Genetic approaches to identifying causal pathways to cardiometabolic diseases

Laura Brigitte Leen Wittemans

MRC-Epidemiology Unit

Clare Hall, University of Cambridge

October 2018

This dissertation is submitted for the degree of

Doctor of Philosophy

Acknowledgements

I want to start this thesis by thanking all the people who have supported me in several ways throughout my PhD journey. First of all, I am very grateful to Dr Claudia Langenberg, my PhD supervisor, for leading me through the PhD with lots of expertise and positive energy, for teaching me how to run a (scientific) marathon instead of a sprint, for pushing me to keep focusing on the bigger picture behind the day-to-day research work, and for giving me the confidence and freedom to become the researcher I want to be.

Secondly, I am very thankful to Dr Felix Day, who has been my mentor throughout my PhD. He has taught me more than I could ever list here, has given me helpful and pragmatic advice on countless occasions during my PhD and has proof-read large parts of this thesis.

I also thank all the members of the aetiology group, for sharing their knowledge and expertise with me and, even more importantly, for being such nice colleagues: Chen, Clara, Ellie, Isobel, Jian'an, Jing Hua, Lina, Luca, Nick B., Nick W., Nicola, Robert, Sara and Vicky. It has been a true pleasure working with you all.

I also thank the PhD students whom I have shared the PhD experience with at the unit, and I particularly thank Sherly and Eirini for their friendship and support.

I thank the entire staff of the MRC-Epidemiology Unit for enabling me to conduct my PhD research, in various direct and indirect ways. I especially thank Meriel Smith, for her hard work to make the PhD training for me and all students in the department as smooth as possible.

Finally, I am very grateful to Elias, for supporting me and taking care of me all throughout my PhD journey. I am very thankful to my parents, for their relentless support and belief in me, and to Caroline, Steven, Opa, Shohre, Fariborz, Mona, Cele, Carla, Livia and Maud for their support, friendship and kindness.

Declarations

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text

It does not exceed the word limit of 60,000 words prescribed by Degree Committee of the Faculties of Clinical Medicine and Veterinary Medicine.

Summary of the thesis

Background: In the era of large-scale biobanks and detailed multi-omic and clinical phenotyping, it is now possible to simultaneously measure several thousands of biological traits and test their associations with health outcomes. The unprecedented scale of the available epidemiological data has led to a surge in the number of proposed risk factors for cardiometabolic and other diseases identified through observational research. However, prominent failures of randomised controlled trials to replicate observational findings highlight that new approaches are urgently required to prioritise risk factors for interventional studies based on their likelihood of causality.

Aim: Genetic evidence provides the opportunity to obtain unconfounded associations in an observational setting and can help to efficiently identify causal, aetiological pathways to cardiometabolic diseases. This work aims to identify and prioritise causal pathways to cardiometabolic diseases by integrating genetic data with detailed metabolic phenotypes, including blood metabolites and objective measures of body size and composition.

Methods: The first chapter summarises and critically evaluates existing genetic strategies for assessing causality in an observational setting and reviews the literature on reported associations between blood metabolites and incident type 2 diabetes (T2D). As a proof-of-concept, I then adopt a genetic approach to a) generate genetic predictors and b) assess evidence of causality for T2D and coronary heart disease (CHD) risk for a selected metabolite (glycine), for which consistent observational evidence suggests protective associations with T2D and CHD development. The approach is then extended to move beyond studying simple indices of obesity and enable causal assessment of refined anthropometric traits. For this, I have led collaborative efforts to develop genetic instruments for overall and regional fat and lean mass.

Findings: Genetic approaches to studying causality in an observational setting rely on important assumptions that are sensitive to violations, particularly in the context of highly correlated 'omic' measures. Existing methods generally consider overall risk factor "levels", rather than assessing the causality of the distinct mechanisms contributing to levels. I provide genetic evidence for a causal cardio-protective effect of the glycine pathway in men and women, with blood pressure as a potential mediating factor. In contrast, no strong evidence for a causal link between glycine and T2D was found, with evidence suggesting that the inverse glycine-to-T2D association is the consequence of a glycine-lowering effect of insulin resistance. Despite total body fat percentage

(BF%) and fat-free mass index (FFMI) being strongly observationally and genetically correlated with BMI, I identify 16 loci associated with higher BF% and lower FFMI but not BMI. Based on observational and genetic studies of regional fat and lean mass, I identify patterns of fat distribution which may not be captured by traditional anthropometric phenotypes, such as fat mass in the arms and in the subcutaneous android region. Fat mass in the arms and subcutaneous android region were observationally only weakly correlated with BMI, WHR and other fat compartments, and genetic loci specific to arm and subcutaneous android fat mass were identified. Conversely, no evidence was found that genetic loci associated with lean mass have heterogeneous effects on lean mass in different regions of the body.

Conclusion: This thesis demonstrates how the integration of large-scale genetic, metabolomic and clinical data can not only prioritise novel aetiological pathways to cardiometabolic conditions, but also formulate hypotheses regarding the underlying physiological mechanisms. The genetic factors identified for refined anthropometric traits, such as a high relative body fat mass in the absence of overweight, allow for a causal assessment of novel and specific body size and composition traits. In summary, this thesis demonstrates and utilises the opportunities that arise from integrating genetic data with refined phenotypes at scale to identify novel targets for the treatment and prevention of cardiometabolic diseases.

Publications

Published or submitted publications related to the thesis:

L. B. L Wittemans, L. A. Lotta & C. Langenberg. **Prioritising Risk Factors for Type 2 Diabetes: Causal Inference through Genetic Approaches**. *Current Diabetes Reports* 18:40 (2018).

L. B. L. Wittemans, L. A. Lotta, C. Olliver-Williams, [19 authors], C. Langenberg. **Assessing the causal association of glycine with risk of coronary heart disease.** *Nature Communications* 10:1060 (2019). doi: 10.1038/s41467-019-08936-1.

M. J. Neville, L. B. L. Wittemans (co-first author), K. E. Pinnick, M. Todorčević, R. Kaksonen, K. H. Pietiläinen, J. Luan, R. A. Scott, N. J. Wareham, C. Langenberg, F. Karpe**. Regional fat depot masses are influenced by protein-coding gene variants.** *Under review at PLoS ONE.*

L. A. Lotta, L. B. L. Wittemans, [15 authors], C. Langenberg. **Specific genetic determinants of gluteofemoral and abdominal fat distribution and risk of type 2 diabetes and cardiovascular disease.** *JAMA* 320(24), 2553-2563 (2018).

C. Hilton, M. J. Neville, L. B. L. Wittemans, [12 authors], F. Karpe. **MicroRNA-196a links human body fat distribution to adipose tissue extracellular matrix composition.** *Under review at EBioMedicine.*

Papers related to the thesis in preparation for submission:

Genome and exome-wide studies of total body fat mass and lean mass index.

Ongoing consortium project

Other publications:

L. A. Lotta, [7 authors], L. B. L. Wittemans, [9 authors]. **Variation in the *LPL* gene, low density lipoprotein-cholesterol lowering alleles and risk of coronary disease and type 2 diabetes.** *JAMA Cardiology* 3(10):957–966 (2018).

Presentations

Oral presentations

- **EPIC metabolomics meeting** – London (September 2017)
- "Integration of metabolomics and genetics to identify metabolic pathways as causal risk factors for complex diseases"
- **Leena Peltonen School of Human Genomics** – Hinxton (August 2017)
- "The metabolome and type 2 diabetes: A systematic literature review and Mendelian randomisation analyses"
- **Conference of the European Diabetes Epidemiology Group** – Dublin, Ireland (April 2016)
- "A genetic approach to characterising the role of the serum metabolome in type 2 diabetes"
- **Postgraduate Conference of the Nutrition Society** – Cambridge (September 2015)
- "Serum metabolomic patterns associated with future type 2 diabetes"

Poster presentations

- **Conference of the American Society of Human Genetics** – San Diego, US (October 2018)
- "Assessing and characterising the causal association of glycine with risk of coronary heart disease"
- **European Association for the Study of Diabetes – Study Group on Genetics of Diabetes (EASD-SGGD)** – Leiden, The Netherlands (May 2017)
- "The metabolome and cardiometabolic disease: a systematic literature review and Mendelian randomisation experiments"
- **The Danish Diabetes Academy Winterschool** – Málaga, Spain (March 2017)
- "The metabolome and type 2 diabetes: a systematic literature review and Mendelian randomisation experiments"
- **ICAN Conference Series on Omics** – Paris, France (December 2015)
- "Assessing and characterising the causal association of glycine with risk of coronary heart disease"

Commonly used abbreviations

BF%: total body fat percentage

BIA: bio-impedance analysis

Bp: base-pair position

BMI: body mass index

CHD: coronary heart disease

Chr: chomosome

CI: confidence interval

DEXA: dual-energy X-ray absorptiometry

EA: effect allele

EAF: effect allele frequency

FFM: fat-free mass

FFMI: fat-free mass index

GWAS: genome-wide association study

LD: linkage desiquilibrium

MAF: minor allele frequency

MR: Mendelian randomisation

OA: other allele

OR: odds ratio

PC: principal component

QC: quality control

SD: standard deviation

SE: standard error

SNP: single nucleotide polymorphism

T2D: type 2 diabetes

WHR: waist-to-hip ratio

WHRadjBMI: waist-to-hip ratio adjusted for BMI

List of Figures

List of Tables

Table of Contents

Introduction

Aims

The central aim of this PhD thesis is to identify novel aetiological pathways to cardiometabolic diseases based on the integration of large-scale genome-wide data with detailed metabolic phenotypes obtained with -omics methods and refined body composition analysis techniques. The primary focus is on the aetiology of type 2 diabetes (T2D), but this focus is extended to coronary heart disease (CHD) in some sections of the thesis.

I will address the following sub-aims and objectives:

- **Sub-aim 1:** To review the current state of genetic and omics-based research on the aetiology of T2D.
  - Objective 1.A: To review and critically evaluate the literature on genetic approaches to identify aetiological pathways to T2D
  - Objective 1.B: To conduct a systematic literature review of longitudinal epidemiological studies on metabolomics and incident T2D.

- **Sub-aim 2:** To develop and implement a genetics-based approach to investigate metabolic pathways as aetiological factors for cardiometabolic conditions, using the glycine pathway as an example.
  - Objective 2.A: To construct and validate genetic instruments for the glycine pathway
  - Objective 2.B: To assess the observational and genetic, unconfounded associations of glycine levels with T2D and CHD
  - Objective 2.C: To identify the biological mechanisms driving the associations of the glycine pathway with T2D and CHD.

- **Sub-aim 3:** To construct genetic instruments associated with total body composition phenotypes.
  - Objective 3.A: To conduct exome chip meta-analyses for total body fat percentage (BF%) and fat-free mass index (FFMI)
  - Objective 3.B: To conduct genome-wide association studies (GWAS) for BF% and FFMI in a large British cohort
  - Objective 3.C: To identify loci for BF%, FFMI and BMI which disproportionately affect fat or lean mass.

- **Sub-aim 4:** To build genetic instruments for fat and lean mass compartments and distribution patterns.
    - Objective 4.A: To conduct meta-analyses of GWAS for regional fat and lean mass and fat and lean mass distribution
    - Objective 4.B: To characterise the fat and lean mass distribution patterns associated with the identified loci.

Outline of the thesis

Chapter 1 gives an overview of the relevant literature. I commence with a brief outline of general concepts of T2D, including its definition, pathophysiology and global impact. Next, I provide an overview and critical evaluation of recent genetic and -omics approaches to identifying causal risk factors for T2D, including a systematic literature review of longitudinal epidemiological studies of metabolomics and T2D.

Chapter 2 describes an approach to assessing the causality of metabolic pathways as aetiological risk factors for cardiometabolic illnesses, based on large-scale genetic, metabolomic and clinical data, using the glycine pathway as an example. An in-depth study of the glycine pathway as a potential aetiological factor for T2D and CHD is presented. I construct genetic instruments for glycine based on a meta-analysis of GWAS for glycine in up to 80,000 participants, conduct Mendelian randomisation (MR) analyses to assess the causality of low glycine levels for T2D and CHD in both sexes combined and by sex, and use the MR framework to generate hypotheses about the biological factors driving associations of the glycine pathway with cardiometabolic diseases.

Chapter 3 focuses on the identification of genetic instruments associated with refined phenotypes of overall body composition, i.e., BF% and FFMI. I describe a consortium project, in which I took a leading role, which aims to identify the genetic determinants of BF% and FFMI through exome chip meta-analyses including up to 112,443 participants and GWAS on UK Biobank participants of European ancestry. I conduct quantitative assessments of the proportionality of the effect sizes of genetic loci for BF%, FFMI and BMI on fat and lean mass in order to identify loci which disproportionately affect fat or lean mass.

Chapter 4 describes genetic analyses on fat and lean mass compartments and distribution, based on dual-energy X-ray absorptiometry (DEXA) imaging. I perform meta-analyses of GWAS for 9 fat compartments and 7 lean mass compartments, including up to 25,452 participants from 5 cohorts. I characterise the fat and lean mass distribution patterns associated with each of the identified loci, and investigate the overall with loci known for standard anthropometric traits.

The concluding discussion summarises the empirical findings, outlines the general limitations of the adopted research approach and discusses the implications and potential avenues for future research.

Rationale

Technological innovations and reduced costs of high-throughput and refined phenotyping methods have led to the availability of phenotypic and clinical data of unprecedented granularity and scale in epidemiological cohorts and biobanks. "Omics" technologies, such as metabolomics and proteomics, have enabled the hypothesis-free measurement of thousands to tens of thousands of molecular markers simultaneously, while imaging techniques such as DEXA imaging have made it possible to obtain detailed body composition measurements at a relatively low cost and high speed.

The surge in the scale and depth of available phenotypic and clinical data in epidemiological cohorts has created the opportunity to conduct large-scale investigations, in which up to several thousands of potential risk factors are tested simultaneously for multiple outcomes in a hypothesis-free manner. This emerging new approach to risk factor discovery is rapidly gaining territory in epidemiological research, in which for a long time the predominant strategy has been to conduct studies of one or a limited set of pre-selected risk factors for one particular outcome and in isolation of other risk factors.

As the hypothesis-free strategy to risk factor discovery has become increasingly popular, this has led to an exponential increase in the number of proposed risk factors for disease outcomes. However, this surge in proposed risk factors has thus far not been met by a similar increase in the identification of confirmed *causal* risk factors, for two main reasons. First, the increasing coverage of the phenome comes with more pronounced correlation patterns between the proposed risk factors, which makes it challenging to distinguish causal factors from the many phenotypes with which they are correlated. Secondly, several high-profile failures of intervention studies set up to validate the causal role of risk factors consistently reported by epidemiological research have highlighted that causality cannot be reliably assessed in observational settings. New strategies are therefore needed to prioritise the proposed risk factors that are likely to be causal.

Genetics can be used to prioritise proposed risk factors that are likely to be causal for disease risk. Genetics-based causality assessment is being facilitated by the recent, drastic increase in the availability of genome-wide data obtained through array-based genotyping or whole-genome or exome sequencing in large population-based cohorts and disease-specific consortia. This facilitates causal inference in observational settings using genetic approaches such as Mendelian randomisation, which has gained tremendous popularity in assessing the causality of traditional risk factors but has not yet been widely adopted for more refined phenotypes, such as omics-based molecular traits.

This availability of phenotypic and genome-wide data at an unprecedented scale offer a timely opportunity to investigate metabolic pathways as aetiological pathways to cardiometabolic diseases and to deepen our understanding of the relationship between body composition and cardiometabolic health. Although metabolic disturbances are key processes in the pathophysiology of T2D and CHD, most research on metabolic risk factors has focused on a handful of markers, including fasting glucose and blood lipid fractions. The emergence of metabolomics in epidemiological studies on cardiometabolic diseases has created the opportunity to study virtually every metabolic pathway in relation to cardiometabolic disease risk. While overweight and obesity, based on simple, crude measures such as body mass index (BMI) and waist-to-hip ratio (WHR), are widely accepted as major culprits for the secular increase in incidence of cardiometabolic diseases, the aetiological relevance of overall and regional body composition is far less understood. The adoption of more advanced anthropometric assessment techniques such as bio-impedance analysis and imaging-based methods in the baseline health assessment protocol of biobanks and epidemiological cohorts enable research on specific body composition phenotypes, such as overall and regional fat and lean mass, as aetiological factors for cardiometabolic disease risk.

Chapter 1: Overview of the literature

Publications related to this chapter

Parts of this literature overview has been published as a review article:

L. B. L. Wittemans, L. A. Lotta & C. Langenberg. **Prioritising Risk Factors for Type 2 Diabetes: Causal Inference through Genetic Approaches.** *Current Diabetes Reports* (2018) 18: 40.

Abstract

Purpose of the literature review: Causality has been demonstrated for few of the many putative risk factors for type 2 diabetes (T2D) emerging from observational epidemiological research. Genetic approaches are increasingly being used to infer causality. In this literature review, I discuss how genetic discoveries have shaped our understanding of the causal role of factors associated with T2D.

Outline and recent findings: I start with a brief introduction which outlines the definitions, global impact, risk factors and pathophysiology of diabetes. Next, I highlight the need to prioritise proposed risk factors for T2D and discuss genetic approaches that allow for causal assessment in an observational setting.

