://doi.org/10.1016/J.PHRS.2021.105976.

[46] C. Carnovale, G. Mosini, M. Gringeri, V. Battini, F. Mazhar, M. Pozzi, E. Clementi, S. Radice, Interaction between paracetamol and lamotrigine: new insights from the FDA Adverse Event Reporting System (FAERS) database, Eur J Clin Pharmacol. 75 (2019) 1323–1325. https://doi.org/10.1007/S00228-019-02691-4.

Part V

Discussion, Future perspective and Appendices



Chapter 7

Discussion and Future Perspective

The overarching aim of this thesis is to gain insight into the etiology and pathophysiology of cardiometabolic disease(s) by investigating potential (causal) risk factors and the interplay between genetic and environmental factors in cardiometabolic diseases. In this thesis, we employed cutting-edge statistical and epidemiological methodologies like clustered Mendelian Randomization (MR) and exploited multiple omics data (genomics and metabolomics) to unravel the mechanisms underlying the development of cardiometabolic diseases using established high-quality databases. In this chapter, I will summarize, interpret and discuss the findings from the different chapters of this thesis. In addition, limitations and future perspectives derived from those findings will be discussed to support progress for future prevention and treatment of cardiometabolic diseases.

MAIN FINDINGS

Insulin Sensitivity and Metabolomics

Insulin resistance plays a critical role in the development of type 2 diabetes (T2D) [1, 2], and is a risk factor for cardiovascular disease [1, 3]. Multiple studies have investigated metabolomic profiles in relation to insulin resistance [4, 5]. However, few studies have focused on the physiological effects of different doses of exogeneous administered insulin on metabolomic measures in individuals without diabetes. In **Chapter 2**, we examined the response of metabolomic measures (e.g., lipoprotein sub-particles and amino acids) to different hyper-insulinemic levels under euglycemic conditions in non-diabetic individuals. We found that the majority of metabolomic measures were sensitive to insulin, and a large fraction of them were differentially insulin-sensitive. This indicates that metabolites are likely to be differentially affected by the degree of hyperinsulinemia, due to insulin resistance. In particular, branched-chain amino acids (BCAAs), showing the highest reduction following high-dose insulin infusion in comparison to low-dose insulin infusion, were the most sensitive metabolomic measures to insulin in this study. Our results are in line with multiple studies showing elevated BCAAs in obese and insulin-resistant individuals [6-8]. In addition, multiple metabolomic measures observed as insulin-sensitive in this study have been described before to be associated with risk of development of cardiometabolic diseases. For instance, increased concentrations of BCAAs are associated with increased risk of coronary artery disease and T2D development [9, 10]. Especially, there is evidence from Mendelian randomization studies that the association between BCAAs and T2D is causal

[11]. As the majority of the studies focused on the pathophysiological effect of insulin resistance on metabolomic profiles, our study filled a gap by assessment of direct effects of hyperinsulinemia effect on metabolomic measures in individuals without diabetes.

Pathophysiology of Type 2 Diabetes Mellitus

Technological development of measuring techniques has made it possible to directly and indirectly quantify mitochondrial DNA more readily and affordably than ever before, which has led to wide availability of mitochondrial DNA related data [12]. As a result, an increasing number of epidemiological studies on the relationship between T2D and mitochondrial function have been published [13-15]. Most investigations have found an association between blood mitochondrial DNA copy number (mtDNA-CN) and T2D [16-19], but no study has investigated whether or not this association is potentially causal. In **Chapter 3** of this thesis, we investigated the potential causal association between blood mtDNA-CN and T2D and body mass index (BMI) using bi-directional two-sample MR analyses. Overall, the results did not provide evidence for a causal association between blood mtDNA-CN and T2D and BMI in either direction, despite being studied in multiple very large study samples (e.g., UK Biobank). Additional analyses indicated that blood mtDNA-CN may not reflect mtDNA-CN in T2D related organs like muscle and liver, which could explain the lack of a potential causal association. The mtDNA-CN in blood, serving as an indirect indicator of hematopoiesis and immunological processes, may not reflect the mitochondrial function in muscle or adipose tissues [12]. Consequently, exploring mtDNA-CN in muscle and adipose tissue, which do reflect mitochondrial function in T2D-related tissues, may offer a more robust way for investigating its potential role as a risk factor for T2D. To the best of our knowledge, this study is the first and largest investigation on the potential causal association and its direction between blood mtDNA-CN and T2D. Although blood mtDNA-CN may not be a biomarker for T2D, it might be valuable for studying the pathophysiology of other diseases. For instance, mtDNA-CN was found to be associated with cardiovascular diseases [20], aging [21] and Parkinson’s disease [22], which might suggest its potential as a biomarker for elucidating the underlying pathological mechanisms of these diseases.

Insulin-like growth factor-1 (IGF-1) as a risk factor of T2D has been studied for years, but the epidemiological findings of their association are

inconsistent. We hypothesized that there might be distinct mechanisms underlying the association of high levels of IGF-1 with disease. Some of these mechanisms may contribute to an increased risk of developing T2D, while others might have no association with T2D or could potentially decrease the risk of T2D development. We hypothesized that these mechanisms might interact and can cancel out the overall impact of high IGF-1 levels on T2D. In **Chapter 4** of this thesis, we employed a novel method referred to as clustered MR, and identified six clusters of genetic variants representing IGF-1 levels, each of which showed a different association with T2D. Notably, three of those clusters suggested a positive causal effect of IGF-1 levels on T2D risk, while the remaining three clusters suggested a negative causal effect of IGF-1 levels on T2D risk. This result indicates that high level of IGF-1 likely affects the development of T2D in a context-dependent manner, depending on the biological pathways that increase IGF-1 levels in the blood. Before investigation of the potential causal association between IGF-1 and T2D, we also performed multivariable-adjusted regression analyses and observed a J-shape association between IGF-1 and T2D, with particularly low levels of IGF-1 being associated with an increased risk of developing T2D. Our observational findings are partly in line with several other studies showing a U-shape association between IGF-1 and T2D [23-25], but contrast with studies showing no evidence for this association [26, 27]. This difference in observational studies, to some extent, is in line with our finding in the clustered MR that the effect of high levels of IGF-1 on T2D is driven by different biological mechanism and that their association is context-dependent. Our study provides a possible reason for the inconsistent observational findings of the association between IGF-1 and type 2 diabetes. Given the use of diverse cohorts consisting of heterogenous participants in investigating the association between IGF-1 and T2D, the fluctuating participant numbers contributing to either increased or decreased risk of T2D development possibly result in inconsistent observational findings concerning the association between IGF-1 and T2D. Understanding the role of IGF-1 in T2D can help us to become more aware of the heterogeneity between individuals and direct the development of personalized medicine like targeted therapy or lifestyle modification.

