Genetics and life course epidemiology of cardiometabolic disease: towards personalized medicine

Ibi, D.

General Introduction

GENERAL INTRODUCTION

Cardiometabolic disease

The global rise in cardiovascular disease (CVD) and type 2 diabetes (T2D), which is driving much of the global noncommunicable disease (NCD) burden, has substantially increased over the past decades (1-4). In 2019, the worldwide number of cases was 523 million for CVD and 463 million for T2D, more than double compared to 1980-1990 (1,3). In the same year, an estimated 17.9 million people died from CVD, accounting for one-third of all deaths worldwide (4). CVD and T2D share several cardiometabolic risk factors including dyslipidemia, obesity, hypertriglyceridemia and hypertension, and thus the term cardiometabolic disease is used to encompass both of them (5,6).

These alarming rates of cardiometabolic disease place a heavy burden not only on the individuals affected and their families, but also on the whole health care system and economy (7). Yet, endeavors to successfully address cardiometabolic disease are challenging due to its complex nature. Although cardiometabolic disease usually manifests itself at middle age or beyond, it is the result of a multifactorial disease process, where the effects of unhealthy lifestyle factors and genetic predisposition interact and accumulate with ageing (8–10). This emphasizes the importance of research into the effects of lifestyle and genetics as well as their interaction in determining cardiometabolic risk factors throughout the life course. Gaining insight into the dynamics of cardiometabolic risk factors at different time points in life can provide better preventive as well as curative strategies for CVD. In this thesis, we will focus on several aspects of lipid metabolism/dyslipidemia and obesity two major players in cardiometabolic disease , at midlife as well as at different time points in adult life.

Lipid metabolism

Lipids including triglycerides (TG) and cholesterol are transported in the blood in lipoproteins. Lipoproteins consist of a hydrophobic core of non-polar lipids, containing primarily cholesterol esters and TG, surrounded by a hydrophilic membrane consisting of phospholipids, free cholesterol, and apolipoproteins (11). Based on their composition and origin, lipoproteins can be divided into five major classes e.g. chylomicrons, very low-density lipoprotein (VLDL), intermediate-density lipoprotein (IDL), low-density lipoprotein (LDL) and high-density lipoprotein (HDL). The two later ones primarily transport cholesterol, whereas chylomicron and VLDLs transport dietary and liver-derived TG, respectively, to peripheral tissues (12).

Lipoprotein metabolism is highly coordinated, and disturbances in any processes involved can lead to dyslipidaemia and ectopic fat deposition (13). Lipoprotein metabolism is partitioned in an exogenous and endogenous pathway (Figure 1), depending weather the TG source of origin is dietary or hepatic (14). After a meal, dietary lipids are emulsified by bile acids and hydrolysed into mono-acyl-glycerol and free fatty acids (FAs) by the action of pancreatic lipase. The FAs, mono-acyl-glycerol and cholesterol are then absorbed into enterocytes of the small intestine, where FAs are re-esterified into TG. Within these enterocytes, the TG and cholesterol are packaged in chylomicrons and are secreted into the lymphatic system and then enter into the blood circulation. Chylomicrons mainly supply dietary TG to white adipose tissue and are characterized by the apolipoprotein B48 (apoB48), which forms the backbone of the lipoprotein (15). During fasting, the liver synthesizes TG-rich VLDL particles, which supply TG to energy demanding tissues such as muscle, and are characterized by the apolipoprotein B100 (apoB100)(endogenous pathway) (16). Once chylomicrons and VLDLs enter the blood, their core TG are hydrolysed in capillaries into glycerol and free fatty acids (FFAs), primarily by the enzyme lipoprotein lipase (LPL), which is bound to the capillary endothelium of metabolically active tissues including adipose tissue, skeletal muscle and heart. The liberated FAs are consequently taken up by underlying parenchymal cells, which either use them directly as energy source or store them for later use. The TG-depleted and cholesterol-enriched chylomicron and VLDL remnants are subsequently taken up by the liver (17).

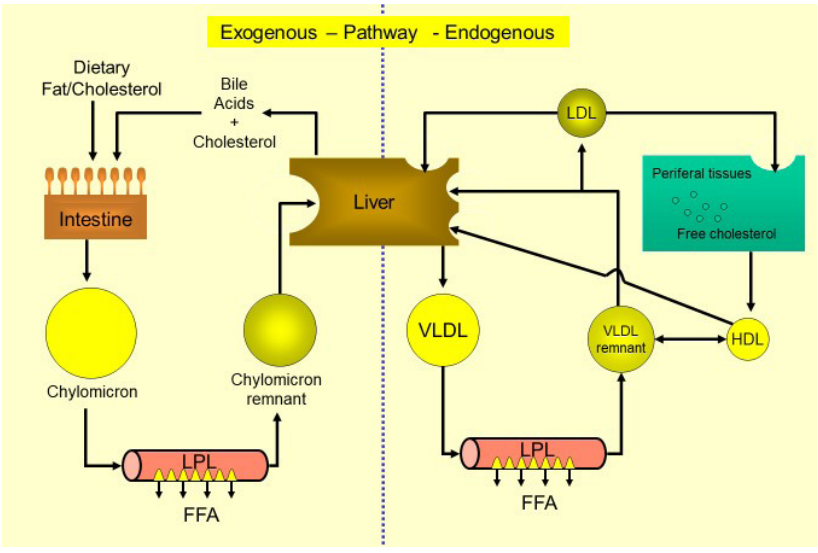

Figure 1. Schematic overview of lipoprotein metabolism. FFA, free fatty acids; LDL, low-density lipoprotein; LPL, lipoprotein lipase; VLDL, very low-density lipoproteins

Environmental risk factors of cardiometabolic disease

The environment we live in predisposes to many factors including unhealthy diet, lack of physical activity and other lifestyle choices that promote obesity and eventually lead to the development of CVD. Often labelled as "obesogenic", these environments are the major drivers of the obesity epidemic that has spread rapidly worldwide in the recent decades (18,19). For example, the increased availability, accessibility and affordability of energy-dense foods, as well as the advanced technologies that have made housework and daily chores less demanding compared to past decades, have contributed to an excess caloric intake and a less active lifestyle, respectively (20). This results in an increased energy intake and decreased energy expenditure, respectively, which creates a positive energy balance, consequently leading to weight gain, (21) and subsequently influencing lipid levels, blood pressure and inflammation (21,22).