I then use three illustrative examples to highlight how genetics can be used to prioritise novel aetiological factors for T2D and refine our understanding of established risk factors:

- I outline how observational studies using a multitude of genetic approaches, have revealed a complex role of lipid metabolism in the aetiology of T2D.
- I describe how the integration of genetic and refined phenotypic data has led to a more nuanced understanding of the effect of overall adiposity on T2D risk and has proposed peripheral fat storage capacity as a protective factor for T2D.
- Finally, I highlight how "-omics" methods have led to a surge in proposed risk factors for T2D by conducting a systematic literature review of metabolomics studies on T2D, and, using examples from the literature, I describe how genetics can help to identify and prioritise causal pathways among the multitude of established and emerging biomarkers.

Summary: Genome-wide association studies of T2D and metabolic traits coupled with high-throughput molecular phenotyping and in-depth characterisation and follow-up of individual loci have provided better understanding of aetiological factors contributing to T2D.

Introduction

Definition of diabetes mellitus

Diabetes Mellitus is a chronic condition characterised by elevated blood glucose levels caused by a lack of insulin production or the inability of the body to adequately respond to insulin[1,2]. According to the World Health Organisation's (WHO) guidelines, diabetes can be diagnosed if at least one of the 3 criteria pointing to hyperglycaemia are met (Table 1.1)[3].

Table 1.1: The World Health Organisation's (WHO) diagnostic criteria for diabetes. Only one of the 3 criteria needs to be met for a diabetes diagnosis. HbA1c: haemoglobin A1c. Adapted from the Global Report on Diabetes (2017) published by the WHO[3].

Marker	Threshold
Fasting plasma glucose	≥7.0 mmol/L (126 mg/dL)
Two-hour plasma glucose	≥ 11.1 mmol/L (200 mg/dL)
HbA1C	≥ 6.5%

Type 2 diabetes (T2D) is the most prevalent of three common types of diabetes, with 90% of all cases being type 2 diabetic[1]. In T2D patients, hyperglycaemia arises from a relative insufficiency in insulin secretion in the context of varying degrees of insulin resistance. It is often related to overweight as a consequence of a long-term positive energy balance and is most commonly diagnosed in middle-aged individuals. In contrast, type 1 diabetes is predominantly diagnosed in children and adolescents and results from the autoimmune-mediated destruction of the insulin-producing β cells of the pancreas, leading to an incapability of the body to produce insulin[4]. Gestational diabetes is the third common type of diabetes and is characterised by hyperglycaemia during pregnancy, possibly caused by the interference of hormones secreted by the placenta with insulin action[5].

The global impact of type 2 diabetes

Estimates by the International Diabetes Federation show that in 2017, 8.8% of the population, or 425 million people worldwide, were affected by diabetes, and prevalence is expected to increase to 9.9% by 2045[1]. Nearly 11% of the global all-cause mortality among people aged 20-79 years could be attributed to diabetes in 2017, and 46% of these deaths occurred before the age of 60[1]. With 80% of the total burden of diabetes lying in low to middle-income countries[1], diabetes can no longer be considered a disease of affluence and poses a serious threat to public health worldwide.

Most disability and mortality associated with T2D is due to complications caused by hyperglycaemia-induced macro- or microvascular damage, of which cardiovascular, renal and eye conditions are the most common chronic complications. Patients with T2D are at 2 to 3 times higher risk of developing a cardiovascular disease, including peripheral artery disease, coronary heart disease, cardiac failure and ischemic stroke[6]. Risk of a cardiovascular disease among T2D patients is closely correlated with glucose control[7]. T2D is also the most common cause of chronic kidney disease worldwide and, together with hypertension, accounts for 80% of all cases. Diabetic retinopathies caused by damage to the microvasculature of the retina due to high glucose are the leading cause of reduced vision among the adult population of working age and affect around 1 in 3 diabetic patients[1,8]. Diabetes has also been shown to have large economic impacts at the level of the individual patient and their family[9], as well as on health care systems and national economies[10,11]. Altogether, this emphasises the need for a better understanding of T2D, and the development of better prevention and treatment strategies worldwide, and the importance of strategies targeted to low and middle-income settings.

Glucose homeostasis, risk factors for type 2 diabetes and pathophysiology

The peptide hormone insulin, which is synthesised by and secreted from the islet β cells of the pancreas, is the master regulator of glucose homeostasis[12]. In healthy individuals who are in a fasted state, insulin secretion is low, and glucose is produced endogenously, primarily through glycogenolysis and gluconeogenesis in the liver, to maintain blood glucose levels within the normal range and provide glucose as a source of energy for the brain. Glucose uptake by the muscles and adipose tissue is low during fasting to preserve glucose for the brain. After carbohydrate ingestion and uptake through the gut, blood glucose levels will rise and signal the islet β cells to start secreting insulin. Insulin travels through the blood stream to insulin-sensitive tissues, where it will stimulate the uptake of glucose by the muscles and adipose tissue, inhibit glucose and lipid secretion by the liver and downregulate lipolysis in adipose tissue. This integrated insulin-induced response ensures that glucose levels return to its fasted concentration within two hours of a meal[12].

Insulin resistance is a pathological state in which the insulin-sensitive tissues become less responsive to normal insulin levels[13]. The development of insulin resistance often co-occurs with, or is preceded by a long-term positive energy balance, due to the overconsumption of energy-dense foods, lack of physical activity, or both. In the fed state, insulin resistance is characterised by a reduced uptake of glucose in the muscles and adipose tissue for a given level of insulin, a relative incomplete inhibition of glucose and lipid production by the liver and incomplete

suppression of lipolysis[12]. To maintain normal glucose levels in an insulin-resistant state, the β cells will respond by increasing insulin secretion, which will lead to hyperinsulinaemia. If the β cell mass starts to reduce and/or β cell function becomes compromised, insulin levels will at some point become insufficient to maintain glucose levels within the normal range after a meal and/or in a fasted state. Because of this mismatch between the levels of insulin required to maintain euglycemia and the capacity of the β cells to keep producing sufficient levels of insulin, T2D eventually develops. The degree of insulin resistance at which the β cells start to decompensate is believed to be highly variable between individuals[13]. For example, maturity onset diabetes of the young (MODY) is a heterogeneous group of monogenic forms of diabetes, in which diabetes develops by β cell dysfunction in the absence of insulin resistance[14]. On the other side of the spectrum, the Donohue's syndrome, caused by severe loss-of-function mutations in the insulin receptor, is characterised by extreme insulin resistance and hyperinsulinemia[15]. Common forms of T2D are characterised by a relative insulin insufficiency[12,13], and genetic evidence indicates that the level of insulin resistance at which β cell decompensation starts to occur may be largely genetically determined[16].

Although unmodifiable risk factors such as family history of T2D[17] and ethnic background[18] explain a substantial proportion of the variance in incidence, several lifestyle-related risk factors, including overweight, obesity and abdominal adiposity[19,20], lack of physical exercise[19,21], sedentary lifestyle[22,23], unbalanced diet[4,15] and smoking[21] have been identified through epidemiological observational research as important predictors of disease incidence. Multiple lifestyle intervention studies have confirmed that acting on these risk factors can be effective in reducing risk of T2D[19,25,26].

Genetic approaches to study the aetiology of type 2 diabetes

The urgency to prioritise proposed risk factors

While causality has been demonstrated for a limited set of lifestyle-related risk factors[19,25–28], the causal relevance for most risk factors proposed for T2D, especially for those emerging from -omics and other deep phenotyping-based studies, remains unknown[29–31]. Randomised controlled trials are the "gold-standard" for causality assessment but are expensive and time-consuming, and may not be feasible (e.g., birth weight, epigenetic markers) or ethical (e.g., alcohol consumption, smoking) for some exposures. Examples of high-profile failures to replicate interventions that were prioritised on the basis of observational evidence highlight the limited ability of even rigorously conducted observational studies to limit the influence of bias and confounding and to draw causal inferences about observed statistical associations[32,33]. While risk factors that are not causally associated with a disease, for example markers of the subclinical disease process, can be useful for disease prediction, this is not the case if the aim is to identify targets for intervention. Hence, new ways to prioritise risk factors on the basis of their causal likelihood are needed.

Genetic evidence through genotyping and sequencing of patients is expected to reduce the high rate of late-stage failures in drug development to translate laboratory models into the clinic[34–36]. Similarly, genetic insights may help to generate evidence about causal relevance and prioritise risk factors for testing in trials. The two most commonly adopted genetic approaches to identifying aetiological pathways to disease are genome-wide association studies (GWAS), based on either array-based genotyping or whole-genome sequencing, and Mendelian randomisation. The next two sections describe GWAS and Mendelian randomisation in the context of T2D.

Genome-wide association studies have linked more than 200 genetic regions to T2D susceptibility

Hypothesis-free testing for associations of germline genetic variants across the genome with T2D can identify genes involved in disease risk, and therefore has the potential to improve our understanding of the aetiological mechanisms of T2D and facilitate the prioritisation of novel targets for treatment and prevention. Such large-scale genetic studies including up to tens of thousands T2D cases have been conducted based on array-based genome[37–39] and exome-wide[40] genotype data and on whole-genome or exome sequence data[41].

Over the past decade, several consortium efforts have combined array-based genome-wide association data from large numbers of studies to identify the genetic determinants of T2D, with sample sizes and genome coverage continuously increasing over time[16,37–39,42–45]. In one of the latest and thus far largest meta-analysis of GWAS, 27 million genetic variants in up to 74,000 T2D cases

and 824,000 controls of European ancestry were tested[16] and more than 240 independent genetic loci reached genome-wide significance.

Sequencing and array-based genome and exome-wide association studies have taught us that the genetic predisposition to T2D is predominantly driven by common variants, of which most are located in intergenic regions and have a small effect size on disease risk[16,40,41]. Furthermore, cross-trait association analyses for genetic loci identified for T2D have corroborated the importance of established aetiological pathways, in particular islet β cell function, adiposity and insulin resistance[16,40,46]. Cross-trait analyses in combination with the integration of islet regulatory annotations suggest that a large proportion of the T2D loci is associated with β cell dysfunction and that only a small subset acts through insulin resistance, which suggests that the genetic predisposition to T2D is strongly driven by insulin secretion defects[16].

Although a few GWAS on individuals of non-European ancestry have been conducted [39,42,44,45], the current genetic research on T2D is still vastly dominated by studies on European individuals, both in terms of sample sizes and the number of conducted studies. This bias towards European ancestry forms a prominent limitation of current genetic research on T2D. The lack of understanding of potential ancestry differences in the genetic architecture of T2D limits investigations into aetiological pathways that are specific to, or play a more prominent role in non-European ancestries. As a disproportionate share of the total burden of T2D is carried by non-European ancestries, larger-scale GWAS on non-European ancestries are urgently needed.

The identification of the causal variant(s) and effector gene(s) corresponding to GWAS loci forms another important challenge in the field of GWAS for T2D and other complex phenotypes. Lack of understanding of the causal genes and variants has been one of the main limiting factors for the translation of GWAS findings to novel biological insights and targets for disease treatment or prevention. In recent years, however, tremendous progress has been made in this area. Genetic fine-mapping techniques[47] have allowed narrowing down the likely causal window at GWAS loci for T2D[16] and other traits[48,49], and ongoing trans-ethnic efforts are expected to lead to further progress in this area of research, as ancestry-specific linkage disequilibrium (LD) structures at the T2D loci can be leveraged for genetic fine-mapping. Substantial progress has also been made in the identification of effector genes corresponding to GWAS loci using a range of strategies, including statistical colocalisation of GWAS signals and cis-expression (eQTL)[50] and protein-quantitative trait loci (pQTL)[51], the integration of tissue-specific regulatory and epigenetic annotations[16,52], and bioinformatics tools that implement some of these aspects and can be run on summary-level GWAS data, such as DEPICT[53] and MetaXScan[54].

Mendelian randomisation to prioritise proposed risk factors for type 2 diabetes

Genetic approaches can support causal inference because genetic variants are less correlated with many of the measured and unmeasured factors that can confound observational exposure-to-disease associations, provided important assumptions are met. The genetic approach to causal inference visualised in Figure 1.3, referred to as "Mendelian randomisation" (MR), was originally proposed to investigate the association between fibrinogen and CHD by Keavney[55] and has since gained popularity and been the subject of numerous studies and reviews[56–59]. MR uses a genetic score, or in some cases a single genetic variant, as an instrumental variable for the exposure in order to assess the unconfounded effect size of an exposure-outcome association. In its simplest form, MR estimates the causal effect size of the exposure on the outcome by taking the ratio of the effect size of the genetic score for the exposure on the outcome, and the effect size of the genetic score on the exposure itself. This estimate represents the long-term causal influence of the exposure on the outcome, under the condition that the genetic instrument for the exposure fulfils the so-called instrumental variable assumptions, which are the following:

1. The instrument is associated with the exposure.
2. The instrument is not associated with any of the confounders of the exposure-outcome association.
3. There is no conditional association between the instrument and the outcome, given the exposure and all confounders.

As the instrumental variable assumptions are not fully empirically testable, one has to be cautious when interpreting the results of an MR analysis, especially if the biological mechanisms through which the genetic variants influence the exposure are unknown. Robust MR methods which rely on a set of more relaxed assumptions have been developed, but these methods come at the cost of reduced power[58–60]. For example, MR-Egger still produces consistent causal estimates when all genetic variants in the score violate the IV assumptions but instead meet the so-called "Instrument Strength Independent of Direct Effect" (InSIDE) assumption, which states that the pleiotropic effect sizes of the genetic variants in the score have to be independent of their effect sizes on the exposure[58]. Weighted median MR methods allow for consistent causal estimates even if up to 50% of the weight contributed comes from genetic variants that violate the IV assumptions[59].

Genetic approaches to estimating causal effect sizes require estimates of genetic associations with exposures of interest and outcomes. For many risk factors, these data have already been generated through GWAS, and results for T2D are publicly available via the online type 2 diabetes knowledge portal (http://www.type2diabetesgenetics.org)[61]. Therefore, the cost and time investments of

genetics-based causal investigations are small, especially compared to the costs, duration and participant burden of clinical trials.

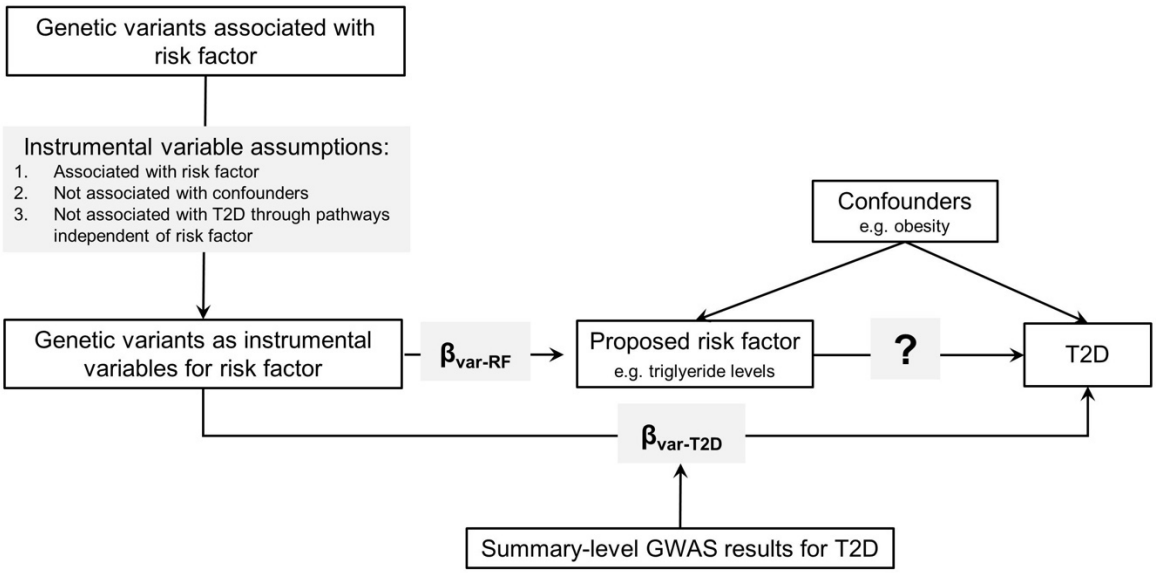


Figure 1.3: Schematic overview of the Mendelian randomisation framework. $\beta_{var\text{-}RF}$: Per-allele effect size of genetic variant on risk factor; $\beta_{var\text{-}T2D}$: Per-allele effect size of genetic variant on T2D. Figure published in Wittemans, Lotta and Langenberg, *Current Diabetes Reports* (2018)[62].