Pathophysiology of Sleep-Associated Dyslipidemia

Habitual sleep is becoming increasingly recognized as a risk factor contributing to the development of cardiometabolic diseases and associated traits (i.e., dyslipidemia) [28]. Given the poorly understood biomolecular

mechanisms underpinning the association between habitual sleep duration and dyslipidemia, **Chapter 5** of this thesis examined the effect of genetic factors and habitual sleep duration on lipid traits including high-density lipoprotein cholesterol (HDL), low-density lipoprotein cholesterol (LDL) and triglycerides (TG). Through the analyses on approximately 732,564 (predominantly European) individuals from multiple population groups across 55 cohorts, our study revealed that the associations of several genetic variants with lipids were modified by either long or short habitual sleep duration. Among the identified lead variants, some were found to be located within the *ASPH* gene which is known to modulate cardiovascular risk by interacting with aspartic and succinic acid [29]. Notably, some additionally identified variants were mapped to the *SLC8A1* gene which has emerged as a therapeutic target for ischemic damage reduction following acute myocardial infarction [30, 31]. Despite the large sample size, especially in comparison with the previous effort [32], we only identified a relatively small number of genetic variants that were modified by differences in habitual sleep duration. Furthermore, most variants had low allele frequencies (i.e., below 1%). Therefore, sleep duration might not be the optimal phenotype to investigate the relationship between genetic variants and lipid levels as sleep is a highly multidimensional trait [33]. Nevertheless, this study offers novel insights into biological mechanism underlying the association between habitual sleep duration and lipid levels. Furthermore, the identified targets for treatment can provide novel insight into the prevention of atherosclerotic cardiovascular disease in individuals with disturbances in habitual sleep duration.

A novel framework - Integration of Pharmacometrics and Epidemiology

A novel methodological framework, PHARMACOM-EPI, was introduced in **Chapter 6** of this thesis. This framework integrated pharmacometrics and epidemiology, which enabled the prediction of drug concentration at the time of occurrence of clinical outcomes (i.e., mortality). In early 2021, a warning issued by U.S. Food and Drug Administration on the antiseizure drug lamotrigine, which claims that lamotrigine has the potential to increase risk of arrhythmias and related sudden cardiac death. Our hypothesis is that the increase risk of arrhythmias and related sudden cardiac death results from drug toxicity. In **Chapter 6**, we applied the PHARMACOM-EPI framework to real-world data to investigate the association between plasma concentration of lamotrigine and the risk of death in older patients. The results showed that a toxic plasma concentration of lamotrigine was

associated with increased risk of all-cause and cardiovascular death in the older lamotrigine users, which supports our hypothesis. Our study is the first observational study using nationwide real-world data to predict plasma concentration to investigate the relationship between lamotrigine plasma concentrations and its related toxicity [34, 35]. This study introduced the PHARMACOM-EPI framework, which is capable of handling real-world data, thereby showcasing its potential applications in future research. By leveraging this methodology, observational data could be harnessed more effectively. Given ethical consideration and cost constrains, clinical trial data containing plasma concentration is not always available. This approach enables observational data to function as a valuable alternative under specific circumstances. When drug dosages are available in observational data, drug plasma concentrations can be predicted and applied in the pharmaco-epidemiological studies or as better-defined confounders that usually ignored in the epidemiological studies. Simultaneously, this study highlights the importance of personalized medicine within distinct population groups (older and pediatric populations) especially when dealing with drugs characterized by narrow therapeutic windows.

Limitations and Future Perspective

Most of the results described in this thesis were limited to mainly European ancestry populations, and the results in this thesis should therefore be interpreted with caution for populations of non-European backgrounds. Future studies within non-European populations are warranted for those findings.

With the advent of new methods for estimation of causal effects, has two-sample Mendelian Randomization become less important?

Considering the limitations of conventional two-sample MR, more and more advanced methods have been developed to investigate the potential causal association between exposures and outcomes. Two-sample MR assumes that the exposure-outcome relation is linear. To address violations of this assumption, non-linear MR was put forward to investigate non-linear causal relationship between exposures and outcomes [36]. Two-step MR and multivariable MR was developed with the requirement to investigate the potential causal association between an exposure and an outcome with mediators [37, 38]. To address the pleiotropic effects limitation in two-sample MR, clustered MR was developed to explore different biological pathways between the exposure and outcome [39], especially in the case of large heterogeneity between the estimated causal effects of individual instruments on the study outcome. With the development of these advanced methods developed to estimate causal effects, does this mean that the conventional two-sample MR is no longer relevant? First of all, it is clear that all those advanced methods were developed based on the original principles of two-sample MR. This indicates conventional MR is the pillar of all the advanced methods. Besides, it is more logical to first perform two-sample MR before moving to some advanced methods. For example, in **Chapter 4**, we used the clustered MR to investigate the association between IGF-1 and type 2 diabetes. The rationale of using the advanced method, clustered MR, is that we observed heterogeneity of the estimated causal effects of instrumental variables in two-sample MR. Basic studies are always the foundation for more in-depth exploration. In summary, although two-sample MR analyses have their limitations, the role of this method in estimation of causal effect cannot be replaced.

Is habitual sleep duration as a lifestyle factor enough to capture the disturbance in lipid profiles?

In **Part III** of this thesis, **Chapter 5** focused on the effect of interaction between genetics and sleep duration on lipid traits in a large multiple-ethnic population. In this study, we solely examined the effect of sleep duration on lipids, and did not account for the complexity of sleep, which is a multidimensional trait that covers a wide spectrum of phenotypes including sleep duration, napping, insomnia, chronotype, etc. A significant amount of research has been focused on the effect of different sleep traits on cardiometabolic diseases. Some studies provided evidence that sleep duration [40, 41], napping [42-44], insomnia [45, 46] and chronotype [47, 48] were associated with risk of cardiometabolic diseases. MR analyses provided evidence that sleep duration, insomnia and daytime napping are causally associated with cardiometabolic diseases [42]. Although many studies have addressed the effect of sleep traits on cardiometabolic diseases, limited studies have taken into account the interaction between different sleep phenotypes [49-51]. These studies suggested that sleep duration is not the only sleep trait contributing to the risk of cardiometabolic diseases. In future studies, we may need to focus more on other sleep traits and on the combination of different sleep phenotypes affecting cardiometabolic diseases. In addition to sleep, it is also important to examine the effect on cardiometabolic diseases from other related lifestyle factors such as exercise, diet, and explore their potential interactions.

Is medication use overlooked in epidemiological observational studies? Can we benefit from multi-disciplinary approaches to handle this confounder?

One of the major challenges in epidemiological observational studies of cardiometabolic disease is medication use by the study participants. According to the paper published by *Jungyeon, et.al.*, many epidemiological studies have not considered or given sufficient attention to medication use, which may have impacted the results leading to bias or wrong interpretation [52]. In this paper, etiological studies from high-ranked journals published between 2015 and 2019 focusing on blood pressure, glucose or lipid measurements were assessed. The results indicated that a large proportion of studies (47%) did not report information on medication use or only provided medication use for part of the variables affected (14%) [52]. One of the main reasons for not considering medication use is absence or insufficiency of data

for medication use in the respective databases. To address this issue, we need to raise awareness about the importance of recording medication use during data collection. **Chapter 6** in **Part IV** of this thesis introduced a novel framework integrating pharmacometrics and epidemiological disciplines to predict plasma concentrations based on dosages and then associated predicted drug concentration with clinical outcomes of interest. Given that the dosage information is accessible, **Chapter 6** may provide the opportunity to address the issue of the lack of medication concentration data. Further efforts are required to determine how to deal with insufficient medication data and how to handle medication use in observational research.

Does multi-omics pave the road toward achieving precision medicine?