Another important factor that influences body's energy balance is smoking (23–27). Research suggests that smoking supresses appetite and thus decreases energy intake (24). In addition, smoking has been shown to have a direct effect on resting energy expenditure, which makes the largest contribution to total energy expenditure (25–27). However, despite these potentially beneficial effects on body weight, smoking has been shown to have adverse effects on cardiometabolic health and increase risk of CVD (28–32). Not only do the chemicals in cigarette smoke cause endothelial dysfunction (31), which leads to many cardiovascular conditions, but smoking also multiplies the risk of CVD by interacting with several metabolic abnormalities such as blood pressure, insulin resistance, and dyslipidemia (28).

Altogether, the environmental factors explain more than 70% of the chronic disease burden including cardiometabolic disease (33,34). This large contribution to disease and the fact that most environmental factors are modifiable make lifestyle interventions the cornerstone of obesity and CVD prevention.

Genetic predisposition

Even though cardiometabolic disease is driven largely by environmental factors, there is a considerable part of inter-individual variability that is explained by genetics. Therefore, it is of great importance for future intervention strategies to focus on unravelling the genes and related biological pathways involved in cardiometabolic disease in addition to lifestyle intervention.

Substantial evidence from twin and family studies has suggested that genetics contribute 40–70% of inter-individual variability in BMI (35). In addition, dyslipidemia

has long been known to have a strong genetic basis (36,37). From monogenic dyslipidemias caused by rare mutations to more complex polygenic forms of dyslipidemia determined usually by common genetic variants, genetic studies have identified numerous genes involved in lipid metabolism (36-38). The archetypal is familial hypercholesterolemia (FH), an autosomal dominant monogenic disorder of elevated LDL-C levels that is caused by mutations in the LDL receptor (LDLR) (39). It was the discovery of the underlying mechanism of FH that led to the development of the statin drugs, which reduce LDL-C and CVD outcomes (40).

In addition, the rapid advances in genome sequencing techniques that became available in the beginning of 21th century (41–43) have contributed to identifying numerous genetic loci involved in cardiometabolic disease as described in the upcoming sections.

Human genetic variation and disease: Single nucleotide polymorphisms and GWAS

The human genome consists of 3.2 billion base pairs, roughly 99.5% to 99.9% of which is identical between any two human beings (44,45). Yet, with the exception of identical twins, every human genome is unique due to variations in the remaining 0.1% to 0.5% of the DNA sequence. The most common type of genetic variations are single nucleotide polymorphisms, (SNPs), with (minor allele) frequencies of ≥1 % in the population by definition (46,47). At least 10 million SNPs have been identified in the human genome that account for 90 % of diversity in the world (42,45). SNPs are located in coding and non-coding regions of the genome, and are classified as neutral and functional (48,49). Neutral SNPs have no effect, whereas the functional ones affect different biological processes and contribute to health and disease (49). Up to date, a considerable number of SNPs have been linked to multifactorial diseases, including cardiometabolic disease, mainly via two approaches: candidate gene association studies and genome-wide association studies (GWAS).

In the candidate gene association studies, SNPs are chosen based on an a priori hypothesis about the role of a gene or a set of genes in a certain phenotype or disease (50). One of the advantages of this approach is that the SNPs are chosen on the basis of biological plausibility of the genes of interest for a disease (51). Candidate gene studies can be exploited to identify the cumulative effects of several genes on phenotype/disease where the effects of individual genes are small. However, being dependent on genes with known involvement in the disease of interest, the candidate gene approach is unable to identify novel genes and underlying pathways that influence the disease (51,52).

The second approach, the GWAS, became available with the tremendous advances in genome sequencing technologies. In 2003, the DNA sequencing of the entire human genome was completed (43). This made it possible to develop within the same year the International Haplotype Mapping (HapMap) Project, which aimed to elucidate the common patterns of human DNA sequence variation, by using whole-genome sequencing data from different populations across the world (42). GWAS are a hypothesis free approach that identifies associations between SNPs across the entire genome and a disease or trait of interest (53). The GWAS are based on the haplotype block structure of the genome, where each haplotype represents a set of nearby SNPs that are co-inherited together, and thus allowing only a few SNPs (tag SNPs) to uniquely identify each haplotype block. This co-inheritance leads to a non-random correlation between the haplotype SNP alleles in the population, known as linkage disequilibrium (LD) (42,54). Using the LD concept, it is sufficient to genotype around 500,000 tag SNPs out of the estimated 10 million identified in the human population to identify common SNPs associated with diseases/traits (55). The consequence of LD is that the SNPs identified through a GWAS may not be a true casual SNP in relation to the trait/disease of interest. Instead the tag SNP may be in LD with a true biologically relevant SNP, which was previously not associated with the outcome of interest. Thus, differently from the hypothesis-driven candidate gene studies, the GWAS approach can lead to the discovery of novel genes and underlying pathways involved in a disease, which can in turn assist in development of potential therapies and drug targets (56). However, successfully discerning a true biological determinant is inherently challenging as most of the identified SNPs through a GWAS map to non-coding regions. Consequently, in addition to replication studies performed to confirm the findings, follow-up validation steps are needed to ascertain and to functionally characterize the potential casual variants (56). Another limitation is that GWAS are often underpowered to detect common SNPs with modest effects on the complex traits, and thus explain only a small proportion of heritability (56,57). One solution to overcome this limitation is to use large sample size cohorts, e.g. in large international consortia. In addition, in order to explain more of the missing heritability of the complex traits/diseases, the interactions between genetic susceptibility and environmental risk factors should be studied (58).

Genetic variants as instruments in Mendelian Randomization Studies

In conventional observational epidemiology, where the association of a modifiable environmental exposure with a disease or trait outcome is studied, discerning a true causation is a major challenge for two main reasons. First, confounding factors, which influence both the exposure and the outcome, can mask an actual causal association. Although confounding can be controlled by several analytical and statisti-

cal methods, residual confounding due to unmeasured or not accurately measured confounding factors often remains a concern (59). Second, in most observational studies (especially cross-sectional studies) reverse causality can occur, where the disease outcome influences the exposure (60). In longitudinal studies, which are observational studies that monitor a study population over an extended period of time, it is often possible to analyse exposures in relation to diseases that occur much later. Thus, the longitudinal approach is likely to suggest cause-and-effect relationships. However, these studies require a large amount of time and resources, and yet are prone to residual confounding (61).