The complex interplay between blood lipid fractions and T2D

Genetics-based causality assessments of blood lipids as diabetes risk factors

After meta-analyses of randomised controlled trials of statins revealed that this major class of low-density lipoprotein cholesterol-lowering drugs modestly increases risk of T2D[63], a series of genetic studies focusing on the contribution of lipid fractions to diabetes risk was conducted. Early MR studies investigating the causality of the blood lipid fractions focused on the separate contributions of three major lipid fractions: total triglycerides, high-, and low-density lipoprotein (HDL, LDL) cholesterol[64–66]. Because levels of these fractions are regulated by partially overlapping physiological processes, genetic scores were not specific to one lipid fraction only, and therefore specific causal inference is limited. To disentangle the causal effect of genetically related risk factors, so-called multivariable MR methods have been developed[67], which estimate independent effect sizes of genetically correlated risk factors on the outcome in one model by taking the genetic correlation between them into account. High HDL cholesterol has been consistently associated with lower incidence of T2D, both observationally[68–70] and in multivariable MR studies[71,72], but recent evidence highlights that the extent to which LDL cholesterol and total triglycerides levels influence diabetes risk may be mechanism-dependent.

Complex relationships between LDL cholesterol and diabetes risk

While there is no doubt that *higher* LDL cholesterol increases the risk of diseases of the heart and vasculature, strong evidence is now available that some mechanisms *lowering* LDL cholesterol increase the risk of T2D[71,72]. Using multivariable MR, two recent publications suggested that higher levels of LDL cholesterol are causally related to lower diabetes risk, in line with evidence of statin trials[63] and a study of patients with familial hypercholesterolaemia, reporting a lower incidence of T2D, compared to their unaffected family members[73]. In contrast, familial hypercholesterolaemia was not associated with lower diabetes risk in an Amish population[74]. Differences in the genetic background, identification, and management of these Dutch and Amish patient populations may have contributed to these discrepant findings.

The extent to which levels of LDL cholesterol influence diabetes risk may depend on the physiological mechanism that contributes to differences in lipid levels[75,76]. For a given effect on LDL, the effect size of the associations between genetically predicted LDL cholesterol and T2D was shown to differ depending on the gene in which LDL variants are located when variants in current or prospective LDL-lowering drug targets, including *PCSK9*, *HMGCR*, *NPC1L1, ABCG5/G8* and *LDLR*, were compared[75]. This is in contrast to the effect of LDL cholesterol on risk of coronary heart disease, which follows a clear log-linear pattern, in observational, genetic[75,77] and intervention studies[78].

The role of total triglycerides may be mechanism-dependent

Although total triglycerides have been consistently associated with higher incidence of T2D[68–70], results for total triglycerides from genetics-based causal assessments are inconsistent[71,72]. Genetic evidence suggests that triglycerides may be ambiguously associated with diabetes risk, with protective or detrimental effects depending on the physiological mechanism driving triglyceride levels. Triglyceride-lowering genes involved in intravascular lipolysis, including intravascular lipoprotein lipase (*LPL*) and its inhibitor angiopoietin-like 4 (*ANGPTL4*), have been associated with lower risk of T2D[76,79,80], while triglyceride-lowering variants in genes involved in the hepatic production of triglyceride-rich lipoprotein particles are linked to higher risks of diabetes and fatty liver[76]. In contrast, there is strong genetic evidence that several triglyceride-raising mechanisms are causally related to risk of CHD[76,81,82].

Moving beyond Mendelian randomisation

These recent findings demonstrate how different genetic approaches can help to investigate and better understand the complex relationships between dyslipidaemia and diabetes risk by looking not only at risk factor levels, but specific mechanisms underlying these differences. Investigating combined genetic scores that encompass a large range of biological mechanisms affecting overall lipid concentrations may help general causal inference about the exposure in question but cannot reveal insight into the complexity underlying regulation of lipid metabolism. Given that LDL loci (for a given effect on LDL) differ in their association with diabetes risk, locus-specific results may also help to better estimate the effect of interventions that target levels through different mechanisms will have on different outcomes. Pleiotropy and heterogeneity of genetic loci is considered a limitation of MR but studies on the role of lipid metabolism for diabetes aetiology highlight that biologically relevant insights can be obtained by investigating individual genetic loci that are pleiotropic or have, given their effect on the exposure, a disproportionate effect size on the outcome.

The central role of peripheral fat in the aetiology of T2D

Adiposity and body composition are major risk factors for T2D

Overweight and obesity are strong modifiable risk factor for T2D[19,25–28] and responsible for the secular increase in the diabetes prevalence. Weight loss interventions have been shown to reduce diabetes risk[19,25–28], and are even able to revert glycaemia in patients with T2D to non-diabetic levels[83]. Genetic evidence supports the causal role of overall adiposity. The major BMI locus near *FTO* was the first found to exert its diabetes risk-raising effect entirely by increasing BMI[84]. More recently, the largest European-ancestry focused GWAS of T2D, which tested 27 million variants and included more than 74,000 cases, showed that 26 of the more than 400 conditionally independent variants at the identified loci for T2D predominantly increase risk through their effect on BMI, including established adiposity loci such as *FTO*, *MC4R* and *TMEM18*[16]. Independent causal roles of overall adiposity and abdominal fat accumulation has also been supported by MR studies[85–87].

Genetic loci link low adiposity to high diabetes risk

Evidence is emerging that not all mechanisms leading to reduced fat accumulation are uniformly beneficial to metabolic health. Loci near *IRS1* and *COBLL1/GRB14* identified for lower total body fat percentage were found to be associated with higher risk of T2D[88,89]. Similarly, a coding variant in the *CREBRF* gene was found to increase BMI by 1.32-1.46 kg/m^2 per allele but reduce diabetes risk by 40% in a Polynesian population[90]. In line with this, the most recent T2D GWAS showed that for a subset of the identified loci the associations with T2D risk were strengthened following adjustment for BMI. Among these were signals previously shown to be associated with lower capacity of adipose tissue generation and expansion[79], including *PPARG* – a gene involved in Mendelian forms of partial lipodystrophy[91]. Together, these results suggest that, when focusing on distinct pathways connecting adiposity and diabetes, certain biological processes may reduce fat accumulation while increasing risk of T2D, while others will reduce risk.

Lipodystrophy-like mechanisms may contribute to insulin resistance

Genetic research on metabolic risk factors demonstrates that distinct pathways linking low adiposity to diabetes risk may be driven by insulin resistance[79,92]. It was recognised early on that discovery of loci associated with insulin resistance is facilitated by accounting for differences in obesity levels[93]; GWAS of insulin resistance have since been conducted based on fasting insulin levels adjusted for BMI[79,93–95]. Research on the relationship between insulin resistance and refined measures of body composition measured by DEXA shows associations between genetic determinants of insulin resistance and reduced fat accumulation in the legs and other peripheral

compartments[79,95]. Positive relationships of insulin resistance with visceral fat mass and liver enzymes[79,95,96] suggests that genetic predisposition to insulin resistance, not primarily driven by a long-term positive energy balance, is linked to a body fat distribution pattern that favours visceral and ectopic fat deposition over deposition in peripheral body compartments. Questions remain over whether impaired peripheral fat and enhanced central fat distribution are independent mechanisms.

Lotta *et al.* found that the 53 loci associated with hallmarks of common insulin resistance are also enriched in patients with familial partial lipodystrophy type 1, which is a rare condition characterised by extreme insulin resistance and the limited ability to store fat peripherally[79]. These loci identified for common insulin resistance are enriched with causal genes involved in monogenic forms of lipodystrophy, including *PPARG*, *PIK3R1* and *INSR* [15]. This suggests the presence of shared mechanisms between common and rare forms of insulin resistance, and that lipodystrophy-like mechanisms may contribute to common forms of insulin resistance, and metabolic and cardiovascular disease.

The majority of genetic studies has been conducted in participants of European ancestry. Many people who develop T2D in South-East Asia, the middle East and sub-Saharan Africa are not overweight or obese[97–99]. It is likely that fat distribution is important, but the causal role of the relative contributions of greater abdominal and lack of gluteo-femoral fat have not been investigated.

The adipose tissue expandability hypothesis

Altogether, these findings support the so-called adipose tissue expandability hypothesis, which states that each individual has a limited capacity to store excess energy in adipose tissue; once that threshold has been reached, excess energy in the form of lipids may accumulate ectopically in diabetes-relevant tissues such as the liver, muscles and the pancreas, where they may contribute to tissue dysfunction and diminished insulin action[100,101]. Individuals with a larger capacity to peripherally accumulate fat may therefore be more protected from the cardiometabolic consequences of a long-term positive energy balance. In support of this hypothesis, longitudinal data on weight change have demonstrated that people with a higher burden of insulin resistance-increasing alleles tend to expand their hip fat depot less as they gain weight than people at low genetic risk to develop insulin resistance[79]. Refined imaging methods, including DEXA and magnetic resonance imaging (MRI), will help to further confirm this hypothesis and the specific contribution of these mechanisms to the broader relationship between fat distribution and metabolic risk in men and women and across ethnic groups.

Using genetics to prioritise causal pathways amongst the multitude of reported biomarker associations

Biomarkers as causal candidates

Interest into the potential causality of biomarkers for T2D has grown, as they can point to novel disease pathways and help to identify potential targets for intervention, provided they are causal. A systematic literature review has shown that more than 160 biomarkers measured in blood or urine have been associated with T2D, but for a small minority has an assessment of causality been attempted or demonstrated[102]. No evidence for causality has been found for the majority of the biomarkers that have been investigated using genetic approaches, including vitamin D[103] , adiponectin[104,105], uric acid[106], C-reactive protein[107] and gamma-glutamyl transferase[108–110]. However, for a few biomarkers, including sex hormone-binding globulin[111,112], B and A-type natriuretic peptide systems[113,114] and bilirubin[115], suggestive evidence for a causal relationship with T2D has been reported. Historically, some of the older studies investigating one biomarker at a time were based on comparatively small sample sizes and a limited understanding of the specificity of the genetic instrument. Replication of these findings is needed using genetic prediction models that maximise exposure variance explained, utilise the largest available GWAS summary statistics, test generalisability across different ancestries where possible and investigate specificity of genetic variants in adequately powered studies with comprehensive coverage of relevant biological pathways.

New wave of omics-based biomarkers

Omics methods have enabled the high-throughput assessment of molecular traits in large epidemiological studies, including the metabolome, proteome and lipidome. This has led to a new wave and scale of observational diabetes "biomarker" discoveries, including metabolites[29,116–119], lipid species[120–123], proteins[124] and methylation markers[125].

Metabolomics is the –omics method that focuses on measuring the entire collection of small molecules or metabolites, or a subset thereof, in a biological sample, which in epidemiological studies is often plasma or serum. Metabolomic profiling thereby enables a comprehensive insight into the metabolic state of an individual[126,127]. Liquid and gas chromatography, mass spectrometry (MS), or a combination of both, and hydrogen nuclear magnetic resonance (^{1}H-NMR) are the most commonly used analytical techniques in metabolomics[126].

Although the hallmark of T2D is disrupted glucose homeostasis – meaning T2D is in essence a metabolic condition – the relationship of metabolic pathways other than glucose and lipid metabolism with T2D is largely unknown. With metabolomics and lipidomics being increasingly

implemented in population-based and T2D-specific epidemiological studies, novel insights on metabolic disturbances other than disruption of glucose homeostasis and dyslipidaemia to T2D are starting to emerge. Besides numerous reported associations of the metabolome with T2D, several metabolomics studies focusing on T2D-related metabolic risk factors, including insulin resistance[128–132] and overweight, obesity and weight gain[133–135], revealed multiple associations across the metabolome.

Metabolomics and type 2 diabetes: a systematic literature review

In order to create an overview of the available literature on blood-based metabolites linked to incidence of T2D, I conducted a systematic literature search of PubMed up to and including 23 April 2018. I searched for epidemiological studies investigating the link between the serum or plasma metabolome and T2D, using a combination of three search term strings: one for T2D, one for epidemiological studies and one for metabolomics and the two most commonly used measurement techniques for metabolomics, i.e., proton nuclear magnetic resonance and mass spectrometry (Supplementary Table 1.1). From the 1,805 titles initially found in PubMed, 85 studies, of which 43 were cross-sectional and 42 were longitudinal, were retrieved (Figure 1.4).

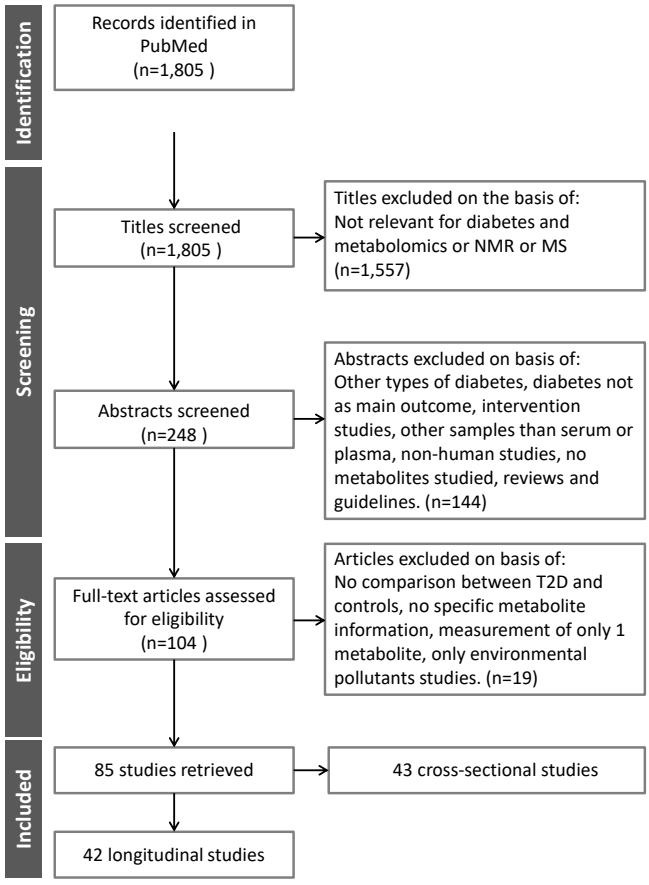


Figure 1.4: PRISMA diagram of systematic literature review on the metabolomic signature of type 2 diabetes. NMR: nuclear magnetic resonance, MS: mass spectrometry, T2D: type 2 diabetes.

The 42 retrieved longitudinal studies on metabolomics and T2D varied greatly in terms of study design, adopted measurement techniques, studied metabolite classes and statistical methods (Table 1.2). Among the 42 studies, there were 24 prospective cohort studies, 10 case-cohort studies, 13 nested case-control studies and 1 retrospective study. A variety of metabolomics techniques was used. In the majority of the studies, a combination of a chromatographic method (gas (GC), liquid (LC), high performance liquid (HPLC) or ultra performance liquid chromatography (UPLC)) with mass spectrometry (MS) or tandem mass spectrometry (MS/MS) was adopted, while 3 studies used chromatography only and one study used tandem mass spectrometry only. Seven other studies relied on proton nuclear magnetic resonance (^{1}H-NMR). In six studies, metabolite profiling was done using commercial metabolomics platforms, more specifically the Biocrates Absolute*IDQ*® platform (Biocrates Life Sciences, Innsbruck, Austria)[29,136,137] and the Metabolon DiscoveryHD4® platform (Metabolon, Inc., Durham, USA)[138–140] (Table 1.2).

Studied metabolite classes and species varied considerably between studies. Nine studies[117–119,128,138–143] used an untargeted approach based on one or multiple measurement techniques, in order to cover the broadest possible spectrum of metabolites and avoid bias towards certain pre-selected metabolite classes (Table 1.2). Four other studies used a targeted approach to cover a wide range of metabolic pathways to achieve near-metabolome-wide coverage[29,137,144,145], while 7 studies focused on aqueous metabolites only[146–152] and 8 studies covered a broad range of lipid classes[31,121–123,144,153–155]. Of the studies on lipid metabolism, 4 studies used NMR-based refined measures of lipoprotein particle size and composition[31,144,153,154]. Other studies were set up to investigate the role of a specific metabolic pathway or a very limited set of metabolites for which there was a prior interest, such as specific nutritional biomarkers[156], saturated fatty acids in phospholipids[120], testosterone metabolites[157,158], ceramides[159], sphyngoid bases[160,161] and omega 6 fatty acids[162] (Table 1.2).

A wide range of statistical methods were used in the 42 identified studies. The vast majority of studies adopted univariate methods in which one metabolite at a time was tested for a statistical association with incident T2D. The logistic regression model was the most commonly adopted univariate method, and a smaller proportion of the studies used time-to-event analyses based on Cox proportional hazards model. The main limitation of using univariate models is that these implicitly assume that each metabolite influences the outcome independently of all other metabolites. In reality, however, groups of metabolites are likely to jointly influence the outcome and effect sizes obtained from simple regression models are therefore likely to be overestimated. An alternative approach to assess associations between the metabolome and T2D is to use regularised multiple regression models in which the associations of all metabolites are tested in one

model while taking into account the correlations between the metabolites. Effect size estimates for each metabolite obtained from this type of model are adjusted for all other metabolites. In some of the more recently published studies, regularised multivariate methods, such as least absolute shrinkage and selection operator (LASSO) and regularised least squares regression, have been used to identify subsets of metabolites significantly and independently associated with T2D[118,139,144,147].

In total, circulating levels of more than 230 unique metabolites have been associated with future onset of T2D (Figure 1.5). In virtually every studied metabolic pathway, metabolite species have been associated with incidence of T2D. Amino acids, carbohydrates, di- and triglycerides, phospholipids, acyl carnitines, fatty acids and lipoproteins are among the most commonly studied and reported metabolite classes for T2D (Figure 1.5).

