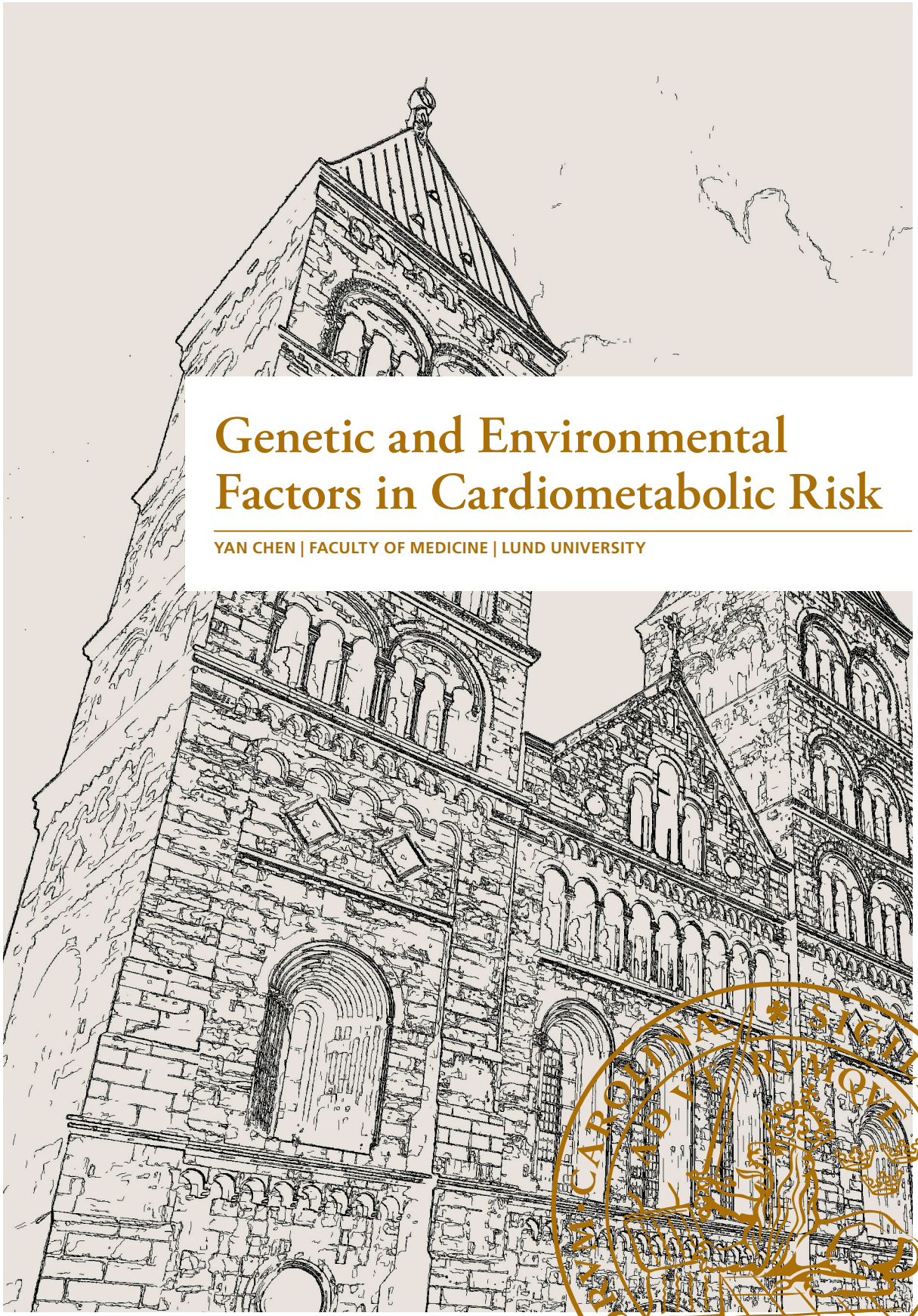
Genetic and Environmental Factors in Cardiometabolic Risk

Chen, Yan

2018

Document Version:
Publisher's PDF, also known as Version of record



Genetic and Environmental Factors in Cardiometabolic Risk

YAN CHEN | FACULTY OF MEDICINE | LUND UNIVERSITY

Genetic and Environmental Factors in Cardiometabolic Risk

Yan Chen

LUND
UNIVERSITY

DOCTORAL DISSERTATION
by due permission of the Faculty of Medicine, Lund University, Sweden.
To be defended at Aulan, Clinical Research Centre, Jan Waldenströms gata 35,
SE-21428 Malmö, Sweden at 9:00, 17th December, 2018.

Faculty opponent
Professor Ronald CW Ma, The Chinese University of Hong Kong

Organization LUND UNIVERSITY	Document name: DOCTORAL DISSERTATION
Author(s) Yan Chen	Date of issue: 17th December, 2018 Sponsoring organization

Genetic and Environmental Factors in Cardiometabolic Risk

Abstract:

Cardiovascular diseases and diabetes mellitus are closely linked by sharing common risk factors, such as obesity, insulin resistance, hypertension and dyslipidaemia. These intermediate risk factors are affected by the joint effects of environment and genetics. Although a multitude of studies aimed to disentangle the effects of gene-environment interactions, many of the published gene-environment interaction analyses are inadequately powered and lack replication due to the small magnitude of interaction effect sizes. Meanwhile, many more potential gene-environment interactions affecting cardiometabolic risk are yet to be discovered. This thesis investigated the role of genetic and environmental factors in a range of cardiometabolic traits (e.g. body mass index (BMI), blood lipids, and blood pressure), using three complementary study designs.

In Paper I, two Swedish cohorts, the Gene-Lifestyle Interactions and Complex Traits Involved in Elevated Disease Risk Study (GLACIER, N=4902) and the Malmö Diet and Cancer Study (MDCS, N=21,824) were analyzed. In the meta-analysis, a nominally significant interaction (P_{int}=0.03) was observed between sugar sweetened beverage (SSB) intake and a genetic risk score based on BMI related single nucleotide polymorphisms (SNPs). With SSB consumption defined as four categories, one SSB intake category increase was associated with 0.18 kg/m^2 mean change of BMI (P= 1.7×10^{-20}; n = 26,726).

In Paper II, interactions between dietary polyunsaturated fatty acid (PUFA) intake and variation at the *fatty acid desaturase* (*FADS*) gene cluster were investigated in the GLACIER cohort (N=5,160). In summary, SNP-, haplotype-, and gene-level interaction signals were observed in relation to serum triglyceride concentrations. Through functional annotation, the *FADS2* rs5792235 SNP was identified as the probable causal variant in the region (owing to its high functionality score).

In Paper III, using repeated-measures data from >18,000 adults in a subcohort of the Västerbotten Health Survey, an environment-wide association study (EWAS) was performed employing linear mixed-models, assuming different intercepts for each individual. A varying number (12-75) of exposure variables showed environmental-wide-significant associations with nine cardiometabolic traits. For the first time, we showed that heptadecanoic acid (C17:0) is strongly associated with a range of cardiometabolic traits.

In conclusion, in this thesis I report novel and confirmatory evidence of environmental risk factors, as well as gene-environment interactions for major intermediate cardiometabolic risk factors.

Key words: sugar-sweetened beverage; genetic risk score; obesity; cardiometabolic traits; gene-environment interaction; *fatty acid desaturase* (*FADS*); haplotype; rare variants; EWAS; longitudinal analysis.

Classification system and/or index terms (if any)

Supplementary bibliographical information		Language: English
ISSN and key title: 1652-8220 Lund University, Faculty of Medicine Doctoral Dissertation Series 2018:155		ISBN: 978-91-7619-724-0
Recipient's notes	Number of pages 295 Security classification	Price

Signature  Date 2018-11-09

Genetic and Environmental Factors in Cardiometabolic Risk

Yan Chen

LUND
UNIVERSITY

Coverphoto by Yan Chen

ISSN 1652-8220

ISBN 978-91-7619-724-0

ISSN 1652-8220
ISBN 978-91-7619-724-0
Lund University, Faculty of Medicine Doctoral Dissertation Series 2018:155

Printed in Sweden by Media-Tryck, Lund University
Lund 2018

MADE IN SWEDEN

For we live by faith, not by sight.

Corinthians 5:7, Bible (New International Version)

Table of Contents

List of Publications

1. Brunkwall L, **Chen Y**, Hindy G, Rukh G, Ericson U, Barroso I, Johansson I, Franks PW, Orho-Melander M, Renström F*. Sugar-sweetened beverage consumption and genetic predisposition to obesity in 2 Swedish cohorts. Am J Clin Nutr. 2016;104(3):809-15.

2. **Chen Y**, Estampador AC, Keller M, Poveda A, Dalla-Riva J, Johansson I, Renström F, Kurbasic A, Franks PW*, Varga TV*. The combined effects of *FADS* gene variation and dietary fats in obesity-related traits in a population from the far north of Sweden: the GLACIER Study. Int J Obes (Lond), 2018 May 24. doi: 10.1038/s41366-018-0112-3.

3. **Chen Y**, Kurbasic A, Patel CJ, Hallmans G, Johansson I, Renström F, Poveda A*, Franks PW*. Environment-wide association study to prioritize lifestyle risk factors for cardiometabolic disease. (manuscript)

Publications not included in this thesis

1. Kurbasic A, Poveda A, **Chen Y**, Agren A, Engberg E, Hu FB, Johansson I, Barroso I, Brändstrom A, Hallmans G, Renström F, Franks PW*. Gene-Lifestyle Interactions in Complex Diseases: Design and Description of the GLACIER and VIKING Studies. Curr Nutr Rep. 2014;3(4):400-11.

2. Brøns C, Saltbæk PN, Friedrichsen M, **Chen Y**, Vaag A*. Endocrine and metabolic diurnal rhythms in young adult men born small vs appropriate for gestational age. Eur J Endocrinol. 2016;175(1):29-40.

3. Varga TV, Kurbasic A, Aine M, Eriksson P, Ali A, Hindy G, Gustafsson S, Luan J, Shungin D, **Chen Y**, Schulz CA, Nilsson PM, Hallmans G, Barroso I, Deloukas P, Langenberg C, Scott RA, Wareham NJ, Lind L, Ingelsson E, Melander O, Orho-Melander M, Renström F, Franks PW*. Novel genetic loci associated with long-term deterioration in blood lipid concentrations and coronary artery disease in European adults. Int J Epidemiol. 2016;46(4):1211-22.

4. Poveda A, **Chen Y**, Brändström A, Engberg E, Hallmans G, Johansson I, Renström F, Kurbasic A, Franks PW*. The heritable basis of gene-environment interactions in cardiometabolic traits. Diabetologia. 2017;60(3):442-52.

Abbreviations

ARA - arachidonic acid

ASB - artificially sweetened beverages

apoA1 - apolipoprotein A1

apoB - apolipoprotein B

BMI - body mass index

CVD - cardiovascular diseases

CHD – coronary heart disease

DHA - docosahexaenoic acid

DIAGRAM - DIAbetes Genetics Replication and Meta-analysis Consortium

ENGAGE - European Network for Genetic and Genomic Epidemiology

EPA - eicosapentaenoic acid

EWAS - environment-wide association study

FADS - fatty acid desaturase

FFQ - food frequency questionnaire

FTO - fat mass and obesity-associated gene

GDM – gestational diabetes mellitus

GIANT - The Genetic Investigation of ANthropometric Traits

GLGC - Global Lipids Genetics Consortium

GLACIER - Gene-Lifestyle Interactions and Complex Traits Involved in Elevated Disease Risk

GRS - genetic risk score

GWAS - genome-wide association study

HDL - high density lipoprotein

HDL-C - high density lipoprotein cholesterol

IVW - inverse variance-weighted average method

IDL - intermediate-density lipoprotein

LD - linkage disequilibrium

LDL - low density lipoprotein

LDL-C - low density lipoprotein cholesterol

MAGIC - Meta-Analyses of Glucose-and Insulin-related traits Consortium

MetS - metabolic syndrome

MDCS - Malmö Diet and Cancer Study

MODY - Maturity-Onset Diabetes of the Young

NF-κB - nuclear factor kappa-light-chain-enhancer of activated B cells

OCS-FA - odd-chain fatty acid

OGTT - oral glucose tolerance test

OR - odds ratio

PAI-1 - plasminogen activator inhibitor 1

PUFA - polyunsaturated fatty acid

SSB - sugar-sweetened beverages

SMCs – smooth muscle cells

T1DM - type 1 diabetes mellitus

T2DM - type 2 diabetes mellitus

TCF7L2 - transcription factor 7-like 2

TFBS - transcription factor binding sites

WHO - World Health Organization

VLDL - very low-density lipoprotein

VHU - Västerbottens Hälsoundersökning

wGRS - weighted genetic risk score

Introduction

Cardiometabolic diseases have become a global epidemic affecting more than one billion people worldwide. According to the World Health Organization (WHO), in 2015, approximately 17.9 million people died from cardiovascular diseases (CVD) and 1.6 million people died from diabetes; many of whom died from cardiovascular complications.[1] Cardiometabolic risk is a condition in which the chances of developing atherosclerosis, coronary heart disease (CHD), stroke, and diabetes mellitus are significantly elevated. A wide-range of intermediate risk factors, including obesity,[2] insulin resistance,[3]index,[4] hypertension,[5] and dyslipidaemia[6,7] are considered to play key roles in the disease process. A complex interrelation network between these intermediate risk factors exists and interrelations among these factors are often bidirectional. Owing to their value in risk prediction, these intermediate risk markers are examined routinely in primary and secondary health care, constituting important data for prognostication and diagnosis.

Cardiometabolic diseases can be heavily affected by a person's genetic background,[8,9] yet the substantial increase in disease prevalence in the modern society, characterized with sedentary behaviors and highly processed food, suggests our environment and lifestyle likely being the trigger.[10] Gene-environment interactions, defined as a phenomenon in which the joint effects of one or more genes with one or more environmental factors cannot be readily explained by their marginal effects, could also play a significant role in the increase of the prevalence of cardiometabolic diseases.[11] The search for genes related to cardiometabolic disorders has been dominated by two approaches: linkage analysis and genetic association studies. Linkage analysis has been successful when diseases exhibit Mendelian patterns of inheritance, such as familial hypercholesterolemia and Maturity-Onset Diabetes of the Young (MODY).[12-14] However, most cardiometabolic traits and hard disease endpoints like hypertension, type-2 diabetes (T2DM) and CHD are caused by a combination of genetic, environmental and lifestyle factors, with each of the components contributing a small phenotypic effect.

Genetic association studies in the general population yield a more comprehensive knowledge of the molecular basis of complex diseases. This can be achieved by either testing for phenotypic associations with genetic variants from a putative candidate gene, or a Genome-wide Association Study (GWAS) microarray, or even the full genome using whole-genome or whole-exome sequencing. In particular,

GWAS has proven to be a highly effective way to discover large number of genetic variants associated with complex traits. By 2015, 755 SNPs located at 366 independent loci were identified as contributors to cardiometabolic diseases.[15] This number has since then dramatically grown owing to the recent wave of GWAS studies; these additional variants help explained some of the missing heritability for cardiometabolic diseases.[16]

Physical activity, behavioural factors, dietary factors, pollution and drugs, among others, can confer an effect on the development of cardiometabolic diseases. The established risk factors for cardiometabolic disorders include sugar-sweetened beverages (SSB),[17] red meat,[18] processed meat,[19,20] tobacco,[21] and sedentary behaviours[22] *etc*. However, as most epidemiological studies focus on environmental factors based on prior hypotheses, a considerable number of factors associated with cardiometabolic disorders may have not been identified.

This thesis explores the independent and joint effects of genetic and environmental risk factors on the following cardiometabolic traits: weight, body mass index (BMI), high- density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), triglycerides, total cholesterol, fasting glucose, 2-hour glucose, and systolic and diastolic blood pressures.

Brief introduction to diabetes and CVD

Diabetes is characterized by chronically raised blood glucose levels and it occurs due to insufficient insulin secretion from the pancreas (termed "insulin deficiency") or the body's inadequate response to insulin (termed "insulin resistance"). The current WHO diagnosis criteria for diabetes is fasting glucose ≥ 7.0 mmol/L (126 mg/dl), or 2-h glucose ≥ 11.1 mmol/L (200mg/dl) or HbA1c ≥ 6.5% (48 mmol/mol).[23] It is widely accepted that there are three main types of diabetes: type 1 diabetes mellitus (T1DM), T2DM and gestational diabetes (GDM).[24] Using a data driven clustering approach, Ahlqvist *et al* identified five subtypes of diabetes, which are essentially different in patient characteristic and risk of complication among diabetes patients from Finland and Sweden.[25] In addition, there are other rare forms of diabetes, including monogenetic diabetes (e.g. Maturity-Onset Diabetes of the Young (MODY), neonatal diabetes) and secondary diabetes (e.g. diabetes raised as a complication of glucagonoma, and drug-induced diabetes).

The main focus of this thesis are the risk factors that lead to T2DM, which accounts for over 90% of all diabetes cases world-wide.[24] T2DM is considered modifiable and most people with T2DM do not depend on exogenous insulin to sustain life. Accompanied with medication, lifestyle change is often recommended to patients; this typically includes eating a healthy diet, being physically active, and losing weight.