One alternative solution is the randomized controlled trial (RCT), in which participants are randomly allocated to a control and treatment group and are followed over time to observe the effect of the treatment. Although RCTs are straightforward to interpret and provide strong evidence for causality, large-scale long-term RCTs are costly (62) and are not always feasible due to ethical issues and non-compliance (59). In addition, RCTs are relatively short term, while chronic diseases arise from lifelong cumulative effects of multiple exposures. An approach to overcome these limitations of RCTs is the Mendelian Randomization (MR) study. In parallel with RCTs, MR uses genetic variants associated with an exposure of interest as genetic instruments to establish causality between the exposure and a disease/trait outcome. MR studies are often considered as ''natural RCTs'' due to the random distribution in the population of the risk and non-risk alleles during gamete formation (Mendel's Law of Independent Assortment), which are used to define an exposed and control group, respectively (63,64). This random (independent) allocation means that any other confounders should be equally present in the exposed (risk alleles) and control (non-risk alleles) group, making the MR approach in theory free of confounding. In addition, MR overcomes the reverse causation limitation of the conventional observational epidemiology since genotypes are determined at conception, and therefore are not influenced by disease. Finally, utilizing the open access to GWAS data, MR can be used as a quicker and less expensive yet powerful method to establish causality as compared to large-scale RCTs (65,66). Accordingly, MR can be used to anticipate the results of RCTs, and thus could spare unnecessary RCTs when a causal association is not observed. For example, contrary to results from observational studies (67), MR analyses showed that genetically influenced higher HDL-C levels did not protect against CVD (68), suggesting that a pharmacological agent boosting HDL-C levels would be futile in CVD interventions. On the other hand, MR can provide supporting evidence for the beneficial role of a potential drug in disease, as illustrated by Lotta et al. (69), which showed that genetically-influenced lower TG levels via LPL pathway lower CVD risk independently of LDL-C lowering agents. However, when

translating genetic findings into pharmacological strategies, it should be realized that the effects of lifelong exposure measured by MR may differ from the relatively short-term pharmacological effects assessed by RCTs.

In order to have valid MR analyses, three strong assumptions must be met: 1) the genetic instrument must be associated with the exposure of interest; 2) there is no common causation of the genetic instrument and the outcome; 3) the genetic instrument affects the outcome only through the exposure (70). A robust association of the SNP with the exposure can verify assumption 1. As mentioned before, random distribution of alleles in a population can control for confounding. However, when the ancestry is related to both the genotype and the outcome, genetic confounding by population substructure violates assumption 2. This can be typically controlled by excluding ethnic outliers and adjusting for population substructure using genetic principal components (71). Assumption 3 is violated when pleiotropy occurs, which means that the SNP affects the outcome through other phenotypes in addition to the exposure. Therefore, a selection of a genetic instrument that is associated with the outcome via the exposure only is critical to satisfy this assumption.

Genetics of obesity

GWAS have enabled the identification of numerous susceptibility loci for obesity, with *FTO* gene being identified as the first and the top common identified BMI and obesity-susceptibility gene (72). The number of discovered obesity-susceptibility loci has grown dramatically since the discovery of the *FTO* in 2007. Among European populations, around 200 loci for BMI (73,74), 49 loci for waist to hip ratio (WHR) (75), and 10 loci for body fat percentage (76) have been identified. In the Genetic Investigation of ANthropometric Traits (GIANT) consortium, an international collaboration that seeks to identify genetic loci for anthropometric traits, a total of 97 loci for BMI have been identified (74). In addition, these studies have used genetic risk scores (GRS) constructed using these 97 loci to estimate a cumulative effect of these obesity-susceptibility loci on BMI, and have shown that combining loci for BMI explains approximately 3 % of the phenotypic differences (,74). Another GWAS by the GIANT consortium in individuals with European ancestry identified 49 loci for abdominal obesity, which was defined by the WHRafter adjustment for BMI (75). Similar to GRS for BMI, GRS for abdominal obesity were generated on the basis of these findings, which have been related to increased risks of various cardiometabolic diseases including T2D and coronary heart disease (CHD) (77,78).

In the search for a better understanding of genetic basis of adiposity and its links to cardiometabolic disease risk, a meta-analysis GWAS of body fat percentage was

conducted in up to 100,716 individuals (76). This study identified a total 12 loci, of which eight were previously related to overall adiposity and four were novel.. Seven of the 12 loci (*TOMM40/APOE, IRS1, SPRY2, COBLL1/GRB14, IGF2BP1, PLA2G6, and CRCT1*) had larger effects on body fat percentage than on BMI, suggestive of a primarily association with body fat, whereas the remaining five loci (*FTO, TMEM18, MC4R, SEC16B, and TUFM/SH2B1*) showed stronger associations on BMI than body fat percentage, suggesting that these variants have potential effects on both body fat and lean mass.

Adult body size and shape change over time, with more progressive changes occurring from 20-60 years of age and following a sex-specific trend (79). Evidence from sex- and age-specific analyses has shown that the genetic factors play an important role in the age-related weight change and sexual dimorphism of body shape (80). For instance, 15 BMI loci had different effects in older (50 years and above) compared to younger (below 50 years), but showed no sex-specific effects. The opposite was observed for body shape: the effects of 44 WHR loci differed between men and women, but not between younger and older adults.

Genetics of lipid metabolism

Similar to obesity, GWAS on plasma lipid levels were performed in order to gain more insights into genetic etiology of lipid metabolism. Worldwide collaborative efforts in the Global Lipids Genetics Consortium (GLGC) have revealed a total of 157 loci (including 62 novel loci) associated with lipid levels in individuals of European ancestry (81). Numerous genes have been linked to TG levels, among which the *LPL*, encoding the LPL protein (82), has a major role in TG metabolism. More specifically, increased LPL activity was associated with lower TG levels and reduced risks for Coronary Artery Disease (CAD) and T2D (83). In addition, in a large population-based Swedish cohort variants in the *apolipoprotein E (APOE)* and in the *APOA1/A4/C3/A5* cluster were robustly associated with changes in plasma total cholesterol and TG levels, respectively (84). More importantly, these identified genes had previously been associated with myocardial infarction and/or CAD, which clearly illustrates the added value of these novel lipid associations in providing additional insight into the pathogenesis of CVD.

Taking advantage of the MR approach, genetic studies have further investigated weather lipid levels have a causal role in the development of CVD. Genetically elevated LDL-C and TG levels have been shown to be causally associated with increased risk of cardiovascular outcomes including ischemic stroke and CAD (85–87).

Altogether this evidence suggests that genetic variants of lipid metabolism may contribute the cardiovascular outcomes by moderating lipid traits.