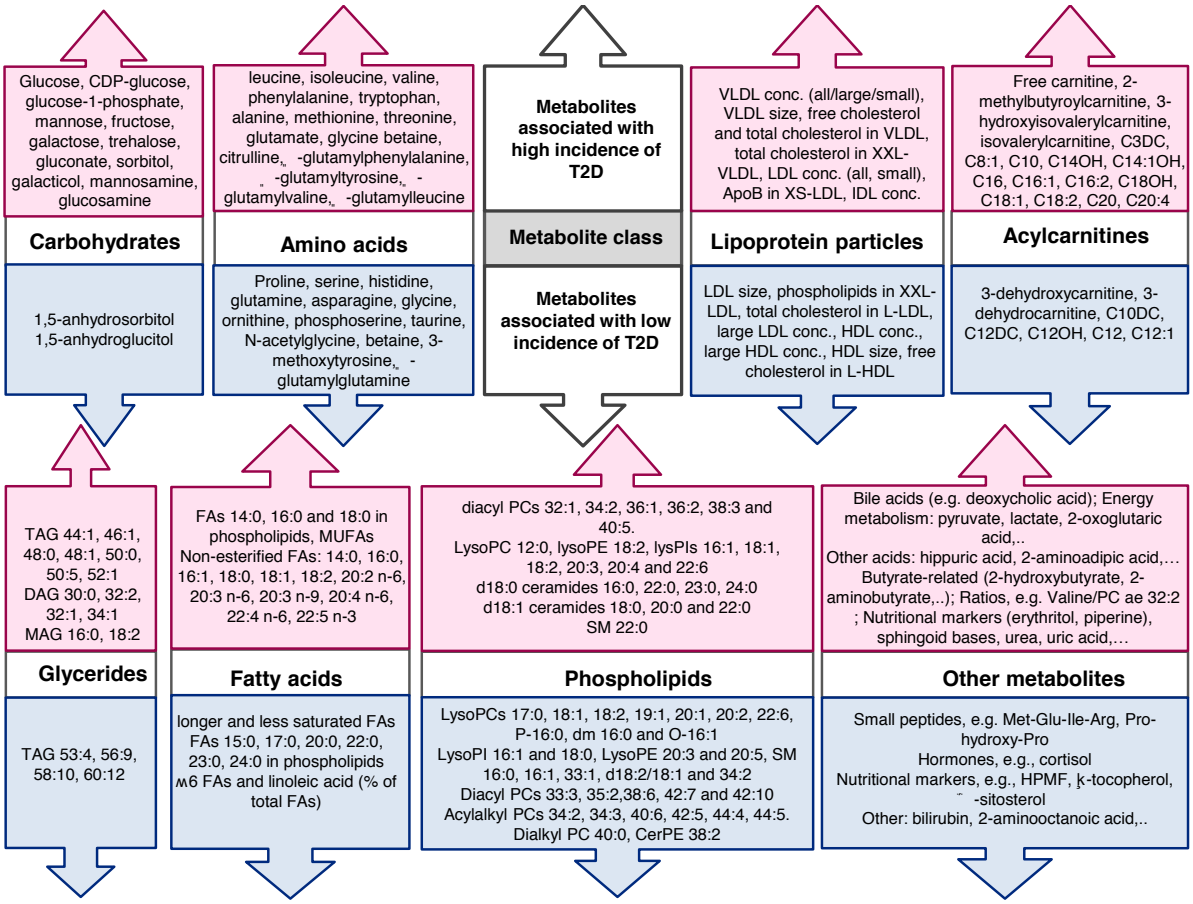


Figure 1.5: Overview of metabolites that have been reported for incidence of type 2 diabetes. Metabolites reported for incident T2D were reported by metabolite class. The pink boxes contain metabolites which have been positively associated with incidence of T2D, whereas the blue boxes contain metabolites inversely associated with T2D incidence.

Table 1.2: Longitudinal studies on metabolomic signatures of incident type 2 diabetes

Publication	Study design	Study population	Incident cases	Follow-up	Measurement technique or platform	Measured metabolite classes
[146]Chen 2016	Nested case-control and case-subcohort	Shanghai Diabetes Study	51	10 years	UPLC-MS	Amino acids
[163]de Mello 2015	Prospective cohort	Finnish Diabetes Prevention Study	130	8 years	GC-MS	Non-cholesterol sterols
[141]Drogan 2015	Nested case-control	EPIC-Potsdam	300	6 years	UPLC-MS	Untargeted approach
[128]Fall 2016	Prospective cohort and nested case-cohort	ULSAM, PIVUS TwinGene, KORA S4	358	Not reported	UPLC-MS	Untargeted approach
[158]Fenske 2015	Prospective cohort	Study of Health in Pomerania	226	5 years	LC-MS	androstenedione, total tostesterone
[153]Festa 2005	Prospective cohort	Insulin resistance atherosclerosis study (IRAS)	130	5.2 years	^{1}H-NMR	Lipoprotein particles
[29]Floegel 2013	Case-cohort	EPIC-Potsdam	800	7 years	AbsoluteIDQ p150 Biocrates	Hexose, acylcarnitines, amino acids, hexose and phosphatidylcholines
[120]Forouhi 2014	Case-cohort	EPIC-InterAct	12,132	7.8 years	GC	Saturated fatty acids in plasma phospholipids
[154]Hodge 2009	Case-cohort	Melbourne Collaborative Cohort Study	59	4 years	^{1}H-NMR	Lipoprotein particles
[157]Joyce 2017	Prospective cohort	Cardiovascular Health Study	112	9.8 years	LC-MS	Testosterone and dihydrotestosterone
[144]Liu 2017	Prospective cohort	Erasmus Rucphen Family study	137	11.3 years	LC-MS, NMR	triglycerides, glycerophospholipids, ceramides, amino acids and other low-weight molecules, lipoprotein particle size sub-fractions
[119]Lu 2016	Nested case-control	Singapore Chinese Health Study	197	6 years	LC-MS, GC-MS	Untargeted approach
[123]Lu 2018	Nested case-control	Singapore Chinese Health Study	160	6 years	HPLC-MS	lysophosphatidylinositols, non-esterified fatty acids, acylcarnitines
[155]Mahendran 2013	Prospective cohort	METSIM	176	4.5 years	^{1}H-NMR	Glycerol and fatty acids
[147]Merino 2018	Prospective cohort	Framingham Heart Study Offspring cohort	95	20 years	LC-MS	Amino acids, amines and other polar metabolites
[136]Molnos 2018	Prospective cohort, nested case-control	KORA and EPIC-Potsdam	910	7 years	AbsoluteIDQ p150 Biocrates	Metabolite ratios
[31]Mora 2010	Prospective cohort	Women's Health Study	1,687	13 years	^{1}H-NMR	Lipoprotein particles
[160]Mwinyi 2017	Nested case-control	CoLaus	251	5 years	LC-MS	Sphingoid bases

Publication	Study design	Study population	Incident cases	Follow-up	Measurement technique or platform	Measured metabolite classes
[161]Othman 2015	Prospective cohort	CAD patients	32	8 years	LC-MS	Sphingoid bases
[142]Padberg 2014	Retrospective study	Blood donors from the Bavarian Red Cross	28	6 years	MxPTM Broad Profiling Platform	Untargeted approach and targeted measurement of eicosanoids
[129]Palmer 2015	Prospective cohort	IRAS	76	5 years	MS/MS	Amino acids and acylcarnitines
[118]Peddinti 2017	Nested case-cohort	Botnia Prospective Study	146	10 years	UHPLC-MS, GC-MS	Untargeted approach
[145]Qiu 2016	Nested case-control	Dongfeng-Tongji and Jiangsu Non-communicable Disease cohorts	1,039 +520	4.6 and 7.6 years	HPLC-MS	Amino acids, acylcarnitines, cholines, amines, purine and indole derivatives, B vitamins
[138]Rebholz 2018	Prospective cohort	ARIC	1,126	20 years	DiscoveryHD4 platform Metabolon	Untargeted approach
[122]Rhee 2011	Nested case-control	Framingham Heart Study	189	12 years	LC-MS	Triglycerides, cholesterol esters, phosphatidylcholines, lysophosphatidylethanolamines, diglycerides and sphingomyelines
[164]Savolainen 2017	Prospective cohort	Cohort of Swedish women	69	5.5 years	GC-MS	α- and δ-tocopherol, alkylresorcinols, β-alanine, CMPF, fatty acids
[156]Savolainen 2017	Prospective cohort	Cohort of Swedish women	69	5.5 years	GC-MS	Not specified
[117]Shi 2018	Nested case-control	Västerbotten Intervention Programme cohort	503	7 years	HPLC-MS	Untargeted approach
[148]Stančáková 2012	Prospective cohort	METSIM	151	4.7 years	^{1}H-NMR	Amino acids
[165]Sun 2016	Prospective cohort	Nutrition and Health of Aging Pop- ulation in China study	507	6 years	LC-MS	Acylcarnitines
[121]Suvitaival 2018	Prospective cohort	METSIM	107	5 years	UPLC-MS	glycerophospholipids, triacylglycerols, cholesterolesters, sphingomyelins and ceramides
[149]Svingen 2016	Prospective cohort	Norwegian cohort of patients undergoing coronary angiography for stable angina pectoris	233	7.5 years	LC-MS, GC-MS	TMAO, choline, betaine, dimethylglycine and sarcosine
[166]Tillin 2015	Prospective cohort	SABRE	340	19 years	^{1}H-NMR	Amino acids
[150]Wang 2011	Nested case-control	Framingham Heart Study	189	12 years	LC-MS	Amino acids, amines and other polar metabolites

Publication	Study design	Study population	Incident cases	Follow-up	Measurement technique or platform	Measured metabolite classes
[151]Wang 2013	Nested case-control	Framingham Heart Study	188	12 years	LC-MS/MS	Organic acids, pyrimidines, purines and other metabolites
[137]Wang-Sattler 2012	Prospective cohort	Cooperative Health Research in the Region of Augsburg (KORA)	91	7 years	AbsoluteIDQ p180 Biocrates	Hexose, acylcarnitines, amino acids, biogenic amines, sphingomyelins and phosphatidylcholines
[159]Wigger 2017	Nested case-cohort, nested case-control	D.E.S.I.R. and CoLaus	189 and 147	9 and 5 years	UPLC-MS	Ceramides and dihydroceramides
[162]Wu 2017	Prospective cohort, case-cohort, nested case-control	20 studies	4,347	3.4-21.4 years	GC, GLC	Linoleic acid and arachidonic acid
[152]Yamakado 2015	Prospective cohort	Japanese population-based cohort	174	4 years	HPLC-MS	Amino acids
[139]Yengo 2016	Case-cohort	D.E.S.I.R.	231	9 years	DiscoveryHD4 platform Metabolon	Untargeted approach
[140]Yu 2016	Prospective cohort	Shanghai Women's and Men's Health studies	73	not reported	DiscoveryHD4 platform Metabolon	Untargeted approach
[143]Zhao 2015	Case-cohort	Strong Heart Family Study	133	5.5 years	LC-MS	Untargeted approach

Challenges of observational metabolomics and other -omics studies

Large-scale metabolomic profiling in observational epidemiological studies has led to a surge in the number of potential novel biomarkers for T2D. However, these findings have only marginally improved our understanding of biological mechanisms leading to T2D, due to several limitations from which the currently available studies suffer. First, very few of the metabolites that have been associated with incident T2D have been reported in more than one study. Moreover, substantial heterogeneity between studies in terms of study design, adopted measurement techniques and statistical methods limit the comparability of the same metabolite reported in different studies. These limitations are particularly applicable to lipid species, as no study has systematically investigated lipids across all different lipid classes, and because limited resolution in the measured molecular structure of lipids makes comparison between different studies even more challenging. Because of this lack of replication and heterogeneity among the studies, a comprehensive overview of the metabolomic signature of incident T2D is still lacking, despite the considerable number of conducted studies.

A second important limitation of the current studies on metabolomic patterns of T2D is that the causality of the associations is not assessed. Although only longitudinal associations between metabolite levels and T2D were considered in this literature review, the reported associations may not reflect causal relationships for reasons of unmeasured confounding, reverse causality or chance. In order for these metabolite patterns to contribute to a better understanding of pathophysiological processes related to T2D, it is essential to identify from this new pool of biomarkers those metabolites that reflect causal pathways.

Molecular traits, such as metabolites, often represent the same biological pathway. Large subsets of metabolites can therefore be strongly correlated with each other. Cohort studies assessing different outcomes have shown that individual metabolites can be associated with multiple disease outcomes [150,167,168]. The lack of specificity with regard to the exposure and outcome highlights that large-scale parallel assessment of metabolites and other -omics-based biomarkers in an observational setting is even more prone to false positive discovery by multiple testing than traditional biomarker studies. Integration of genetic data can be a helpful tool to prioritise potential causal pathways from the dense network of molecular trait-to-disease associations.

Omics integration

Several GWAS on the metabolome have been conducted[169–171], with larger-scale efforts underway. The results of these studies show that, due to close biological proximity, genetic loci explain large proportions of the variance of a given metabolite and relevant biological mechanisms can be identified in relatively small sample sizes. For example, in a study on the genetic architecture of

the human plasma metabolome, GWAS were conducted for 529 metabolites, based on 7,824 participants[51]. Despite the modest sample size, associations of more than 145 genetic loci with 400 metabolites were found. The median variance in metabolite levels explained by the genetic variants was 6.9% (range 1-62%), which is higher than the total variance in BMI explained by over 550 signals identified in a GWAS of around 700,000 participants[172].

Integration of genetic markers with large-scale molecular phenotyping can enable the identification of novel molecular aetiological pathways. For example, GWAS in combination with metabolomics was recently used to assess the causality of the widely reported associations between branched-chain amino acids and incidence of T2D[150]. Based on a genetic score including variants in the *PPM1K* gene, which encodes the activator for the enzyme catalysing the rate-limiting step in branched-chain amino acid breakdown, evidence was found that branched-chain amino acid catabolism and T2D may be bidirectionally causally related[173].

As -omics technologies are becoming more widely implemented in epidemiological studies of increasing size, opportunities to obtain an even more comprehensive insight into disease mechanisms may arise by integrating multiple layers of -omics data, such as proteomics.

Conclusion of the literature overview

Rapid progress in the identification of the genetic basis of T2D and related phenotypes, together with technological advances that facilitate high-throughput and refined phenotyping at scale, provide important opportunities for epidemiological research to identify novel aetiological pathways and enhance causal understanding of potential risk factors that are amenable to intervention. Better-powered studies now provide greater sensitivity to investigating pleiotropy that may have previously been overlooked and may have confounded effect sizes of genetically predicted exposures. Traditional genetic approaches to inferring causality have generally focused on average effects, sometimes estimated from large polygenic scores. Recent examples have highlighted opportunities that arise from distinguishing the effects of specific mechanisms represented by a locus or a subset of loci, which may otherwise be overlooked.

Chapter 2: Assessing the causal associations of glycine metabolism with type 2 diabetes and coronary heart disease

Contributions and collaborations

I performed all analyses described in this chapter, with the exception of the GWAS for glycine levels in the INTERVAL study, which was conducted by Dr Clare Oliver-Williams and Dr Adam Butterworth (Cardiovascular Epidemiology Unit, University of Cambridge) and the GWAS for blood pressure traits in the UK Biobank, which were conducted by Dr Joanna Howson, Dr Praveen Surendran and Savita Karthikeyan (Cardiovascular Epidemiology Unit, University of Cambridge).

Publications related to this chapter

The analyses on the role of glycine as an aetiological factor for coronary heart disease presented in this chapter has been submitted to Nature Communications:

L. B. L. Wittemans, L. A. Lotta, C. Olliver-Williams, [19 authors], C. Langenberg. **Assessing the causal association of glycine with risk of coronary heart disease.** *Nature Communications* 10:1060 (2019). doi: 10.1038/s41467-019-08936-1.

Abstract

Background: Through high-throughput metabolic profiling, virtually every metabolic pathway has been linked to T2D, but the causality of these proposed risk factors remains largely unknown. Identifying those metabolic pathways which play an aetiological role in cardiometabolic diseases may point to novel strategies to treat or prevent disease in individuals who are at high risk. Multiple studies have reported that a higher serum concentration of glycine is associated with lower incidence of T2D and CHD, but the causality of these observational findings and the underlying biological mechanisms are incompletely understood.

Aims: This chapter aims to develop an MR-based approach to assess the aetiological role of metabolic pathways for cardiometabolic diseases and to characterise the biological mechanisms underlying causal relationships between the metabolome and disease incidence. Using glycine metabolism as an example, I a) construct and validate genetic instruments for glycine levels, b) assess the role of glycine as a causal pathway to T2D and CHD and c) identify the physiological mechanisms that drive the associations between glycine levels and incidence of T2D and CHD.

Methods: Meta-analyses of GWAS on serum levels of glycine were conducted in up to 80,003 participants. MR analyses for glycine as the exposure and CHD and T2D as outcomes were performed using 4 genetic scores with different levels of specificity to glycine. Reverse MR experiments were conducted to assess the causal associations of 3 T2D risk factors with glycine levels, and the role of CHD risk factors as mediators of the effect of glycine on CHD was assessed using multivariable MR.