CVD refers to a group of disorders including stroke, CHD, and peripheral artery disease. CVD often relates to a process called atherosclerosis in which plaque accumulates inside the arteries. Plaque formation is a series of processes involving lipoprotein retention, inflammatory cell recruitment, foam cell formation, apoptosis and necrosis, smooth muscle cells (SMCs) proliferation and matrix synthesis, calcification, angiogenesis, and so on.[26] As time goes on, the plaque may partially or totally block the blood flow thus reducing the oxygen supply in the heart, brain, pelvis, legs, arms or kidneys. CVD causes 17.9 million death worldwide every year, which makes CVD the largest contributor to global mortality.[1] CVD is also the leading cause of mobidity and mortality among people with diabetes. High blood glucose levels over a long period can damage blood vessels by releasing oxygen free radicals, glycation, inactivation of proteins within vascular cells, and stimulating endothelial cell apoptosis.[27-29] In high- and middle- income countries, 14.8% to 40.5% of the middle-aged population with diabetes develops some kind of CVD eventually.[30] There is a lack of data showing the development of CVD complications among people with diabetes in low-income countries. However, CVD is expected to cause more deaths in developing countries.[31]

From Reaven's hypothesis to cardiometabolic risk

The comorbidity of diabetes and cardiovascular disease has been known for more than a century.[32] The underlying mechanisms on how T2DM and cardiovascular complications are connected have been vigorously debated and are controversial. In Gerald Reaven's Banting lecture in 1988, he highlighted the central role of insulin resistance in the pathology of T2DM, hypertension and CHD.[33] He also suggested that insulin resistance causes a cluster of clinical abnormalities including hyperinsulinemia, impaired glucose tolerance, decreased HDL-C and increased triglyceride levels. Reaven described this cluster of abnormalities as 'Syndrome X' which tends to increase the risk of both T2DM and CVD.[33] A range of methods for quantifying insulin sensitivity have been developed and validated since then, with various advantages and limitations.[34] The gold standard method for assessing insulin sensitivity is the hyperinsulinemic-euglycemic clamp, but this is an invasive, costly and time-consuming method, that is not feasible in large research studies or in clinical practice. Other intermediate methods exist, some of which have been applied in epidemiological settings. Recognizing the importance of insulin resistance as a central feature of dysmetabolism, the concept of 'Metabolic Syndrome' (MetS) was developed and has been used extensively in research and clinical practice.[35-38] The initial aims of this concept were to bridge the field of diabetology and cardiology and to assist clinicians to treat its components and initiate preventive measures before the onset of cardiometabolic diseases. The key components of MetS are almost the same in different definitions, including obesity, glucose intolerance, dyslipidaemia, blood pressure, proinflammatory state and prothrombotic state, but they differ in details and criterion.[38]

Using the Medical Subject Heading (MeSH) term "Metabolic Syndrome" to search the related literature in PubMed since Reaven's Banting lecture was published, ~ 22,500 research articles can be identified (in humans). Several studies have shown that MetS, characterized by US National Cholesterol Education Programme Adult Treatment Panel III (NCEP ATP VIII) criteria, is associated with 1.5- to 2-fold increased risk of CVD and 3- to 5-fold increased risk of T2DM.[39-43] Of note, however, the value of MetS in clinical practice remains controversial. Kahn *et al* argued that the construct of MetS does not give a better prediction of future diabetes or CVD than its individual components.[44] A study conducted by Wilson *et al* showed that fasting plasma glucose is a far better single predictor of diabetes than the combination of other MetS components, and there was no striking difference in predicting CVD when including 1, 3 or 5 MetS components.[39] Another concern raised by Kahn and others is that no specific treatment for MetS exists and treatment of metabolic abnormalities is still based on its component parts.[44] Kahn proposed the concept of 'cardiometabolic risk' to describe factors that predict CVD and diabetes mellitus which has been widely accepted by the medical research

community.[45] As shown in Figure 1, cardiometabolic risk is similar to MetS but with a much broader meaning. Lifestyle factors such as smoking or physical inactivity are also included. Unlike the definition of MetS, cardiometabolic risk includes people with diagnosed chronic disease as well. Other, newly emerged cardiometabolic risk factors include inflammation and hypercoagulation. These cardiometabolic risk factors are interlinked and sometimes bidirectionally associated with one another. Treatment of modifiable cardiometabolic factors has shown to be an effective way to reduce CVD-related mortality.[46]

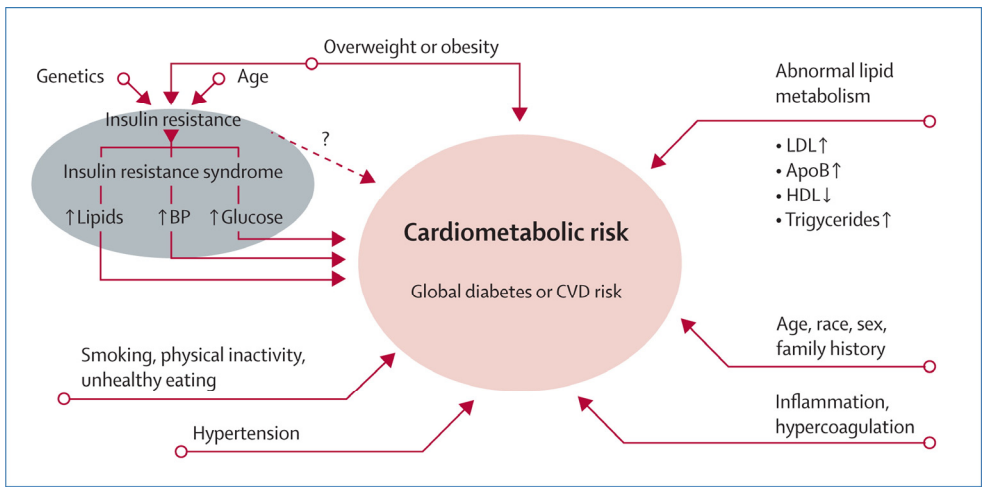


Figure 1. Factors contributing to diabetes mellitus and CVD.
BP: blood pressure; LDL: low-density lipoprotein; Apo B: apolipoprotein B; HDL: high-density lipoprotein. Ref: Kahn R. Metabolic syndrome is it a syndrome? Does it matter? Circulation 2007; 115(13): 1806-10.

The cardiometabolic risk factors

Obesity

Obesity is a medical condition in which excess body fat (adipose tissue) accumulates in the body and may impair health. In 2016, it was estimated that more than 1.9 billion adults were overweight and 650 million were obese.[47] Sufficient evidence has shown that obesity is associated with increased risk of T2DM, CVD, certain types of cancer (endometrial, breast, colon, kidney *etc.*), and premature death.[48] The WHO has made obesity management the top priority for disease prevention in its public health agenda.[49] However, we need to be aware of the fact that about 30% of overweight or obese individuals do not develop diabetes or other metabolic

diseases.[50] It is also noteworthy that fat distribution is more important than the total quantity of fat: subcutaneous fat accounts for 80-90% of total body fat and it is mainly stored in the abdominal, subscapular, and femoral areas; about 5-20% of total fat is stored surrounding the intra-abdominal organs, mainly in the mesentery and omentum, and this part of the adipose tissue is termed visceral fat.[51] Compared with subcutaneous fat, visceral fat is considered more pathogenic. It releases fatty acids, inflammatory agents, and hormones that ultimately lead to higher LDL-C, triglycerides, blood glucose, and blood pressure.[52] Fox *et al* conducted regression analyses to compare the effects of visceral and subcutaneous fat on multiple metabolic risk factors in 3,000 individuals of the Framingham Heart Study.[53] Both visceral and subcutaneous fat showed association with continuous metabolic factors (e.g. blood lipids, systolic- and diastolic- blood pressure, fasting glucose) and dichotomous risk factors (e.g. diabetes, MetS). Notably, the magnitude of association with all risk factors was consistently stronger for visceral fat.[53]

In clinical practice and large epidemiological studies, BMI has been most frequently used as a surrogate measure of adiposity. It is calculated as a person's weight in kilograms divided by the square of the height in meters. Because calculation only requires weight and height, BMI is an inexpensive and easy-to-perform measurement and can be standardized across different populations.[54] Table 1 shows the WHO classification of different BMI groups among adults (age ≥18 years) based on its association with mortality. Obesity is defined as BMI ≥30 kg/m^2.[49] Mokdad *et al* investigated the association between obesity, and diabetes and other health risk factors; they found that adults with BMI over 40 kg/m^2 have substantially higher risk of diabetes (odds ratio [OR]: 6.39-8.50), high blood pressure (OR: 5.67-7.17), high cholesterol level (OR: 1.67-2.13) compared with normal weight participants.[55]

Table 1 Classification of adults according to BMI proposed by a WHO expert committee

WHO Classification	BMI	Population description
Underweight	<18.5	Thin
Normal range	18.5-24.9	'Healthy', 'normal','acceptable'
Overweight	25.0-29.9	Pre-obese
Obesity class I	30.0-34.9	Moderate obese
Obesity class II	35.0-39.9	Severe obese
Obesity class III	≥ 40.0	Very severe obese

Note: The criteria for obesity is different for Asians. For people with the same age, sex, and BMI, Asians tend to have a higher percentage of body fat and a higher risk to develop inverse health consequence than Caucasians.[56] In a later report, a WHO expert committee suggested the cut-off point for overweight in Asians to be 23 kg/m2 and for obesity to be 27.5 kg/m2.[57]

Alternative methods such as waist circumference, waist:hip ratio, and skinfold thickness can also be applied for estimating body fatness. Waist circumference and waist:hip ratio, which are strong indicators of abdominal adiposity, have been

suggested to be superior to BMI in predicting CVD risk and mortality in recent studies.[58-60] Skinfold thickness estimates the subcutaneous fat deposit at different locations. With recent technological developments, several commercialized body fat analysers have been developed. Computed tomography (CT) and magnetic resonance imaging (MRI) have shown to effectively distinguish subcutaneous and visceral adipose tissue.[61] Dual-energy X-ray absorptiometry (DXA) applies the X-ray technology to measure the bone mineral density and can be also used to measure body composition with high accuracy.[62] However, these technology-based approaches are used mainly as a validation of the traditional anthropometry-based measurements or in small studies due to their high cost.

Dyslipidaemia

Common lipids, such as triglycerides, phospholipids, cholesterol and cholesterol esters, are minimally soluble in aqueous media. Cholesteryl esters, a group of lipids formed by an ester bond formed between a long chain fatty acid and the hydroxyl group of a cholesterol molecule, are the predominant form of cholesterol transport and storage. Cholesteryl esters and triglycerides bind to a various type of apolipoproteins to form complex lipid-protein capsules, termed as lipoproteins, which can travel freely in the blood and tissue fluid. For lipoproteins, the greater their protein:lipid ratio, the higher their density and the smaller their size and *vice versa*. Chylomicrons, very low-density lipoproteins (VLDLs) and intermediate-density lipoproteins (IDLs) are larger in size and contain mainly triglycerides and within their core. LDLs and HDLs are relatively smaller and contain mainly cholesterol.[63]

Dyslipidaemia is a disorder of lipoprotein metabolism. In most circumstances, for each lipoprotein classes, particle numbers are highly correlated with the concentrations of the lipids they cumulatively carry.[64] Therefore, plasma levels of HDL-C and LDL-C can be used as surrogate estimates of HDL and LDL concentrations. In the context of cardiometabolic risk, dyslipidaemia is most often manifested by elevated levels of total cholesterol, LDL-C and/or triglycerides, and decreased levels of HDL-C in blood. Ample evidence has demonstrated that elevated LDL-C is the primary risk factor for atherosclerosis.[65] LDL particles invade and accumulate in the artery walls following endothelium dysfunction.[66] Then LDL undergoes a series of oxidation and enzymatic reaction facilitating the internalization of LDL particles by macrophages.[67] Over time, the macrophages become cholesterol-laden foam cells and accumulate in the artery wall (intima) and form a plaque. At the same time, SMCs proliferate and migrate from the tunica media into the intima. The SMCs contributes to the formation of a firm, fibrous cap covering the plaque. As the fibrous cap grows, it can eventually rupture, releasing the plaque into the blood stream, thereby forming a blood clot.[68] Consistent

epidemiological studies have identified HDL particles as cardioprotective. To be precise, HDL participates in both atherogenic and atheroprotective processes. HDL particles remove nonesterified cholesterol from the lipid-laden macrophages back to the liver, which is thought as the main HDL mechanism of protection. However, HDLs can lose their protective capabilities during inflammation. In this case, the high level of dysfunctional HDLs is associated with an increased risk of CVD. Another possible mechanism of HDL being atherogenic is due to the oxidation of its major protein component apolipoprotein A1 (apoA1).[69] The mechanism of how triglycerides influence cardiovascular health is not fully understood, however it is generally accepted that a high concentration of blood triglycerides directly leads to the enrichment of triglycerides in LDL and HDL particles. Experimental studies suggested that triglyceride-rich HDL particles may lose their function, and triglyceride enriched LDLs show an increased atherogenic property.[70]

Insulin resistance

Insulin resistance is a condition in which insulin-mediated glucose disposal from blood into muscle and other tissues is impaired. Just as T2DM and CVD, insulin resistance is a heterogenous metabolic disorder. Multiple mechanisms that may cause insulin resistance have been proposed, of which obesity-induced inflammation has been considered to play a key role.[3] In the short term, the pancreas can fight against insulin resistance by releasing more insulin to compensate the attenuated effects of insulin and maintain a normal glucose level.[71] However, this compensation eventually reaches a point where no matter how much insulin is released, blood glucose is constantly above the threshold of prediabetes which in the end may lead to the development of diabetes.[72] In the clinic, fasting glucose and oral glucose tolerance tests (OGTT) are often used to diagnose prediabetes and diabetes. Fasting glucose is measured after an overnight fast. For OGTT, the participants are requested to fast for 8-12 hours prior to the test. Then glucose is measured before and after 2 hours of a 75g oral glucose load is administered. WHO/IDF defines diabetes as a fasting glucose level $\geq$7.0 mmol/L or 2h-glucose $\geq$11.1 mmol/L, impaired glucose tolerance as fasting glucose <7.0mmol/L and 2h-glucose 7.8-11.1 mmol/L, and impaired fasting glucose as fasting plasma glucose 6.1-6.9 mmol/L, and 2-h glucose <7.8 mmol/L.[73]

Glucose serves as a major source of energy for all cells and organs. The circulating glucose mainly derives from three sources: intestinal absorption from diet, release of glucose from glycogen (glycogenolysis) and synthesis of glucose from non-carbohydrate precursors (gluconeogenesis).[74] The liver plays a major role in glucose homeostasis as it controls various pathways of glucose utilization and endogenous glucose production. Apart from the liver, the pancreas also exerts a key role in glucose homeostasis by secreting various digestive hormones. Insulin is secreted by

the pancreatic β-cells in response to increased blood glucose and amino acid levels in the postprandial state. Insulin binds to insulin receptors on many cells to promote glucose disposal in peripheral tissues. Insulin accelerates glycogen synthesis in muscle, adipose tissue, and liver. Meanwhile, insulin inhibits glucagon secretion from islet α-cells to stop the liver from producing glucose through glycogenolysis and gluconeogenesis. Insulin can also act on lipid and protein metabolism, for example, insulin reduce lipolysis in adipose tissue, increase uptake of triglycerides by adipose tissue and muscle, and increase VLDL synthesis in the liver.[74,75] A second β-cell hormone called amylin acts as a complementary hormone to insulin.[76]

Insulin resistance contributes to the development of CVD through several pathways. Insulin resistance at adipocytes, working through hyperinsulinemia, enhances hydrolysis of circulating triacylglycerol and free fatty acid influx to the liver.[77] The excess free fatty acids stabilize apolipoprotein B (apoB) production in the liver, a major component of VLDL particles, and further causes increased hepatic VLDL release.[78] Meanwhile, impaired insulin signalling through PI3-kinase /Akt pathway, together with reduced lipoprotein lipase activity, slows down the clearance of ApoB. Thus, a combination of all the above factors leads to raised blood triglyceride and VLDL levels.[79] Hypertension often coexists with insulin resistance. However, the link between insulin resistance and hypertension is less apparent. Current evidence suggests that inflammatory cytokines such as tumour necrosis factor α (TNF-α), interleukin-6 (IL-6) and nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) control signalling of insulin through prohibiting or activating phosphorylation of insulin receptor substrate-1 at the tyrosine residue.[80] Although limited evidence exists in humans, Zhou *et al* have shown in Dahl salt-sensitive rats that inhibition of NF-κB inflammatory pathway also reduce blood pressure and vascular inflammation,[81,82] indicating insulin resistance may be associated with hypertension through inflammation. Another possible link to CVD is that insulin resistant individuals tend to have increased levels of fibrinogen,[83] factor VII,[84] and plasminogen activator inhibitor 1 (PAI-1),[85] which are indispensable to blood clotting and fibrinolysis. The Emerging Risk Factors Collaboration (ERFC) consortium studied ~25,000 individuals from 52 cohorts and found that adding fibrinogen to the conventional CVD risk factors such as age, sex, smoking, history of diabetes, hypertension, and total cholesterol improved the predictability of CVD among low- and intermediate- risk populations.[83] Raised factor VII activity has shown to be independently associated with insulin resistance among CVD patients.[84] PAI-1 has shown to prohibit fibrinolytic activity of plasminogen activator and an elevated high PAI-1: plasminogen activator ratio is an early sign of CVD.[86] Insulin itself may be atherogenic. Evidence has shown that impaired insulin signalling may stimulate proatherogenic pathways in vascular SMCs.[87]

Hypertension

Hypertension, also known as "high blood pressure", is a chronic condition in which blood pressure is persistently elevated. In the circulation, blood always flows from the region with higher pressure to the one with lower pressure. As the blood moves, it keeps pushing against the side of blood vessels. Blood pressure is the force per unit area exerted by circulating blood on the wall of arteries. Two kinds of blood pressure are quantified in millimetres of mercury (mmHg): systolic and diastolic blood pressure. Systolic blood pressure refers to the peak arterial pressure reached during the ventricular contraction. Diastolic blood pressure refers to the minimal arterial pressure just before the ventricular contraction begins. Normally, in absence of disease, the two types of blood pressure go hand-in-hand and rise and fall for equal proportion. WHO defines hypertension as a systolic blood pressure ≥140 mmHg and/or a diastolic blood pressure ≥90 mmHg.[88] This definition is based on the assumption that people with blood pressures above these levels are at increased risk of atherogenic CVDs, kidney disease, microvascular complications and premature death.[89-96] Globally, hypertension causes 45% of deaths among heart disease patients and 51% of deaths among stroke patients.[97]

There is a complex interrelation amongst insulin resistance, hypertension, and diabetes. Approximately 25-47% of people with hypertension have developed insulin resistance or diabetes.[98] Treatment of blood pressure with antihypertensive drugs is often accompanied with improved insulin resistance.[99] Conversely, metformin, as the first line of antidiabetic drugs, can lower blood pressure among a non-diabetic hypertensive population.[100] Insulin regulates blood pressure during hyperinsulinemia by stimulating nitric oxide release in the endothelium.[101] Insulin resistance can increase blood volume by inducing sodium and potassium imbalance, or cause vasoconstriction by controlling calcium and magnesium imbalance.[102] Hypertension can damage the cardiovascular system and kidney slowly without a noticeable symptom.