Gene-environment interactions

The majority of cardiometabolic diseases involve multiple disease-causing factors, which makes research for preventive and curative strategies one of the most significant challenges that medical research faces today. In such multifactorial diseases, genetic and environmental factors rarely act on isolation from each other, but rather interact and influence each other (88). Accumulating evidence suggests that environmental risk factors may modify the effects of genetic factors on disease and on the other hand genetic variants can increase the risk of diseases in high-risk environments (88–91). For example, the noted rise of obesity worldwide in the recent decades might reflect an interaction between genetics and the changing environmental exposures that together alter the penetrance (the proportion of individuals with a specific genetic variant who express the related phenotype) of genetic variants. Genetic factors associated with obesity may influence behavioural responses to environmental context by influencing appetite and stimulating food intake (92,93). One classical example is food intake regulation by *MC4R* gene, which controls appetite and food intake (92,94). On the other hand, environmental factors as well can alter the effects of genetic variants. For example, it has been shown that the relationship between the *FTO* genotype and obesity is attenuated among physically active individuals (95,96). Thus, the environmental changes and genetic variants associated with obesity may not affect all individuals equally.

In the last years new approaches, study designs and methods are emerging for exploring the gene-environment interactions in large-scale human populations. Nevertheless, more future efforts are needed to clarify the interplay between genetic and environmental factors on cardiometabolic disease and translate these findings into prevention and treatment practices.

Metabolomics: A powerful tool in cardiometabolic disease

A key factor in the fight against cardiometabolic disease is gaining insights into the pathophysiological processes that underlie this complex disease. Metabolomics, which is the study of small-molecule metabolites within cells, biofluids, tissues or organisms, is a powerful tool for delineating mechanistic links between risk factors (exposures) and disease (97). Metabolites are the intermediates and end products of cellular metabolism, and plasma metabolite levels reflect a response to normal or pathological conditions (98). In this last case, the metabolites provide valuable

information on the metabolic disturbances linked to disease process as well as disease progression (99), and, thus, make it possible to be used as disease biomarkers.

The most commonly used analytical technologies used in metabolomics include gas chromatography coupled to mass spectrometry (GC-MS), liquid chromatography coupled with single-stage mass spectrometry (LC-MS) and nuclear magnetic resonance (NMR) spectroscopy (100). The latter one has emerged as the preferred platform for large-scale metabolomics studies due to its fast, inexpensive, non-destructive, highly automatable and exceptionally reproducible quantification of circulating metabolites (101,102). NMR is based on the detection of certain atomic nuclei when exposed to an external magnetic field. Protons (1H) are often the nuclei of choice due to their abundance (100) . A commonly used high-throughput 1H NMR metabolomics platform is the Nightingale Health metabolic biomarker platform (Nightingale Health Ltd., Helsinki, Finland), which quantifies lipoprotein subclass profiling with lipid concentrations within 14 subclasses, abundant proteins and various low molecular-weight metabolites, covering both routine biomarkers and emerging medically relevant biomarkers (103). The biomarkers include detailed measures of cholesterol metabolism, fatty acid compositions, and various low-molecular weight metabolites, such as amino acids, ketones and glycolysis metabolites. For 14 lipoprotein subclasses, the lipid concentrations and composition are measured in terms of TG, phospholipids, total cholesterol, cholesterol esters, and free cholesterol, and total lipid concentration within each subclass.

Cohort Studies and genetic consortia

The studies described in this thesis were performed with data from cohort studies and genetic consortia. The study populations and study design of these cohorts and consortia are summarized below.

NEO Study

To address the aims of the studies described in **Chapters 2, 3 and 4** we used data from the Netherlands Epidemiology of Obesity (NEO) study (104). The NEO study is a population-based prospective cohort study including men and women aged between 45 and 65 years. This cohort was designed to study pathways that lead to disease in persons with overweight or obesity. Therefore, from the greater area of Leiden, The Netherlands, all inhabitants with a self-reported body mass index (BMI) of 27 kg/m2 or higher were eligible to participate. In addition, inhabitants from one nearby municipality (Leiderdorp, The Netherlands) in the same age group were invited to participate regardless of their BMI, forming a reference population for BMI distribution. In total, 6,671 participants were included from September 2008 until

September 2012. At baseline, participants completed questionnaires with respect to demographic, lifestyle, and clinical information and went an extensive physical examination, including anthropometric measurements and blood sampling.

Doetinchem cohort

The studies in **Chapters 5 and 6** were performed in the The Doetinchem Cohort Study (DCS). DCS is a prospective population-based cohort of 7769 men and women aged 20-59 years living in Doetinchem, a small town in the north-eastern part of the Netherlands (105). At baseline (1987–1991: round 1), 20,155 people aged 20–59 years were invited to visit the municipal health centre to participate in the 'Monitoring Project on Cardiovascular Disease Risk Factors'. This project, which is the origin of the DCS, was aimed at providing prevalence estimates of cardiovascular disease risk factors, such as smoking, blood pressure and serum cholesterol levels. The protocol was extended, because the focus broadened from cardiovascular to other chronic diseases such as cancer, diabetes, musculoskeletal disorders and COPD. Due to extension of the protocol, with similar budget, not all 12 405 participants in the MP-CVDRF could be re-invited. Instead, a random sample of in total 7769 of the respondents at baseline was invited for second examination (1993–1997: round 2, mean age: 46 years). This random sample, which is considered the basis of the DCS, was invited for follow-up examinations every 5 years. The response rates varied between 75% and 80% in all rounds. Demographic characteristics, medical history of chronic diseases, medication use, and lifestyle factors were collected using standardized questionnaires. Trained staff performed standardized measurements of anthropometric traits (height, weight, waist circumference), blood pressure and blood sampling during a visit to the municipal health service.

Oxford Biobank study

In addition to the NEO study, in **Chapters 2, 3 and 4** we used individual data from the Oxford BioBank (OBB) study. The OBB is a population-based cohort of randomly selected healthy participants of 7,185 individuals aged 30 to 50 years from Oxfordshire (UK), included between 1999 and May 2015 (106). The primary goal of the OBB is investigating the physiological consequences of genetic mechanisms of chronic non-communicable diseases, including T2D and CVD, and to explore potential therapeutic targets in the prevention and treatment of these diseases. Participants with a history of several diseases including myocardial infarction, diabetes mellitus, heart failure, untreated malignancy, other ongoing systemic diseases or ongoing pregnancy were not eligible for study inclusion. At baseline (Stage 1), all participants had questionnaire-based assessments on demographic characteristics, medical history of chronic diseases and potential disease risk factors or confounders in

disease pathology, such as physical activity, smoking and alcohol intake. In addition, participants underwent an extensive baseline physical examination including a broad range of metabolic-, CVD- and obesity-related phenotypes based on blood plasma phenotyping, genetic biomarkers, anthropometric measurements and body composition assessment using dual-energy X-ray absorptiometry (DXA). Subsequently, based on a research question of interest, selected participants wereinvited for a second visit (recall) (Stage 2).