Findings: 27 genetic loci reached genome-wide significance ($p \leq 5 \times 10^{-8}$) for glycine levels, including 6 near genes with enzymatic functions in glycine metabolism, and of which 22 have not previously been linked to glycine. Significant heterogeneity between the effect sizes for men and women was found for 3 loci, including *CPS1*. Genetically predicted levels of glycine were not associated with T2D risk (sexes combined: OR [95% CI] per 1 SD genetically predicted increase in glycine = 1.00 [0.97, 1.04]) based on a genetic score comprised of all 24 common loci for glycine, despite the consistently reported inverse observational association. However, when using a genetic score comprised of two loci which were highly specific to the glycine pathway and were located in genes of the glycine cleavage system, a significant inverse association of genetically predicted glycine with T2D risk was found (OR for T2D per SD glycine= 0.79 [0.68,0.92], p=0.002).

Genetically higher fasting insulin was significantly associated with lower glycine levels, but BMI and insulin secretion were not.

Conversely, significant associations of genetically predicted glycine with CHD risk were found (OR [95% CI] per 1 SD genetically predicted glycine=0.95 [0.92,0.98]) based on 4 distinct genetic scores with different levels of specificity to the glycine pathway, and no evidence was found that this effect size differs between men and women (p for heterogeneity =0.74). The inverse genetic relationship between glycine levels and CHD was attenuated when adjusting for genetically predicted blood pressure.

Conclusion: This chapter provides new insight into the genetic architecture of glycine metabolism and the findings indicate a potential protective role of glycine levels for CHD, with blood pressure as a potential mediating factor. Conversely, we found evidence that the causal effect of glycine on T2D may be pathway-specific and that the strong inverse observational glycine-to-T2D association is a consequence of insulin resistance.

Introduction

Glycine – the metabolically versatile, smallest amino acid

Glycine is the simplest amino acid in the human body and is non-essential, as it can be synthesised from serine and betaine, but it is also taken up through the diet[174]. Glycine is an intermediate of a wide range of metabolic pathways[174], including one-carbon metabolism, serine, glutathione and purine biosynthesis, lipid metabolism and detoxification reactions. Over the past decade, multiple studies have ascribed a protective role against cardiometabolic diseases to glycine, but these findings are mostly based on observational epidemiological research[29,175], or experimental findings on rodents[176–178].

Glycine and type 2 diabetes

Observational epidemiological studies have consistently linked low glycine levels to higher prevalence and incidence of T2D[29,129,147] and to diabetes-related pathophysiological processes, including insulin resistance[179,180], reduced insulin secretion[181] and adiposity[180,182]. To assess the causality of the inverse glycine-to-T2D association, Xie *et al.* conducted a genetics-based assessment of the causality of glycine on insulin resistance and on T2D based on one locus for glycine in carbamoyl phosphate synthase 1 (*CPS1*), but did not find conclusive evidence for a causal role of glycine[183]. In a recent study by Merino *et al.*, a protective effect of glycine on T2D risk was suggested based on a significant association of a genetic score comprised of 5 reported loci for glycine, including *CPS1*, with genetic risk of T2D, based on 11,600 T2D cases and 33,000 controls[147]. However, this association was mostly driven by the locus in *CPS1*, which has been reported to be highly pleiotropic[169,184], and the observed association could therefore have been driven by glycine-independent pathways. Altogether, no conclusive evidence on the causality of low glycine levels for T2D risk has been found.

Glycine and coronary heart disease

Glycine was recently found to be inversely associated with risk of acute myocardial infarction in a prospective cohort study of stable angina patients[175] and cardio-protective effects have been reported for glycine based on studies in mammalian model systems[185,186] and *in vitro* experiments[187]. A genetic study suggested a sex-specific genetic association of glycine metabolism with risk of coronary heart disease (CHD)[188] , based on a genetic variant associated with higher levels of glycine and lower risk of CHD in women but not men. However, this study focused on a single variant in *CPS1*, a gene known to be more strongly associated with levels of glycine[189] and other metabolites[184] in women, compared to men. Uncertainty about the role of glycine metabolism in risk of CHD and its potential sex-specific nature therefore remains.

Genome-wide association studies for glycine

GWAS for plasma glycine levels aiming to identify genetic regions involved in glycine metabolism have been previously conducted in up to 25,000 participants, identifying 5 genetic loci in/near *CPS1*, *GLDC*, *GCSH*, *ALDH1L1* and *PPP1R3B*, associated with glycine levels[169–171,184]. A strong sex difference in the effect size of the *CPS1* locus has been reported[189] but potential sex differences in the effect sizes of these loci have not been systematically assessed. A larger GWAS of glycine levels could increase the number of genetic loci robustly associated with glycine, thereby improving our understanding of the genetic determinants of glycine metabolism and provide genetic instruments to assess the causality of glycine on T2D, CHD and their risk factors, using the Mendelian randomisation framework[56].

Objectives

In this chapter, I aim to (1) identify genetic determinants of plasma levels of glycine, (2) investigate the likelihood of a causal role of glycine levels on risk of T2D and CHD in men and women, using novel and sex-specific genetic scores for glycine, and (3) characterise the biological pathways underlying the glycine-to-disease associations, using summary-level GWAS results derived from large-scale epidemiological studies.

Methods

Studies

EPIC

The European Prospective Investigation into Cancer and nutrition (EPIC) is an ongoing multi-centre European cohort study which was originally set up to study the role of nutrition in cancer risk[190]. More than half a million participants were recruited over the 10 participating countries and 26 study centres. Several sub-studies of EPIC have been set up to study the risk factors of other complex diseases, including T2D and cardiovascular diseases.

The **EPIC-Norfolk** study is a cohort of 25,000 middle-aged individuals from the general population of Norfolk (East England)[191], and is one of the 2 EPIC study centres in the UK. Untargeted metabolomics measurements using the DiscoveryHD4® platform (Metabolon, Inc., Durham, USA) on non-fasted plasma samples have been completed in two sub-cohorts of randomly selected participants – 5,989 randomly selected participants in sub-cohort A and 5,977 participants in sub-cohort B – and on a T2D case-sub-cohort study which included all 586 T2D cases and another 746 participants randomly selected from the entire cohort. There was no overlap between sub-cohorts A and B and the T2D case-sub-cohort. In total, 999 metabolites were measured in both sub-cohorts A and B, of which 894 were measured in at least 50% of the combined sample size. Further details about the metabolomics measurements have been described elsewhere[173]. Genome-wide genotyping was done using the Affymetrix UK Biobank Axiom array and genotype data were imputed based on a combined reference panel (Haplotype Reference Consortium[192], UK10K[193] and 1000 Genomes Phase 3[194]) and using IMPUTE2[195]. Data from EPIC-Norfolk were used to conduct a GWAS for glycine, to estimate sex-specific effect sizes of the glycine variants with glycine levels, to test associations of the genetic scores and variants for glycine across the metabolome, and to assess the observational association of glycine levels with incident T2D, CHD, myocardial infarction, stroke and stroke sub-types.

EPIC-CVD, another sub-study of EPIC, is a prospective case-sub-cohort focusing on the risk factors of cardiovascular diseases and includes nearly 14,000 incident CHD cases and sub-cohort comprised of 18,249 randomly selected participants. Details about the participants, phenotype measurements and disease case ascertainment have previously been described[196]. Genotyping was performed using the HumanCoreExome array, the Quad 660 genotyping chip and the Infinium OmniExpressExome array (all from Illumina) and genotyped data were imputed to a joint 1000

Genomes Phase 3[197] (May 2013)-UK10K[193] reference imputation panel. Data on 28,217 participants (33.7% CHD cases) of the EPIC-CVD study were used to assess the effect sizes of the genetic variants for glycine with CHD, for both sexes combined and separately.

The **InterAct** study is a T2D case-cohort study nested within EPIC, which was designed to study the interaction between lifestyle and genetic factors in relation to T2D[198]. The study includes 12,403 incident cases of T2D and a randomly selected sub-cohort of 16,154 individuals from 9 European countries. Genome-wide genotyping was done using the HumanCoreExome array. Genome-wide genotyped data were imputed to the European 1000 Genomes reference panel (March, 2012 release) using IMPUTE2. Data from the InterAct study were used to assess the sex-combined and sex-specific associations of the genetic loci for glycine with risk of T2D.

Fenland

The Fenland study is a longitudinal cohort study including more than 12,400 participants from the general population of Cambridgeshire (UK)[173]. Fasted plasma concentrations of 174 metabolites were measured using the Absolute*IDQ*® p180 Kit (Biocrates Life Sciences, Innsbruck, Austria) and more details on the measurement, processing and QC have been described elsewehere[173]. Genome-wide genotyping of the Fenland participants was done in three waves; the first 1,400 individuals were genotyped using the Affymetrix Genome-Wide Human SNP Array 5.0, the next 9,369 participants on the Affymetrix UK Biobank Axiom Array and the final group of 1,118 participants on the Illumina Infinium CoreExome Array. The same imputation strategy as for the EPIC-Norfolk study was used. GWAS for glycine and assessment of the sex-specific effect sizes of the glycine variants were run in the Fenland study.

INTERVAL

The INTERVAL study is a randomised trial of approximately 50,000 whole blood donors enrolled from all 25 static centres of NHS Blood and Transplant[199]. For the present study, non-fasting serum blood samples were provided by 40,509 unrelated individuals from the INTERVAL trial. The samples were analysed using a high-throughput serum NMR metabolomics platform[200,201], which provided information on 230 metabolites, including glycine. We removed participants with more than >30% of metabolite measures missing, duplicated individuals, and metabolic data more than 10 SDs from the mean. Genotyping was conducted using the Affymetrix UK Biobank Axiom array. Prior to imputation, SNPs missing in more than 1% of the samples or failing in more than

one batch were excluded. Monomorphic and multi-allelic variants and variants that failed to pass the threshold on clustering quality, were intensity outliers or deviated from Hardy-Weinberg equilibrium were omitted. The data were imputed to a joint 1000 Genomes Phase 3 (May 2013)-UK10K reference imputation panel. Data from INTERVAL participants with glycine measures were used to run a GWAS for glycine and to assess the sex-specific effect sizes of the glycine variants on glycine.

UK Biobank

The UK Biobank study is a longitudinal cohort study of more than 500,000 participants from across the UK[202]. Genotyping was performed using the UK Biobank Axiom Array and imputation was based on the reference panel from the haplotype reference consortium, using IMPUTE2. Sex-combined and sex-specific associations of the genetic variants for glycine with CHD, T2D and blood pressure were assessed in the UK Biobank.

Participants of all studies gave written informed consent and all studies were approved by local ethics committees.

Genome-wide association analyses for glycine levels in the EPIC-Norfolk, Fenland and INTERVAL studies

Genome-wide association studies (GWAS) for glycine levels were conducted in the Fenland (N=9,324), EPIC-Norfolk (N=5,840) and INTERVAL (N=40,509) studies. For the Fenland study, separate GWAS were run for the three different genotyping arrays. The GWAS in EPIC-Norfolk only included sub-cohort A as at the time the GWAS were conducted metabolite measurements in sub-cohort B were not yet available. Glycine levels were natural log transformed, winsorized at 5 SDs and transformed to the Z score. Analyses were conducted based on generalised linear mixed models adjusted for age, sex and the first four principal components, using BOLT-LMM[203]. Genetic variants were excluded if the standard error (SE) >10, the absolute value of beta >5, p-value for the Hardy-Weinberg equilibrium $<1\times10^{-6}$ or info score < 0.3.

Meta-GWAS of glycine levels

Results of the Fenland, EPIC-Norfolk and INTERVAL studies were meta-analysed with publicly available summary-level GWAS results from KORA and TwinsUK[204] (downloaded from http://mips.helmholtz-muenchen.de/proj/GWAS/ gwas/index.php?task=download) and an NMR-based multi-cohort discovery[170] (downloaded from http://www.computationalmedicine.fi/data#NMR_GWAS), resulting in a total sample size of up to 80,003 individuals. Meta-analysis of the 5 studies was conducted using METAL[205] and was based

on the p-values, directions of effect and sample sizes, to minimize the effect of heterogeneity due to differences in analytical decisions and between metabolomics platforms. Variants were omitted if they were measured in less than half the total sample size or in fewer than 3 of the 5 studies, or if MAF<0.5%. Pooled effect sizes and SEs were generated through an effect size-based meta-analysis of the Fenland, EPIC-Norfolk (sub-cohort A only) and INTERVAL studies (total N=55,673) in METAL.

Identification of primary and secondary signals

Distance-based clumping using 1Mb windows was used to identify independent loci significantly associated with glycine (p-value<5×10^{-8}). Because of the very strong and wide signal at *CPS1*, a window of 3 Mb on both sides of the lead variant rs715 was used to capture the entire locus. Secondary signals were identified through approximate conditional analyses using GCTA-COJO[206]. To maximise the sample size, this analysis was conducted on the pooled Z scores generated in the p-value-based meta-GWAS. Variants with MAF<1% were omitted from the conditional analyses. HRC-imputed genome-wide data from 19,318 EPIC-Norfolk participants were used as an LD reference panel and a joint p-value threshold of 5×10^{-8} was used to identify secondary signals. Of the 73 variants selected through the approximate conditional analysis, 11 were in LD (R^2>0.05) with another selected variant and were therefore omitted. At the *CPS1* locus, 44 secondary signals remained of which 27 were low frequency variants. A second LD filter to remove variants in high LD with the common sentinel variant (D'<0.05) was applied, after which only 5 variants at the *CPS1* locus remained.

Sex-specific effect sizes of glycine loci

Sex-specific effect sizes for lead SNPs at the 27 loci on glycine levels standardised by sex were estimated in the INTERVAL, Fenland and EPIC-Norfolk studies (sub-cohorts A and B) (in total 30,226 men and 31,957 women) and meta-analysed using the R package 'metafor'[207]. As a sensitivity analysis, sex-specific effect sizes on natural-log transformed instead of within-sex standardised glycine levels were estimated in the Fenland and EPIC-Norfolk studies (9,927 men and 11,284 women).

Associations of glycine variants with T2D, CHD and related phenotypes

Associations of the genetic variants for glycine with T2D for both sexes combined were estimated in the InterAct study (4,712 female and 4,596 male cases; 7,190 female and 4,333 male controls) and the UK Biobank (7,301 female and 12,318 male cases; 181,442 female and 149,249 male controls), and extracted from published summary-level GWAS results by the DIAGRAM

consortium[37] (34,840 cases and 114,981 controls). Sex-combined and sex-specific GWAS for T2D in the InterAct study were run using SNPTEST, based on logistic regression models adjusted for age, sex, assessment centre and the first 4 genetic principal components calculated in PLINK[208]. In the UK Biobank, associations of the SNP dosages with prevalent and incident T2D, identified based on ICD10 codes, were estimated by fitting logistic regression models adjusted for age, sex, 4 genetic principal components and genotyping chip in STATA v15.0 (StataCorp, College Station, Texas, USA). Only UK Biobank participants from the subset of unrelated British ancestry participants were included. The sex-specific associations were estimated in the InterAct study and UK Biobank only, and meta-analyses of the effect sizes were conducted using the R package 'metafor'[207].

Sex-combined and sex-specific effect estimates of the glycine variants on CHD were assessed in the EPIC-CVD study and the UK Biobank, and obtained as lookups from CARDIoGRAMplusC4D[209] for the sex-combined associations (60,801 cases and 123,504 controls) and from the German MI family studies[209–212] for sex-specific associations (994 female and 2,804 male cases; 2,752 female and 2,554 male controls). The associations of the genetic variants for glycine with CHD were tested in EPIC-CVD based on Cox proportional hazards models and using Prentice weighting and robust standard errors (3,712 female and 5,786 male cases; 14,764 female and 13,453 male controls). Models were adjusted for age, genotyping array, testing centre, the first four genetic principle components and sex for the sex-combined analyses. In the UK Biobank, logistic regression models adjusted for age, the first 10 principal components, genotyping chip and sex (for sex-combined analyses) were run on 18,501 incident and prevalent CHD cases (5,147 women) and 333,545 controls (184,608 women) from the unrelated subset of participants of British ancestry, identified based on a combination of self-reported and genetically determined ancestry (UK Biobank data field 22006)[213].

The effects of the glycine variants on SBP and DBP were assessed using data from 203,943 male and 241,417 female European participants of the UK Biobank. Participants of European ancestry were identified based on K-means clustering of the first 4 principal components calculated from genome-wide SNP genotypes. Individuals who fell within the European cluster but self-identified as non-European were omitted from the genetically-defined European ancestry group. GWAS on rank-based inverse normally transformed SBP and DBP were conducted within sex and using BOLT-LMM. Analyses were adjusted for age, age^2, BMI and genotyping array. For variants both genotyped and imputed, imputed probabilities were used if the variant was imputed well (INFO>0.7) and the genotyping call rate was less than 98%. Sex-combined estimates were generated through fixed-effect meta-analyses of the sex-specific estimates.

Associations of the glycine variants with triglycerides, LDL, HDL and total cholesterol were obtained from publicly available GWAS summary-results based on up to 188,577 participants from the Global Lipids Genetics Consortium[214] (downloaded from http: //csg.sph.umich.edu/abecasis/public/lipids2013/). Look-ups for 13 blood cell traits came from publicly available GWAS summary results based on a genetic discovery in 173,480 participants[215] (downloaded from http://www.bloodcellgenetics.org/).