The findings presented in **Chapter 4** of this thesis provide evidence for considerable individual heterogeneity, and **Chapter 6** stressed the importance of personalized medicine especially for medication with a narrow therapeutic window. These observations provide arguments to further develop the discipline of precision medicine. According to US Food and Drug Administration (FDA), precision medicine is an approach to provide a precise strategy for prevention, diagnosis and treatment of disease taking into account people's differences in genes, lifestyles and environments [53]. In the current thesis, genomics and metabolomics were used to identify genetic variants that were associated with cardiometabolic disease, and to unravel underlying biological pathways of cardiometabolic disease to increase our awareness of personalized medicine. However, neither genomics alone nor other single omics cannot capture the complex holistic biological processes underlying most diseases [54]. With the advancement of different omics techniques, high-throughput data in the disciplines of genomics, epigenomics, transcriptomics, proteomics, and metabolomics can now be relatively easily and cheaply obtained [55]. Integration of multiple omics data (which is called multi-omics) needs to be applied in future studies to provide a more comprehensive view of the etiology and pathophysiology of cardiometabolic diseases [56]. With the increasing availability of multi-omics data and awareness of the importance of biological interconnections, network-based approaches are rapidly emerging as a promising method which represent biological systems as interconnected networks of bio-molecules (including genes, proteins, metabolites, etc.) [57]. Employing network-based approaches for analyzing multi-omics data enables the integration of biochemical process which includes gene regulation of protein production, protein interaction with each other to carry out cellular functions

and the interconnectivity of metabolites across different metabolic pathways [58]. By comprehensive elucidation of disease pathophysiology, this approach has the potential to significantly advance both personalized prevention and treatment of diseases [59]. The realization of personalized or stratified medicine can be achieved through patient stratification or disease subtyping based on identified mechanisms and identification of risk factors for distinct populations.

Reference

[1]Ginsberg HN (2000) Insulin resistance and cardiovascular disease. J Clin Invest 106(4): 453-458. 10.1172/jci10762
[2]Goldstein BJ (2002) Insulin resistance as the core defect in type 2 diabetes mellitus. The American Journal of Cardiology 90(5, Supplement 1): 3-10. https://doi.org/10.1016/S0002-9149(02)02553-5
[3]Kosmas CE, Bousvarou MD, Kostara CE, Papakonstantinou EJ, Salamou E, Guzman E (2023) Insulin resistance and cardiovascular disease. J Int Med Res 51(3): 3000605231164548. 10.1177/03000605231164548
[4]Gu X, Al Dubayee M, Alshahrani A, et al. (2020) Distinctive Metabolomics Patterns Associated With Insulin Resistance and Type 2 Diabetes Mellitus. Front Mol Biosci 7: 609806. 10.3389/fmolb.2020.609806
[5]Palmer ND, Stevens RD, Antinozzi PA, et al. (2015) Metabolomic Profile Associated With Insulin Resistance and Conversion to Diabetes in the Insulin Resistance Atherosclerosis Study. The Journal of Clinical Endocrinology & Metabolism 100(3): E463-E468. 10.1210/jc.2014-2357
[6]Tai ES, Tan ML, Stevens RD, et al. (2010) Insulin resistance is associated with a metabolic profile of altered protein metabolism in Chinese and Asian-Indian men. Diabetologia 53(4): 757-767. 10.1007/s00125-009-1637-8
[7]Newgard CB, An J, Bain JR, et al. (2009) A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. Cell Metab 9(4): 311-326. 10.1016/j.cmet.2009.02.002
[8]Kim JY, Park JY, Kim OY, et al. (2010) Metabolic profiling of plasma in overweight/obese and lean men using ultra performance liquid chromatography and Q-TOF mass spectrometry (UPLC-Q-TOF MS). J Proteome Res 9(9): 4368-4375. 10.1021/pr100101p
[9]Chen X, Yang W (2015) Branched-chain amino acids and the association with type 2 diabetes. J Diabetes Investig 6(4): 369-370. 10.1111/jdi.12345
[10]Flores-Guerrero JL, Osté MCJ, Kieneker LM, et al. (2018) Plasma Branched-Chain Amino Acids and Risk of Incident Type 2 Diabetes: Results from the PREVEND Prospective Cohort Study. Journal of Clinical Medicine 7(12): 513
[11]Lotta LA, Scott RA, Sharp SJ, et al. (2016) Genetic Predisposition to an Impaired Metabolism of the Branched-Chain Amino Acids and Risk of Type 2 Diabetes: A Mendelian Randomisation Analysis. PLoS Med 13(11): e1002179. 10.1371/journal.pmed.1002179
[12]Picard M (2021) Blood mitochondrial DNA copy number: What are we counting? Mitochondrion 60: 1-11. https://doi.org/10.1016/j.mito.2021.06.010
[13]Poulton J, Luan J, Macaulay V, Hennings S, Mitchell J, Wareham NJ (2002) Type 2 diabetes is associated with a common mitochondrial variant: evidence from a population-based case-control study. Hum Mol Genet 11(13): 1581-1583. 10.1093/hmg/11.13.1581
[14]Park KS, Chan JC, Chuang LM, et al. (2008) A mitochondrial DNA variant at position 16189 is associated with type 2 diabetes mellitus in Asians. Diabetologia 51(4): 602-608. 10.1007/s00125-008-0933-z
[15]Jiang W, Li R, Zhang Y, et al. (2017) Mitochondrial DNA Mutations Associated with Type 2 Diabetes Mellitus in Chinese Uyghur Population. Scientific Reports 7(1): 16989. 10.1038/s41598-017-17086-7
[16]Memon AA, Sundquist J, Hedelius A, Palmér K, Wang X, Sundquist K (2021) Association of mitochondrial DNA copy number with prevalent and incident type 2 diabetes in women: A population-based follow-up study. Sci Rep 11(1): 4608. 10.1038/s41598-021-84132-w