UK Biobank Cohort

In addition to the analyses in the NEO and OBB cohort, in **Chapter 4** we performed some analyses using individual data from the UK BioBank cohort. The UK Biobank cohort is a prospective general population cohort. Baseline assessments took place between 2006 and 2010 in 22 different assessment centers across the United Kingdom (107). A total of 502,628 participants between the age of 40 and 70 years were recruited from the general population. Invitation letters were sent to eligible adults registered to the National Health Services (NHS) and living within a 25 miles distance from one of the study assessment centers. The UK Biobank study was approved by the North-West Multi-center Research Ethics Committee (MREC). Access for information to invite participants was approved by the Patient Information Advisory Group (PIAG) from England and Wales. All participants in the UK Biobank study provided written informed consent.

Genetic consortia

For the MR studies used in **Chapters 3, 4 and 6**, we used publically available summary data from the genetic consortia GLGC and GIANT. In general, the aim of genetic consortia is to collect data from a large number of genotyped cohorts to perform GWAS on the outcome of interest, and subsequently combine the results in a meta-analysis to find robust effect estimates. GLGC was set up to perform a GWAS meta-analyses on concentrations of LDL-C, HDL-C, TG and total cholesterol (81). This meta-analysis was based on 188,577 individuals of European ancestry from 60 different studies. The GIANT consortium is an international collaboration that seeks to identify genetic loci associated with anthropometric traits including BMI, height, and traits related to waist circumference (such as WHR adjusted for BMI, or BMI adjusted for WHR). The largest BMI GWAS performed by GIANT comprised up to 339,224 individuals from 125 studies of mainly European ancestry populations (74).

Objectives and outline of this thesis

A better understanding of the underlying pathophysiology of cardiometabolic disease and the long-term and cumulative exposure of its risk factors over the life

course can lead to new insights in treatment and prevention approaches of this disease. Therefore, in the first part of this thesis we focus on the genetic determinants of lipid metabolism as atherogenic dyslipidemia is a major component of cardiometabolic disease and consequently of CVD. In the second part of the thesis, we study the age-related changes of cardiometabolic risk factors over the life course across four generations.

Part I –GENETIC DETERMINANTS OF LIPID METABOLISM

The first objective of this thesis is to gain insight into the genetics and thus underlying mechanisms of lipid metabolism during both fasting and postprandial states. Specifically, in **Chapter 2** we aim to identify genetic variants that determine the postprandial TG response to a mixed meal, independently from fasting TG levels. In order to gain further insight into the pathophysiology of the postprandial TG response, in **Chapter 2** we also describe the effects of the identified genetic variants on postprandial responses of NMR-based metabolomics. As shown in **Chapter 1**, in addition to statins that lower LDL-C, new therapies that enhance LPL-mediated clearance of TG are in development for CVD prevention. Using MR approaches, recent studies have suggested that drugs that enhance LPL-mediated lipolysis are likely to provide additional cardiovascular benefits in addition to existing LDL-C–lowering agents. Using a similar approach, in **Chapter 3**, we aim to gain insights into the underlying mechanisms behind these observed effects, by assessing the causal associations between genetically-influenced TG levels via *LPL* alleles and fasting NMR-based lipoprotein and metabolite measures on a background of genetically-determined lower LDL-C. Several LPL modulators have been shown to affect circulating TG levels the risk of CVD, but the effects of LPL and its modulators on the detailed lipoprotein profile have never been studied in depth. Elucidation of these specific effects can provide insight in the therapeutic potential of LPL modulators in reducing CVD risk. Therefore, in **Chapter 4,** we aim to investigate the role of apolipoprotein A-V , a natural activator of LPL and potent TG regulator, on lipid metabolism. We provide in this chapter detailed insight into the effects of genetically-influenced TG levels by *apolipoprotein A5 (APOA5)* gene alone as well as in combination with genetically-influenced TG levels via *LPL alleles* on the NMR-based lipoprotein profile in the NEO study.

Part II – CARDIOMETABOLIC AND GENETIC RISK PROFILES OVER THE LIFE COURSE IN DIFFERENT GENERATIONS OF MEN AND WOMEN

Utilizing the long follow-up of nearly 30 years of the Doentichem cohort, in the second part of the thesis we take the opportunity to assess the age-related changes of anthropometric measures of obesity and associated cardiometabolic risk factors

across four generations (**Chapter 5**). As introduced in **Chapter 1**, obesity results from a complex interplay of genetics, environment and behavior. Thus, in order to fully understand the mechanisms that underpin this complex trait, in **Chapter 6** we aim to investigate the temporal aspect of the genetic basis of obesity and its interaction with the changing obesogenic environment, by performing GWAS analyses on longitudinal measures of body weight at specific age groups from 30-70 years.

Finally, in **Chapter 7**, the general discussion, the main findings of this thesis and their implications for future preventive and therapeutic strategies for cardiometabolic disease are discussed.