Mendelian randomisation methods

Mendelian randomisation (MR) analyses to assess the causal effect estimate for glycine on cardio-metabolic disease risk were based on four genetic scores for glycine with different levels of power versus specificity to glycine. MR analyses were based on summary-level data and four different methods: inverse variance-weighted MR[216], MR-Egger[58], weighted median MR and penalised weighted median MR[59]. Heterogeneity in the effects of the genetic variants on the outcome were assessed based on the Cochran's Q statistic, and directional pleiotropy was estimated based on the MR-Egger intercept. Priority was given to the results of weighted median MR if there was evidence for heterogeneity and/or directional pleiotropy (p-value for Cochran's Q <0.05 and/or p-value for Egger's intercept <0.05). The advantage of the weighted median MR method is that it can produce unbiased causal estimates, even if up to 50% of the information comes from invalid genetic instruments. The MR-Egger method was not used for the 2 SNP score.

Reverse MR analyses were conducted to assess the causality of risk factors for T2D on glycine levels using the same analytical approach as for the forward MR analyses. Previously published and validated genetic scores by the GIANT and MAGIC consortia were used for BMI[217] (97 genetic variants), fasting insulin adjusted for BMI as a marker of insulin resistance[95] (10 genetic variants) and insulin levels at 30 minutes during an oral glucose tolerance test as a marker for early-phase insulin secretion[95,218] (21 genetic variants). Effect sizes of the variants in the genetic scores on glycine levels were based on the effect size-based meta-analysis of the GWAS for glycine in INTERVAL, Fenland and EPIC-Norfolk.

Two-sample summary level data multivariable MR analyses were run to obtain the effect size of glycine levels adjusted for blood pressure on CHD as previously described by Day *et al.*[219]. In brief, weighted multilinear models were fitted with the effect sizes of the glycine SNPs on glycine and on the SBP and/or DBP as the explanatory variables and the effect sizes of the glycine SNPs on CHD as the independent variable. Weighting was based on the inverse variance of the effect sizes of the genetic variants for glycine on CHD.

Cox proportional hazards models to test the associations of observational glycine levels with CHD, MI and stroke

Cox proportional hazards models with robust standard errors were fitted to estimate the hazard ratios for T2D based on 1 SD increase in glycine levels. Age at recruitment was used as the underlying timescale, and the model was adjusted for sex, BMI, waist-hip ratio, educational attainment, smoking, alcohol consumption, physical activity, blood pressure and blood lipids. The models for T2D were run in the T2D case-sub-cohort (586 T2D cases and 746 random cohort participants); therefore prentice weighting was applied.

We tested if glycine levels at baseline were associated with incidence of CHD, MI and stroke in 11,966 participants of the EPIC-Norfolk study (sub-cohorts A and B). During follow-up, 2,053 participants developed CHD; 659 and 1,163 participants had a myocardial infarction and stroke, respectively. 212 stroke events were confirmed to be haemorrhagic whereas 569 were confirmed to be ischemic. Cox proportional hazards models were fitted to estimate the hazard ratios for CHD, myocardial infarction and stroke based on 1 SD increase in glycine levels. Age at recruitment was used as the underlying timescale, and models were adjusted for sex. Sex-specific analyses were conducted on within-sex standardised glycine levels. Analyses were performed in sub-cohorts A and B separately and meta-analysed using a fixed-effects model using the R package 'metafor'[207].

Pooling observational evidence for glycine to incident T2D

We meta-analysed risk ratio estimates for T2D based on 1 SD increase in glycine from the EPIC-Norfolk T2D case-cohort study with previously reported risk ratios estimated for KORA[137], EPIC-Potsdam[29], SABRE[166], IRAS[129] and a Japanese population-based cohort study[152]. We conducted a fixed and random-effects meta-analysis and estimated the heterogeneity between studies based on the I^2 statistic in Stata v14.0 using Metan (StataCorp, College Station, Texas, USA).

Results

Meta-GWAS identifies 22 novel loci for glycine levels

We conducted a meta-analysis of GWAS (meta-GWAS) for glycine levels in up to 80,003 participants, including 55,673 participants from the Fenland[173], EPIC-Norfolk[191] and INTERVAL[199] studies, and two publicly available summary-level GWAS datasets[169,170]. We identified 27 genetic loci for glycine levels at genome-wide significance level (p-value<5×10^{-8}), of which 22 have not previously been reported for glycine (Figure 2.1, Table 2.1). A total of 20 secondary signals at 8 loci were identified through approximate conditional association analyses (Supplementary Table 2.1).

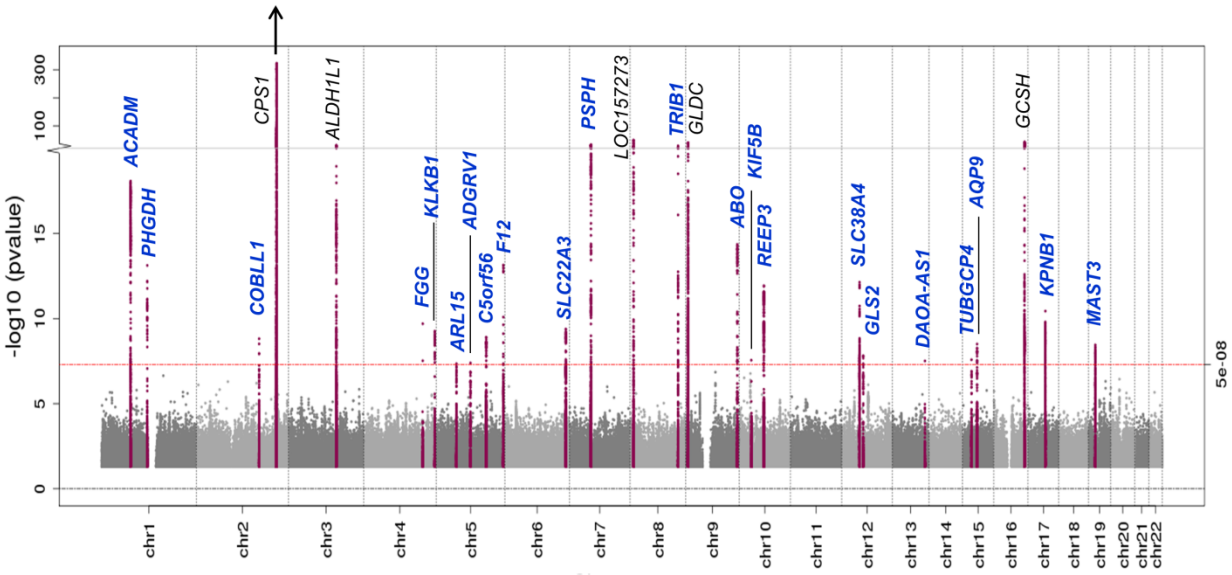


Figure 2.1: Manhattan plot of meta-GWAS for glycine levels. Loci in blue have not previously been reported for glycine. P-value for *CPS1* locus = 3×10^{-1632}

Six of the 27 loci are in (or near) genes encoding enzymes involved in glycine metabolism (Figure 2.2): glycine decarboxylase (*GLDC,* intronic single nucleotide polymorphism (SNP) rs17591030), glycine cleavage system protein H (*GCSH*, intergenic variant rs9923732 9 kb downstream of gene), phosphoserine phosphatase (*PSPH*, rs4947534 in 3' UTR region), phosphoglycerate dehydrogenase (*PHGDH*, rs561931 in 5' UTR region), carbamoyl phosphate synthase 1 (*CPS1*, rs715 in 3' UTR region) and aldehyde dehydrogenase 1 family member L1 (*ALDH1L1*, rs9862438, intronic variant in anti-sense non-coding RNA of *ALDH1L1*).

Table 2.1: 27 genetic loci reaching genome-wide significance for glycine levels in a p-value-based meta-analysis of GWAS in up to 80,003 participants. Chromosome (chr) and base pair position (pos) according to genome build GRCh37. [a] p-value-based meta-analysis of all 5 studies, [b]meta-analysis of the sex-combined betas and SEs from the INTERVAL, EPIC-Norfolk and Fenland studies, [c] meta-analysis of the sex-specific betas and SEs from the INTERVAL, EPIC-Norfolk and Fenland studies. † loci previously reported for glycine; * loci not reported for any metabolite. EA: effect allele, OA: other allele, EAF: effect allele frequency from p-value-based meta-analysis, N: sample size, SE: standard error.

Sentinel variant	Chr:pos	EA	OA	EAF	Variant type and nearest gene	P-value-based meta-analysis[a]			Meta-analyses of effect sizes and SEs					
									sex-comb[b]		men[c]		women[c]	
						p	N	HetPval	Beta	SE	Beta	SE	Beta	SE
rs4646961	1:76,217,169	A	G	0.297	intronic variant in *ACADM*	8.4E-19	79,999	0.62	0.048	0.006	0.045	0.007	0.050	0.009
rs561931	1:120,254,506	G	A	0.593	5' UTR variant of *PHGDH*	7.6E-14	79,998	0.65	0.033	0.006	0.033	0.007	0.028	0.009
rs10184004	2:165,508,389	T	C	0.400	Intergenic variant near *COBLL1* (28 kb) and *GRB14* (30 kb)	1.5E-09	80,002	0.20	0.036	0.006	0.020	0.007	0.056	0.009
rs715 †	2:211,543,055	C	T	0.313	3'UTR variant of *CPS1*	3E-1632	80,000	3.8E-32	0.444	0.006	0.233	0.007	0.691	0.009
rs9862438 †	3:12,591,0381	T	C	0.416	ncRNA intronic variant in *ALDH1L1-AS2*	1.1E-30	80,001	0.10	0.058	0.006	0.057	0.007	0.065	0.009
rs148685782	4:155,533,035	G	C	0.996	Synonymous variant in *FGG*	2.0E-10	55,673	0.02	0.309	0.049	0.170	0.056	0.362	0.072
rs71640034	4:187,161,048	A	G	0.511	intronic variant in *KLKB1*	5.6E-10	74,406	0.19	0.034	0.006	0.022	0.007	0.031	0.009
rs156380	5:53,378,450	C	T	0.807	intronic variant in *ARL15*	4.5E-08	79,998	0.74	0.031	0.007	0.020	0.009	0.036	0.011
rs3105793	5:90,226,061	A	G	0.273	intronic variant in *ADGRV1*	4.0E-08	79,998	0.20	0.028	0.006	0.024	0.008	0.034	0.010

Sentinel variant	Chr:pos	EA	OA	EAF	Variant type and nearest gene	P-value-based meta-analysis[a]			Meta-analyses of effect sizes and SEs					
									sex-comb[b]		men[c]		women[c]	
						p	N	HetPval	Beta	SE	Beta	SE	Beta	SE
rs10900807	5:131,757,480	G	C	0.805	ncRNA intronic variant in *C5orf56*	1.3E-09	79,999	0.81	0.036	0.007	0.029	0.009	0.041	0.011
rs2545801	5:176,841,339	C	T	0.747	intergenic variant near *F12* (5 kb) and *GRK6* (12 kb)	7.2E-14	74,404	0.56	0.042	0.007	0.015	0.008	0.054	0.010
rs543159	6:160,776,017	A	C	0.482	intronic variant in *SLC22A3*	4.2E-10	74,405	0.39	0.035	0.006	0.027	0.007	0.035	0.009
rs4947534	7:56,079,094	C	T	0.760	3' UTR variant of *PSPH*	7.1E-34	79,998	0.08	0.072	0.007	0.061	0.008	0.085	0.010
rs9987289	8:9,183,358	A	G	0.100	ncRNA intronic variant in *LOC157273*	1.7E-49	79,999	0.07	0.124	0.01	0.118	0.012	0.124	0.015
rs28601761	8:126,500,031	G	C	0.416	intergenic variant near *TRIB1* (49kb) and *LINC00861* (435 kb)	8.5E-30	73,832	0.13	0.063	0.006	0.080	0.007	0.060	0.009
rs17591030 †	9:6,550,024	C	T	0.715	intron variant in *GLDC*	1.9E-40	80,000	0.03	0.08	0.006	0.055	0.008	0.118	0.010
rs676996	9:136,146,077	T	G	0.668	intron variant in *ABO*	4.4E-15	74,403	0.32	0.04	0.006	0.034	0.007	0.033	0.009
rs190595610 *	10:32,274,880	A	G	0.997	Intergenic variant near *ARHGAP12* (57 kb) and *KIF5B* (23 kb)	9.0E-09	70,912	0.11	0.253	0.056	0.132	0.066	0.279	0.084
rs10740134	10:65,315,433	T	C	0.515	intron variant in *REEP3*	1.2E-12	80,001	0.06	0.038	0.006	0.016	0.007	0.044	0.009
rs12297321	12:47,109,387	T	C	0.152	Intergenic variant near *SLC38A4* (38 kb) and *LOC100288798* (630 kb)	7.4E-13	79,999	0.67	0.048	0.008	0.036	0.009	0.058	0.012
rs2638314	12:56,866,334	A	T	0.182	intronic variant in *GLS2*	1.5E-08	74,402	0.23	0.042	0.007	0.040	0.009	0.048	0.011

Sentinel variant	Chr:pos	EA	OA	EAF	Variant type and nearest gene	P-value-based meta-analysis[a]			Meta-analyses of effect sizes and SEs					
									sex-comb[b]		men[c]		women[c]	
						p	N	HetPval	Beta	SE	Beta	SE	Beta	SE
rs9514191	13:104,520,138	C	G	0.312	intergenic variant near *LINC01309* (440 kb) and *DAOA-AS1*(159 kb)	3.1E-08	74,402	0.58	0.034	0.006	0.020	0.007	0.040	0.009
rs201393666 *	15:43,677,979	A	C	0.029	intronic variant in *TUBGCP4*	2.6E-08	54,724	0.43	0.097	0.017	0.105	0.020	0.049	0.026
rs2280195	15:58,467,095	A	G	0.441	intronic variant in *AQP9*	3.2E-09	80,000	0.72	0.028	0.006	0.032	0.007	0.015	0.009
rs9923732 †	16:81,110,903	A	G	0.914	Upstream variant of *C16orf46,* 9 kb downstream of *GCSH*	1.2E-41	80,001	0.58	0.119	0.011	0.127	0.013	0.110	0.016
rs8078686	17:45,735,706	C	T	0.509	intron variant in *KPNB1*	3.7E-11	80,001	0.78	0.035	0.006	0.036	0.007	0.021	0.009
rs273510	19:18,223,350	A	G	0.708	intron variant in *MAST3*	3.6E-09	74,403	0.90	0.034	0.006	0.026	0.007	0.040	0.010

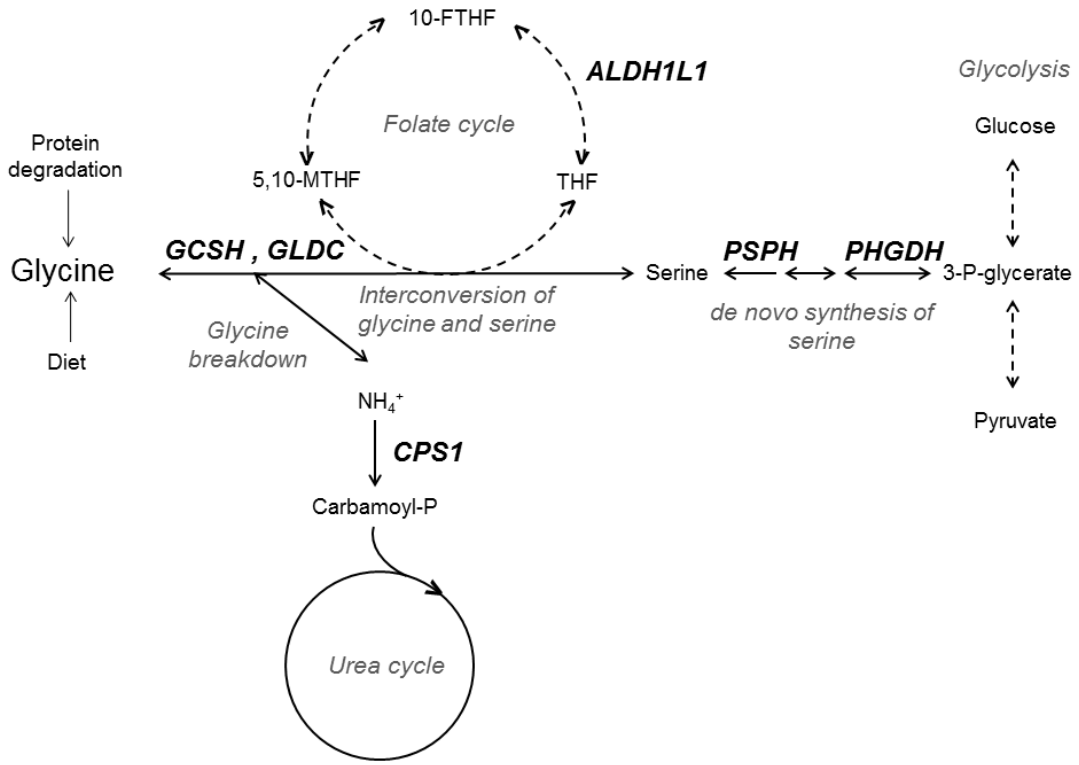


Figure 2.2: Schematic overview of glycine metabolism and 6 glycine loci in or near genes encoding enzymes related to glycine metabolism. *GLDC* and *GCSH* are part of the glycine cleavage system, the major enzyme complex responsible for glycine degradation. *PHGDH* and *PSPH* encode enzymes involved in de novo biosynthesis of serine, which can be converted into and synthesised from glycine. The interconversion of glycine and serine serves as an important source of methyl groups in the folate cycle, a metabolic pathway of which the enzyme encoded by *ALDH1L1* is a component. CPS1 encodes the enzyme responsible for the rate-limiting step of the urea cycle which is responsible for the detoxification of ammonia. Glycine breakdown produces ammonia; changes in the efficiency of the urea cycle are therefore likely to have upstream effects on glycine metabolism. *GCSH*: glycine cleavage system protein H, *GLDC*: glycine decarboxylase, *CPS1*: carbamoyl-phosphate synthase 1, *ALDH1L1*: aldehyde dehydrogenase 1 family member L1, *PSPH*: phosphoserine phosphatase, *PHGDH*: phosphoglycerate dehydrogenase. *THF*: tetrahydrofolate, *5,10-MTHF*: N5-N10-methylene-tetrahydrofolate, *10-FTHF*: 10-formyl-tetrahydrofolate.