[17]Liu X, Longchamps RJ, Wiggins KL, et al. (2021) Association of mitochondrial DNA copy number with cardiometabolic diseases. Cell Genomics 1(1): 100006. https://doi.org/10.1016/j.xgen.2021.100006
[18]Reiling E, Ling C, Uitterlinden AG, et al. (2010) The association of mitochondrial content with prevalent and incident type 2 diabetes. J Clin Endocrinol Metab 95(4): 1909-1915. 10.1210/jc.2009-1775
[19]Fazzini F, Lamina C, Raftopoulou A, et al. Association of mitochondrial DNA copy number with metabolic syndrome and type 2 diabetes in 14 176 individuals. Journal of Internal Medicine n/a(n/a). https://doi.org/10.1111/joim.13242
[20]Luo J, Noordam R, Jukema JW, et al. (2023) Low leukocyte mitochondrial DNA abundance drives atherosclerotic cardiovascular diseases: a cohort and Mendelian randomization study. Cardiovasc Res 119(4): 998-1007. 10.1093/cvr/cvac182
[21]Mengel-From J, Thinggaard M, Dalgård C, Kyvik KO, Christensen K, Christiansen L (2014) Mitochondrial DNA copy number in peripheral blood cells declines with age and is associated with general health among elderly. Human Genetics 133(9): 1149-1159. 10.1007/s00439-014-1458-9
[22]Pyle A, Anugrha H, Kurzawa-Akanbi M, Yarnall A, Burn D, Hudson G (2016) Reduced mitochondrial DNA copy number is a biomarker of Parkinson's disease. Neurobiology of Aging 38: 216.e217-216.e210. https://doi.org/10.1016/j.neurobiolaging.2015.10.033
[23]Schneider HJ, Friedrich N, Klotsche J, et al. (2011) Prediction of incident diabetes mellitus by baseline IGF1 levels. Eur J Endocrinol 164(2): 223-229. 10.1530/eje-10-0963
[24]Sandhu MS, Heald AH, Gibson JM, Cruickshank JK, Dunger DB, Wareham NJ (2002) Circulating concentrations of insulin-like growth factor-I and development of glucose intolerance: a prospective observational study. Lancet 359(9319): 1740-1745. 10.1016/s0140-6736(02)08655-5
[25]Friedrich N, Thuesen B, Jørgensen T, et al. (2012) The association between IGF-I and insulin resistance: a general population study in Danish adults. Diabetes Care 35(4): 768-773. 10.2337/dc11-1833
[26]Drogan D, Schulze MB, Boeing H, Pischon T (2016) Insulin-Like Growth Factor 1 and Insulin-Like Growth Factor–Binding Protein 3 in Relation to the Risk of Type 2 Diabetes Mellitus: Results From the EPIC–Potsdam Study. American Journal of Epidemiology 183(6): 553-560. 10.1093/aje/kwv188
[27]Similä ME, Kontto JP, Virtamo J, et al. (2019) Insulin-like growth factor I, binding proteins -1 and -3, risk of type 2 diabetes and macronutrient intakes in men. British Journal of Nutrition 121(8): 938-944. 10.1017/S0007114519000321
[28]St-Onge MP, Grandner MA, Brown D, et al. (2016) Sleep Duration and Quality: Impact on Lifestyle Behaviors and Cardiometabolic Health: A Scientific Statement From the American Heart Association. Circulation 134(18): e367-e386. 10.1161/cir.0000000000000444
[29]Pell VR, Chouchani ET, Frezza C, Murphy MP, Krieg T (2016) Succinate metabolism: a new therapeutic target for myocardial reperfusion injury. Cardiovascular Research 111(2): 134-141. 10.1093/cvr/cvw100
[30]Jang IK, Weissman NJ, Picard MH, et al. (2008) A randomized, double-blind, placebo-controlled study of the safety and efficacy of intravenous MCC-135 as an adjunct to primary percutaneous coronary intervention in patients with acute myocardial infarction: Evaluation of MCC-135 for left ventricular salvage in acute myocardial infarction (EVOLVE). Am Heart J 155(1): 113.e111-118. 10.1016/j.ahj.2007.08.020
[31]Kawasumi H, Satoh N, Kitada Y (2007) Caldaret, an intracellular Ca2+ handling modulator, limits infarct size of reperfused canine heart. J Pharmacol Sci 103(2): 222-233. 10.1254/jphs.fp0060765

[32]Noordam R, Bos MM, Wang H, et al. (2019) Multi-ancestry sleep-by-SNP interaction analysis in 126,926 individuals reveals lipid loci stratified by sleep duration. Nature Communications 10(1): 5121. 10.1038/s41467-019-12958-0
[33]Chung J, Goodman M, Huang T, Bertisch S, Redline S (2021) Multidimensional sleep health in a diverse, aging adult cohort: Concepts, advances, and implications for research and intervention. Sleep Health 7(6): 699-707. 10.1016/j.sleh.2021.08.005
[34]Christensen J, Trabjerg BB, Dreier JW (2022) Cardiac morbidity and mortality associated with the use of lamotrigine. Epilepsia 63(9): 2371-2380. 10.1111/epi.17339
[35]Diederich S, Hemmeter U, Paulmann M, Mockenhaupt M (2023) Effects of dosage in new users of lamotrigine inducing epidermal necrolysis: Results of the German Registry of Severe Skin Reactions. Epilepsia 64(5): 1259-1265. https://doi.org/10.1111/epi.17563
[36]Staley JR, Burgess S (2017) Semiparametric methods for estimation of a nonlinear exposure-outcome relationship using instrumental variables with application to Mendelian randomization. Genetic epidemiology 41(4): 341-352. 10.1002/gepi.22041
[37]Burgess S, Thompson SG (2015) Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. Am J Epidemiol 181(4): 251-260. 10.1093/aje/kwu283
[38]Carter AR, Sanderson E, Hammerton G, et al. (2021) Mendelian randomisation for mediation analysis: current methods and challenges for implementation. European Journal of Epidemiology 36(5): 465-478. 10.1007/s10654-021-00757-1
[39]Foley CN, Mason AM, Kirk PDW, Burgess S (2020) MR-Clust: Clustering of genetic variants in Mendelian randomization with similar causal estimates. Bioinformatics. 10.1093/bioinformatics/btaa778
[40]Yan B, Yang J, Zhao B, Fan Y, Wang W, Ma X (2021) Objective Sleep Efficiency Predicts Cardiovascular Disease in a Community Population: The Sleep Heart Health Study. J Am Heart Assoc 10(7): e016201. 10.1161/jaha.120.016201
[41]Yiallourou SR, Maguire GP, Carrington MJ (2021) Sleep quantity and quality and cardiometabolic risk factors in Indigenous Australians. J Sleep Res 30(2): e13067. 10.1111/jsr.13067
[42]Jia Y, Guo D, Sun L, et al. (2022) Self-reported daytime napping, daytime sleepiness, and other sleep phenotypes in the development of cardiometabolic diseases: a Mendelian randomization study. Eur J Prev Cardiol 29(15): 1982-1991. 10.1093/eurjpc/zwac123
[43]Lin D, Sun K, Li F, et al. (2014) Association between habitual daytime napping and metabolic syndrome: a population-based study. Metabolism 63(12): 1520-1527. 10.1016/j.metabol.2014.08.005
[44]Leng Y, Cappuccio FP, Surtees PG, Luben R, Brayne C, Khaw KT (2016) Daytime napping, sleep duration and increased 8-year risk of type 2 diabetes in a British population. Nutr Metab Cardiovasc Dis 26(11): 996-1003. 10.1016/j.numecd.2016.06.006
[45]Laugsand LE, Strand LB, Platou C, Vatten LJ, Janszky I (2014) Insomnia and the risk of incident heart failure: a population study. Eur Heart J 35(21): 1382-1393. 10.1093/eurheartj/eht019
[46]Laugsand LE, Vatten LJ, Platou C, Janszky I (2011) Insomnia and the risk of acute myocardial infarction: a population study. Circulation 124(19): 2073-2081. 10.1161/circulationaha.111.025858
[47]Li T, Xie Y, Tao S, et al. (2023) Prospective study of the association between chronotype and cardiometabolic risk among Chinese young adults. BMC Public Health 23(1): 1966. 10.1186/s12889-023-16902-2
[48]Baldanzi G, Hammar U, Fall T, et al. (2022) Evening chronotype is associated with elevated biomarkers of cardiometabolic risk in the EpiHealth cohort: a cross-sectional study. Sleep 45(2). 10.1093/sleep/zsab226