REFERENCES

1. Roth GA, Mensah GA, Johnson CO, et al; GBD-NHLBI-JACC Global Burden of Cardiovascular Diseases Writing Group. Global burden of cardiovascular diseases and risk factors, 1990-2019: update from the GBD 2019 Study. *J Am Coll Cardiol*. 2020;76:2982-3021.
2. Lin X, Xu Y, Pan X, et al. Global, regional, and national burden and trend of diabetes in 195 countries and territories: an analysis from 1990 to 2025. *Sci Rep*. 2020;10:14790.
3. Saeedi P, Petersohn I, Salpea P, et al; IDF Diabetes Atlas Committee. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the International Diabetes Federation Diabetes Atlas, 9th edition. *Diabetes Res Clin Pract*. 2019;157:107843.
4. World Health Organization. June 11, 2021. "Cardiovascular diseases (CVDs)".
5. Pradeepa R, Nazir A, Mohan V. Type 2 diabetes and cardiovascular diseases: do they share a common soil? The Asian Indian experience. *Heart Asia*. 2012;4:69-76.
6. Canoy, D., Tran, J., Zottoli, M., Ramakrishnan, R., Hassaine, A., Rao, S., Li, Y., Salimi-Khorshidi, G., Norton, R., & Rahimi, K. Association between cardiometabolic disease multimorbidity and all-cause mortality in 2 million women and men registered in UK general practices. *BMC medicine*. 2021; 19:258.
7. Miller CE, Quayyum Z, Mcnamee P, Salman RAS. Economic burden of intracranial vascular malformations in adults: Prospective population-based study. *Stroke*. 2009 Jun 1;40:1973–9.
8. Yeates K, Lohfeld L, Sleeth J, Morales F, Rajkotia Y, Ogedegbe O. A Global Perspective on Cardiovascular Disease in Vulnerable Populations. *Canadian Journal of Cardiology*. 2015; 31:1081–93.
9. Gidding SS, Rana JS, Prendergast C, McGill H, Carr JJ, Liu K, et al. Pathobiological Determinants of Atherosclerosis in Youth (PDAY) Risk Score in Young Adults Predicts Coronary Artery and Abdominal Aorta Calcium in Middle Age. *Circulation*. 2016;133:139–46.
10. Berenson GS, Srinivasan SR, Bao W, Newman WP, Tracy RE, Wattigney WA. Association between Multiple Cardiovascular Risk Factors and Atherosclerosis in Children and Young Adults. N Engl J Med. 1998;338:1650–6.
11. Rensen PCN, De Vrueh RLA, Kuiper J, Bijsterbosch MK, Biessen EAL, Van Berkel TJC. Recombinant lipoproteins: Lipoprotein-like lipid particles for drug targeting. *Adv Drug Deliv*. 2001;47(2–3):251–76.
12. Williams KJ. Molecular processes that handle - and mishandle - dietary lipids. *J Clin Invest*. 2008;118 :3247–59.
13. Bartelt A, Merkel M, Heeren J. A new, powerful player in lipoprotein metabolism: Brown adipose tissue. *Journal of Molecular Medicine*. 2012;90:887–93.
14. Feingold KR. Introduction to Lipids and Lipoproteins. In: Feingold KR, Anawalt B, Boyce A, et al., eds. *Endotext*. South Dartmouth (MA): MDText.com, Inc.; January 19, 2021.
15. Kindel T, Lee DM, Tso P. The mechanism of the formation and secretion of chylomicrons. *Atheroscler Suppl*; 2010;11:11–6.
16. Tessari P, Coracina A, Cosma A, Tiengo A. Hepatic lipid metabolism and non-alcoholic fatty liver disease. *Nutr Metab Cardiovasc Dis*; 2009;19: 291–302.
17. Preiss-Landl K, Zimmermann R, Hämmerle G, Zechner R. Lipoprotein lipase: The regulation of tissue specific expression and its role in lipid and energy metabolism. *Curr Opin Lipidol*. 2002;13: 471–81.

18. Hruby A, Manson JE, Qi L, et al. Determinants and Consequences of Obesity. *Am J Public Health*. 2016;106:1656-1662.
19. Swinburn B, Egger G, Raza F. Dissecting obesogenic environments: The development and application of a framework for identifying and prioritizing environmental interventions for obesity. *Prev Med (Baltim)*. 1999 ;29:563–70.
20. Hobbs M, Radley D. Obesogenic environments and obesity: a comment on 'Are environmental area characteristics at birth associated with overweight and obesity in school-aged children? Findings from the SLOPE (Studying Lifecourse Obesity PrEdictors) population-based cohort in the south of England'. *BMC Med*. 2020;18:59.
21. Hill JO, Commerford R. Physical activity, fat balance, and energy balance. *Int J Sport Nutr*. 1996;6:80-92.
22. Kojta, I., Chacińska, M., & Błachnio-Zabielska, A. Obesity, Bioactive Lipids, and Adipose Tissue Inflammation in Insulin Resistance. *Nutrients*. 2020;12:1305.
23. Audrain-Mcgovern J, Benowitz NL. Cigarette smoking, nicotine, and body weight. *Clin Pharmacol Ther*. 2011;90:164–8.
24. Jo YH, Talmage DA, Role LW. Nicotinic receptor-mediated effects on appetite and food intake. *J Neurobiol*. 2002;53:618–32.
25. Martínez De Morentin PB, Whittle AJ, Fernø J, Nogueiras R, Diéguez C, Vidal-Puig A, et al. Nicotine induces negative energy balance through hypothalamic AMP-activated protein kinase. *Diabetes*. 2012;61:807–17.
26. Hofstetter A, Schutz Y, Jéquier E, Wahren J. Increased 24-Hour Energy Expenditure in Cigarette Smokers. *N Engl J Med*. 1986;314:79–82.
27. Collins LC, Walker J, Stamford BA. Smoking multiple high- versus low-nicotine cigarettes: Impact on resting energy expenditure. *Metabolism*.1996;45:923–6.
28. Roy A, Rawal I, Jabbour S, Prabhakaran D. Tobacco and Cardiovascular Disease: A Summary of Evidence. In: Disease Control Priorities, Third Edition: Cardiovascular, Respiratory, and Related Disorders. The World Bank; 2017;5: 57–77. Available from:
29. Yang H, Negishi K, Otahal P, Marwick TH. Clinical prediction of incident heart failure risk: A systematic review and meta-analysis. *Open Heart*. 2015;2(1):e000222.
30. Njølstad I, Arnesen E, Lund-Larsen PG. Smoking, serum lipids, blood pressure, and sex differences in myocardial infarction: A 12-year follow-up of the Finnmark study. *Circulation*. 1996; 93:450–6.
31. Owens AP, Mackman N. Tissue factor and thrombosis: The clot starts here. *Thromb Haemost*. 2010;104: 432–9.
32. Roth GA, Nguyen G, Forouzanfar MH, Mokdad AH, Naghavi M, Murray CJL. Estimates of global and regional premature cardiovascular mortality in 2025. *Circulation*. 2015;132:1270–1.
33. Willett WC. Balancing life-style and genomics research for disease prevention. *Science*. 2002; 296:695–8. /
34. Lichtenstein P, Holm N V., Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and Heritable Factors in the Causation of Cancer — Analyses of Cohorts of Twins from Sweden, Denmark, and Finland. *N Engl J Med*. 2000; 343:78–85.
35. Allison DB, Kaprio J, Korkeila M, Koskenvuo M, Neale MC, Hayakawa K. The heritability of body mass index among an international sample of monozygotic twins reared apart. *Int J Obes*. 1996;20:501–6.