Nine of the 27 loci (*COBLL1-GRB14*, *CPS1*, *SLC22A3*, *LOC157273*, *TRIB1*, *ABO*, *REEP3* and *KPNB1*) have previously been reported for blood lipid fractions[214] (low and high-density lipoprotein (LDL and HDL)-cholesterol, total cholesterol and total triglycerides), but the direction of association with lipid traits relative to the glycine-raising allele differed between loci (Supplementary Table 2.2). Six loci are known to be associated with coagulation-related traits, such as plasma fibrinogen levels[220–222] (*CPS1, FGG, C5orf56*) and activated partial thromboplastin time[223–225] (*KLKB1*, *F12*, *ABO*). Glycine-raising alleles at *KLKB1*, *C5orf5* and *F12* were associated with higher coagulation while the glycine-raising alleles at *CPS1*, *FGG* and *ABO* were associated with lower coagulation. Three loci are established loci for T2D or glycaemic traits; the signal at *LOC157273* has been reported for higher fasting glucose and insulin[93], *GLS2* for higher fasting glucose[94] and *COBLL1-GRB14* has been reported for lower risk of T2D[37] and lower fasting insulin

levels[94]. Finally, *TRIB1* is an established CHD risk-raising locus[226] (Supplementary Table 2.2). Apart from *CPS1*, none of the lipid, coagulation or cardio-metabolic disease-associated loci have a known role in glycine metabolism and effect sizes of these loci on glycine levels were generally small compared to the effects of loci near genes encoding enzymes of glycine metabolism (Figure 2.6A).

Per-allele effect sizes of the common loci (minor allele frequency (MAF) > 5%) were between 0.03 and 0.12 standard deviations (SDs) of glycine levels, with the exception of the association near *CPS1*, of which each allele of the lead variant rs715 was associated with 0.444±0.006 higher SDs of glycine. This variant explained 13.7% of the variance in glycine and is in high linkage disequilibrium (LD) (r^2=0.904) with a missense variant (rs1047891, 4217C>A, Thr1406Asn, 2.5 kb upstream of rs715). Conditioning on rs1047891 in EPIC-Norfolk showed that the effect of rs715 was entirely driven by rs1047891 (beta±SE on glycine before adjustment = 0.565±0.013, p-value<2×10^{-302}; after adjustment for rs1047891: beta±SE=0.032±0.049, p-value=0.52).

Previous studies highlighted sex differences in the effect size of the *CPS1* locus on glycine levels[189]; we therefore compared sex-specific effect sizes across all loci on glycine levels standardised by sex (Bonferroni-corrected p-value threshold for sex difference <0.002) (Figure 2.3A, Table 2.1). We confirmed that the effect of the *CPS1* locus is nearly three times stronger in women than in men (women: beta±SE=0.691±0.009, men: beta±SE=0.233±0.007, p-value for sex difference <2×10^{-302}) and found similar significant sex differences in effect size for rs17591030 (*GLDC*, women: beta±SE=0.118±0.010, men: beta±SE=0.055±0.008, p-value for sex difference=3.3×10^{-7}) and rs10184004 (*TRIB1*, women: beta±SE=0.060±0.009, men: beta±SE=0.080±0.007, p-value=0.001). Suggestively stronger effect sizes in women were observed for rs148685782 (*FGG*, p-value=0.04) and rs2545801 (*F12*, p-value=0.0023), while a suggestively stronger effect in men was found at rs10740134 (*REEP3*, p-value=0.01) (Figure 2.3A, Table 2.1). As the distribution of glycine levels itself is sex-specific (women: mean±SD=1.137±0.321, men: mean±SD=0.966±0.195 (arbitrary units); p-value for sex difference<2.2×10^{-16}), we assessed differences in effect sizes on log-transformed in place of within-sex standardised glycine levels as a sensitivity analysis, which gave similar results (Supplementary Table 2.3). Due to larger effect size of rs715 in women, the cumulative variance in glycine levels explained by the 27 variants was nearly 2.5 times higher in women (25.1%) than in men (10.6%) (Figure 2.3B). The sex difference in effect sizes of rs715 was also observed for 65 other metabolites significantly associated with rs715 in the EPIC-Norfolk study (Supplementary Figure 2.1).

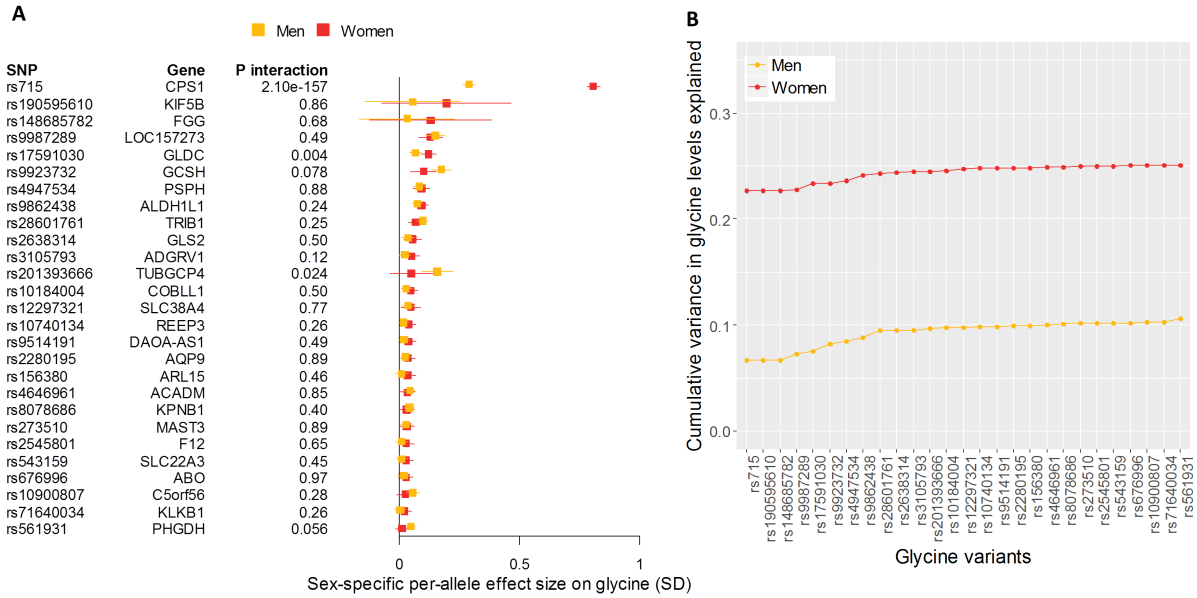


Figure 2.3: Sex-specific effect sizes and cumulative variance explained by lead SNPs at the 27 genetic loci associated with glycine. A| Sex-specific effect sizes of lead SNPs at the significant loci for glycine levels, based on a meta-analysis of the Fenland, EPIC-Norfolk and INTERVAL studies, including 30,226 men and 31,957 women. **B|** Cumulative variance in glycine levels explained by lead SNPs at all 27 loci in men (N=5,086) and women (N=5,706) of the EPIC-Norfolk study.

Genetic scores predict glycine levels: trade-off between power and specificity

We generated a genetic score comprising lead SNPs at the loci significantly associated with glycine levels and tested its specificity for glycine metabolism. Three variants — rs201393666 in *TUBGCP4*, rs148685782 in *FGG* and rs190595610 near *KIF5B* — were excluded from the score as look-ups for these were not available in all summary-level GWAS datasets for CHD. The effect size-weighted genetic score including the 24 remaining loci was robustly associated with glycine levels, explaining 15.6% of the total variance after adjustment for age, sex and batch effects, and 10.0% and 24.9% in men and women, respectively.

The 24 loci were within 1 Mb of loci previously reported for other metabolites, for 9 of which the lead variant we identified for glycine was not in high LD ($r^2 \geq 0.8$) with the reported lead variant(s) (Supplementary Table 2.2). We therefore tested if the glycine score affected metabolic pathways unrelated to glycine metabolism. Among the 894 metabolites measured in at least half of the sample size of the EPIC-Norfolk study (Bonferroni-corrected p-value threshold $<5.6\times10^{-5}$), the glycine score was most strongly associated with glycine but also with 70 other metabolites, of which the majority were either strongly correlated with glycine levels (several glycine-conjugated metabolites) or metabolically related to glycine, including urea cycle and choline metabolites, serine, creatine, and glycine-conjugated fatty acids. The glycine score was also associated with 33

metabolites not directly related to glycine, including unknown metabolites, phospholipids and acyl carnitines (Supplementary Table 2.4).

We constructed three additional genetic scores for glycine with decreasing numbers of loci but increasing specificity to glycine metabolism. The 'glycine-related' score was comprised of the 6 loci near genes encoding enzymes related to glycine metabolism (Figure 2.2), explained 14.9% of the glycine variance and remained associated with 66 metabolites (Supplementary Table 2.4). The *CPS1* locus by itself was significantly associated with all but four of these 66 metabolites; we therefore generated a third score that excluded the pleiotropic *CPS1* locus. This score explained 1.2% of the glycine variance and was associated with 10 metabolites, of which 7 were linked to glycine (γ-glutamylglycine, N-acetylglycine, propionylglycine, N-palmitoylglycine, isovalerylglycine, serine and cinnamoylglycine) and three were unknown metabolites. Finally, we constructed a 2 SNP score, based on the two loci at *GCSH* and *GLDC* encoding enzymes of the glycine cleavage system. The 2 SNP score explained 0.6% of the variance in glycine and was associated with γ-glutamylglycine, N-isovalerylglycine, 3-methylglutaconate and its carnitine conjugate, and one unknown metabolite.

Glycine levels are inversely associated with incidence of T2D

The strength of association between standard deviations (SDs) of serum levels of glycine and incidence of T2D was available in 5 of the 42 longitudinal studies identified in the systematic literature review on metabolomics studies for T2D, of which one reported the association for Europeans and South-Asians separately. In the 4 cohorts including participants of European ancestry, glycine levels were strongly and inversely associated with incident T2D, but this association was not significant in the Japanese[152] and South-Asian cohorts[166]. The lack of association in the latter could have been due to the adjustment for HOMA-IR, which was not included as a covariate in the other studies. We investigated the association between glycine and incident T2D in the EPIC-Norfolk study (N=1,332 of which 586 incident T2D cases) and found a 42% lower incidence of T2D per SD higher glycine levels (hazard ratio (HR) [95% CI] for T2D per SD increase in glycine = 0.58 [0.47, 0.70]). We next pooled effect estimates of circulating glycine levels on incident T2D of seven cohorts and found a significant inverse association between glycine levels and disease incidence (fixed-effect pooled OR per 1 SD higher glycine =0.78 [0.72, 0.84]; random effect estimate [95% CI] = 0.77 [0.66, 0.90], based on 2,067 T2D cases), although there was substantial heterogeneity between the studies (I^2=28.0%, p=0.001). As a sensitivity analysis, we excluded non-European cohorts, which led to a stronger association of glycine with incident T2D and non-significant heterogeneity (I^2=73.9%, p=0.235, OR=0.70 [0.64,0.77], 1,840 T2D cases) (Figure 2.4).

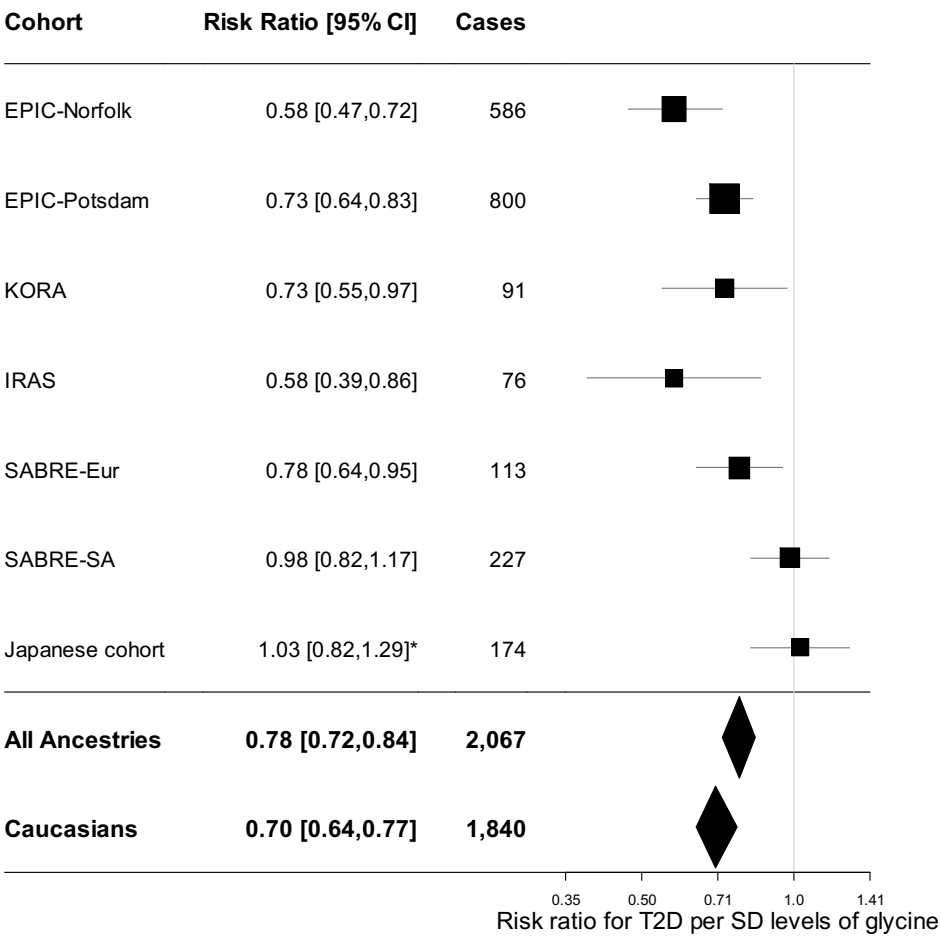


Figure 2.4: Forest plot of pooled and study-level observational risk ratio estimates for T2D per 1 SD higher levels of glycine based on 7 studies. Eur: Europeans, SA: South-Asians. CI: confidence intervals. SD: standard deviation, * Adjusted for HOMA-IR.

The association of genetically predicted glycine levels with T2D risk is pathway-specific

Based on Mendelian randomisation analyses including up to 63,767 T2D cases and 457,195 controls and the full genetic score for glycine comprised of 24 loci, genetically predicted glycine levels were not associated with T2D risk (OR [95% CI] for T2D per genetically predicted SD higher levels of glycine = 1.00 [0.97, 1.04], p= 0.84) (Figure 2.5 and 2.6A, Supplementary Table 2.5). Sex-specific analyses in up to 12,013 female and 16,914 male T2D cases indicated that genetically predicted glycine levels were also not associated with T2D risk in men or women only (OR [95% CI] in women = 1.01 [0.94, 1.09]; OR [95% CI] in men = 1.04 [0.98, 1.10], p-value for sex difference=0.54) (Figure 2.5, Supplementary Figures 2.2A and 2.3A for sex-specific dosage plots, Supplementary Table 2.5 for full results). As there was evidence for moderate heterogeneity as well as directional pleiotropy (Sex-combined analyses: Cochran's Q p-value=0.02; p-value for

Egger's intercept =0.03), the median weighted method was used for the main analyses. Causal effect estimates were very similar across all four MR methods and when using the 6 SNP score, which only included loci in/near loci encoding enzymes related to glycine metabolism (Figure 2.6B, Supplementary Figure 2.2B, and 2.3B, Supplementary Table 2.5). No evidence was found for heterogeneity nor for directional pleiotropy when using the 6 SNP score.