Search for the genetic factors affecting cardiometabolic risk

Cardiometabolic risk is heritable

During the past few decades, multiple lines of evidence have shown a significant genetic basis on T2DM, CVD and their related traits. Family members share a genetic heritage, as well as environment, lifestyle and behaviour. Vassy and colleagues found statistically significant association between the number of parents

with diabetes and a T2DM genetic risk score (GRS) in the population-based PPP-Botnia Study (P=0.03) and a similar trend in the Framingham Offspring Study (P=0.07).[103] Twins provide a valuable source of information on the heritability of cardiometabolic disorders. Monozygotic twins share almost 100% of their genetic material and dizygotic twins, share about 50% of their genetic material, similarly to "regular" siblings. Besides genetics, twins also share much of the intrauterine and postnatal environments. In classical twin studies, by comparing the phenotypic resemblance between monozygotic twins and dizygotic twins, researchers can estimate how much of a trait's variation is explained by genetic variation.[104] The concordance rate for T2DM and abnormal glucose tolerance in monozygotic twins was constantly higher than in dizygotic twins indicating that genetics play a significant role in T2DM.[105] The estimated heritability from twin studies for obesity is between 40%-70%,[106-108] for blood pressure is 30%-60%,[109] for plasma cholesterol and triglycerides 56%-77%.[110,111]

Mapping the causal genes

Mapping the causal genes for cardiometabolic disorders has gone through two main stages: linkage analyses and genetic association studies. **Linkage studies** are based on the principle that genes that are sufficiently close on a chromosome tend to be inherited together during meiosis. Linkage analysis is most powerful for identifying genes for traits or diseases caused by mutations in a single-gene or a few genes.[112] Almost all cardiometabolic disorders are caused by variations in multiple genes with each of them conferring a small effect. Only a few studies have identified convincing linkage results among complex traits. Chromosome 10p demonstrated a strong linkage signal of T2DM in Mexican Americans[113] and Icelanders[114]. To find the causal region, Grant *et al* conduced single-marker association analysis with microsatellite genotype data and identified *transcription factor 7-like 2* (*TCF7L2*) significantly associated with T2DM.[115] Homozygous risk allele carriers have almost doubled risk of developing diabetes than the non-carriers.[115] Following the original finding, the effects of *TCF7L2* on T2DM risk were replicated in several other populations.[116-118]

Genetic association studies that investigate polymorphisms in candidate genes or across the whole genome are more efficient in identifying cardiometabolic disorder-related genes. Candidate gene studies require previous knowledge of disease aetiology in order to select the gene candidates. The most prominent examples of candidate gene studies are the discovery of the *peroxisome proliferator-activated receptor gamma* (*PPARG*) and *potassium voltage-gated channel subfamily J member 11* (*KCNJ11)*, both of which harbour missense mutations associated with T2DM and have been used as targets for developing antidiabetic drugs.[119-122] One of the greatest breakthroughs in genetic research was the completion of the Human

Genome Project (HGP) in 2003, which mapped ~20,500 genes in the human genome and resulted in the complete DNA sequence for all chromosomes.[123] Following this achievement, the International HapMap Project Phase I built a haplotype map of the human genome with 500,000 tags SNPs (out of 10 million SNPs),[124] which guided the design and prioritization of SNP for genotyping assays. Illumina and Affymetrix, the two largest competitors in sequencing and genotyping technology, have developed several GWAS arrays with ~550,000 to ~1 million genetic markers. At the early stage of GWAS, case-control design was most commonly used, where allele frequencies of the panel SNPs were compared between the healthy controls and cases using either logistic regression or a contingency table method. For quantitative traits, generalized linear regression models were often used to test associations.[125] Unlike the candidate gene studies, GWAS does not require prior knowledge of the genes and their roles in the disease aetiology. Due to its agnostic nature, GWAS has dramatically accelerated the pace of gene discovery. To date, the GWAS catalog has collected over 3300 publications with nearly 60,000 unique SNP-trait associations, with several thousand SNPs being associated with cardiometabolic risk.[126] The most strongly associated SNPs in GWAS are often in non-coding or intronic regions of genes. Fine-mapping of loci in large samples is often needed to identify functional variant(s). To do this cost-efficiently, the Illumina CardioMetaboChip array (in short, Metabochip) was developed to target the regions associated with a wide-range of metabolic disorders (e.g. T2DM, CVD, dyslipidaemia, obesity). The customized Metabochip array selected ~220,000 SNPs combining results of large meta-analyses of GWAS with the catalog of HapMap and 1000 Genome Project.[127] Exome genotyping arrays were also developed using the same technology as Metabochip to target the exonic regions and rare variants.[128] With the drastic decrease in cost, whole genome sequencing and exome sequencing, targeting both common and rare variants, have also become possible in large samples. The GWAS design is based on the *common diseases-common variants* hypothesis. Recently, sequencing-based association studies from the GoT2D and T2D-GENES consortia showed that rare and low-frequency variants may only play a minor role in the development of T2DM.[129]

GWAS discovery of cardiometabolic loci

Over 900 SNPs have been identified to be associated with BMI, waist-hip ratio and other obesity-related traits.[130] The *Fat Mass and Obesity-associated* gene (*FTO*) was the first robust GWAS-identified obesity gene[131] and so far the one with the largest effect on BMI and obesity risk in outbred populations.[132] *FTO* encodes an enzyme in the AlkB family of proteins which is involved in DNA repair, fatty acid metabolism, and posttranslational modification.[133] Because of its robust association with BMI, the *FTO* locus variants can be used as instruments for BMI and obesity in Mendelian randomization studies, which is an epidemiological study design for making causal inference of a risk factor on clinical outcomes.[134] Fall *et al* used

rs9939609 at the *FTO* locus as the instrumental variable for BMI and tested its association with 24 cardiometabolic traits in nearly 200,000 individuals. They found a positive association between BMI-increasing A allele of rs9939609 and multiple cardiometabolic outcomes, including T2DM, dyslipidaemia, hypertension, 2-h glucose, fasting glucose, systolic and diastolic blood pressure. Most of these associations were mediated through *FTO*'s effect on BMI,[135] supporting Kahn's hypothesis showed in Figure 1.

As the effect sizes of GWAS-identified genetic variants are generally small, a large sample size is always ideal to gain sufficient statistical power. Large international GWAS consortia have been established to facilitate intellectual collaboration and data-sharing. The Genetic Investigation of ANthropometric Traits (GIANT) consortium is an international collaboration with a focus on identifying genetic loci that modulate height and obesity-related traits. The GIANT consortium has published a series of highly cited GWAS papers on body height, BMI, waist, and waist-hip ratio.[136-147] The European Network for Genetic and Genomic Epidemiology (ENGAGE) consortium published the first comprehensive lipid GWAS study in 2008, in which they reported 22 loci affecting total cholesterol, HDL-C, LDL-C, and triglyceride levels.[148] The following year, Kathiresan *et al* reported 30 loci associated with lipoprotein levels, of which 11 signals were novel.[149] Two big waves of gene discovery for plasma lipids and lipoproteins were led by the Global Lipids Genetics Consortium (GLGC). They reported 95 loci and 159 loci independently associated with blood lipids, respectively.[150,151] Recently, Hoffmann *et al* conducted a GWAS using longitudinal electronic health records and reported 121 novel SNP associations. The explained variance of GWAS-identified SNPs to date for lipids traits ranged from 17.2% to 27.1%.[152]

Sladek *et al* conducted the first GWAS for T2DM in a French case-control cohort and identified four novel loci with modest effects on T2DM.[153] Later on, three GWAS reported similar top findings.[154-156] A second wave of GWAS was led by two large diabetes consortia: DIAbetes Genetics Replication and Meta-analysis Consortium (DIAGRAM) and Meta-Analyses of Glucose-and Insulin-related traits Consortium (MAGIC) discovering over 100 loci associated with glycaemic traits, insulin metabolism or diabetes.[157-166] A recent noticeable GWAS discovery was a meta-analysis of 62,892 T2DM cases with 596,424 controls from three large European GWAS datasets: DIAGRAM, GERA, and UK Biobank, which reported 139 common variants and four rare variants associated with T2DM.[167]

The earliest GWAS on blood pressure was not as successful as for other cardiometabolic traits. The first comprehensive GWAS study led by Wellcome Trust Case Control Consortium failed to detect any statistically significant association signals with hypertension.[156] In 2009, the Global Blood Pressure Genetics Consortium (GBPGEN) identified a number of regions, *CYP17A1*,

CYP1A2, *FGF5*, *SH2B3*, *MTHFR*, *c10orf107*, *ZNF652* and *PLCD3* associated with systolic or diastolic pressure in a GWAS study with 34,433 individuals of European ancestry.[168] The CHARGE Consortium identified four loci *ATP2B1*, *CYP17A1*, *PLEKHA7*, *SH2B3* associated with systolic blood pressure and six loci *ATP2B1*, *CACNB2*, *CSK-ULK3*, *SH2B3*, *TBX3-TBX5*, *ULK4* associated with diastolic blood pressure.[169] The CHARGE and Global Blood Pressure Genetics Consortium (GBPGEN) consortium merged into the International Consortium for Blood Pressure (ICBP). ICBP published several key GWAS papers on blood pressure in the following years.[170-172] Two large GWAS studies using ExomeChip genotyping data were published. Liu *et al* conducted a two-stage association study in 327,288 individuals and identified 32 novel loci.[173] Surendran *et al* identified 31 new blood pressure loci in nearly 350,000 individuals.[174] Using UK Biobank data, Warren *et al* identified and validated 107 loci associated with blood pressure (systolic, diastolic or pulse pressure).[175] Hoffmann *et al* identified ~320 independent loci in relation with blood pressure using electronic health records from three large studies.[176]

Search for the environmental triggers

Evidence of the environmental effect on cardiometabolic risk

Epidemiological studies support the view that environmental factors, mainly nutrition and physical activity, have contributed to the rising epidemic of cardiometabolic disorders in the modern society. Pima Indians are a group of native Americans who have been settled in the region now called southern Arizona for centuries. The Arizona Pima have transitioned from traditional farming to a typical American rural lifestyle, with high intakes of energy-dense food and a sedentary lifestyle since the 1900s. Correspondingly, the Arizona Pima Indians develop obesity and have the highest prevalence of diabetes (50%, age- and sex- adjusted) in the U.S.. Meanwhile, the genetically-similar Mexican Pima Indians continue with traditional farming involving intensive physical activity. The Mexican Pima are relatively leaner and have only 1/5 of the prevalence (6.9%) of diabetes, which is not different from the non-Pima Mexicans living in the same region. The difference in diabetes prevalence of these two groups is very likely caused by the differences of lifestyle and environments.[177]

The cardiometabolic risk increases when individuals move from low-risk to high-risk countries. Japanese migrants living in the US showed a higher prevalence of heart diseases, stroke, glucose tolerance, diabetes than the native Japanese.[178-180] Punjabi Indians living in West London demonstrated a higher BMI, systolic blood

pressure, serum cholesterol, apoB, fasting glucose and lower HDL levels compared with their siblings in Punjab, India.[181] CVD is rare in traditional African societies, but the African American have a higher risk of CVD than other ethnic groups living in the US.[182] The noticeable change in the disease rate among migrants indicates that environmental factors play a dominant role in driving the increase of prevalence in cardiometabolic disorders.

The global shifts in nutritional environments have been contributing to the outbreak of cardiometabolic diseases. Remarkable changes have occurred, especially in developing countries. Benefiting from international trade and free market, people in low- and middle-income countries have access to a wider variety of food products at a lower cost than ever before. A typical example is edible oil consumption; By 2010, inexpensive edible oil has been available for both developed and developing countries. Meanwhile, edible oil consumption has increased 3-6 fold in all populations compared to 1980s.[183] Another noticeable diet change is added sugar; In the U.S., SSB are the top source of calories and largest source of added sugar.[184] In Mexico, 1/5 of total energy intake is from SSB in both adolescents and adults.[185] SSB have high sugar content, produce low satiety levels, and provide incomplete compensation of total energy.[186] Malik *et al* reviewed 30 studies from 1996 to 2005 and concluded that SSB consumption was associated with weight gain and obesity.[187] The high content of rapidly absorbable sugar in SSB may lead to insulin resistance, β-cell dysfunction, inflammation, hypertension, dyslipidaemia, and finally increase the risk of T2DM, MetS and CVD.[188] Schulze *et al* suggested that the effect of SSB in T2DM and MetS is modulated through its association with obesity.[189] Another important change in diet is animal source food intake, including eggs, meat (pork and all other red meat), poultry, dairy, and fish. The most significant change in animal source food consumption occurred in middle- and low-income countries. There was a 10-fold increase in animal source food intake in China in 2000 compared to 1960s.[183] A meta-analysis has shown that processed meat is associated with an increased risk of CHD and diabetes. However, there is no clear evidence of association between diabetes and CVD and red meat.[190] Consumption of legumes and coarse grains have also been reduced. Although there has been a significant increase on fruit and vegetable intake, the change has been less marked than changes in edible oil and animal source food. Total energy intake in developing countries increases too, topped by the Middle East, followed by China, Latin America.[191] Reduced physical activity due to motorized travel, labor-saving devices at home and work are also seen. Leisure time activities are becoming more sedentary; television watching and video games are becoming the main entertainments among adults and adolescents. It is estimated that physical inactivity contributes 6% to the burden of CHD and 7% to T2DM.[192]

Exposome and environment-wide association studies (EWAS)

For chronic diseases, genetics only has a modest influence of the disease risk.[193] Thus, identifying specific environmental risk factors associated with cardiometabolic risk and designing/implementing interventions in the general population becomes critical for disease prevention. Some of these environmental factors have been intensively investigated and replicated for their association with cardiometabolic traits. Nevertheless, there are many other environmental factors that have not been investigated or have an unclear role in disease development. The majority of environmental studies tend to focus on one exposure or one category of exposures, such as air pollution, diet or physical activity, which leads to a highly fragmented literature on epidemiological associations.[194] The concept of exposome was proposed by Wild to encourage researchers to evaluate the effects of environmental exposures in a systematical manner.[195] A complete exposome refers to the totality of life-course environmental exposures (including lifestyle factors), from the prenatal period onwards.[195] Unlike the genome, the exposome is highly dynamic. At different timepoints of life, an individual will have a particular profile of exposures.

To clarify the concept, four layers of environmental exposures may be considered (Figure 2): social and economic determinants, behavior and environmental risk factors, extracellular environmental signals, and intrinsic environmental factors. Social and economic factors are on the outmost layer of all environmental exposures. The impact of social and economic factors can be penetrated to the inner layers and finally influence the innermost cellular environment which directly interacts with the genome or impose epigenetic changes. For example, with the social and economic growth in the past 30 years, SSB become more affordable specially in middle- and low-income countries. Consumption of SSB can induce dramatic increase of glucose and insulin in the extracellular environment – blood, and further suppress the expression of the appetite hormone - ghrelin.[196]

The concept ‘exposome’ refers to the complete assessment of environmental exposures from conception onwards. A practical strategy to implement this concept in research is to take snap-shots of a person’s exposure data at different stages of life. Many large existing cohort studies with multiple time-points of data are available and constitute a precious resource to study the exposome in a comprehensive way. As discussed above, the external exposures can exert influence on the internal environment. Thus, it is reasonable to measure the internal chemical environment as a surrogate of external exposures.[193] With the development of omics technologies, such as transcriptomics, proteomics, metabolomics or epigenomics, a large amount of data can be generated to characterize downstream biological events, which in principle can be examined for association with the disease endpoint, in an analogous manner to GWAS.[197]

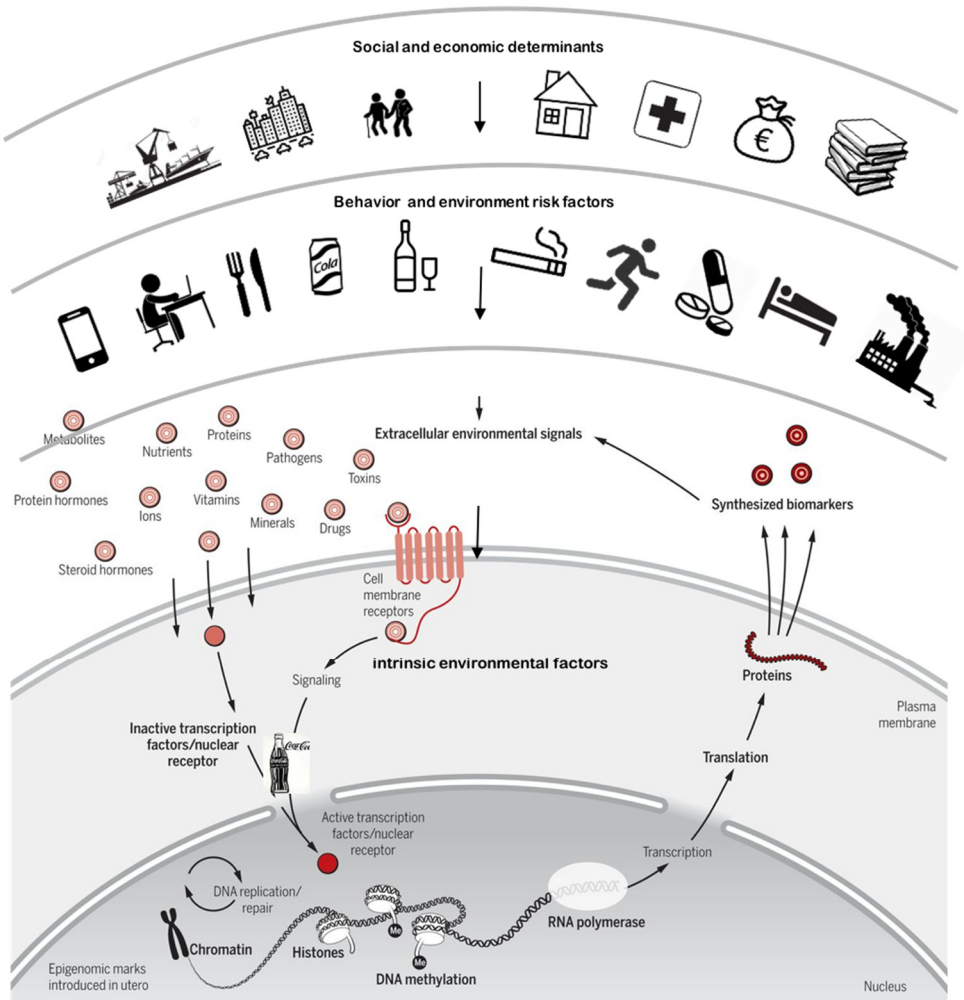


Figure 2. The concept of exposome, adapted from Franks, PW *et al*, Exposing the exposures responsible for type 2 diabetes and obesity. Science. 2016 Oct 7;354(6308):69-73.

Patel *et al* formalized the exposome concept and termed it as EWAS.[198] The purpose of EWAS is to comprehensively screen environmental factors for their association with disease (traits). In the first EWAS study, Patel *et al* used 266 unique environmental agents measured in blood as exposures and examine their association with T2DM.[198] Since then, EWAS have been widely applied to investigate whether nutrients, contaminants, prescribed drugs, lifestyle, and socio-economic factors are associated with diseases and disease complications.[199-203] Tzoulaki *et al* performed

EWAS analyses on 82 nutrients and 3 urine electrolytes with systolic and diastolic blood pressure: they confirmed some previously reported associations, such as the inverse association between non-hem iron, phosphorus, magnesium and blood pressure; more importantly, they discovered that B vitamins (folacin, riboflavin and thiamin) were negatively associated with blood pressure, which was poorly studied before.[199] By merging the Swedish Cancer Register with Prescription Drug Register, Patel *et al* tested associations of 552 pharmaceutical prescriptions with different type of cancer risk (any, breast, colon, or prostate cancer) during a 5.5 years of follow-up in 9,014,975 individuals; they found that 26% of the studied drugs were associated with any cancer in a time-to-event analysis and also identified the drugs associated with different type of cancers.[201]

Gene-environment (G-E) interaction

G-E interaction can be understood as "a different effect of an environmental exposure on disease risk in persons with different genotypes," or alternatively, "a different effect of a genotype on disease risk in persons with different environmental exposures".[204] Figure 3 shows a simplified scenario of G-E interaction for a quantitative traits, in which the effects of the environmental exposures on a quantitative trait are plotted across genotypes. If the estimated effect of the environmental exposure significantly differs across genotypes, we can conclude that a G-E interaction may exist. This type of interaction is usually tested using a multiplicative term in a linear regression model. Many other approaches for modeling interactions exist (see paper by Thomas[11] for detailed overview of G-E interaction methods).

The earliest G-E interaction studies relied on hypothesis-driven approaches, in which the combined effects of genetic loci (often identified from animal studies or *in vitro* studies) and specific candidate environmental exposures (e.g., dietary fat, physical activity, smoking, education level) in a given disease or trait were tested. More recently, the candidate variants for interaction studies have been these identified through GWAS meta-analyses for their main effects. These large studies have identify thousands of loci associated with relevant traits and researchers have often used the most strongly associated variants in studies of G-E interactions.[205] SNPs identified by GWAS can be tested for interaction directly, or by summing the risk alleles to construct a GRS. Qi *et al* calculated GRS on the basis of 32 BMI-associated loci in three large US cohorts and analyzed its interaction with the intake of SSB in relation with BMI or obesity.[206] They showed that the genetic association with adiposity was stronger among individuals with higher consumption of SSB.