36. Saadatagah S, Jose M, Dikilitas O, et al. Genetic basis of hypercholesterolemia in adults [published correction appears in *NPJ Genom Med*. 2021 Jun 29;6:56]. *NPJ Genom Med*. 2021;6:28.
37. Dron JS, Hegele RA. Genetics of Lipid and Lipoprotein Disorders and Traits. *Curr Genet Med Rep*. 2016;4:130–41.
38. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. . 2013;45:1274–85.
39. Brown MS, Goldstein JL. A receptor-mediated pathway for cholesterol homeostasis. *Science*. 1986 ;232:34–47.
40. Hegele RA. Monogenic dyslipidemias: Window on determinants of plasma lipoprotein metabolism. *American Journal of Human Genetics*. 2001;69:1161–77.
41. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409(6822):860–921.
42. Belmont JW, Hardenbol P, Willis TD, Yu F, Yang H, Ch'Ang LY, et al. The international HapMap project. *Nature*. 2003;426:789–96.
43. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome [published correction appears in Science 2001 Jun 5;292(5523):1838]. *Science*. 2001;291:1304-1351.
44. Robert F, Pelletier J. Exploring the Impact of Single-Nucleotide Polymorphisms on Translation. *Front Genet*. 2018;9:507.
45. Kruglyak L, Nickerson DA. Variation is the spice of life. *Nature Genetics*. 2001;27: 234–6.
46. Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, et al. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet*. 1999;22(3):231–8.
47. Orr N, Chanock S. Chapter 1 Common Genetic Variation and Human Disease. *Adv Genet*; 2008;62: 1–32.
48. Hrdlickova B, de Almeida RC, Borek Z, Withoff S. Genetic variation in the non-coding genome: Involvement of micro-RNAs and long non-coding RNAs in disease. Biochimica et Biophysica Acta - Molecular Basis of Disease. *Elsevier B.V.*; 214;1842: 1910–22.
49. Albert PR. What is a functional genetic polymorphism? Defining classes of functionality. *Journal of Psychiatry and Neuroscience*. 2011;36:363–5.
50. Tabor HK, Risch NJ, Myers RM. Candidate-gene approaches for studying complex genetic traits: Practical considerations. *Nature Reviews Genetics*. 2002;3: 391–7.
51. Candidate Gene - an overview | ScienceDirect Topics. Available from: https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/candidate-gene.
52. Daly AK, Day CP. Candidate gene case-control association studies: Advantages and potential pitfalls. *British journal of clinical pharmacology*. 2001;52: 489–99.
53. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JPA, et al. Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nature Reviews Genetics*. 2008;52: 356–69.
54. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature*. 2007; 449(7164):851–61.
55. Hardy J, Singleton A. Genomewide Association Studies and Human Disease. *N Engl J Med*. 2009; 360:1759–68.

56. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics*. 2019;20: 467–84.
57. Hirschhorn JN. Genomewide Association Studies — Illuminating Biologic Pathways. *N Engl J Med*. 2009;360:1699–701.
58. Frazer KA, Murray SS, Schork NJ, Topol EJ. Human genetic variation and its contribution to complex traits. *Nature Reviews Genetics*. 2009;10: 241–51.
59. Qi L. Mendelian randomization in nutritional epidemiology. *Nutrition Reviews*. 2009;67: 439–50.
60. Smith GD, Ebrahim S. Mendelian Randomization: Genetic Variants as Instruments for Strengthening Causal Inference in Observational Studies. 2008.
61. Caruana EJ, Roman M, Hernández-Sánchez J, Solli P. Longitudinal studies. *J Thorac Dis*. 2015;7(11):E537-E540.
62. Speich B, von Niederhäusern B, Schur N, Hemkens LG, Fürst T, Bhatnagar N, et al. Systematic review on costs and resource use of randomized clinical trials shows a lack of transparent and comprehensive data. *Journal of Clinical Epidemiology*. 2018;96:1–11.
63. Thanassoulis G, O'Donnell CJ. Mendelian randomization: Nature's randomized trial in the post-genome era. *JAMA*. 2009;301: 2386–8.
64. Smith GD, Ebrahim S. What can mendelian randomisation tell us about modifiable behavioural and environmental exposures? *British Medical Journa*l. 2005;330: 1076–9.
65. Lee K, Lim C-Y. Mendelian Randomization Analysis in Observational Epidemiology. *J Lipid Atheroscler*. 2019;8(2):67.
66. Storm CS, Kia DA, Almramhi M, Wood NW. Using Mendelian randomization to understand and develop treatments for neurodegenerative disease. *Brain Commun*.2020];2
67. Gordon DJ, Probstfield JL, Garrison RJ, Neaton JD, Castelli WP, Knoke JD, et al. High-density lipoprotein cholesterol and cardiovascular disease. Four prospective American studies. *Circulation*.1989;79:8–15.
68. Voight BF, Peloso GM, Orho-Melander M, Frikke-Schmidt R, Barbalic M, Jensen MK, et al. Plasma HDL cholesterol and risk of myocardial infarction: A mendelian randomisation study. *Lancet*. 2012;380:572–80.
69. Lotta LA, Stewart ID, Sharp SJ, Day FR, Burgess S, Luan J, et al. Association of Genetically Enhanced Lipoprotein Lipase–Mediated Lipolysis and Low-Density Lipoprotein Cholesterol–Lowering Alleles With Risk of Coronary Disease and Type 2 Diabetes. 2018;1–10.
70. Vanderweele TJ, Tchetgen Tchetgen EJ, Cornelis M, Kraft P. Methodological challenges in Mendelian randomization. *Epidemiology*. 2014 ;25(3):427–35.
71. Dixon P, Hollingworth W, Harrison S, Davies NM, Davey Smith G. Mendelian Randomization analysis of the causal effect of adiposity on hospital costs. *J Health Econ*. 2020;70:102300.
72. Frayling TM, et al. A Common Variant in the FTO Gene Is Associated with Body Mass Index and Predisposes to Childhood and Adult Obesity. *Science*. 2007;316:889–894.
73. Day FR, Loos RJ. Developments in obesity genetics in the era of genome-wide association studies. *J Nutrigenet Nutrigenomics*. 2011;4:222–38.
74. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518:197–206.
75. Shungin D, Winkler T, Croteau-Chonka DC, Ferreira T, Locke AE, Mägi R, et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature*. 2015;518:187–96.