[49]Goodman MO, Dashti HS, Lane JM, et al. (2023) Causal Association Between Subtypes of Excessive Daytime Sleepiness and Risk of Cardiovascular Diseases. J Am Heart Assoc 12(24): e030568. 10.1161/jaha.122.030568
[50]Li J, Yin J, Luo Y, et al. (2022) Association of healthy sleep pattern with the risk of cardiovascular disease and all-cause mortality among people with diabetes: A prospective cohort study. Diabetes Res Clin Pract 186: 109822. 10.1016/j.diabres.2022.109822
[51]He L, Ma T, Cheng X, Bai Y (2023) The association between sleep characteristics and the risk of all-cause mortality among individuals with cardiometabolic multimorbidity: a prospective study of UK Biobank. J Clin Sleep Med 19(4): 651-658. 10.5664/jcsm.10404
[52]Choi J, Dekkers OM, le Cessie S (2022) How measurements affected by medication use are reported and handled in observational research: A literature review. Pharmacoepidemiol Drug Saf 31(7): 739-748. 10.1002/pds.5437
[53]FDA US (2018) Precision Medicine. Available from https://www.fda.gov/medical-devices/in-vitro-diagnostics/precision-medicine
[54]Karczewski KJ, Snyder MP (2018) Integrative omics for health and disease. Nat Rev Genet 19(5): 299-310. 10.1038/nrg.2018.4
[55]Dai X, Shen L (2022) Advances and Trends in Omics Technology Development. Front Med (Lausanne) 9: 911861. 10.3389/fmed.2022.911861
[56]Kreitmaier P, Katsoula G, Zeggini E (2023) Insights from multi-omics integration in complex disease primary tissues. Trends Genet 39(1): 46-58. 10.1016/j.tig.2022.08.005
[57]Zhou G, Li S, Xia J (2020) Network-Based Approaches for Multi-omics Integration. Methods Mol Biol 2104: 469-487. 10.1007/978-1-0716-0239-3_23
[58]Yan J, Risacher SL, Shen L, Saykin AJ (2018) Network approaches to systems biology analysis of complex disease: integrative methods for multi-omics data. Briefings in Bioinformatics 19(6): 1370-1381. 10.1093/bib/bbx066
[59]Han J-DJ (2008) Understanding biological functions through molecular networks. Cell Research 18(2): 224-237. 10.1038/cr.2008.16

Chapter 8

Appendices

Summary

Samenvatting

PhD Portfolio

Publications and Manuscript

Curriculum Vitae

Acknowledgements

Summary

Despite extensive research, significant gaps remain in fully understanding and preventing the onset of cardiometabolic disease. In this thesis, we aimed to unravel part of the etiology and pathophysiology of cardiometabolic disease by investigating potential (causal) risk factors and the interplay between genetic and environmental factors in cardiometabolic diseases using cutting-edge statistical and epidemiological methods.

In **Chapter 1**, I provided an overview of the disease background, explained the relevant biological concepts, described the methodology employed and the study populations utilized in this thesis.

In **Chapter 2** of this thesis, we focused on insulin sensitivity of metabolomic measures. Although the association between insulin sensitivity and metabolomic measures has been extensively studied, the actual response of these measures to changes in insulin concentration in non-diabetic individuals has been understudied. We examined this response of metabolomic measures under different hyper-insulinemic levels while maintaining euglycemic conditions in middle-aged non-diabetic individuals. In this study, we compared the changes in the concentrations of 151 metabolomic measures after low- and high-dose insulin infusion compared to baseline levels by using linear mixed-effect models for repeated measures. The results showed that 90 out of the 151 metabolomic measures changed in concentration after low-dose insulin infusion compared with baseline, and 121 metabolomic measures changed in concentration after high-dose insulin infusion compared with baseline. Additionally, we compared the changes in concentrations of metabolomics measures after high-dose insulin infusion with those after low-dose insulin infusion. We found that some metabolomic measures seemed to reach their maximum response already at low-insulin infusion, while 99 out of 151 metabolomic measures showed an additional response at high-insulin dose infusion. In particular, the amino acids including leucine, isoleucine, valine, glutamine and tyrosine had a large decrease in concentration after high-dose insulin infusion when compare to baseline and low dose insulin infusion. In conclusion, metabolomic measures are differentially insulin sensitive and may thus be differentially affected by the development of insulin resistance. Those findings also suggested that the response to different insulin levels is likely organ-specific, which is reflected by changes in metabolomic measure after low-dose insulin infusion (primarily affecting metabolomic measures in the liver) and high-dose

insulin infusion (primarily affecting metabolomic measures in peripheral tissues). For patients, this could provide additional insight in the stage or the type of insulin resistance and thus potentially provide opportunities for personalized intervention.

In **Chapter 3** of this thesis, we investigated the relationship between blood mitochondrial function and type 2 diabetes (T2D). Mitochondria are essential organelles for the production of cellular energy and dysfunctional energy production may be cause and/or consequence of T2D. We associated mitochondrial DNA copy number (mtDNA-CN) as a measure for mitochondrial function and associated this measure with T2D incidence. We performed this analysis using multivariable-adjusted Cox proportional hazard models, and we found that a higher level of blood mtDNA-CN was associated with lower risk of developing T2D. To assess whether or not this association is causal, we performed bi-directional two-sample Mendelian Randomization (MR) analyses. The results showed that there was no evidence for a causal association between blood mtDNA-CN and type 2 diabetes in either direction. In order to investigate this relationship further, we analyzed the potential causal association between blood mtDNA-CN and BMI using bi-directional two-sample MR, but no indication for causal associations between these two features was found in either direction. Our hypothesis that blood mtDNA-CN drives the risk of T2D and obesity is based on the assumption that blood mtDNA-CN reflects overall mitochondrial function including mitochondrial function of tissues like liver and muscle that play critical roles in T2D. With some additional analyses, we found that higher blood mtDNA-CN was not associated with lean mass, nor with AST, ALT and ALP as measures of liver function. A slightly lower GGT was observed with an increase in blood mtDAN-CN. Altogether, these results suggest that the observed association between low blood mtDNA-CN and higher risk of type 2 diabetes is likely not causal. This could potentially be explained by our findings that blood mtDNA-CN is not associated with muscle mass and weakly with liver function and may thus not be a marker of muscle mitochondrial function and only a weak marker for liver mitochondrial function.

In **Chapter 4** of this thesis, we investigated the association between blood IGF-1 levels and T2D. Although it is biologically plausible that IGF-1 plays a role in T2D, previous studies have reported inconsistent findings on this association. We first performed Cox proportional hazard analyses on blood IGF-1 levels and T2D, and investigated a possible J-shape association between IGF-1 concentration and T2D, with low levels of IGF-1 being

associated with a higher risk of developing T2D. Subsequently, we performed two-sample MR analyses in the same study population, and observed that higher genetically-influenced IGF-1 was associated with higher risk of T2D. Considering the large heterogeneity in the estimated causal effects of the individual genetic instruments in the two-sample MR, we employed a novel method, clustered MR, to investigate the association between genetically-influenced IGF-1 level with T2D. We identified six clusters of genetically-influenced IGF-1, based on genetic instruments with similar estimated causal effects, to be associated with T2D, among which three clusters were related to a higher risk of T2D while the remaining three clusters were related to a lower risk of T2D. The main clusters in which a higher IGF-1 was associated with a lower risk of type 2 diabetes comprised instruments mapping to genes in the growth-hormone signaling pathway, whereas the main clusters in which a higher IGF-1 was associated with a higher risk of type 2 diabetes comprised instruments mapping to genes in pathways related to amino acid metabolism and genomic integrity. These results provide evidence that the effect of IGF-1 on T2D is dependent on the underlying metabolic/biomolecular pathways.