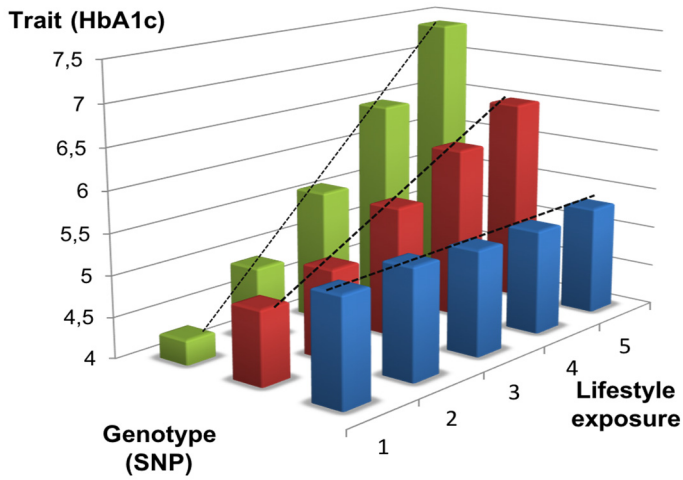


Figure 3. A hypothetic scenario of G-E interaction between a biallelic variant and an environmental risk factor. In this case, a G-E interaction may exist because the magnitute of association between environmental exposures and outcome are significantly different across different genotype subgroups. Ref: Franks PW. Gene × Environment Interactions in Type 2 Diabetes. Curr Diab Rep (2011) 11: 552.

The interaction effect sizes are generally small in magnitude, and it is suggested that a well-powered G-E interaction study requires at least four times larger sample size than that for an association study with a comparable magnitude.[11] Although an extensive literature on G-E interaction exists, it is likely that many of these studies are inadequately powered and only a few have been replicated in independent populations. Li *et al* demonstrated that physical activity diminishes the genetic risk of obesity predisposed by 12 BMI related SNPs in the EPIC-Norfolk cohort.[207] To replicate Li's finding, Ahmad *et al* meta-analyzed over 110,000 adults of European ancestry, almost a six time larger sample size of the original report, to yield a statistically significant GRS-physical activity interaction; this may be caused by a substantial loss of power due to heterogeneity in the individual cohorts.[208]

Aims & Objectives

The overarching objective of this thesis is to investigate how genetic and environmental factors independently and jointly affect cardiometabolic traits.

The Specific Aims are:

Paper I: to investigate whether the risk of obesity associated with SSB and artificially sweetened beverages (ASB) intake is modified by genetic predisposition to obesity in two large Swedish cohorts: GLACIER and MDCS.

Paper II: to systematically assess whether genetic variation in the *FADS* genes region modifies the association between dietary polyunsaturated fatty acid (PUFA) intake and cardiometabolic traits in the GLACIER cohort and to functionally annotate top-ranking signals to estimate their regulatory potential.

Paper III: to conduct EWAS analyses in a longitudinal setting to screen modifiable lifestyle factors associated with cardiometabolic risk in a sub-cohort of the Västerbottens Hälsoundersökning study (in English "Västerbotten Health Survey", VHU; also known as the Västerbotten Intervention Study)

Materials and Methods

Below is the brief summary of study materials and methods. Detailed information about cohort description, clinical data measurements, genotyping information, and statistical methods is provided in each manuscript.

Study populations

In this thesis, we tested different hypotheses mainly using sub-cohorts of the VHU. A sub-cohort of VHU called GLACIER (cohort registration number: ISRCTN35275922) has genotype information available.

VHU is a prospective, population-based cohort study from a sub-Arctic population in Västerbotten county in northern Sweden. In the 1970s and early 1980s, Västerbotten had the highest CVD mortality in Sweden with 720/100,0000 inhabitants per year among adults.[209] VHU was designed to monitor risk factors and ultimately reduce the local morbidity and mortality from CVD and diabetes. Since 1984, all adult residents in Västerbotten were invited to participate in a comprehensive health survey at the age of 30, 40, 50, and 60 years. Recruitment of people at age 30 years was subsequently discontinued owing to resource constraints. Participants were requested to fast overnight before visiting their primary care center. Blood samples were drawn by trained nurses and stored at the Northern Swedish Biobank in Umeå. **Paper I** and **Paper II** used the sub-cohort GLACIER in which around 6,000 participants were genotyped with the MetaboChip array.[210] Another sub-cohort of VHU, including 69,765 participants with 94,991 health examinations, was used for the analyses in **Paper III**, among which 18,493 participants underwent a 10-year follow-up examination and a further 1,793 participants also had 20-year follow-up examinations. The Regional Ethical Review Board in Umeå approved the study protocol and all study participants provided written informed consent as part of VHU.

The MDCS is a prospective, population-based cohort study from the southern Swedish city of Malmö.[211] The study was a joint effort between the International Agency for Research on Cancer (IARC), the Swedish Cancer Society, the Swedish Medical Research Council and the Faculty of Medicine, Lund University, Sweden.

Between 1991 and 1996, all residents born between 1926 and 1945 living in Malmö were invited to participate in the study through letters and advertisements in local newspapers and public places.[212] The primary aim of the MDCS was to investigate the plausible associations between dietary factors and certain types of cancers.[211] As time went on, diabetes and cardiovascular diseases research has also been intensively conducted within MDCS. In the analyses for **Paper I,** we selected participants with complete dietary data regarding SSB and ASB consumption, genetic information and excluded those with T2DM, CVD, and cancer. In total, 21,824 participants in MDCS were eligible for the analyses. The ethics committee at Lund University approved the MDCS protocols and all participants provided written informed consent.

Study specific methods

Clinical characteristics

In both GLACIER and MDCS, body weight was measured with a balance-beam scale to the nearest 0.1 kg and height to the nearest 1 cm with a wall-mounted stadiometer, allowing participants to wear indoor clothing and no shoes. BMI was then calculated as weight in kilograms divided by the square of height in meters. In VHU (**Paper II** and **Paper III)**, capillary blood was drawn following an overnight fast and a second draw was carried two hours after a 75g oral glucose load. Blood glucose and serum lipid levels were then measured with the Reflotron bench-top analyzer (Roche Diagnostics Scandinavia AB). LDL-C levels were calculated using the Friedewald formula.[213] Blood pressures were measured once using mercury-gauge sphygmomanometer. From 2009 onwards, serum lipids were analyzed using clinical chemical analysis and blood pressure was measured twice in a sitting position and the average was used for analyses. More details about the clinical data sampling are given in each paper.

Diet and lifestyle measurement

In the VHU study, socio-economic factors, general health, psychosocial factors, tobacco/alcohol consumption of the participants were obtained using standard questionnaires. Physical activity was assessed through a modified version of the International Physical Activity Questionnaire.[214] A semi-quantitative food frequency questionnaire (FFQ) was used to quantify the dietary habits during the past year.[215] The FFQ initially consisted of 84 independent or aggregated food items and beverages and had been used until the mid-1990s. Since 1996, the FFQ was

shortened to 66 items by combining similar items or removing items that provided minimal unique information. Portion size, which ranged from never to ≥ 4 servings/day, was used to quantify the general portion size of staple foods (potato/pasta/rice), vegetables, and primary protein source (meat/fish). Total energy intake was calculated based on the nutritional values available through the Swedish Food Composition Database.[216] In **Paper I**, the 9-level frequency scale of SSB intake (never, occasionally, 1–3 times/month, 1 time/week, 2–3 times/week, 4–6 times/week, 1 time/day, 2–3 times/day, 4 times/day) was recoded into four categories: seldom, ≤ 2 times/year; low, 1–3 times/month; medium, 1–3 times/week; and high, ≥ 4 times/week to align the ranges within the four categories of SSB intake in the MDCS. ASBs consumption data was not available in GLACIER. In **Paper II**, dietary n-3 PUFA was calculated by summing the intakes of α-linolenic, eicosapentaenoic acid (EPA), and docosahexaenoic acid (DHA) intakes. Dietary n-6 PUFA was calculated by summing the intakes of linoleic acid and arachidonic acid (ARA). Total PUFA is the sum of n-3 PUFA and n-6 PUFA intake. In **Paper III**, some of the dietary variables were collapsed into single items according to suggestions offered by Swedish Board of Agriculture.[217] The categorical variables that were not ordinal were further converted into binary variables (described in detail in the paper). A total of 108 numeric and 52 categorical variables were included in the analyses.

In MDCS, dietary intake was assessed using both a 7-day food menu booklet and a 168-item FFQ. A 45-minutes interview was conducted to acquire additional information such as cooking methods and product choice. Mean daily food intake (g/d) was calculated based on the information above. Season, as a variable, was added as a covariate to adjust for its potential confounding effects on dietary habits. Models were further adjusted for dietary assessment methods as a minor change was implemented in September 1994. SSB in MDC included both carbonated and noncarbonated beverages but not juice. ASB were defined as beverages sweetened with low-energy or energy-free sweeteners instead of sugar.

Both VHU and MDCS used the Swedish Food Composition Database as the reference to calculate nutrient and energy content. When necessary, additional harmonization is done by dichotomizing the variables.

Genotype data

In GLACIER (**Paper I** and **II**), genomic DNA was extracted from peripheral leukocytes and diluted to 4 ng/µl.[218] Genotyping was performed using MetaboChip.[210] In MDCS (**Paper I**), genomic DNA was extracted from whole blood samples using Qiagen Maxipreps. Genotyping was performed using a Sequenom MassArray. For SNPs in which the Sequenom genotyping failed, TaqMan or KASParallelic discrimination on an ABI 7900HT (Applied Biosystems) was conducted.

In **Paper I**, the original plan was to use 32 BMI-associated SNPs identified through GWAS[137,138] to construct a GRS. For those SNPs not present in MDCS or GLACIER, we use the web-based tool SNAP[219] to identify highly correlated SNP proxies. To facilitate comparison and subsequent meta-analysis, we excluded the missing SNPs with no proxies available from both cohorts. We replaced the missing genotypes with the mean in both cohorts for participants with >60% of the genotype information available (>18 SNPs genotyped out of 30 SNPs) as previously described.[220] We constructed a GRS for each individual by summing up the number of risk alleles for the 30 BMI-associated loci. A second weighted GRS (wGRS)[221] based on the published effect sizes of the BMI-related SNPs was also constructed.

In **Paper II**, we extracted the genotypic information for all available SNPs (N = 290) in the *FADS* region (Chr.11:61317028-61416099, build 36). Under the hypothesis that the regulatory regions of these genes may be proximal to the exonic sites, we also extracted the genotypes for variants within 200 kb upstream and downstream of the *FADS* region (N = 436). Rare variants with minor allele count <10 were excluded (N = 284). Variants with a Hardy–Weinberg equilibrium *P*-value <0.0001 (0.05/436) were flagged but not excluded from the analyses, as potential deviations from expected genotype frequencies might reflect evolutionary processes, which were of special interest for the *FADS* region. In total, 442 variants were included in the analyses, of which 24% of SNPs had a minor allele frequency ≤5% (Figure 4). Implementing the algorithm developed by Gabriel *et al*[222] in Haploview,[223] 30 haplotype blocks covering the region were derived. Genotype values at each locus were coded as 0, 1, and 2, thereby assuming an additive effect of alleles.

Figure 4. Minor allele frequency distribution of the SNPs on MetaboChip in the *FADS1-FADS2-FADS3* gene cluster

Statistical Methods

Statistical analyses in **Paper I** were performed using SPSS version 20.0 (IBM) in MDCS and SAS version 9.4 (SAS Institute) in GLACIER. Meta-analysis in **Paper I** was performed using Stata Version 12 (StataCorp) Statistical analyses for **Paper II** and **Paper III** were performed using *R* software v3.2.2.[224] Additionally, PLINK v1.07,[225] gPLINK v2.050,[225] and Haploview[223] were used for extraction of genotype data to reconstruct of the haplotypes in **Paper II**

Linear regression

In **Paper I**, we applied generalized linear regression models by fitting BMI as the dependent variable and the interaction term (GRS×SSB or GRS×ASB) together with their marginal effect terms. The basic model was adjusted for age, sex, and study-specific covariates (Equation 1). In GLACIER, we added FFQ version as a study-specific covariate. In MDCS, season was included as a study-specific covariate in the analyses. In a secondary model, we further added physical activity, smoking, alcohol consumption, and total energy intake as covariates to account for their potential confounding effects (Equation 2).

Equation 1. $BMI = \alpha + \beta_{age} + \beta_{sex} + \beta_{study\ specific\ variable} + \beta_{GRS} + \beta_{SSB/ASB} + \beta_{GRS\times SSB/ASB} + \varepsilon$

Equation 2. $BMI = \alpha + \beta_{age} + \beta_{sex} + \beta_{study\ specific\ variable} + \beta_{PA} + \beta_{smoking} + \beta_{alcohol} + \beta_{TEI} + \beta_{GRS} + \beta_{SSB/ASB} + \beta_{GRS\times SSB/ASB} + \varepsilon$

Here, α is the intercept; GRS, genetic risk score; SSB, sugar-sweetened beverages; ASB, artificial sweetened beverages; PA, physical activity; TEI, total energy intake.

In **Paper II**, we performed generalized linear regression modelling to assess the pairwise interaction effects for the six candidate SNPs found to modulate fatty acid composition [226] and dietary PUFA intake with a range of cardiometabolic traits as outcome variables. To control for the confounding effects from total energy intake, we regressed PUFA intake on total energy intake to compute residuals and added residuals along with total energy intake in the model. We further adjusted the models for age, age^2, sex, FFQ version, and the first four genomic principal components to account for population stratification. The model with body weight as the outcome was additionally adjusted for height (Equation 3). As lipid traits (HDL-C, LDL-C, triglycerides, and total cholesterol) and blood glucose (2h-glucose, and fasting glucose) can fluctuate with the fasting states, we added fasting status known (1/0 for yes vs. no) and fasting hours (0/1 for >8h vs. 4-8h) as dummy covariates (Equation 4).

Equation 3. $Weight = \alpha + \beta_{age} + \beta_{age^2} + \beta_{sex} + \beta_{FFQ\ version} + \beta_{SNP} + \beta_{TEI} + \beta_{height} + \beta_{PC1-4} + \beta_{PUFA.residual} + \beta_{SNP\times PUFA.residual} + \varepsilon$

Equation 4. $lipids/glucose = \alpha + \beta_{age} + \beta_{age^2} + \beta_{sex} + \beta_{FFQ\ version} + \beta_{SNP} + \beta_{TEI} + \beta_{fasting\ status} + \beta_{fasting\ hours} + \beta_{PC1-4} + \beta_{PUFA.residual} + \beta_{SNP \times PUFA.residual} + \varepsilon$

Here, α is the intercept; ε is the error; TEI, total energy intake.

Meta-analysis

In **Paper I**, we conducted both fixed-effects and random-effects meta-analysis using the inverse variance-weighted average method (IVW) to pool the results from MDCS and GLACIER. The meta-analysis revealed a low heterogeneity statistic in the random-effect meta-analysis; thus, we only reported the results from the fixed-effect models.

Gene-/haplotype-centric analyses

We performed region-wide and haplotype-wide interaction and joint tests with dietary PUFA intake for different cardiometabolic traits in **Paper II**. Based on the assumption that rare variants may convey larger underlying effects than we expected, we applied a weighting method used in Sequence Kernel Association Test (SKAT)[227] in which the weights of rare variants are greater than the weights of the common variants. We conducted the G-E interaction tests with fixed genetic effects (INT-FIX), with random genetic effects (INT-RAN), and joint association and interaction effects (JOINT), while adjusting for covariates (same covariates as the interaction tests for single-variants). Then we applied Bonferroni correction for adjustment of multiple testing. As the dietary PUFA variables (n-3 PUFA, n-6 PUFA, and total PUFA) are correlated, as are the cardiometabolic traits, we calculated the effective number of environmental factors (1.37 out of 3) and phenotypic traits (7.38 out of 8).[228] Accordingly, $P \leq 0.005$ [0.05/(7.38×1.37)] was considered as statistically significant in gene-centric analysis. $P \leq 0.002$ [0.05/30] was considered statistically significant in haplotype-centric analysis (only one environmental exposure and outcome used here).

Functional annotation

In **Paper II**, most of the original candidate SNPs[226] and variants in the haplotype blocks showing evidence for interaction were located in the non-coding region of the genome. To pinpoint the most likely causal genetic variants, we applied a series of tools to functionally annotate the genetic variants. Firstly, we used CHromHMM[229] to systematically classify the chromatic states of the variants into enhancers, repressors, promoters, insulators or other states in nine human cell lines.

With another online database, 3DSNP,[230] we ranked the candidate variants and the variants in the same linkage disequilibrium (LD) block (r^2>0.8) based on their functionality score. We prioritized the variants with higher functionality score than their original counterparts in the subsequent steps of the analysis. The functionality score was assigned based on six functional parameters: i) evidence for disruption of transcription factor binding sites (TFBS); ii) evolutionary conservation; iii) ability to alter sequence motifs; iv) being located in a promoter region; v) being located in an enhancer region; and vi) number of topological interactions with distant genomic regions (genes and/or variants) via chromatin loops.[230] We illustrated the functional parameters of the top-ranking SNPs in radar charts (one example presented in Figure 7). Topological features of these SNPs were shown in Circos plots (Figure 8). We obtained serum lipids quantitative trait locus (QTL) and expression QTL (eQTL) information for these variants using HaploReg.[231] We performed gene ontology enrichment analysis on distal interacting genes using the PANTHER Overrepresentation Test to identify relevant biological processes and pathways.[232]

Environment-wide association study (EWAS)

In **Paper III**, we performed EWAS analyses to assess the relationships of ~160 environment exposures with cardiometabolic traits from >16,000 adults living in northern Sweden. EWAS is an analogous approach to GWAS focusing on environmental variables. Taking advantage of the repeatedly measured environmental data in the VHU sub-cohort (Figure 5), we ran linear mixed regression to model the average linear effects of environmental variables on cardiometabolic traits (Equation 5). We modeled the individual differences by assuming different random intercepts for each participant. The models for BMI and blood pressure were adjusted for age, age^2, sex, FFQ version (only for dietary variables) and follow-up years. Models for lipid and glucose traits were additionally adjusted for fasting states. We adopted a conventional Bonferroni correction to control for multiple testing: $P \leq 3.47 \times 10^{-5}$[0.05/(9×160)] was considered statistically significant. We further sorted all experiment-wide significant results into 10 discrete categories that might represent targets for lifestyle interventions: i) alcohol consumption; ii) non-alcoholic beverage consumption; iii) food; iv) nutrients; v) general health; vi) physical activity & fitness; vii) psychosocial; viii) sleep; ix) social conditions; and x) tobacco use. Then, we rank-ordered these variables within each category by the variance explained (R^2) in each outcome and, for the sake of simplicity, focus on the top five variables from each category.

Equation 5. $Trait = \alpha + \beta_{exposure} + \beta_{age} + \beta_{age^2} + \beta_{sex} + \beta_{other\ covariates} + (1|participant) + \varepsilon$, where $1|participant$ represents different random intercepts for each participant, and FFQ version is added as additional covariate only for dietary variables.

We additionally tested whether those lifestyle variables were associated with long-term changes of cardiometabolic traits (ΔTrait) using linear regression model (Equation 6).

Equation 6. $\Delta Trait = \alpha + \beta_{exposure} + \beta_{age.B} + \beta_{age.B^2} + \beta_{age.F} + \beta_{sex} + \beta_{other\ covariates} + \varepsilon$, where $age.B$ is the age at baseline and $age.F$ is the age at follow-up, and FFQ version is added as additional covariate for dietary variables.

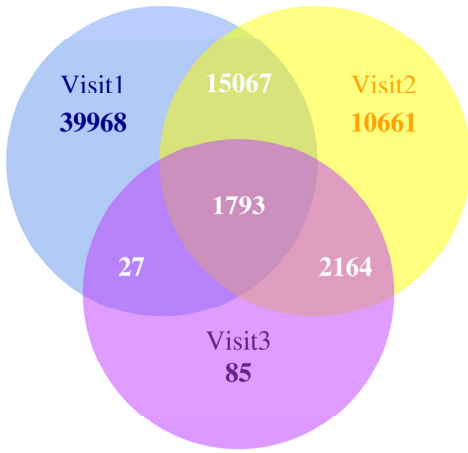


Figure 5. Venn diagram showing the distribution of participants on the three visits. The overlap represents the participants having mutiple visits.