76. Lu Y, Day FR, Gustafsson S, Buchkovich ML, Na J, Bataille V, et al. New loci for body fat percentage reveal link between adiposity and cardiometabolic disease risk. *Nat Commun*. 2016;7(1):1–15.
77. Huang T, Qi Q, Zheng Y, Ley SH, Manson JAE, Hu FB, et al. Genetic predisposition to central obesity and risk of type 2 diabetes: Two independent cohort studies. *Diabetes Care*. 2015; 38:1306–11.
78. Emdin CA, Khera A V., Natarajan P, Klarin D, Zekavat SM, Hsiao AJ, et al. Genetic association of waist-to-hip ratio with cardiometabolic traits, type 2 diabetes, and coronary heart disease. *JAMA*. 2017;317:626–34.
79. Villareal DT, Apovian CM, Kushner RF, Klein S. Obesity in older adults: technical review and position statement of the American Society for Nutrition and NAASO, The Obesity Society. *Am J Clin Nutr*. 2005;82:923–34.
80. Winkler TW, Justice AE, Graff M, Barata L, Feitosa MF, Chu S, et al. The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. *PLoS Genet*. 2015;11:1–42.
81. Global Lipids Genetics Consortium, Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and Refinement of Loci Associated with Lipid Levels. *Nat Genet*. 2013;45(11):1–24.
82. Merkel M, Eckel RH, Goldberg IJ. Lipoprotein lipase: Genetics, lipid uptake, and regulation. *Journal of Lipid Research*. 2002;42:1997–2006.
83. Liu DJ, Peloso GM, Yu H, Butterworth AS, Wang X, Mahajan A, et al. Exome-wide association study of plasma lipids in >300,000 individuals. *Nat Genet*. 2017;49:1758–66.
84. Varga T V., Kurbasic A, Aine M, Eriksson P, Ali A, Hindy G, et al. Novel genetic loci associated with long-term deterioration in blood lipid concentrations and coronary artery disease in European adults. *Int J Epidemiol*. 2017;46:1211–22.
85. Hindy G, Engström G, Larsson SC, Traylor M, Markus HS, Melander O, et al. Role of blood lipids in the development of ischemic stroke and its subtypes: A mendelian randomization study. *Stroke*. 2018;49(4):820–7.
86. Tan Y De, Xiao P, Guda C. In-depth Mendelian randomization analysis of causal factors for coronary artery disease. *Sci Rep*. 2020;10:1–13.
87. Allara E, Morani G, Carter P, Gkatzionis A, Zuber V, Foley CN, et al. Genetic Determinants of Lipids and Cardiovascular Disease Outcomes: A Wide-Angled Mendelian Randomization Investigation. *Circ Genomic Precis Med*. 2019; 12:543–51.
88. Qi L. Gene-Diet Interactions in Complex Disease: Current Findings and Relevance for Public Health. *Current Nutrition Reports*. 2012;1:222–7.
89. Qi L, Cho YA. Gene-environment interaction and obesity. *Nutrition Reviews*. Nutr Rev; 2008;66:684–94.
90. Rosenquist JN, Lehrer SF, O'Malley AJ, Zaslavsky AM, Smoller JW, Christakis NA. Cohort of birth modifies the association between FTO genotype and BMI. *Proc Natl Acad Sci U S A*. 2015;112:354–9.
91. Walter S, Mejia-Guevara I, Estrada K, Liu SY, Glymour MM. Association of a genetic risk score with body mass index across different birth cohorts. *JAMA*. 2016;316:63–9.
92. Qi L, Kraft P, Hunter DJ, Hu FB. The common obesity variant near MC4R gene is associated with higher intakes of total energy and dietary fat, weight change and diabetes risk in women. *Hum Mol Genet*. 2008;17:3502–8.

93. Fredriksson R, Hägglund M, Olszewski PK, Stephansson O, Jacobsson JA, Olszewska AM, et al. The obesity gene, FTO, is of ancient origin, up-regulated during food deprivation and expressed in neurons of feeding-related nuclei of the brain. *Endocrinology*. 2008;149:2062–71.
94. Baldini G, Phelan KD. The melanocortin pathway and control of appetite-progress and therapeutic implications. *Journal of Endocrinology*. 2019;241: R1–33.
95. Kilpeläinen TO, Qi L, Brage S, Sharp SJ, Sonestedt E, Demerath E, et al. Physical activity attenuates the influence of FTO variants on obesity risk: A meta-analysis of 218,166 adults and 19,268 children. *PLoS Med*. 2011;8: e1001116.
96. Graff M, Scott RA, Justice AE, Young KL, Feitosa MF, Barata L, et al. Genome-wide physical activity interactions in adiposity — A meta-analysis of 200,452 adults. *PLoS Genet*. 2017;13:130.
97. Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: Beyond biomarkers and towards mechanisms. *Nature Reviews Molecular Cell Biology*. 2016;17: 451–9.
98. Lu C, Thompson CB. Metabolic regulation of epigenetics. *Cell Metab*. 2012;16: 9–17.
99. Gowda GAN, Zhang S, Gu H, Asiago V, Shanaiah N, Raftery D. Metabolomics-based methods for early disease diagnostics. *Expert Review of Molecular Diagnostics*. 2008;8:617–33.
100. Emwas AH, Roy R, McKay RT, Tenori L, Saccenti E, Nagana Gowda GA, et al. Nmr spectroscopy for metabolomics research. *Metabolites*. 2019;9:123.
101. Takis PG, Ghini V, Tenori L, Turano P, Luchinat C. Uniqueness of the NMR approach to metabolomics. *Trends in Analytical Chemistry*. 2019;120:115300.
102. Guennec A Le, Giraudeau P, Caldarelli S. Evaluation of fast 2D NMR for metabolomics. *Anal Chem*. 2014;86:5946–54.
103. Soininen P, Kangas AJ, Würtz P, Suna T, Ala-Korpela M. Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Cardiovascular Epidemiology and Genetics. *Circ Cardiovasc Genet*. 2015;8:192–206.
104. de Mutsert R, den Heijer M, Rabelink TJ, Smit JWA, Romijn JA, Jukema JW, et al. The Netherlands Epidemiology of Obesity (NEO) study: study design and data collection. *Eur J Epidemiol*. 2013;28:513–23.
105. Verschuren WM, Blokstra A, Picavet HS, Smit HA. Cohort profile: the Doetinchem Cohort Study. *Int J Epidemiol*. 2008;37(6):1236-1241. doi:10.1093/ije/dym292
106. Karpe F, Vasan SK, Humphreys SM, Miller J, Cheeseman J, Dennis AL, et al. Cohort profile: The Oxford Biobank. *Int J Epidemiol*. 2018;47:21-21g.
107. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12:e1001779.