In **Chapter 5** of this thesis, we aimed to understand the effect of genetic factors and exposure to both long and short habitual sleep duration on dyslipidemia traits including high-density lipoprotein cholesterol (HDL-c), low-density lipoprotein cholesterol (LDL-c) and triglycerides (TG). In this study, 732,564 individuals from multiple population groups across 55 cohorts were involved. Through a 1-degree of freedom genome-wide variant-sleep interaction analyses on lipid levels, we identified 9 lead variants of which variants mapping to AMPD3 (LDL-c, LTST), ASPH (TG, LTST) and DLEU1 (TG, STST; HDL-c, STST) genes were most significant. Seven additional lead variants were identified by a 2-degree of freedom joint analyses, among which the lead variant mapping to SLC8A1 (TG, STST) which had been previously identified as a therapeutic target for reduction of ischemic damage following acute myocardial infarction. In conclusion, these results provided insight into the biological pathways underpinning the associations between disturbances in habitual sleep duration and lipid levels.

In **Chapter 6** of this thesis, in order to make best use of observational data and to understand how the drug concentration can associate with the outcome, we introduced a novel methodological framework, PHARMACOM-EPI, which can predict drug concentrations at occurrence of a clinical outcome. A warning was issued by the U.S. Food and Drug Administration in 2021 on the antiseizure drug lamotrigine, which claimed

that lamotrigine has the potential to increase risk of arrhythmias and related sudden cardiac death. Our hypothesis is that the increase risk of arrhythmias and related sudden cardiac death results from drug toxicity. By utilization of the PHARMACOM-EPI framework, lamotrigine concentrations for older patients with seizure were predicted at the time of death using a published population pharmacokinetic model which was selected based on demographic information and model accessibility. Patients were subsequently categorized into toxic and non-toxic groups based on the lamotrigine therapeutic range, after which the association between two groups and all-cause mortality were analyzed. This study found that a toxic plasma concentration of lamotrigine is associated with increased risk of all-cause death in the older lamotrigine users. Application of this novel framework PHARMACOM-EPI provided evidence to support our hypothesis that the increased risk of all-cause and cardiovascular death was associated with a toxic plasma concentration level of lamotrigine among older lamotrigine users. By leveraging this methodology, observational data could be harnessed more effectively.

In **Chapter 7**, I discussed the findings from Chapter 2 through Chapter 6 within the context of previous researches, their potential applications and future directions of research.

Samenvatting

Ondanks uitgebreid onderzoek blijven er aanzienlijke hiaten in het begrip van de ontwikkeling van cardiometabole ziekten en daarmee in het ontwikkelen van preventieve mogelijkheden. In dit proefschrift trachtten we een deel van de etiologie en pathofysiologie van cardiometabole aandoeningen te ontrafelen door potentiële (causale) risicofactoren en de interactie tussen genetische en omgevingsfactoren bij cardiometabole ziekten te onderzoeken met behulp van geavanceerde statistische en epidemiologische methoden.

In **Hoofdstuk 1** wordt een overzicht gepresenteerd van de achtergrond van cardiometabole ziekten, de relevante biologische concepten uitgelegd, en de gebruikte methodologie en de studiepopulaties beschreven die in dit proefschrift zijn gebruikt.

In **Hoofdstuk 2** van dit proefschrift hebben we ons gericht op de insulinegevoeligheid van metabolieten in het bloed. Hoewel veel bekend is over de associaties van insuline niveaus met metabolieten in met name individuen met insuline resistentie, is er relatief weinig bekend over de response van metabolieten op insuline in niet-diabetische individuen. Deze response hebben we onderzocht door de niveaus van metabolieten te vergelijken onder verschillende hyperinsulinemische niveaus, waarbij euglycemische omstandigheden werden gehandhaafd in niet-diabetische individuen van middelbare leeftijd. In deze studie vergeleken we de veranderingen in de concentraties van 151 metabolieten metingen na infusie met lage en hoge dosis insuline ten opzichte van het startpunt van de studie, met behulp van regressie modellen voor herhaalde metingen. De resultaten lieten zien dat, vergeleken met het startpunt van de studie, de concentratie van 90 van de 151 metabolieten veranderden na infusie met lage dosis insuline en de concentraties van 121 metabolieten veranderden na infusie met hoge dosis insuline. Daarnaast vergeleken we de veranderingen in concentraties van metabolieten na infusie met hoge dosis insuline met die na infusie met lage dosis insuline. We vonden dat sommige metabolieten al bij lage dosis insuline-infusie hun maximale respons leken te bereiken, terwijl 99 van de 151 metabolieten een aanvullende respons vertoonden bij hoge dosis insuline-infusie. In het bijzonder de aminozuren leucine, isoleucine, valine, glutamine en tyrosine vertoonden ten opzichte van de basislijn een sterke afname in concentratie na infusie met hoge dosis insuline vergeleken met de infusie met lage dosis insuline. Hieruit kunnen we concluderen dat

de insuline gevoeligheid van metabolieten verschilt en dus dat deze metabolieten verschillend worden beïnvloed bij de ontwikkeling van insulineresistentie. Deze bevindingen suggereren ook dat de respons op verschillende insulineniveaus waarschijnlijk orgaanspecifiek is, wat blijkt uit de veranderingen in metabolieten na infusie met lage dosis insuline (primair van invloed op metabolieten in de lever) en hoge dosis insuline (primair van invloed op metabolieten in perifere weefsels). Voor patiënten kan dit aanvullende inzichten bieden in de fase van, of het type insulineresistentie en aldus kansen bieden voor een gepersonaliseerde interventie.

In **Hoofdstuk 3** van dit proefschrift hebben we de relatie tussen mitochondriën in het bloed en type 2 diabetes (T2D) onderzocht. Mitochondriën zijn cruciaal voor de energieproductie van de cel en verstoring van deze functie kan zowel oorzaak als gevolg zijn van T2D. Als maat voor mitochondriële functie hebben we het aantal mitochondriële DNA kopieën (mtDNA-CN) in het bloed genomen en geassocieerd met T2D. Dit is gedaan met behulp van Cox proportionele regressiemodellen, waarbij we gecorrigeerd hebben voor confounders. We hebben gevonden dat een hoger bloed mtDNA-CN was geassocieerd met een lager risico op het ontwikkelen van T2D. Om te beoordelen of deze associatie causaal is, hebben we bidirectionele Mendeliaanse Randomisatie (MR) analyses uitgevoerd. De resultaten lieten echter geen bewijs zien voor een causale associatie tussen bloed mtDNA-CN en type 2 diabetes in welke richting dan ook. Om deze relatie verder te onderzoeken, hebben we ook de mogelijke causale associatie tussen bloed mtDNA-CN en body mass index (BMI) met behulp van bidirectionele MR geanalyseerd. Maar ook hierbij hebben we geen aanwijzingen gevonden voor causale associaties tussen deze twee kenmerken in welke richting dan ook. Onze hypothese dat bloed mtDNA-CN het risico op T2D en obesitas aandrijft is gebaseerd op de veronderstelling dat bloed mtDNA-CN een weerspiegeling is van de algehele mitochondriale functie. Dit inclusief de mitochondriale functie van weefsels zoals de lever en spier die een cruciale rol spelen in T2D. Met aanvullende analyses hebben we gevonden dat hoger bloed mtDNA-CN niet geassocieerd was met minder spiermassa, noch met AST, ALT en ALP als maten voor leverfunctie. Een iets lagere GGT werd waargenomen bij een verhoging van bloed mtDNA-CN. Al met al suggereren deze resultaten dat de waargenomen associatie tussen laag bloed mtDNA-CN en een hoger risico op type 2 diabetes waarschijnlijk niet causaal is. Dit wordt mogelijk verklaard door onze bevindingen dat bloed mtDNA-CN niet geassocieerd is met spiermassa en slechts zwak geassocieerd is met leverfunctie. Derhalve is bloed mtDNA-CN mogelijk

geen marker van mitochondriale functie in spierweefsel en slechts een zwakke marker voor mitochondriale functie in de lever.