Results and Discussion

Paper I

Overweight and obesity increase the risk of T2DM, hypertension, and CVDs and have become one of the most serious public health challenges in the modern times.[48] Numerous epidemiological studies, including clinical trials, have shown a strong positive associations between SSB intake and BMI.[184,233-235] Trumbo *et al* conducted a systematic review of 59 observational and 17 interventional studies and found inconsistent evidence of the association between SSB intake and BMI after adjusting for total energy intake.[236] The inconsistency of the evidence may be partly attributed to G-E interactions. Qi *et al* analyzed data from three large US cohorts (N~33,000) and observed a more pronounced genetic association with adiposity among individuals with higher consumption of SSB, but not with the consumption of ASB.[206] These exciting results motivated us to seek replication of these findings, as described in **Paper I**.

We constructed a GRS based on 30 BMI-associated SNPs in MDCS and GLACIER. Firstly, we tested the association of SSB, ASB (only in MDCS), and the GRS on BMI. In the meta-analysis, each increment in SSB intake category was associated with 0.18 kg/m^2 average increase of BMI (standard error [SE]=0.02, P=1.7×10^{-20}). Notably, each increment in ASB intake category was shown to be associated with 0.64 kg/m^2 average increase of BMI in MDCS (SE=0.04, P=3.9×10^{-58}). Overweight or obese people may choose artificial sweeteners over sugar as means to lose weight.[237] Another possible mechanism can be that the artificial sweeteners do not stimulate insulin secretion thus cannot suppress appetite, leading to increased total energy intake.[238] In addition, ASB may modulate microbiota composition and microbial metabolic pathway which links to glucose intolerance in the host.[239]

The main focus of this paper was to investigate the interaction effects between SSB/ASB and genetic predisposition to BMI. There was a statistically significant interaction (P_{int}=0.02) between the GRS and SSB intake (defined either as four categories – seldom, low, medium, and high intake –, or as dichotomized intake – seldom-to-low consumption vs medium-to-high consumption). The magnitude of the association between the GRS and BMI was greater at higher levels of SSB consumption. Further adjusting for potential confounding factors (e.g. season, and evaluation method) did not materially change the magnitude of the interaction. In

the lifestyle-adjusted pooled analysis with dichotomized SSB intake, each 10-unit increment of the GRS was associated with a mean 1.31 kg/m^2 higher BMI (SE=0.11) in individuals reporting medium-to-high SSB intake (P=1.2×10^{-33}). Among participants reporting seldom or low SSB intake, each 10-unit increment of the GRS was associated with a 0.83 kg/m^2 (SE=0.09) higher in BMI (P=6.0×10^{-21}). Similar results were obtained using the wGRS. We also reported the results of the association between SSB intake and BMI by GRS categories (Figure 6). In the pooled analyses, each SSB category increment was associated with 0.15 kg/m^2 (SE=0.04) increment of BMI (P=1.3×10^{-4}, n=6835) in the lowest GRS quartile, compared with 0.24 kg/m^2 (SE = 0.04) increment of BMI (P=2.9×10^{-8}, n=6766) in the highest GRS quartile. In accordance with Qi *et al*,[206] we did not observe any interaction between ASB and genetic predisposition.

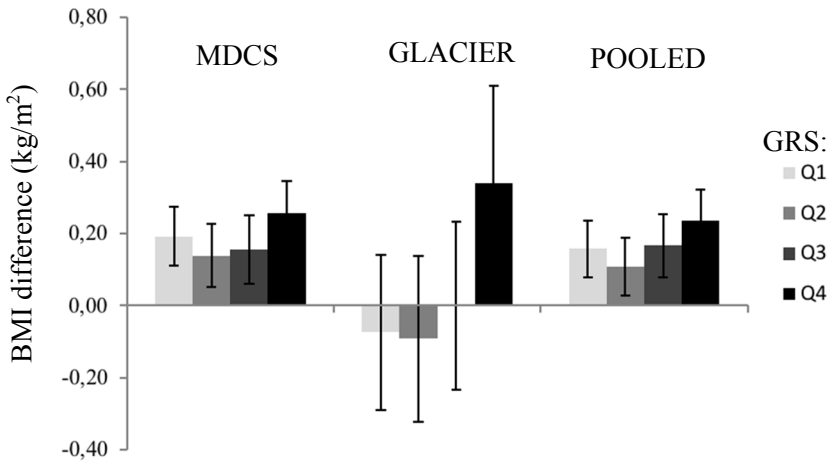


Figure 6. The association between SSB intake and BMI in MDC (N=21,824), GLACIER (N=4,905) and pooled data stratified by quantiles of the GRS. Analysis were adjusted for age, sex, alcohol consumption, smoking, physical activity, total energy intake, and other cohort specific variables. Inverse-variances weighted fixed effect meta-analysis was used for the pooled analysis.

It is important to bear in mind that the current study is a cross-sectional study in nature and SSB and ASB intake were assessed using self-administered questionnaires. We attempted to meta-analyze two cohorts to maximize the statistical power, but the SSB intake was assessed differently in each cohort, which may have increased heterogeneity and thus reduced statistical power. Specifically, a 7-day food diary was used to evaluated SSB intake in the MDCS which may have resulted in a larger number of seldom consumers than in the GLACER cohort, in which an FFQ was implemented. The first quantile of SSB intake group in MDCS

consisted of zero-consumers only, which was not the case in GLACIER. All of these differences may partially explain why the GLACIER population demonstrated a more noticeable volume-dependent trend on BMI for SSB intake.

Paper II

The *FADS* gene cluster encodes the key regulating enzymes in PUFA metabolism. Genetic variants in *FADS* had been previously associated with blood lipids, glycemic traits and clinical endpoints such as T2DM and CVD.[240] In a recent study, Fumagalli *et al* identified six SNPs in the *FADS* region showing strong positive selection among Greenlandic Inuit,[226] a population isolate in the Arctic circle. The extreme, cold weather in the Arctic makes the region unsuitable for agriculture or animal husbandry, which forces the Inuit to subsist predominantly on marine food. The traditional Inuit diet has high levels of monounsaturated and n-3 PUFAs, which may have placed certain genetic variants regulating fat desaturation under selective pressure. Fumagalli *et al* found associations between these six SNPs and multiple quantitative traits, including bloods lipids and glucose, with the strongest effects seen on body weight and height.[226] These associations remained statistically significant after adjusting for other variants in the *FADS* gene cluster, indicating that these were independent associations. Thus, we systematically examined the presence of gene × dietary PUFA interactions at the *FADS1-FADS2-FADS3* gene cluster across multiple cardiometabolic traits in a population isolate from the north of Sweden.

The most robust interaction, from the tests involving the six SNPs above, was shown between rs174602 and n-3 PUFA intake on total cholesterol (P_{int}=0.001). When the participants were stratified by n-3 PUFA intake, the minor C allele at rs174602 was associated with decreased total cholesterol (P=0.02) among those with low n-3 PUFA intake; no such trend was observed among those with high n-3 PUFA intake (P=0.50). The *FADS* gene cluster is well-known for its role in lipid metabolism and SNPs in *FADS1* have also shown to be associated with fasting glucose in GWAS.[241] We observed an interaction between the *FADS2* rs174570 and n-6 PUFA intake on fasting glucose level (P_{int}=0.005). For participants with high n-6 PUFA intake, the T allele at rs174570 was associated with decreased fasting glucose level (P=0.01). No association (P=0.34) was observed in participants who reported low PUFA-6 intake. The possible biological mechanism of *FADS* variants influencing glucose or insulin metabolism is that the modulation of long-chain PUFA composition can influence membrane fluidity which further leads to changes of the number of insulin receptors and their affinity to insulin.[242]

The associations found in Greenlandic Inuit population may not be the same in a northern Swedish population due to different diets and adaption of *FADS* genes.[243] Therefore, we extended our analyses to the entire *FADS1-FADS2-FADS3* gene cluster to search for possible interaction signals. In the gene-centric analyses, we observed a significant interaction effect between the *FADS* gene cluster and n-3 PUFA intake on triglycerides (P_{INT_RAN}=0.005). Subsequently, we divided the *FADS* gene region into haplotypes and refined the interaction signal to haplotypes (haplotype-centric analyses) in relation to blood triglyceride levels. None of the haplotype block interactions surpassed the pre-defined threshold (P≤0.002); however, three haplotype blocks (haplotype block 12, 16, and 21) showed tentative evidence for interactions with n-3 PUFA on triglycerides.

We functionally annotated all the genetic variants in these three haplotypes and the six variants[226] originally reported, using ANNOVAR. We used ChromHMM to classify the variants of interest into enhancers, repressors, promoters, insulators or other chromatin states based on data from nine human cell types. A cluster of variants in haplotype 21 showed enhancer enrichment in chromatin state, two neighboring variants (rs187943834 and rs117518711) showed promoter states and one variant (rs7115739) showed an insulator state. Variants in haplotypes 12 and 16 in HepG2 cells demonstrated much less regulatory states. rs174570 showed weak promoter, active promoter, or strong enhancer states across different cell lines. rs174602 showed weak transcription state in most of the cell lines.

Using the eb-based database 3DSNP, we obtained information about distal interacting genes and prioritized lead variants based on SNP functionality scores. The variant rs174570 showed the highest functionality score (score=96.7, driven by promoter status) among all tagged SNPs. After re-ranking all variants by the leading SNP, the *FADS2* intronic rs5792235 deletion (CA/C) showed the highest functionality score (score=218.8, driven by promoter and transcription binding site [TFBS] status), which is 18X larger than its original proxies, rs174599 (score=12.13; r^2=0.83) and rs174601 (score=11.69; r^2=0.83). We show the proximal and distal interaction genes for rs5792235 using a Circos plot (Figure 8). Many of the distal interacting genes have shown to be related with lipids, glycaemia, and other cardiometabolic traits. These findings are supported by results generated using other functional annotation tools: for example, HaploReg eQTL analyses indicated that rs5792235 and its distal interacting variants are associated with glycemic and lipid traits in previous GWAS. All the evidence above suggests that rs5792235 in *FADS2* is a potential causal variant in the region.

There are a few limitations in the current study. One is that the dietary PUFA intake is derived from FFQs. Although the FFQ used in the VHU cohort has been carefully validated, self-reported method relies on the participants' memory and is less accurate to measure the absolute intake of dietary PUFA. We are also aware of that

in-silico functional validation only provides suggestive evidence of functionality. *In vitro* experiments would offer more conclusive insights into the mechanism, but those analyses are beyond the scope of this study.

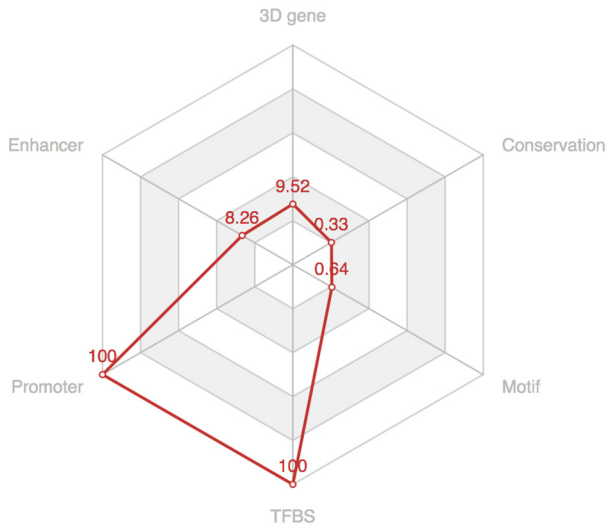


Figure 7. Radar chart of rs5792235. The six axes of the hexagon represent functionality levels (0-100) for enhancer status, promoter status, transcription factor binding site, motifs, evolutionary conservation and 3D interacting genes, as suggested by 3DSNP.

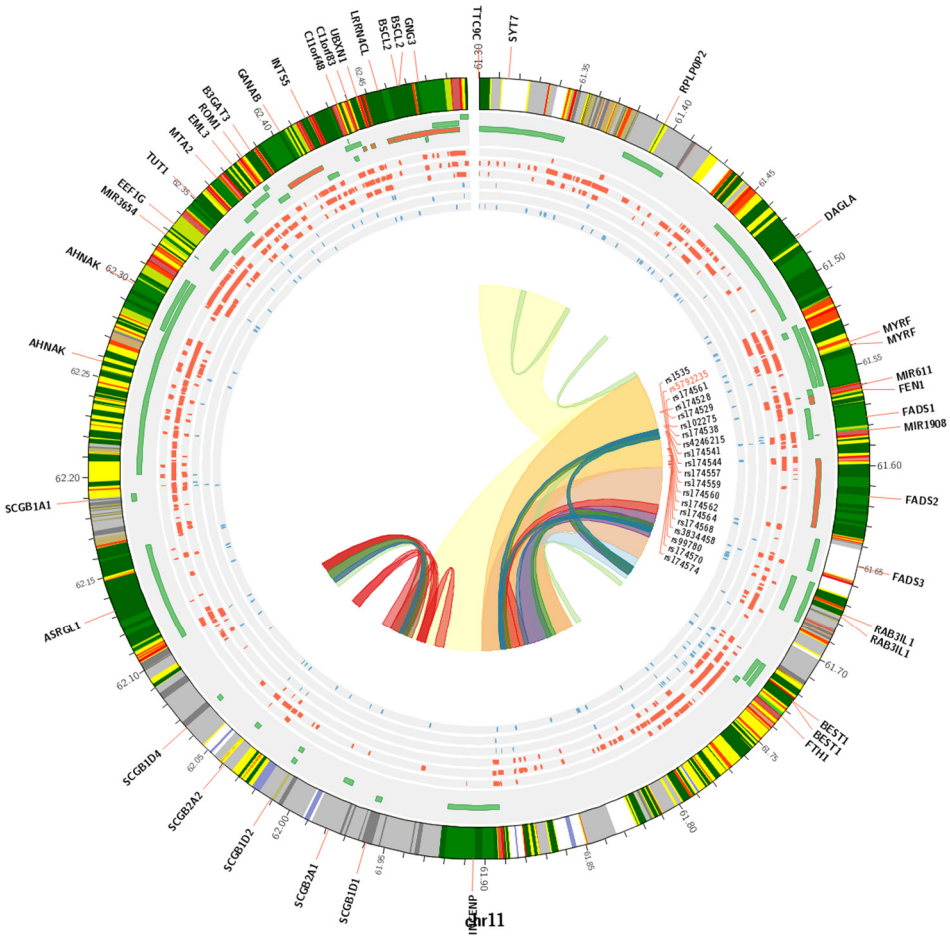


Figure 8. Circos plot based on rs5792235. The plot shows rs5792235 and its proxies (shown in black text around rs5792235). From outer to inner, the circles represent ChromHMM chromatin states, annotated genes (green), histone modification set (red), transcription factor set (blue), current variant (rs5792235) and associated variants, and 3D chromatin interactions, respectively. The three circles in the histone modification set are H3K4me1, H3K4me3, H3K27ac, and the three circles in the transcription factor set are CTCF, CEBPB and CEBPD. Color schemes for the ChromHMM chromatin states and the 3D interaction loops can be found at http://biotech.bmi.ac.cn/3dsnp/documentation/tutorials/ .

Paper III

EWAS are analogous to GWAS, in that they aim to screen many disease-related environmental factors in a hypothesis-free manner. EWAS have been used to

identify nutrients, pollutants, prescribed drugs, and lifestyle factors associated with diseases and quantitative traits.[198,199,201-203,244] However, almost all published EWAS has been undertaken in cross-sectional settings.

Table 2: Top variables associated with cardiometabolic traits

Phenotype	Exposures	Class	Value	DF	Effect	p-Value	R^2
BMI	Fitness	Physical activity	7	16815	-0.05	5.13E-210	0.083
	Training or exercising during the last 3 months	Physical activity	5	19264	-0.01	1.40E-44	0.061
	Processed meat	Food		16195	0.01	1.64E-35	0.058
HDL-C	Alcohol	Alcohol		2906	0.0198	1.40E-60	0.178
	Wine	Alcohol		2828	0.0093	5.32E-43	0.180
	Secoisolariciresinol	Nutrients		2967	0.0537	3.84E-34	0.168
LDL-C	Trans fat	Nutrients		2833	0.0726	6.84E-21	0.042
	Palmitic acid	Nutrients		2952	0.0596	1.57E-13	0.039
	Coffee	Beverage		2978	0.0075	5.28E-13	0.039
Triglyceride	Fitness	Physical activity	7	15363	-0.0915	7.06E-92	0.097
	Training or exercising during the last 3 months	Physical activity	5	16922	-0.0327	3.30E-47	0.073
	Smoking status	Tobacco use	5	17121	0.0275	1.31E-46	0.074
Total cholesterol	Palmitic acid	Nutrients		16548	0.0837	6.09E-71	0.081
	Formic acid	Nutrients		16428	0.0670	1.47E-64	0.081
	Carbohydrate	Nutrients		16624	-0.0866	1.04E-50	0.079
Fasting glucose	Smoking status	Tobacco use	5	19386	0.0106	5.43E-28	0.049
	Training or exercising during the last 3 months	Physical activity	5	19164	-0.0117	4.80E-21	0.050
	Trans fat	Nutrients		15847	0.0308	3.29E-20	0.036
2-h glucose	Smoking status	Tobacco use	5	18272	-0.0214	4.78E-40	0.062
	Coffee	Beverage		15713	-0.0080	3.13E-27	0.058
	Snuff status	Tobacco use	2	17158	-0.0125	4.04E-22	0.058
Systolic blood pressure	Education level	Social	4	19793	-0.0126	2.99E-39	0.135
	Fitness	Physical activity	7	16772	-0.0206	9.80E-25	0.128
	Smoking status	Tobacco use	5	19421	-0.0063	8.45E-13	0.128
Dystolic blood pressure	Fitness	Physical activity	7	16825	-0.0291	3.40E-32	0.129
	Education level	Social	4	19868	-0.0125	7.12E-27	0.129
	Training or exercising during the last 3 months	Physical activity	4	19273	-0.0085	1.03E-21	0,124

Value - category showing the strongest association in linear mixed model; DF – degree of freedom in the linear mixed model; Effect - effect size in the linear mixed model; p - P value in the linear mixed model; R^2 - marginal R^2 in the linear mixed model.

In **Paper III**, we integrated a linear mixed model-based analyses into EWAS to analyze the relationships of ~160 environmental exposures with cardiometabolic traits using repeated measures data from >18,000 participants from a sub-cohort of VHU. Many of the associations were statistically significant at an experiment-wise

level, some of which were significant for all the traits: 75 exposure variables for BMI, 51 exposure variables for HDL, 12 exposure variables for LDL, 60 exposure variables for triglycerides, 50 exposure variables for total cholesterol, 40 exposure variables for fasting glucose, 27 exposure variables for 2-*h* glucose, 27 exposure variables for SBP, and 33 exposure variables for DBP. Table 2 summarizes the top-three exposures associated with each cardiometabolic trait.

The modifiable exposure variables identified through these EWAS analyses may be informative for the design of health-promoting lifestyle interventions. Therefore, we sorted all experiment-wide significant exposure variables into 10 discrete categories (described in Methods). We ranked the index variables by the variance explained (R^2) for their respective outcome traits and summarized the top five variables within each category (the detailed results can be found in the manuscript). 'education level', 'smoking, 'amount of exercise in the last 3 months', 'aerobic fitness', 'self-reported health', 'heptadecanoic acid', and 'soda/soft drinks/juice' demonstrated strong associations with at least five out of eight cardiometabolic traits. However, the exposure variables that are generally on the list of public health recommendations such as consumption of 'animal protein', 'bananas', and 'fiber' and 'iron' were only associated with one of the traits, suggesting that modifying these specific exposures may have minimal impact on overall cardiometabolic health.