In **Hoofdstuk 4** van dit proefschrift hebben we de associatie tussen bloed IGF-1-niveaus en T2D onderzocht. Hoewel het biologisch plausibel is dat IGF-1 een rol speelt in T2D, hebben eerdere studies over deze associatie inconsistente bevindingen gerapporteerd. We hebben eerst confounder-gecorrigeerde Cox proportionele regressiemodellen uitgevoerd op bloed IGF-1-niveaus en T2D. Vervolgens hebben we een mogelijke J-vormige associatie tussen IGF-1 concentratie en T2D onderzocht, waarbij specifiek lagere IGF-1 niveaus werden geassocieerd met een hoger risico op het ontwikkelen van T2D. Vervolgens hebben we MR-analyses uitgevoerd in dezelfde studiepopulatie. Hierbij vonden we dat een hoger genetisch beïnvloed IGF-1 geassocieerd was met een hoger risico op T2D. Echter, de geschatte causale effecten van de individuele genetische instrumenten voor IGF-1 in de MR analyses lieten een grote heterogeniteit zien. Om dit nader te onderzoeken hebben we een nieuwe methode gebruikt, geclusterde MR, om de associatie tussen genetisch beïnvloed IGF-1 niveaus en T2D te onderzoeken. Op basis van genetische instrumenten met vergelijkbare geschatte causale effecten vonden we zes clusters van genetisch beïnvloed IGF-1, die geassocieerd waren met T2D. Hiervan waren drie clusters gerelateerd aan een hoger risico op T2D, terwijl de overige drie clusters gerelateerd waren aan een lager risico op T2D. De belangrijkste clusters waarin een hoger IGF-1 geassocieerd was met een lager risico op type 2 diabetes bestonden uit instrumenten die betrekking hadden op genen in het groeihormoon-signaleringspad. De belangrijkste clusters waarin een hoger IGF-1 werd geassocieerd met een hoger risico op type 2 diabetes bevatten instrumenten die betrekking hadden op genen in paden gerelateerd aan aminozuurmetabolisme en genoom-integriteit. Deze resultaten geven aan dat het effect van IGF-1 op T2D afhankelijk is van de onderliggende metabole/biomoleculaire mechanismen.

In **Hoofdstuk 5** van dit proefschrift hebben we de interacties onderzocht tussen slaap en lipiden die een rol spelen bij het risico op hart- en vaatziekten. Dat er een verband is tussen slaap en hart- en vaatziekten is duidelijk, maar niet of lipiden daar mogelijk een rol bij spelen. We hebben het effect onderzocht van genetische factoren en blootstelling aan zowel lange als korte gebruikelijke slaapduur op dyslipidemie zoals high-density lipoprotein cholesterol (HDL-c), low-density lipoprotein cholesterol (LDL-c) en triglyceriden (TG). In deze studie waren 732.564 individuen uit meerdere bevolkingsgroepen van 55 verschillende cohorten betrokken. Door middel

van een genoomwijde interactieanalyse tussen varianten en slaapduur hebben we we 9 leidende varianten geïdentificeerd waarvan varianten die betrekking hebben op de AMPD3 (LDL-c, LTST), ASPH (TG, LTST) en DLEU1 (TG, STST; HDL-c, STST) genen het meest significant waren. Zeven aanvullende leidende varianten werden geïdentificeerd door een gezamenlijke analyse met 2 vrijheidsgraden, waaronder een variant gelinkt aan SLC8A1 (TG, STST) die eerder is geïdentificeerd als een therapeutisch doelwit voor de vermindering van ischemische schade na een acuut myocardinfarct. Deze resultaten bieden inzicht in de biologische paden die ten grondslag liggen aan de associaties tussen verstoringen in de gebruikelijke slaapduur en lipiden niveaus die een rol spelen bij hart- en vaatziekten.

Bij interventie studies met medicijnen is het belangrijk om te weten bij welke concentratie van werkzame stof in het bloed mogelijke bijwerkingen optreden. Echter deze concentraties worden zelden gemeten en zeker niet op grote schaal. Om optimaal gebruik te maken van de grote hoeveelheden beschikbare observationele gegevens en te begrijpen hoe de medicijnconcentratie kan worden geassocieerd met de uitkomst van een interventiestudie, hebben we een nieuw methodologisch raamwerk geïntroduceerd in **hoofdstuk 6** van dit proefschrift, PHARMACOM-EPI, dat medicijnconcentraties in bloed kan voorspellen bij het optreden van een klinische uitkomst. In 2021 werd een waarschuwing uitgegeven door de U.S. Food and Drug Administration over het anti-epileptisch medicijn lamotrigine, waarin werd aangegeven dat lamotrigine het risico op hartritmestoornissen en gerelateerde plotse hartdood zou kunnen verhogen. We hebben de hypothese getest dat het verhoogde risico op hartritmestoornissen en gerelateerde plotse hartdood voortkomt uit medicijntoxiteit. Door gebruik te maken van het PHARMACOM-EPI raamwerk, werden lamotrigine-concentraties voor oudere patiënten met epileptische aanvallen voorspeld op het moment van overlijden. Dit hebbe we gedaan met behulp van gepubliceerde populatie-farmacokinetische modellen die geselecteerd waren op basis van demografische informatie en modeltoegankelijkheid. Patiënten werden vervolgens gecategoriseerd in toxische en niet-toxische groepen op basis van het therapeutische bereik van lamotrigine, waarna de associatie van deze twee groepen met de totale sterfte werd geanalyseerd. Deze studie vond dat een toxische plasmaconcentratie van lamotrigine geassocieerd is met een verhoogd risico op totale sterfte bij oudere lamotrigine-gebruikers. Toepassing van dit nieuwe PHARMACOM-EPI raamwerk leverde bewijs ter ondersteuning van onze hypothese dat het verhoogde risico op totale en cardiovasculaire sterfte geassocieerd is met een

toxisch plasmaconcentratieniveau van lamotrigine onder oudere lamotrigine-gebruikers. Deze methodologie maakt het mogelijk om observationele gegevens effectiever te benutten.

In **Hoofdstuk 7** heb ik de bevindingen uit Hoofdstuk 2 tot en met Hoofdstuk 6 besproken in de context van eerder onderzoek, de mogelijke toepassingen van de bevindingen, en de toekomstige richtingen van onderzoek.