A noteworthy finding from the EWAS analysis is that the odd-chain fatty acid (OCS-FA) 'heptadecanoic acid' (C17:0) was associated with six out of eight cardiometabolic traits (Table 3). Another OCS-FA 'pentadecanoic acid' (C15:0) was associated with four traits, although explaining a much smaller degree of variance in the traits than did heptadecanoic acid (Table 4). These OCS-FAs are exogenous in origin, mostly deriving from consumption of dairy products, veal, beef, lamb, and mutton. Wolk *et al* evaluated the OCS-FA content in diet for 81 healthy women and found that C17:0 accounts for 10.61% of milk fat and 0.81% of ruminant fat; C15:0 accounts for 1.05% of milk fat and 0.43% of ruminant meat fat. The positive association between OCS-FAs and milk fat makes OCS-FA become biomarkers of dairy intake in epidemiological studies.[245] Khaw *et al* investigated the effect of plasma phospholipid fatty acids on the risk of CHD in ~25,000 individuals from the EPIC-Norfolk study and found OCS-FA was protective from CHD.[246] Another study conducted by the same group also found OCS-FA was inversely associated with the incidence of T2DM.[247] To our knowledge, this is the first time to show that heptadecanoic acid and pentadecanoic acid are associated with a range of intermediate cardiometabolic markers.

Table 3. The association between heptadecanoic acid and cardiometabolic trait

Phenotype	DF.LMM	Effect.LMM	p.LMM	R2.LMM
BMI	16475	-0.006	0.0003	0.056
HDL	2915	0.033	3.27E-08	0.038
LDL	2923	0.038	2.10E-14	0.165
Triglyceride	14484	0.003	0.5868	0.066
Total cholesterol	16386	0.048	1.99E-45	0.079
Fasting glucose	16336	-0.015	9.37E-10	0.043
2-h glucose	15449	-0.027	1.99E-11	0.055
Systolic blood pressure	16355	-0.009	1.23E-05	0.124
Diastolic blood pressure	16355	-0.009	1.23E-05	0.124

Table 4. The association between pentadecanoic acid and cardiometabolic trait

Phenotype	DF.LMM	Effect.LMM	p.LMM	R2.LMM
BMI	16475	-0.005	0.0063	0.056
HDL	2923	0.037	8.29E-13	0.165
LDL	2915	0.035	1.79E-08	0.039
Triglyceride	14484	0.004	0.3638	0.066
Total cholesterol	16386	0.049	1.13E-44	0.079
Fasting glucose	16336	-0.014	4.38E-08	0.043
2-h glucose	15449	-0.027	2.51E-10	0.055
Systolic blood pressure	16355	-0.009	1.02E-05	0.124
Diastolic blood pressure	2882	0.011	0.0364	0.082

DF.LMM, degree of freedome in the linear mixed model; Effect.LMM, effect size in linear mixed model; p.LMM, *p* value in linear mixed model; R^2, variance explained in linear mixed model. The models for BMI and blood pressure were adjusted for age, age2, sex, FFQ version (for dietary variables) and follow-up years. Models for lipids and glucose traits were additionally adjusted for fasting states.

In the longitudinal linear regression model EWAS analyses, we identified several environmental exposures associated with long-term change of cardiometabolic traits. Snuff (a type of oral tobacco popular in Sweden) use was associated with an increase on BMI. Long commute distance to work was negatively associated with triglyceride concentrations. 'Amount of exercise undertaken in the last 3 months', 'trans fat', 'formic acid', 'butter/margarine', 'education', and 'heptadecanoic acid' were statistically associated with long term change of total cholesterol. 'Having only coffee or tea for breakfast' was associated with low fasting glucose. 'Smoking' was associated with increased systolic blood pressure and diastolic blood pressure.

We believe the current analysis is the first of its kind to systematically screen environmental factors in a longitudinal setting. With the EWAS approach, we were able to utilize all available data collected in an epidemiological study to discover novel environmental exposures affecting disease risk. However, we have to keep in mind that the environmental exposures were assessed with self-administered

questionnaires, which are likely prone to bias. Thus, the key findings from these analyses should be validated using independent methods and materials. Moreover, the current analyses include both numeric and categorical variables, which complicates the interpretation of the data. Finally, many of the independent variables are strongly related with each other, and we have not determined which of these is most likely to be causal and which are merely correlates.

Summary and conclusion

In **Paper I**, we aimed to evaluate whether SSB intake modulate the genetic predisposition to obesity using a GRS based on the published BMI-associated SNPs. In **Paper II**, I took the challenge to screen the G-E interaction signals in the *FADS* genetic region comprised of hundreds of SNPs. In **Paper III**, we focused on the environmental risk factors of cardiometabolic traits, where we conducted EWAS analyses using a longitudinal dataset to identify environmental factors useful for designing intervention studies.

The major findings in these papers are:

In **Paper I**, SSB consumption have a more pronounced effect on obesity among individuals genetically predisposed to obesity, to whom a healthy lifestyle specially reduction of SSB consumption should be suggested. The interaction effects we observed in the Swedish population showed a similar magnitude to the previous analysis in the American population.

In **Paper II**, the *FADS* genetic variants demonstrated interaction signals with dietary PUFA intake on cardiometabolic traits. In the single genetic variant analysis, the strongest interactions were observed between rs174602 and n-3 PUFA intake on total cholesterol, and between rs174570 and n-6 PUFA intake on fasting glucose. With the gene- and haplotype-centric analyses, we narrowed down 3 haplotype blocks showing tentative interaction effects with n-3 PUFA on triglycerides. *In silico* function annotation for those candidate SNPs suggested that multiple functional variants might exist and we identified rs5792235 *FADS2* as a potential causal variant in the region.

In **Paper III**, a large number of environmental exposures showed environmental-wide associations with the studied cardiometabolic traits. By sorting the significant variables into discrete intervention categories and ranking them by r^2, education level, amount of exercise undertaken during the last 3 months, aerobic fitness, state of health during the last year, heptadecanoic acid, SSB, commuting to work, and wine consumption showed significant associations with ≥5 cardiometabolic traits. Heptadecanoic acid showed a significant association with cardiometabolic traits for the first time.

Future perspectives

In the past decade, much effort in the area of genetic epidemiology of cardiometabolic diseases has been invested in the identification of genetic variants associated with these diseases. GWAS studies, together with more recent exome- and whole-genome sequencing studies have successfully identified thousands of genetic loci associated with cardiometabolic diseases. However, these identified variants do not explain all of the predisposition to cardiometabolic diseases; much of the remainder is likely attributable to environmental exposures and their interactions with genetic variants, which are the major focus of this thesis.

The study of G-E interactions has undergone several phases, from testing *a priori* biologically-driven hypotheses to testing candidate loci derived from GWAS analyses. In order to maximize statistical power, most of these studies have focused on common variants. A number of sophisticated analytical methods testing the cumulative effects of the rare variants have been developed.[248] However, many of them have only been tested in simulation studies and need to be carefully validated in different populations. In **Paper II**, applied a region-based approach to screen G-E interaction signals in a genetic region with multiple rare variants. After detecting several putative interaction signals, we functionally annotated the top-ranked haplotypes using public available databases and identified the variants most likely to be functional. The validation of our findings in experimental studies would strengthen our conclusions, where the underlying mechanisms can be scrutinized in greater detail.

G-E interaction studies can also be improved by better exposure assessment. So far, a large number of epidemiological studies have primarily relied on self-report methods to quantify the environmental exposures (also a limitation of the current studies). However, the exposome is often highly complex and most existing studies do not account for this. The exposome concept has been gaining attention in the last decade and several EWAS have been conducted with the hope of discovering novel environmental risk factors associated with diseases. With the epidemiological data collected in VHU, we conducted EWAS analyses in **Paper III**. Environmental factors identified by EWAS provide good candidates for further G-E interaction analyses. A number of research groups have developed methods to use metabolites and gut microbial features to derive objective signatures of diet and other exposures.[249,250] Wearable devices offer a complimentary approach for the objective

assessment of environmental exposures such as sleep, stress and physical activity. Mobile phone apps have also been developed to record dietary intake and psychosocial interactions.

Acknowledgement

I praise my God for leading me, inspiring me, and shaping me in the four-year journey. My doctoral studies did not only equip me with the cutting-edge knowledge, but it also helped me understand more about myself and humanity. I learned what my interests are, what I am good at, and more importantly, I got to know what my weaknesses are, what are the areas I should aim to improve. I am lucky enough to have met and interacted with so many lovely people, who have taught me a great deal about both scientific research and life in general.

Paul W Franks, thank you for introducing me to the field of genetic epidemiology. I remember how shy, insecure, and naïve I was when I joined the GAME Unit. I really appreciate your generous support and encouragement during the past years. Your encyclopedic knowledge, great enthusiasm for research, and positive attitude really motivated me to work harder and to strive towards my goals. I am grateful for your patience and for sharing your personal and professional experiences to help me through difficult phases.

Azra Kurbasic, I love your spicy personality. I consider you both a mentor and a friend. Thank you for pushing me so hard to step out of my comfort zone and learn new things. Thank you for taking care of the young people in the group. I hope the best for your new career.

Alaitz Poveda, I admire your problem-solving ability. Your door is always open when I have practical questions related to my research. I enjoy the discussions with you about my projects and I know I can always learn something new from you. I am also thankful for your valuable comments on my thesis.

Frida Renström, thank you for teaching me all the basics when I joined the group. You always consider the feelings of other people. Your kindness and warm personality really helped me adapt to the daily life of the group. I wish I could spend more time working with you.

The past and present GAME unit members that I have the pleasure to work with: **Tibor V Varga**, your passion for science has been always inspiring, it is wonderful to work with you and being friends with you; **Robert Koivula**, special thanks to you for calling me from the UK just to give me advice on thesis writing; **Angela Estampador**, my four-year desk-mate, I am thankful for your understanding, encouragement and generous help in the past; **Pernilla Siming**, I have never heard

you say “No” when I asked for help, and your presence has made our life much easier; **Simon Timpka**, I am grateful to have such a colleague I can always trust and talk to; **Ashfaq Ali**, your continuous support and encouragement are appreciated more than you know**. Adem Yesuf Dawed**, **Adebayo Sanusi**, **Daniel Coral, Dmitry Shungin**, **Enass Elhamalawi**, **Giuseppe Nicola Giordano**, **Hugo Fitipaldi**, **Hugo Pomares-Millan**, **Jeremy Pomeroy**, **Jerneja Mislej**, **Johanna Ekström**, **Jonathan Dalla-Riva**, **Louise Bennet**, **Maria Keller**, **Mi Huang**, **Naeimeh Atabaki**, **Pascal Mutie**, **Sebastian Kalamajski**, **Shafqat Ahmad**, **Sonja Ruhnke**, **Stefan A Escher**, **Åsa Tornberg**, I have learnt so much from all of you. Your presence at the GAME Unit made my PhD journey so memorable!

My main collaborator in the EWAS project **Chirag Patel** at Harvard Medical School: thank you very much for the foundation work in the EWAS project, and for sharing your great knowledge with me.

I would like to thank the collaborators at Umeå University **Ingegerd Johansson** and **Göran Hallmans** for the tremendous support for my research.

I am also hugely appreciative to **Niels Grarup** and **Nils Wierup**. Thanks for being my half-time examiners and for providing constructive feedback about my research.

I thank all past and present member of the DPLU PhD student steering committee: **Anna Lessmark**, **Cajsa Davegårdh**, **Caroline Montén**, **Christina-Alexandra Schulz**, **Eliana Garcia Vaz**, **Jenifer Vallejo**, **Johanna Sernevi Säll**, **Linda Ahlkvist**, **Lovisa Heyman**, **Nina Geidenstam, Oktawia Nilsson-Wolanin**. I just want to say: you guys rock! It was a really enjoyable experience to work with you on organizing events for PhD students.

A big thanks to all the colleagues at **LUDC**, specially **Anna Wendt**, **George Hindy**, **Gull Rukh**, **Leif Groop**, **Lena Eliasson**, **Louise Brunkwall**, **Hindrik Mulder, Isabel Goncalves, Malin Fex**, **Maria Gomez**, **Maria Olofsson**, **Mattias Borell**, **Marju Orho-Melander**, **Ulrika-Blom Nilsson**, **Ulrika Ericson**.

Thanks to all my colleagues at Rigshospitalet where I spent a research visit during my Masters’ study, **Allan Vaag**, **Charllote Brøns**, **Martin Friedrichsen**, **Linn Gillberg**, **Line Hjort**, **Nina Schiøler Hansen**, **Louise Groth Grunnet**, **Jette Bork-Jensen**. It was a fantastic experience!

Last but not least, I owe my deepest gratitude towards my wife **Lin Luo**. Your unfailing love and support have always been my strength. I am thankful to my son **Samuel Lixing Chen** for bringing us so much happiness during my thesis writing. I thank my parents, in-laws, sister and other family members for their immense support. You are my walls of protection.

References

1. World Health Organization. WORLD HEALTH STATISTICS 2018 : monitoring health for the sdgs, sustainable development goals. [S.l.]: WORLD HEALTH ORGANIZATION; 2018.
2. Van Gaal LF, Mertens IL, De Block CE. Mechanisms linking obesity with cardiovascular disease. *Nature* 2006; **444**(7121): 875-80.
3. Defronzo RA, Ferrannini E. Insulin Resistance - a Multifaceted Syndrome Responsible for Niddm, Obesity, Hypertension, Dyslipidemia, and Atherosclerotic Cardiovascular-Disease. *Diabetes Care* 1991; **14**(3): 173-94.
4. Coutinho M, Gerstein HC, Wang Y, Yusuf S. The relationship between glucose and incident cardiovascular events. A metaregression analysis of published data from 20 studies of 95,783 individuals followed for 12.4 years. *Diabetes Care* 1999; **22**(2): 233-40.
5. Sowers JR, Epstein M, Frohlich ED. Diabetes, hypertension, and cardiovascular disease - An update. *Hypertension* 2001; **37**(4): 1053-9.
6. Howard BV, Robbins DC, Sievers ML, et al. LDL cholesterol as a strong predictor of coronary heart disease in diabetic individuals with insulin resistance and low LDL - The Strong Heart Study. *Arterioscl Throm Vas* 2000; **20**(3): 830-5.
7. Austin MA, Hokanson JE, Edwards KL. Hypertriglyceridemia as a cardiovascular risk factor. *Am J Cardiol* 1998; **81**(4a): 7b-12b.
8. Barrettconnor E, Khaw K. Family History of Heart-Attack as an Independent Predictor of Death Due to Cardiovascular-Disease. *Circulation* 1984; **69**(6): 1065-9.
9. Poulsen P, Kyvik KO, Vaag A, Beck-Nielsen H. Heritability of Type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance - a population-based twin study. *Diabetologia* 1999; **42**(2): 139-45.
10. Eckel RH, Jakicic JM, Ard JD, et al. 2013 AHA/ACC Guideline on Lifestyle Management to Reduce Cardiovascular Risk A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation* 2014; **129**(25): S76-S99.
11. Thomas D. Gene--environment-wide association studies: emerging approaches. *Nat Rev Genet* 2010; **11**(4): 259-72.
12. Lehrman MA, Schneider WJ, Sudhof TC, Brown MS, Goldstein JL, Russell DW. Mutation in Ldl Receptor - Alu-Alu Recombination Deletes Exons Encoding Transmembrane and Cytoplasmic Domains. *Science* 1985; **227**(4683): 140-6.

13. Yamagata K, Furuta H, Oda N, et al. Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young (MODY1). *Nature* 1996; **384**(6608): 458-60.
14. Yamagata K, Oda N, Kaisaki PJ, et al. Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset diabetes of the young (MODY3). *Nature* 1996; **384**(6608): 455-8.
15. Atanasovska B, Kumar V, Fu JY, Wijmenga C, Hofker MH. GWAS as a Driver of Gene Discovery in Cardiometabolic Diseases. *Trends Endocrin Met* 2015; **26**(12): 722-32.
16. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature* 2009; **461**(7265): 747-53.
17. Malik VS, Popkin BM, Bray GA, Despres JP, Willett WC, Hu FB. Sugar-sweetened beverages and risk of metabolic syndrome and type 2 diabetes: a meta-analysis. *Diabetes Care* 2010; **33**(11): 2477-83.
18. Qi Q, Chu AY, Kang JH, et al. Sugar-sweetened beverages and genetic risk of obesity. *N Engl J Med* 2012; **367**(15): 1387-96.
19. Abete I, Romaguera D, Vieira AR, Lopez de Munain A, Norat T. Association between total, processed, red and white meat consumption and all-cause, CVD and IHD mortality: a meta-analysis of cohort studies. *Br J Nutr* 2014; **112**(5): 762-75.
20. Larsson SC, Orsini N. Red meat and processed meat consumption and all-cause mortality: a meta-analysis. *Am J Epidemiol* 2014; **179**(3): 282-9.
21. Ambrose JA, Barua RS. The pathophysiology of cigarette smoking and cardiovascular disease: an update. *J Am Coll Cardiol* 2004; **43**(10): 1731-7.
22. Wilmot EG, Edwardson CL, Achana FA, et al. Sedentary time in adults and the association with diabetes, cardiovascular disease and death: systematic review and meta-analysis. *Diabetologia* 2012; **55**(11): 2895-905.
23. World Health Organization. Global report on diabetes. World Health Organization; 2016.
24. International Diabetes Federation. IDF diabetes atlas. Brussels: International Diabetes Federation; 2017.
25. Ahlqvist E, Storm P, Karajamaki A, et al. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endo* 2018; **6**(5): 361-9.
26. Bentzon JF, Otsuka F, Virmani R, Falk E. Mechanisms of plaque formation and rupture. *Circ Res* 2014; **114**(12): 1852-66.
27. Mullarkey CJ, Edelstein D, Brownlee M. Free-Radical Generation by Early Glycation Products - a Mechanism for Accelerated Atherogenesis in Diabetes. *Biochem Bioph Res Co* 1990; **173**(3): 932-9.
28. Qin XB, Goldfine A, Krumrei N, et al. Glycation inactivation of the complement regulatory protein CD59 - A possible role in the pathogenesis of the vascular complications of human diabetes. *Diabetes* 2004; **53**(10): 2653-61.

29. Baumgartnerparzer S, Wagner L, Pettermann M, Gessl A, Grillari J, Waldhausl W. High Glucose Triggered Apoptosis in Cultured Endothelial-Cells. *Diabetes* 1995; **44**: A221-A.
30. International Diabetes Federation. Diabetes and cardiovascular disease. Brussels: International Diabetes Federation; **2016**.
31. Institute of Medicine, Committee on Preventing the Global Epidemic of Cardiovascular Disease: Meeting the Challenges in Developing Countries, Fuster V, Kelly BB. Promoting cardiovascular health in the developing world a critical challenge to achieve global health. Washington: National Academies Press; 2010.
32. Levine SA. Angina pectoris - Some clinical considerations. *J Amer Med Assoc* 1922; **79**: 928-33.
33. Reaven GM. Banting lecture 1988. Role of insulin resistance in human disease. *Diabetes* 1988; **37**(12): 1595-607.
34. Muniyappa R, Lee S, Chen H, Quon MJ. Current approaches for assessing insulin sensitivity and resistance in vivo: advantages, limitations, and appropriate usage. *Am J Physiol Endocrinol Metab* 2008; **294**(1): E15-26.
35. World Health Organization, Department of Noncommunicable Disease Surveillance. Definition, diagnosis, and classification of diabetes mellitus and its complications : report of a WHO consultation. Geneva: World Health Organization, Dept. of Noncommunicable Disease Surveillance; 1999.
36. Balkau B, Charles MA. Comment on the provisional report from the WHO consultation. European Group for the Study of Insulin Resistance (EGIR). *Diabet Med* 1999; **16**(5): 442-3.
37. Expert Panel on Detection E, Treatment of High Blood Cholesterol in A. Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III). *JAMA* 2001; **285**(19): 2486-97.
38. Alberti KG, Zimmet P, Shaw J. Metabolic syndrome--a new world-wide definition. A Consensus Statement from the International Diabetes Federation. *Diabet Med* 2006; **23**(5): 469-80.
39. Wilson PWF, D'Agostino RB, Parise H, Sullivan L, Meigs JB. Metabolic syndrome as a precursor of cardiovascular disease and type 2 diabetes mellitus. *Circulation* 2005; **112**(20): 3066-72.
40. McNeill AM, Schmidt MI, Rosamond WD, et al. The metabolic syndrome and 11-year risk of incident cardiovascular disease in the atherosclerosis risk in communities study. *Diabetes Care* 2005; **28**(2): 385-90.
41. Dekker JM, Girman C, Rhodes T, et al. Metabolic syndrome and 10-year cardiovascular disease risk in the hoorn study. *Circulation* 2005; **112**(5): 666-73.
42. Gami AS, Witt BJ, Howard DE, et al. Metabolic syndrome and risk of incident cardiovascular events and death - A systematic review and meta-analysis of longitudinal studies. *Journal of the American College of Cardiology* 2007; **49**(4): 403-14.
43. Galassi A, Reynolds K, He J. Metabolic syndrome and risk of cardiovascular disease: A meta-analysis. *Am J Med* 2006; **119**(10): 812-9.

44. Kahn R, Buse J, Ferrannini E, Stern M. The metabolic syndrome: time for a critical appraisal. *Diabetologia* 2005; **48**(9): 1684-99.
45. Kahn R. Metabolic syndrome is it a syndrome? Does it matter? *Circulation* 2007; **115**(13): 1806-10.
46. Cannon CP. Cardiovascular disease and modifiable cardiometabolic risk factors. *Clin Cornerstone* 2007; **8**(3): 11-28.
47. World Health Organization. 10 facts on obesity. 2017. http://www.who.int/features/factfiles/obesity/en/ (accessed July 18th 2018).
48. Kopelman PG. Obesity as a medical problem. *Nature* 2000; **404**(6778): 635-43.
49. World Health Organization. Obesity : Preventing and managing the global epidemic. Geneva: World Health Organization; 2000.
50. Karelis AD. Metabolically healthy but obese individuals. *Lancet* 2008; **372**(9646): 1281-3.
51. Ibrahim MM. Subcutaneous and visceral adipose tissue: structural and functional differences. *Obes Rev* 2010; **11**(1): 11-8.
52. Despres JP. Health consequences of visceral obesity. *Ann Med* 2001; **33**(8): 534-41.
53. Fox CS, Massaro JM, Hoffmann U, et al. Abdominal visceral and subcutaneous adipose tissue compartments - Association with metabolic risk factors in the Framingham Heart Study. *Circulation* 2007; **116**(1): 39-48.
54. Willett WC, Dietz WH, Colditz GA. Guidelines for healthy weight. *N Engl J Med* 1999; **341**(6): 427-34.
55. Mokdad AH, Ford ES, Bowman BA, et al. Prevalence of obesity, diabetes, and obesity-related health risk factors, 2001. *Jama-J Am Med Assoc* 2003; **289**(1): 76-9.
56. Gallagher D, Heymsfield SB, Heo M, Jebb SA, Murgatroyd PR, Sakamoto Y. Healthy percentage body fat ranges: an approach for developing guidelines based on body mass index. *Am J Clin Nutr* 2000; **72**(3): 694-701.
57. Who. Appropriate body-mass index in Asian populations and its implications for policy and intervention strategies. (vol 363, pg 157, 2004). *Lancet* 2004; **363**(9412): 902-.
58. Welborn TA, Dhaliwal SS. Preferred clinical measures of central obesity for predicting mortality. *Eur J Clin Nutr* 2007; **61**(12): 1373-9.
59. Janssen I, Katzmarzyk PT, Ross R. Waist circumference and not body mass index explains obesity-related health risk. *Am J Clin Nutr* 2004; **79**(3): 379-84.
60. Bigaard J, Frederiksen K, Tjonneland A, et al. Waist circumference and body composition in relation to all-cause mortality in middle-aged men and women. *Int J Obesity* 2005; **29**(7): 778-84.
61. Cornier MA, Despres JP, Davis N, et al. Assessing adiposity: a scientific statement from the American Heart Association. *Circulation* 2011; **124**(18): 1996-2019.
62. Mazess RB, Barden HS, Bisek JP, Hanson J. Dual-energy x-ray absorptiometry for total-body and regional bone-mineral and soft-tissue composition. *Am J Clin Nutr* 1990; **51**(6): 1106-12.
63. Crook MA. Clinical Biochemistry and Metabolic Medicine. 2013.

64. Kingsbury KJ, Bondy G. Understanding the essentials of blood lipid metabolism. *Prog Cardiovasc Nurs* 2003; **18**(1): 13-8.
65. Ference BA, Ginsberg HN, Graham I, et al. 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. *Eur Heart J* 2017; **38**(32): 2459-72.
66. Bonetti PO, Lerman LO, Lerman A. Endothelial dysfunction - A marker of atherosclerotic risk. *Arterioscl Throm Vas* 2003; **23**(2): 168-75.
67. Greaves DR, Gordon S. The macrophage scavenger receptor at 30 years of age: current knowledge and future challenges. *J Lipid Res* 2009; **50**: S282-S6.
68. Goldstein JL, Brown MS. A Century of Cholesterol and Coronaries: From Plaques to Genes to Statins. *Cell* 2015; **161**(1): 161-72.
69. Navab M, Reddy ST, Van Lenten BJ, Fogelman AM. HDL and cardiovascular disease: atherogenic and atheroprotective mechanisms. *Nature Reviews Cardiology* 2011; **8**(4): 222-32.
70. Miller M, Stone NJ, Ballantyne C, et al. Triglycerides and Cardiovascular Disease A Scientific Statement From the American Heart Association. *Circulation* 2011; **123**(20): 2292-333.
71. Ahren B, Pacini G. Islet adaptation to insulin resistance: mechanisms and implications for intervention. *Diabetes Obes Metab* 2005; **7**(1): 2-8.
72. Lillioja S, Mott DM, Spraul M, et al. Insulin-Resistance and Insulin Secretory Dysfunction as Precursors of Non-Insulin-Dependent Diabetes-Mellitus - Prospective Studies of Pima-Indians. *New Engl J Med* 1993; **329**(27): 1988-92.
73. World Health Organization, International Diabetes Federation. Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia : report of a WHO/IDF consultation. Geneva, Switzerland: World Health Organization; 2006.
74. Cryer PE. Glucose homeostasis and hypoglycemia. Williams textbook of endocrinology. Philadelphia: Saunders/Elsevier; 1992: 1223-53.
75. Dimitriadis G, Mitrou P, Lambadiari V, Maratou E, Raptis SA. Insulin effects in muscle and adipose tissue. *Diabetes Res Clin Pr* 2011; **93**: S52-S9.
76. Butler PC, Chou J, Carter WB, et al. Effects of Meal Ingestion on Plasma Amylin Concentration in Niddm and Nondiabetic Humans. *Diabetes* 1990; **39**(6): 752-6.
77. Villena JA, Roy S, Sarkadi-Nagy E, Kim KH, Sul HS. Desnutrin, an adipocyte gene encoding a novel patatin domain-containing protein, is induced by fasting and glucocorticoids. *J Biol Chem* 2004; **279**(45): 47066-75.
78. Ginsberg HN. Review: Efficacy and mechanisms of action of statins in the treatment of diabetic dyslipidemia. *J Clin Endocr Metab* 2006; **91**(2): 383-92.
79. Schneider JG, Ren J, Feng C, et al. ATM haploinsufficiency produces vascular insulin resistance, impaired glucose tolerance, hypertension, and accelerated atherosclerosis in apoE null mice. *Arterioscl Throm Vas* 2004; **24**(5): E5-E.
80. Arkan MC, Hevener AL, Greten FR, et al. IKK-beta links inflammation to obesity-induced insulin resistance. *Nature Medicine* 2005; **11**(2): 191-8.

81. Zhou MS, Schulman IH, Raij L. Role of angiotensin II and oxidative stress in vascular insulin resistance linked to hypertension. *Am J Physiol-Heart C* 2009; **296**(3): H833-H9.
82. Zhou MS, Schulman IH, Raij L. Vascular inflammation, insulin resistance, and endothelial dysfunction in salt-sensitive hypertension: role of nuclear factor kappa B activation. *Journal of Hypertension* 2010; **28**(3): 527-35.
83. Kaptoge S, Di Angelantonio E, Pennells L, et al. C-Reactive Protein, Fibrinogen, and Cardiovascular Disease Prediction. *New Engl J Med* 2012; **367**(14): 1310-20.
84. Karatela RA, Sainani GS. Interrelationships of factor VII activity and plasma leptin with insulin resistance in coronary heart disease. *Atherosclerosis* 2010; **209**(1): 235-40.
85. Tofler GH, Massarob J, O'Donnell CJ, et al. Plasminogen activator inhibitor and the risk of cardiovascular disease: The Framingham Heart Study. *Thromb Res* 2016; **140**: 30-5.
86. Ridker PM, Cushman M, Stampfer MJ, Tracy RP, Hennekens CH. Inflammation, aspirin, and the risk of cardiovascular disease in apparently healthy men. *New Engl J Med* 1997; **336**(14): 973-9.
87. Wang CC, Gurevich I, Draznin B. Insulin affects vascular smooth muscle cell phenotype and migration via distinct signalling pathways. *Diabetes* 2003; **52**(10): 2562-9.
88. World Health Organization. A global brief on hypertension : silent killer, global public health crisis: World Health Day 2013. Geneva, Switzerland: World Health Organization; 2013.
89. de Ferranti SD, de Boer IH, Fonseca V, et al. Type 1 Diabetes Mellitus and Cardiovascular Disease A Scientific Statement From the American Heart Association and American Diabetes Association. *Circulation* 2014; **130**(13): 1110-U156.
90. Arauz-Pacheco C, Parrott MA, Raskin P. The treatment of hypertension in adult patients with diabetes. *Diabetes Care* 2002; **25**(1): 134-47.
91. de Boer IH, Bangalore S, Benetos A, et al. Diabetes and Hypertension: A Position Statement by the American Diabetes Association. *Diabetes Care* 2017; **40**(9): 1273-84.
92. Tuttle KR, Bakris GL, Bilous RW, et al. Diabetic Kidney Disease: A Report From an ADA Consensus Conference. *Diabetes Care* 2014; **37**(10): 2864-83.
93. Chobanian AV, Bakris GL, Black HR, et al. Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *Hypertension* 2003; **42**(6): 1206-52.
94. Eryd SA, Gudbjornsdottir S, Manhem K, et al. Blood pressure and complications in individuals with type 2 diabetes and no previous cardiovascular disease: national population based cohort study. *Bmj-Brit Med J* 2016; **354**.
95. Forouzanfar MH, Liu P, Roth GA, et al. Global Burden of Hypertension and Systolic Blood Pressure of at Least 110 to 115 mm Hg, 1990-2015. *Jama-J Am Med Assoc* 2017; **317**(2): 165-82.

96. Rapsomaniki E, Timmis A, George J, et al. Blood pressure and incidence of twelve cardiovascular diseases: lifetime risks, healthy life-years lost, and age-specific associations in 1.25 million people. *Lancet* 2014; **383**(9932): 1899-911.
97. World Health O. Global health risks mortality and burden of disease attributable to selected major risks. 2009.
98. Lind L, Berne C, Lithell H. Prevalence of insulin resistance in essential hypertension. *Journal of Hypertension* 1995; **13**(12): 1457-62.
99. Pollare T, Lithell H, Selinus I, Berne C. Application of Prazosin Is Associated with an Increase of Insulin Sensitivity in Obese Patients with Hypertension. *Diabetologia* 1988; **31**(7): 415-20.
100. Snorgaard O, Kober L, Carlsen J. The effect of metformin on blood pressure and metabolism in nondiabetic hypertensive patients. *Journal of Internal Medicine* 1997; **242**(5): 407-12.
101. Sowers JR. Insulin resistance and hypertension. *Am J Physiol-Heart C* 2004; **286**(5): H1597-H602.
102. DeFronzo RA, Cooke CR, Andres R, Faloona GR, Davis PJ. The effect of insulin on renal handling of sodium, potassium, calcium, and phosphate in man. *J Clin Invest* 1975; **55**(4): 845-55.
103. Vassy JL, Shrader P, Jonsson A, et al. Association between parental history of diabetes and type 2 diabetes genetic risk scores in the PPP-Botnia and Framingham Offspring Studies. *Diabetes Res Clin Pr* 2011; **93**(2): E76-E9.
104. Boomsma D, Busjahn A, Peltonen L. Classical twin studies and beyond. *Nature Reviews Genetics* 2002; **3**(11): 872-82.
105. Kyvik KO, Green A, Becknielsen H. Concordance Rates of Insulin-Dependent Diabetes-Mellitus - a Population-Based Study of Young Danish Twins. *Brit Med J* 1995; **311**(7010): 913-7.
106. Maes HHM, Neale MC, Eaves LJ. Genetic and environmental factors in relative body weight and human adiposity. *Behavior Genetics* 1997; **27**(6): 600-.
107. Stunkard AJ, Foch TT, Hrubec Z. A Twin Study of Human Obesity. *Jama-J Am Med Assoc* 1986; **256**(1): 51-4.
108. Stunkard AJ, Harris JR, Pedersen NL, Mcclearn GE. The Body-Mass Index of Twins Who Have Been Reared Apart. *New Engl J Med* 1990; **322**(21): 1483-7.
109. Kupper N, Ge DL, Treiber FA, Snieder H. Emergence of novel genetic effects on blood pressure and hemodynamics in adolescence - The Georgia cardiovascular twin study. *Hypertension* 2006; **47**(5): 948-54.
110. Pietilainen KH, Soderlund S, Rissanen A, et al. HDL Subspecies in Young Adult Twins: Heritability and Impact of Overweight. *Obesity* 2009; **17**(6): 1208-14.
111. Elder SJ, Lichtenstein AH, Pittas AG, et al. Genetic and environmental influences on factors associated with cardiovascular disease and the metabolic syndrome. *J Lipid Res* 2009; **50**(9): 1917-26.
112. Pericak-Vance MA. Analysis of genetic linkage data for Mendelian traits. *Curr Protoc Hum Genet* 2001; **Chapter 1**: Unit 1 4.

113. Duggirala R, Blangero J, Almasy L, et al. Linkage of type 2 diabetes mellitus and of age at onset to a genetic location on chromosome 10q in Mexican Americans. *American Journal of Human Genetics* 1999; **64**(4): 1127-40.
114. Reynisdottir I, Thorleifsson G, Benediktsson R, et al. Localization of a susceptibility gene for type 2 diabetes to chromosome 5q34-q35.2. *American Journal of Human Genetics* 2003; **73**(2): 323-35.
115. Grant SFA, Thorleifsson G, Reynisdottir I, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nature Genetics* 2006; **38**(3): 320-3.
116. Damcott CM, Pollin TI, Reinhart LJ, et al. Polymorphisms in the transcription factor 7-like 2 (TCF7L2) gene are associated with type 2 diabetes in the Amish - Replication and evidence for a role in both insulin secretion and insulin resistance. *Diabetes* 2006; **55**(9): 2654-9.
117. Cauchi S, Meyre D, Dina C, et al. Transcription factor TCF7L2 genetic study in the French population - Expression in human beta-cells and adipose tissue and strong association with type 2 diabetes. *Diabetes* 2006; **55**(10): 2903-8.
118. Zhang CL, Qi L, Hunter DJ, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene and the risk of type 2 diabetes in large cohorts of US women and men. *Diabetes* 2006; **55**(9): 2645-8.
119. Altshuler D, Hirschhorn JN, Klannemark M, et al. The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet* 2000; **26**(1): 76-80.
120. Gloyn AL, Weedon MN, Owen KR, et al. Large-scale association studies of variants in genes encoding the pancreatic beta-cell K-ATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. *Diabetes* 2003; **52**(2): 568-72.
121. Florez JC, Burtt N, de Bakker PIW, et al. Haplotype structure and genotype-phenotype correlations of the sulfonylurea receptor and the islet ATP-sensitive potassium channel gene region. *Diabetes* 2004; **53**(5): 1360-8.
122. Clausen JO, Hansen T, Bjorbaek C, et al. Insulin-Resistance - Interactions between Obesity and a Common Variant of Insulin-Receptor Substrate-1. *Lancet* 1995; **346**(8972): 397-402.
123. Collins FS, Lander ES, Rogers J, Waterston RH, Conso IHGS. Finishing the euchromatic sequence of the human genome. *Nature* 2004; **431**(7011): 931-45.
124. International HapMap C. A haplotype map of the human genome. *Nature* 2005; **437**(7063): 1299-320.
125. Bush WS, Moore JH. Chapter 11: Genome-Wide Association Studies. *Plos Computational Biology* 2012; **8**(12).
126. GWAS Catalog. 10th Anniversary of the NHGRI-EBI GWAS Catalog. 2018. https://www.ebi.ac.uk/gwas/anniversary (accessed 10th August 2018).
127. Voight BF, Kang HM, Ding J, et al. The Metabochip, a Custom Genotyping Array for Genetic Studies of Metabolic, Cardiovascular, and Anthropometric Traits. *Plos Genetics* 2012; **8**(8).

128. Huyghe JR, Jackson AU, Fogarty MP, et al. Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. *Nature Genetics* 2013; **45**(2): 197-201.
129. Fuchsberger C, Flannick J, Teslovich TM, et al. The genetic architecture of type 2 diabetes. *Nature* 2016; **536**(7614): 41-7.
130. Loos RJF. The genetics of adiposity. *Current Opinion in Genetics & Development* 2018; **50**: 86-95.
131. Frayling TM, Timpson NJ, Weedon MN, 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**(5826): 889-94.
132. Lu YC, Loos RJF. Obesity genomics: assessing the transferability of susceptibility loci across diverse populations. *Genome Med* 2013; **5**.
133. Gerken T, Girard CA, Tung YCL, et al. The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science* 2007; **318**(5855): 1469-72.
134. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology* 2003; **32**(1): 1-22.
135. Fall T, Hagg S, Magi R, et al. The Role of Adiposity in Cardiometabolic Traits: A Mendelian Randomization Analysis. *Plos Med* 2013; **10**(6).
136. Thorleifsson G, Walters GB, Gudbjartsson DF, et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nature Genetics* 2009; **41**(1): 18-24.
137. Willer CJ, Speliotes EK, Loos RJF, et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nature Genetics* 2009; **41**(1): 25-34.
138. Speliotes EK, Willer CJ, Berndt SI, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genetics* 2010; **42**(11): 937-U53.
139. Yang J, Loos RJF, Powell JE, et al. FTO genotype is associated with phenotypic variability of body mass index. *Nature* 2012; **490**(7419): 267-+.
140. Lindgren CM, Heid IM, Randall JC, et al. Genome-Wide Association Scan Meta-Analysis Identifies Three Loci Influencing Adiposity and Fat Distribution. *Plos Genetics* 2009; **5**(6).
141. Allen HL, Estrada K, Lettre G, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 2010; **467**(7317): 832-8.
142. Heid IM, Jackson AU, Randall JC, et al. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nature Genetics* 2011; **43**(11): 1164-.
143. Randall JC, Winkler TW, Kutalik Z, et al. Sex-stratified Genome-wide Association Studies Including 270,000 Individuals Show Sexual Dimorphism in Genetic Loci for Anthropometric Traits. *Plos Genetics* 2013; **9**(6).