PhD portfolio

Education and Courses	Years	Hours
Code and data management with git (MGC course)	2020	6
Practical Linux	2020	8
Basic method and reasoning in biostatistics	2020	42
Epidemiology	2020	84
Python programming	2020	24
Multi-omics data integration using R	2020	40
PhD introductory course	2021	5
Meta analysis	2021	24
Survival analysis (Advanced Biostatistics)	2021	42
Regression analysis	2021	40
Causal inference	2022	84
Job Orientation Training	2022	24
(Inter)national conference and meetings	**Years**	**Hours**
European and international congress on obesity	2020	32
Annual Dutch Diabetes Research Meeting 2020	2020	16
EAS Congress	2021	32
Mendelian Randomization Conference 2021	2021	24
ECO / IFSO-EC Congress on Obesity	2022	32
Meeting of the Study Group on the Genetics of Diabetes (SGGD)	2022	24
28th MGC workshop	2022	32
Awards	**Years**	**Hours**
Leiden University Fund	2023	-
Young Investigator Fellowship for the 89th EAS Virtual Congress Helsinki, Finland	2021	-
Traineeship abroad	**Years**	**Hours**
Cambridge Ellis Machine Learning Summer School	2022	56
Visiting PhD in Copenhagen University	2023	240
Teaching activities	**Years**	**Hours**
Mendelian Randomization	2022	8
FOS course Cardiometabolic Disease	2022	2

Publications and Manuscript

- **Wang W**, Battini V, Carnovale C, Noordam R, van Dijk KW, Kragholm KH, van Heemst D, Soeorg H, Sessa M. A novel approach for pharmacological substantiation of safety signals using plasma concentrations of medication and administrative/healthcare databases: A case study using Danish registries for an FDA warning on lamotrigine. Pharmacol Res. 2023 Jul;193:106811. doi: 10.1016/j.phrs.2023.106811. Epub 2023 Jun 1. PMID: 37268178.
- **Wang W**, Luo J, Willems van Dijk K, et al. (2022) Assessment of the bi-directional relationship between blood mitochondrial DNA copy number and type 2 diabetes mellitus: a multivariable-adjusted regression and Mendelian randomisation study. Diabetologia. 10.1007/s00125-022-05759-6
- **Wang W**, Tesfay EB, van Klinken JB, et al. (2022) Clustered Mendelian randomization analyses identify distinct and opposing pathways in the association between genetically influenced insulin-like growth factor-1 and type 2 diabetes mellitus. Int J Epidemiol. 10.1093/ije/dyac119
- **Wang W**, van Dijk KW, Wijsman CA, et al. (2021) Differential insulin sensitivity of NMR-based metabolomic measures in a two-step hyperinsulinemic euglycemic clamp study. Metabolomics 17(6): 57. 10.1007/s11306-021-01806-2
- Loh NY, **Wang W**, Noordam R, Christodoulides C (2022) Obesity, Fat Distribution and Risk of Cancer in Women and Men: A Mendelian Randomisation Study. Nutrients 14(24): 5259
- Xu H, Schwander K, Brown MR, **Wang W**, et al. (2021) Lifestyle Risk Score: handling missingness of individual lifestyle components in meta-analysis of gene-by-lifestyle interactions. Eur J Hum Genet 29(5): 839-850. 10.1038/s41431-021-00808-x
- Nagarajan P, Winkler TW, Bentley AR, Miller CL, Kraja AT, Schwander K, Lee S, **Wang W**, et al. A Large-Scale Genome-Wide Study of Gene-Sleep Duration Interactions for Blood Pressure in 811,405 Individuals from Diverse Populations. medRxiv [Preprint]. 2024 Mar 8:2024.03.07.24303870. doi: 10.1101/2024.03.07.24303870. PMID: 38496537; PMCID: PMC10942520.Sleep- lipids
- Noordam R, **Wang W**, Nagarajan P, et al. A Large-Scale Genome-Wide Gene-Sleep Interaction Study in 732,564 Participants Identifies Lipid Loci Explaining Sleep-Associated Lipid Disturbances. Under submission

Curriculum Vitae

Wenyi Wang (王文怡) was born in Guizhou, China, in January 1995. After completing high school in 2013, she pursued a Bachelor's degree in Pharmacy at Wenzhou Medical University and graduated in 2017. In the same year, she began her master's research training in quantitative pharmacology at the Division of Pharmacology of Leiden University, Netherlands. During her master's program, she performed an internship at Leiden Academic Center for Drug Research (LACDR) under the supervision of Prof. Dr. Elizabeth de Lange and Dr. Michiel van Esdonk, focusing on population pharmacokinetics (PK) modeling and semi-mechanistic physiologically based pharmacokinetic (PBPK) modeling. In 2019, she completed her second internship at the department of Pharmacometrics of Uppsala University, Sweden, where she worked on pharmacokinetic-pharmacodynamic (PKPD) modeling under the guidance of Prof. Dr. Lena E. Friberg, Prof. Dr. Elisabet I. Nielsen and Dr. Anders Thorsted. After obtaining her Master's degree in Bio-pharmaceutical Sciences in 2019, she continued her education as a PhD candidate in Genetic Epidemiology at the Department of Human Genetics, Leiden University Medical Center, under the supervision of Prof. dr. Ko Willems van Dijk, Dr. Diana van Heemst and Dr. Raymond Noordam. Part of her PhD research was performed abroad in Department of Drug Design and Pharmacology at Copenhagen University, Denmark, under supervision of Dr. Maurizio Sessa and Dr. Hiie Soeorg, working on a project to combine the fields of pharmacometrics and epidemiology. Currently, she works as a Scientist of Clinical Pharmacology and Quantitative Science at GENMAB. Looking ahead, she will continue her journey in exploring possibilities in integrating pharmacometrics and genetic epidemiology.

Acknowledgements

Completing this thesis was a challenging journey, particularly as it involved exploring a new area for me, with parts of it undertaken during the pandemic. I could not have done it without the help and support of many people, to whom I am deeply grateful.

First and foremost, I would like to express my sincere gratitude to my promoter, Prof.dr. Ko Willems van Dijk, and my co-promoters, Dr. Diana van Heemst and Dr. Raymond Noordam. Your guidance, expertise, and encouragement have been invaluable throughout this journey. Ko, I want to express my heartfelt gratitude for your unwavering support not only in academic matters but also in providing emotional support during challenging times. Diana, your deep biological knowledge and your ability to come up with creative ideas whenever we were stuck have been invaluable during this journey. Raymond, I truly appreciate your daily supervision, your willingness and patience to answer my numerous questions. Your thoughtful perspectives on data analysis and project development have been crucial to my research.

I would like to extend my heartfelt gratitude to my supervisors abroad, Dr. Maurizio Sessa and Dr. Hiie Soeorg. Maurizio, thank you for hosting me in your research group and giving me the opportunity to work on a project that merges the two fields I am passionate about. I deeply appreciate not only the professional insights you shared but also the invaluable career and life guidance you provided. To Hiie, I am grateful for your assistance with mathematics and your patience in helping me resolve the issues I encountered in the model. Additionally, I want to express my sincere thanks to my colleagues at Copenhagen University. Though our time together was short, I truly enjoyed it and am thankful for the experience.

To my colleagues at Human Genetics, Endocrinology and Epidemiology departments, with whom I had the opportunity to share my projects and receive valuable feedback during regular research meetings. Special thanks to Linjun, Jiao and Dorina. Linjun, thank you for your companionship, for the shared moments of joys and tears, and for the cherished connection we've built. Jiao, thanks for warmly welcoming me in both Leiden and Copenhagen, for sharing your experience, and for providing help with various matter throughout this journey. Dorina, thanks for the enriching conversations we

shared about research and life. The time spent with you has been truly enjoyable.

I would also like to thank my parents for their selfless dedication, unwavering support and their deep trust in me. Without you, I can never have grown into the person I am today. 我要感谢我的父母， 感谢你们无私的付出, 坚定不移的支持和对我的信任。没有你们，我不可能成长为今天的我。

Lastly, I would like to thank my husband, Chenlin. Thank you for embracing all my emotions, for always supporting my decisions, and for being by my side through the most challenging times of my PhD. I've often said that the greatest blessing of my doctoral years has been meeting you.