144. Berndt SI, Gustafsson S, Magi R, et al. Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nature Genetics* 2013; **45**(5): 501-U69.

145. Wood AR, Esko T, Yang J, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics* 2014; **46**(11): 1173-86.

146. Shungin D, Winkler TW, Croteau-Chonka DC, et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* 2015; **518**(7538): 187-U378.

147. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* 2015; **518**(7538): 197-U401.

148. Aulchenko YS, Ripatti S, Lindqvist I, et al. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nature Genetics* 2009; **41**(1): 47-55.

149. Kathiresan S, Willer CJ, Peloso GM, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nature Genetics* 2009; **41**(1): 56-65.

150. Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010; **466**(7307): 707-13.

151. Willer CJ, Schmidt EM, Sengupta S, et al. Discovery and refinement of loci associated with lipid levels. *Nature Genetics* 2013; **45**(11): 1274-+.

152. Hoffmann TJ, Theusch E, Haldar T, et al. A large electronic-health-record-based genome-wide study of serum lipids. *Nature Genetics* 2018; **50**(3): 401-+.

153. Sladek R, Rocheleau G, Rung J, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 2007; **445**(7130): 881-5.

154. Saxena R, Voight BF, Lyssenko V, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007; **316**(5829): 1331-6.

155. Scott LJ, Mohlke KL, Bonnycastle LL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 2007; **316**(5829): 1341-5.

156. Burton PR, Clayton DG, Cardon LR, et al. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007; **447**(7145): 661-78.

157. Zeggini E, Scott LJ, Saxena R, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nature Genetics* 2008; **40**(5): 638-45.

158. Prokopenko I, Langenberg C, Florez JC, et al. Variants in MTNR1B influence fasting glucose levels. *Nature Genetics* 2009; **41**(1): 77-81.

159. Dupuis J, Langenberg C, Prokopenko I, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nature Genetics* 2010; **42**(2): 105-U32.

160. Saxena R, Hivert MF, Langenberg C, et al. Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nature Genetics* 2010; **42**(2): 142-U75.

161. Soranzo N, Sanna S, Wheeler E, et al. Common Variants at 10 Genomic Loci Influence Hemoglobin A(1C) Levels via Glycemic and Nonglycemic Pathways. *Diabetes* 2010; **59**(12): 3229-39.
162. Voight BF, Scott LJ, Steinthorsdottir V, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nature Genetics* 2010; **42**(7): 579-U155.
163. Strawbridge RJ, Dupuis J, Prokopenko I, et al. Genome-Wide Association Identifies Nine Common Variants Associated With Fasting Proinsulin Levels and Provides New Insights Into the Pathophysiology of Type 2 Diabetes. *Diabetes* 2011; **60**(10): 2624-34.
164. Manning AK, Hivert MF, Scott RA, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nature Genetics* 2012; **44**(6): 659-U81.
165. Scott RA, Lagou V, Welch RP, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nature Genetics* 2012; **44**(9): 991-+.
166. Morris AP, Voight BF, Teslovich TM, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature Genetics* 2012; **44**(9): 981-+.
167. Xue A, Wu Y, Zhu ZH, et al. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat Commun* 2018; **9**.
168. Newton-Cheh C, Johnson T, Gateva V, et al. Genome-wide association study identifies eight loci associated with blood pressure. *Nature Genetics* 2009; **41**(6): 666-76.
169. Levy D, Ehret GB, Rice K, et al. Genome-wide association study of blood pressure and hypertension. *Nature Genetics* 2009; **41**(6): 677-87.
170. Wain LV, Verwoert GC, O'Reilly PF, et al. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nature Genetics* 2011; **43**(10): 1005-U122.
171. Ehret GB, Munroe PB, Rice KM, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* 2011; **478**(7367): 103-9.
172. Ehret GB, Ferreira T, Chasman DI, et al. The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. *Nature Genetics* 2016; **48**(10): 1171-84.
173. Liu CY, Kraja AT, Smith JA, et al. Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. *Nature Genetics* 2016; **48**(10): 1162-70.
174. Surendran P, Drenos F, Young R, et al. Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nature Genetics* 2016; **48**(10): 1151-61.
175. Warren HR, Evangelou E, Cabrera CP, et al. Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk (vol 49, pg 403, 2017). *Nature Genetics* 2017; **49**(10): 1558-.

176. Hoffmann TJ, Ehret GB, Nandakumar P, et al. Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nature Genetics* 2017; **49**(1): 54-64.

177. Schulz LO, Bennett PH, Ravussin E, et al. Effects of traditional and western environments on prevalence of type 2 diabetes in Pima Indians in Mexico and the U.S. *Diabetes Care* 2006; **29**(8): 1866-71.

178. Robertson TL, Kato H, Gordon T, et al. Epidemiologic studies of coronary heart disease and stroke in Japanese men living in Japan, Hawaii and California. Coronary heart disease risk factors in Japan and Hawaii. *Am J Cardiol* 1977; **39**(2): 244-9.

179. Fujimoto WY, Boyko EJ, Hayashi T, et al. Risk Factors for Type 2 Diabetes: Lessons Learned from Japanese Americans in Seattle. *J Diabetes Investig* 2012; **3**(3): 212-24.

180. Worth RM, Kato H, Rhoads GG, Kagan K, Syme SL. Epidemiologic studies of coronary heart disease and stroke in Japanese men living in Japan, Hawaii and California: mortality. *Am J Epidemiol* 1975; **102**(6): 481-90.

181. Bhatnagar D, Anand IS, Durrington PN, et al. Coronary Risk-Factors in People from the Indian Subcontinent Living in West London and Their Siblings in India. *Lancet* 1995; **345**(8947): 405-9.

182. Yusuf S, Reddy S, Ounpuu S, Anand S. Global burden of cardiovascular diseases - Part II: Variations in cardiovascular disease by specific ethnic groups and geographic regions and prevention strategies. *Circulation* 2001; **104**(23): 2855-64.

183. Popkin BM, Adair LS, Ng SW. Global nutrition transition and the pandemic of obesity in developing countries. *Nutr Rev* 2012; **70**(1): 3-21.

184. Hu FB. Resolved: there is sufficient scientific evidence that decreasing sugar-sweetened beverage consumption will reduce the prevalence of obesity and obesity-related diseases. *Obes Rev* 2013; **14**(8): 606-19.

185. Barquera S, Hernandez-Barrera L, Tolentino ML, et al. Energy Intake from Beverages Is Increasing among Mexican Adolescents and Adults. *J Nutr* 2008; **138**(12): 2454-61.

186. Decastro JM. The Effects of the Spontaneous Ingestion of Particular Foods or Beverages on the Meal Pattern and Overall Nutrient Intake of Humans. *Physiol Behav* 1993; **53**(6): 1133-44.

187. Malik VS, Schulze MB, Hu FB. Intake of sugar-sweetened beverages and weight gain: a systematic review. *American Journal of Clinical Nutrition* 2006; **84**(2): 274-88.

188. Malik VS, Popkin BM, Bray GA, Despres JP, Hu FB. Sugar-sweetened beverages, obesity, type 2 diabetes mellitus, and cardiovascular disease risk. *Circulation* 2010; **121**(11): 1356-64.

189. Schulze MB, Manson JE, Ludwig DS, et al. Sugar-sweetened beverages, weight gain, and incidence of type 2 diabetes in young and middle-aged women. *Jama-J Am Med Assoc* 2004; **292**(8): 927-34.

190. Micha R, Wallace SK, Mozaffarian D. Red and Processed Meat Consumption and Risk of Incident Coronary Heart Disease, Stroke, and Diabetes Mellitus A Systematic Review and Meta-Analysis. *Circulation* 2010; **121**(21): 2271-U52.

191. Food and Agriculture Organization of the United Nations. Globalization of food systems in developing countries : impact on food security and nutrition. Rome: Food and Agriculture Organization of the United Nations; 2004.
192. Lee IM, Shiroma EJ, Lobelo F, et al. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *Lancet* 2012; **380**(9838): 219-29.
193. Rappaport SM, Smith MT. Epidemiology. Environment and disease risks. *Science* 2010; **330**(6003): 460-1.
194. Patel CJ, Ioannidis JP. Studying the elusive environment in large scale. *JAMA* 2014; **311**(21): 2173-4.
195. Wild CP. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev* 2005; **14**(8): 1847-50.
196. Shearrer GE, O'Reilly GA, Belcher BR, et al. The impact of sugar sweetened beverage intake on hunger and satiety in minority adolescents. *Appetite* 2016; **97**: 43-8.
197. Wild CP. The exposome: from concept to utility. *Int J Epidemiol* 2012; **41**(1): 24-32.
198. Patel CJ, Bhattacharya J, Butte AJ. An Environment-Wide Association Study (EWAS) on type 2 diabetes mellitus. *PLoS One* 2010; **5**(5): e10746.
199. Tzoulaki I, Patel CJ, Okamura T, et al. A nutrient-wide association study on blood pressure. *Circulation* 2012; **126**(21): 2456-64.
200. Lind PM, Riserus U, Salihovic S, Bavel B, Lind L. An environmental wide association study (EWAS) approach to the metabolic syndrome. *Environ Int* 2013; **55**: 1-8.
201. Patel CJ, Ji J, Sundquist J, Ioannidis JP, Sundquist K. Systematic assessment of pharmaceutical prescriptions in association with cancer risk: a method to conduct a population-wide medication-wide longitudinal study. *Sci Rep* 2016; **6**: 31308.
202. McGinnis DP, Brownstein JS, Patel CJ. Environment-Wide Association Study of Blood Pressure in the National Health and Nutrition Examination Survey (1999-2012). *Sci Rep* 2016; **6**: 30373.
203. Patel CJ, Bhattacharya J, Ioannidis JPA, Bendavid E. Systematic identification of correlates of HIV infection: an X-wide association study. *Aids* 2018; **32**(7): 933-43.
204. Ottman R. Gene-environment interaction: Definitions and study designs. *Preventive Medicine* 1996; **25**(6): 764-70.
205. Franks PW. Gene x Environment Interactions in Type 2 Diabetes. *Curr Diabetes Rep* 2011; **11**(6): 552-61.
206. Qi QB, Chu AY, Kang JH, et al. Sugar-Sweetened Beverages and Genetic Risk of Obesity. *New Engl J Med* 2012; **367**(15): 1387-96.
207. Li SX, Zhao JH, Luan JA, et al. Physical Activity Attenuates the Genetic Predisposition to Obesity in 20,000 Men and Women from EPIC-Norfolk Prospective Population Study. *Plos Med* 2010; **7**(8).

208. Ahmad S, Rukh G, Varga TV, et al. Gene x Physical Activity Interactions in Obesity: Combined Analysis of 111,421 Individuals of European Ancestry. *Plos Genetics* 2013; **9**(7).
209. Norberg M, Wall S, Boman K, Weinehall L. The Vasterbotten Intervention Programme: background, design and implications. *Global Health Action* 2010; **3**.
210. Voight BF, Kang HM, Ding J, et al. The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet* 2012; **8**(8): e1002793.
211. Berglund G, Elmstahl S, Janzon L, Larsson SA. The Malmo Diet and Cancer Study. Design and feasibility. *J Intern Med* 1993; **233**(1): 45-51.
212. Manjer J, Carlsson S, Elmstahl S, et al. The Malmo Diet and Cancer Study: representativity, cancer incidence and mortality in participants and non-participants. *Eur J Cancer Prev* 2001; **10**(6): 489-99.
213. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem* 1972; **18**(6): 499-502.
214. Hallal PC, Victoria CG. Reliability and validity of the International Physical Activity Questionnaire (IPAQ). *Med Sci Sport Exer* 2004; **36**(3): 556-.
215. Johansson I, Hallmans G, Wikman A, Biessy C, Riboli E, Kaaks R. Validation and calibration of food-frequency questionnaire measurements in the Northern Sweden Health and Disease cohort. *Public Health Nutr* 2002; **5**(3): 487-96.
216. Livsmedelsverket. The food database. http://www.livsmedelsverket.se/en/food-and-content/naringsamnen/livsmedelsdatabasen/ (accessed October 12th 2016).
217. Eidstedt M. Food consumption and nutritive values, data up to 2016. In: Jordbruksverk S, editor.: Helen Stoye; 2017.
218. Renstrom F, Shungin D, Johansson I, et al. Genetic Predisposition to Long-Term Nondiabetic Deteriorations in Glucose Homeostasis Ten-Year Follow-Up of the GLACIER Study. *Diabetes* 2011; **60**(1): 345-54.
219. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PIW. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 2008; **24**(24): 2938-9.
220. Fontaine-Bisson B, Renstrom F, Rolandsson O, et al. Evaluating the discriminative power of multi-trait genetic risk scores for type 2 diabetes in a northern Swedish population. *Diabetologia* 2010; **53**(10): 2155-62.
221. Renstrom F, Payne F, Nordstrom A, et al. Replication and extension of genome-wide association study results for obesity in 4923 adults from northern Sweden. *Hum Mol Genet* 2009; **18**(8): 1489-96.
222. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. *Science* 2002; **296**(5576): 2225-9.
223. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**(2): 263-5.
224. R Development Core Team. R: A language and environment for statistic computing. Vienna, Austria: R Foundation for Statistic Computing; 2015.

225. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**(3): 559-75.

226. Fumagalli M, Moltke I, Grarup N, et al. Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science* 2015; **349**(6254): 1343-7.

227. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* 2011; **89**(1): 82-93.

228. Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet* 2004; **74**(4): 765-9.

229. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* 2012; **9**(3): 215-6.

230. Lu Y, Quan C, Chen H, Bo X, Zhang C. 3DSNP: a database for linking human noncoding SNPs to their three-dimensional interacting genes. *Nucleic Acids Res* 2017; **45**(D1): D643-D9.

231. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 2012; **40**(Database issue): D930-4.

232. Gene Ontology C. Gene Ontology Consortium: going forward. *Nucleic Acids Res* 2015; **43**(Database issue): D1049-56.

233. Fagherazzi G, Vilier A, Sartorelli DS, Lajous M, Balkau B, Clavel-Chapelon F. Consumption of artificially and sugar-sweetened beverages and incident type 2 diabetes in the Etude Epidemiologique aupres des femmes de la Mutuelle Generale de l'Education Nationale-European Prospective Investigation into Cancer and Nutrition cohort. *Am J Clin Nutr* 2013; **97**(3): 517-23.

234. Hu FB, Malik VS. Sugar-sweetened beverages and risk of obesity and type 2 diabetes: Epidemiologic evidence. *Physiol Behav* 2010; **100**(1): 47-54.

235. de Ruyter JC, Katan MB, Kuijper LDJ, Liem DG, Olthof MR. The Effect of Sugar-Free Versus Sugar-Sweetened Beverages on Satiety, Liking and Wanting: An 18 Month Randomized Double-Blind Trial in Children. *Plos One* 2013; **8**(10).

236. Trumbo PR, Rivers CR. Systematic review of the evidence for an association between sugar-sweetened beverage consumption and risk of obesity. *Nutr Rev* 2014; **72**(9): 566-74.

237. Mackenzie T, Brooks B, O'Connor G. Beverage intake, diabetes, and glucose control of adults in America. *Ann Epidemiol* 2006; **16**(9): 688-91.

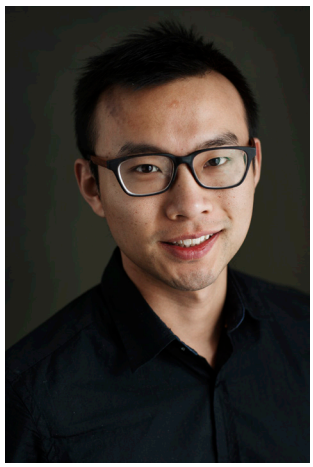
238. Yang Q. Gain weight by "going diet?" Artificial sweeteners and the neurobiology of sugar cravings: Neuroscience 2010. *Yale J Biol Med* 2010; **83**(2): 101-8.

239. Suez J, Korem T, Zeevi D, et al. Artificial sweeteners induce glucose intolerance by altering the gut microbiota. *Nature* 2014; **514**(7521): 181-6.

240. Murff HJ, Edwards TL. Endogenous Production of Long-Chain Polyunsaturated Fatty Acids and Metabolic Disease Risk. *Curr Cardiovasc Risk Rep* 2014; **8**(12).

241. Manning AK, Hivert MF, Scott RA, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet* 2012; **44**(6): 659-69.
242. Cormier H, Rudkowska I, Thifault E, Lemieux S, Couture P, Vohl MC. Polymorphisms in Fatty Acid Desaturase (FADS) Gene Cluster: Effects on Glycemic Controls Following an Omega-3 Polyunsaturated Fatty Acids (PUFA) Supplementation. *Genes (Basel)* 2013; **4**(3): 485-98.
243. Ye K, Gao F, Wang D, Bar-Yosef O, Keinan A. Dietary adaptation of FADS genes in Europe varied across time and geography. *Nat Ecol Evol* 2017; **1**: 167.
244. Hall MA, Dudek SM, Goodloe R, et al. Environment-wide association study (EWAS) for type 2 diabetes in the Marshfield Personalized Medicine Research Project Biobank. *Pac Symp Biocomput* 2014: 200-11.
245. Wolk A, Vessby B, Ljung H, Barrefors P. Evaluation of a biological marker of dairy fat intake. *Am J Clin Nutr* 1998; **68**(2): 291-5.
246. Khaw KT, Friesen MD, Riboli E, Luben R, Wareham N. Plasma Phospholipid Fatty Acid Concentration and Incident Coronary Heart Disease in Men and Women: The EPIC-Norfolk Prospective Study. *Plos Med* 2012; **9**(7).
247. Forouhi NG, Koulman A, Sharp SJ, et al. Differences in the prospective association between individual plasma phospholipid saturated fatty acids and incident type 2 diabetes: the EPIC-InterAct case-cohort study. *Lancet Diabetes Endo* 2014; **2**(10): 810-8.
248. Bomba L, Walter K, Soranzo N. The impact of rare and low-frequency genetic variants in common disease. *Genome Biol* 2017; **18**.
249. Garcia-Perez I, Posma JM, Gibson R, et al. Objective assessment of dietary patterns by use of metabolic phenotyping: a randomised, controlled, crossover trial. *Lancet Diabetes Endo* 2017; **5**(3): 184-95.
250. David LA, Maurice CF, Carmody RN, et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 2014; **505**(7484): 559-+.

Printed by Media-Tryck, Lund 2018 NORDIC SWAN ECOLABEL 3041 0903



Yan Chen is a graduate in nutrition & food science from China. He obtained his Master degree in Food Technology and Nutrition from Lund University. The focus of his doctoral thesis was to investigated the role of genetic and environmental factors in a range of cardiometabolic traits (e.g. BMI, blood lipids, and blood pressure), using three complementary study designs.

LUND UNIVERSITY

FACULTY OF MEDICINE

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2018:155
ISBN 978-91-7619-724-0
ISSN 1652-8220