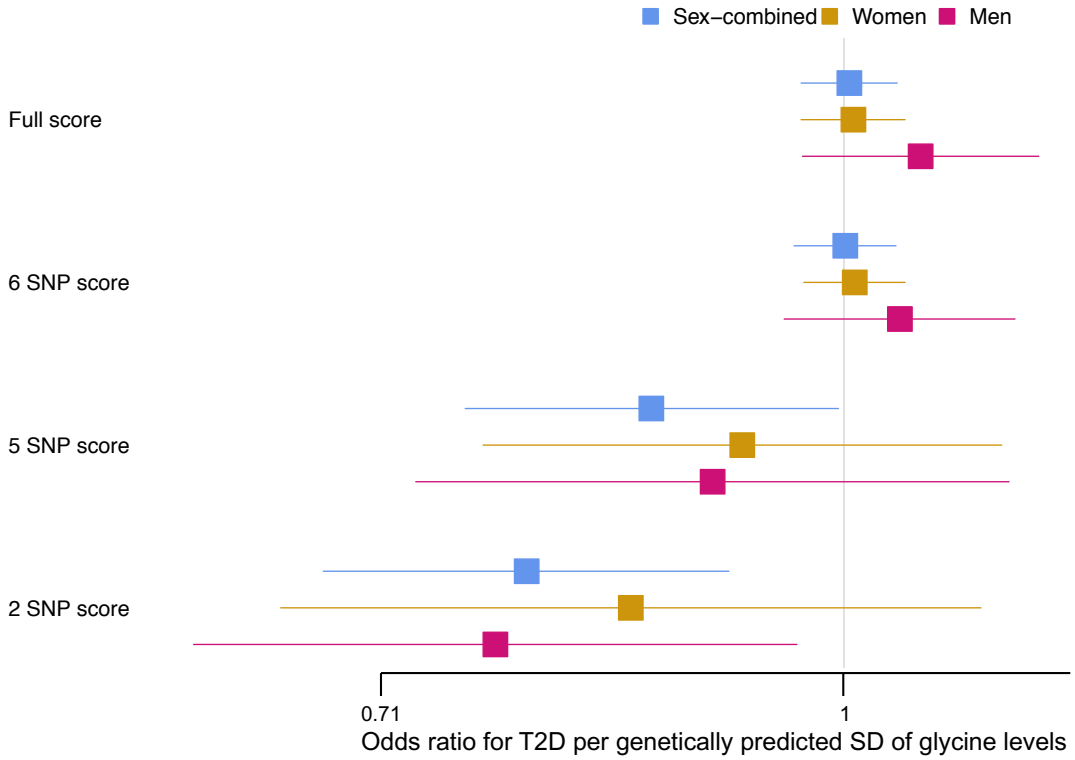


Figure 2.5: Forest plot of the odds ratios ± 95% confidence intervals for type 2 diabetes per standard deviation genetically predicted higher levels of glycine using 4 different genetic instruments for glycine. Genetic estimates are based on 63,767 T2D cases and 457,195 controls from InterAct, UK Biobank and DIAGRAM for sex-combined analyses. Sex-specific genetic analyses included 12,013 female and 16,914 male T2D cases and 202,124 female and 164,944 male controls from InterAct and UK Biobank. For the sex-specific analyses, the standard deviations of the sex-specific glycine distributions were used, which were 0.321 and 0.195 (arbitrary units) for women and men in the EPIC-Norfolk study, respectively. The full score was comprised of all 24 common loci for glycine, the 6 SNP score only included loci in/near genes encoding enzymes related to glycine metabolism, the 5 SNP score excluded the *CPS1* locus and the 2 SNP score only included loci near *GLDC* and *GCSH*.

However, after removing the strong but non-specific *CPS1* locus from the 6 SNP score for glycine, the association of genetically predicted glycine with T2D reached significance in the sex-combined analyses (sex-combined: OR=0.868 [0.757,0.996], p=0.044; women: OR=0.928 [0.767,1.123], p=0.45; men: OR=0.908 [0.730,1.129], p=0.39) (Figures 2.5 and 2.6C, Supplementary Figures 2.2C and 2.3C). When using the 2 SNP score which only included the two loci near *GCSH* and *GLDC* the effect size of genetically predicted glycine on T2D risk became stronger (sex-combined:

OR=0.792 [0.682,0.919], p=0.002) (Figure 2.5 and 2.6D) and no heterogeneity by sex was found (women: OR=0.855 [0.661,1.106], p=0.23; men: OR=0.774 [0.620,0.966][, p=0.024; p for heterogeneity by sex=0.57) (Supplementary Figures 2.2D and 2.3D).

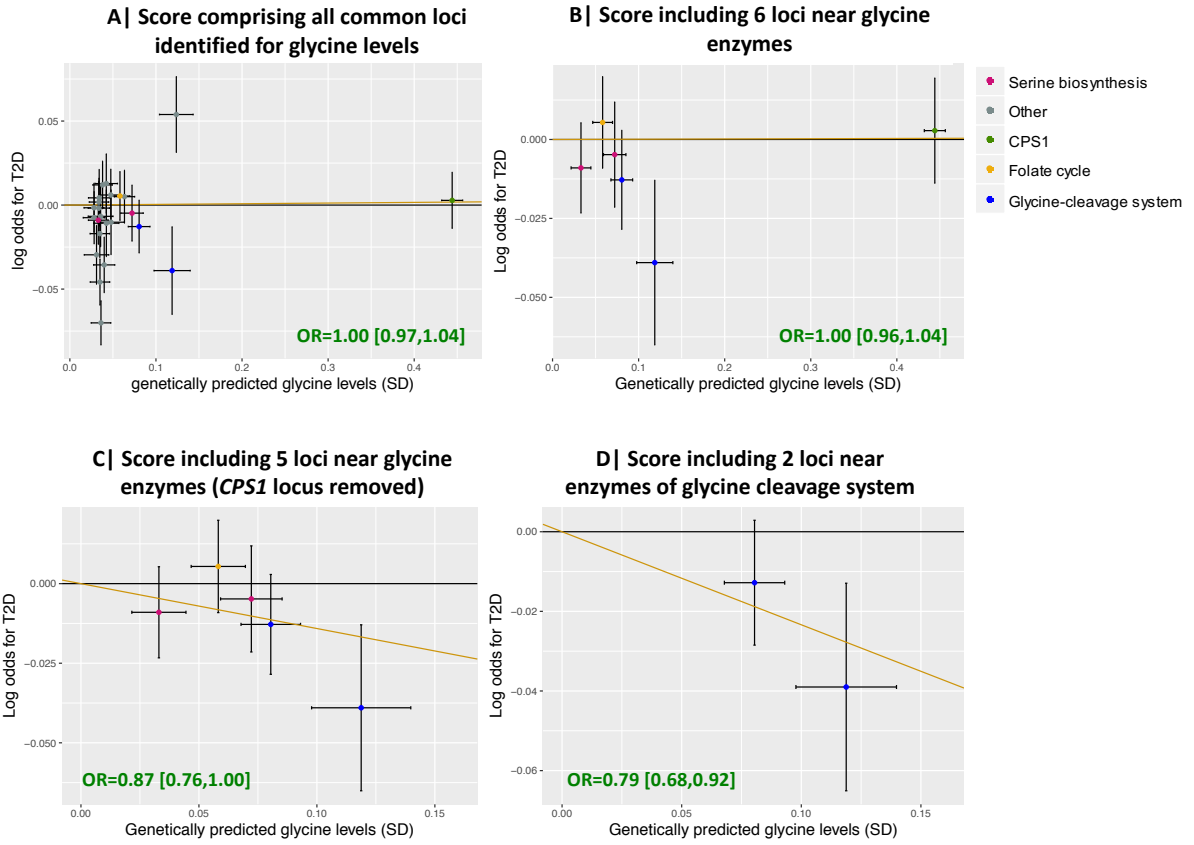


Figure 2.6: Dosage plots of the effect sizes of genetic variants for glycine on standard deviations of glycine levels versus the log odds for type 2 diabetes for sex-combined analyses. A| For the full glycine score, **B|** the score comprising 6 loci in/near enzymes related to glycine metabolism, **C|** a score comprising 5 glycine-related loci without the CPS1 locus and, **D|** a 2 SNP score including *GCSH* and *GLDC* genes encoding enzymes part of the glycine cleavage system. The orange line represents the slope estimated using the weighted median method. SD: standard deviation, T2D: type 2 diabetes.

Genetically higher fasting insulin is inversely associated with genetically predicted glycine levels

To investigate if the association between low levels of glycine and high incidence of T2D could be a consequence of early disease processes of T2D, we assessed the causality of three risk factors for T2D – elevated body mass index (BMI), reduced early-phase insulin secretion (IS) and insulin resistance (IR) – on circulating levels of glycine. We found that genetically higher IR was associated with genetically predicted lower levels of glycine (beta [95%CI] on SDs of glycine levels per genetically predicted unit of fasting insulin levels = -0.960 [-1.376, -0.544], p =5.98×10^{-6}), while genetically predicted BMI (beta per SD genetically predicted BMI on SDs of glycine levels =-0.04

[-0.10,0.02], p=0.169) and IS (beta per genetically predicted units of insulin at 30 minutes during an oral glucose tolerance test on SDs of glycine levels =0.02 [-0.04,0.09], p=0.52) were not associated with glycine levels (Figure 2.7). No evidence for heterogeneity or directional pleiotropy was found. The inverse-variance weighted MR analyses were therefore used as the main analyses, but similar effect sizes were found using the MR-EGGER, median weighted and penalized median weighted MR (Supplementary Table 2.6).

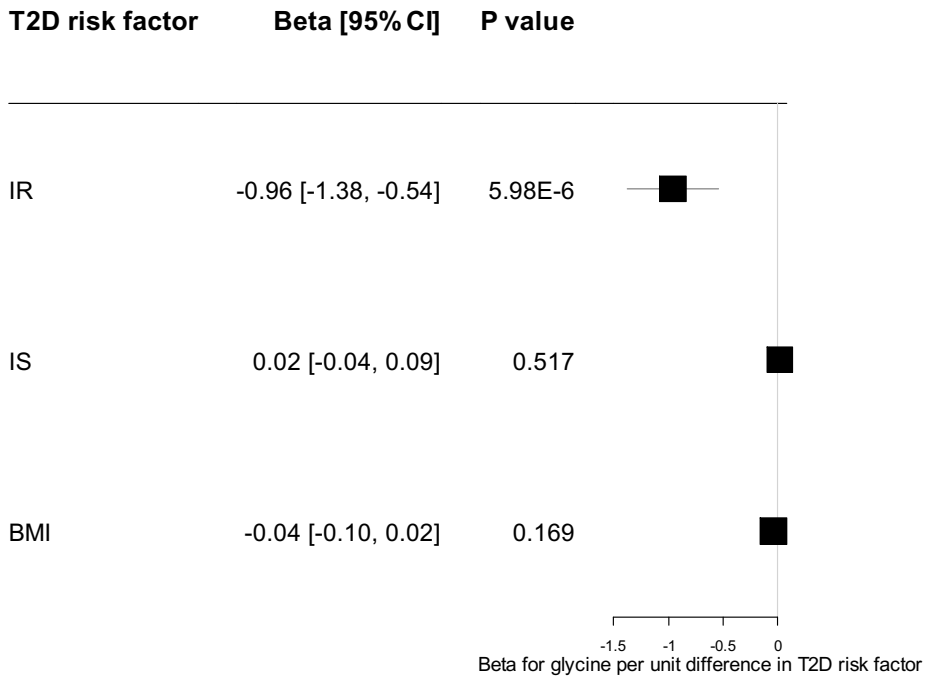


Figure 2.7: Forest plot of the genetic associations of insulin resistance (IR), early-phase insulin secretion (IS) and body mass index (BMI) with glycine levels. Analyses were based on the inverse variance-weighted MR method and previously identified genetic scores for BMI (96 SNPs, Locke et al.[217]) , fasting insulin adjusted for BMI (10 SNPs, Scott et al.[95]) and 30-minute insulin during an oral glucose tolerance test (21 SNPs, Scott et al.[95]).

High glycine levels are genetically and observationally associated with lower incidence of CHD in both women and men

Based on Mendelian randomisation analyses in up to 88,800 CHD cases and 485,266 controls and using the genetic score for glycine comprised of 24 variants, we estimated that each SD increase in genetically predicted glycine levels was associated with a 5.2% reduction in risk of CHD (odds ratio (OR) [95% confidence intervals (CI)] =0.948 [0.918, 0.979], p-value=0.001) (Figure 2.8 and 2.9A, Supplementary Table 2.7). Sex-specific analyses in up to 9,852 female and 21,994 male CHD cases showed that the effect sizes of glycine on CHD risk were similar between women and men (OR [95% CI] = 0.95 [0.91,1.00] and 0.93 [0.84,1.02], respectively; p-value for sex difference=0.60) (Figure 2.8 and Supplementary Figures 2.4A and 2.5A). After taking into account the sex-specific effect on glycine, the genetic glycine-CHD association based on *CPS1* only did also not differ by

sex (women: OR [95% CI] =0.96 [0.91,1.00], men: OR [95% CI] =0.94 [0.85,1.03], p-value for sex difference=0.74).

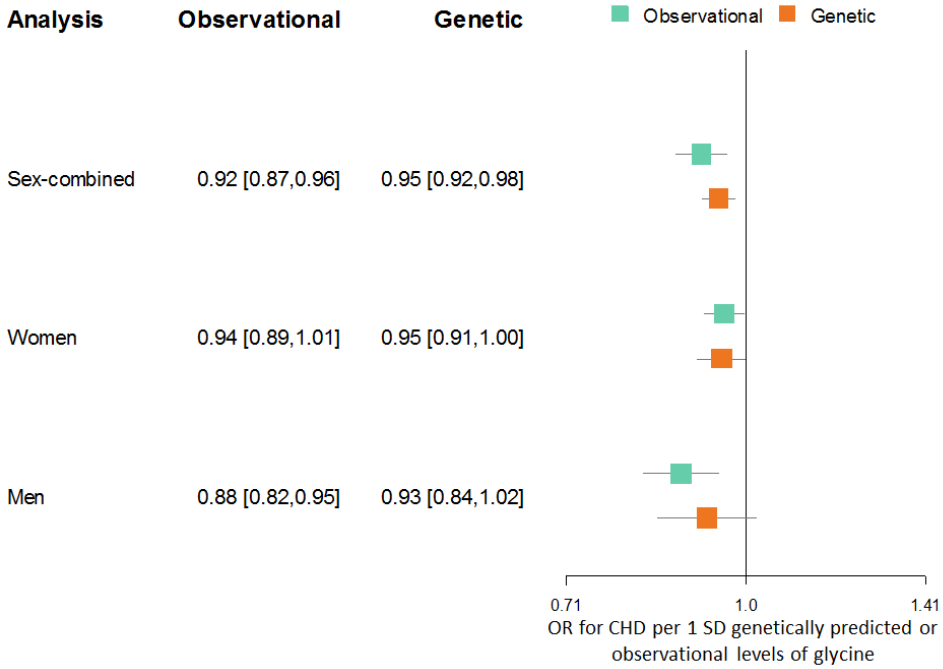


Figure 2.8: Forest plot of the odds ratios for CHD per standard deviation observationally or genetically predicted higher levels of glycine. Genetic estimates are based on 88,800 CHD cases and 485,266 controls from EPIC-CVD, UK Biobank and CARDIoGRAMplusC4D for sex-combined analyses. Sex-specific genetic analyses included 9,852 female and 21,994 male CHD cases and 202,124 female and 164,944 male controls for from UK Biobank, EPIC-CVD and the German Family MI studies. For the sex-specific analyses, sex-specific standard deviations were used, which were 0.321 and 0.195 (arbitrary units) for women and men in EPIC-Norfolk, respectively. Observational analyses were based on 11,147 participants (4,989 men and 6,158 women) from the EPIC-Norfolk study, of which 2,053 (1,223 men and 830 women) were incident CHD cases.

Visual inspection of the scatter plots (Figure 2.9), statistical tests for heterogeneity and pleiotropy (Egger's intercept p-value=0.003; men-only: Cochran's Q p-value=0.09) (Supplementary Table 2.6) and the presence of several variants within the score known to be associated with other traits, suggest that the effect sizes of the 24 glycine-raising genetic variants on CHD risk may be heterogeneous. To reduce the influence of potential pleiotropic variants on the causal estimate, we used the weighted median method for the main analyses. To verify if the inverse association of genetically predicted glycine with genetic risk of CHD was driven by metabolites unrelated to glycine metabolism or other pleiotropic mechanisms, we conducted a series of sensitivity analyses using the three genetic scores that are subsets of the full score with increasing specificity to glycine metabolism. When using the glycine-related score which included the 6 loci with a known biological link to glycine, similar effect estimates for glycine on CHD risk were found and the heterogeneity decreased (Egger's intercept p-value=0.44; men-only: Cochran's Q p-value=0.80)

(Figure 2.9B, Supplementary Figures 2.4B and 2.5B, Supplementary Table 2.7). Removing the strong but non-specific *CPS1* locus from the glycine-related score increased the causal effect estimate of glycine on CHD risk (sex-combined: OR=0.80 [0.71,0.91], p-value=4.8×10^{-4}) (Figure 2.9C), and a similar effect estimate was obtained when using the 2 SNP score (*GCSH* and *GLDC*) (sex-combined: OR=0.80 [0.69,0.92], p-value=1.7×10^{-3}) (Figure 2.9D, Supplementary Table 2.7, Supplementary Figures 2.4 C-D and 2.5 C-D).

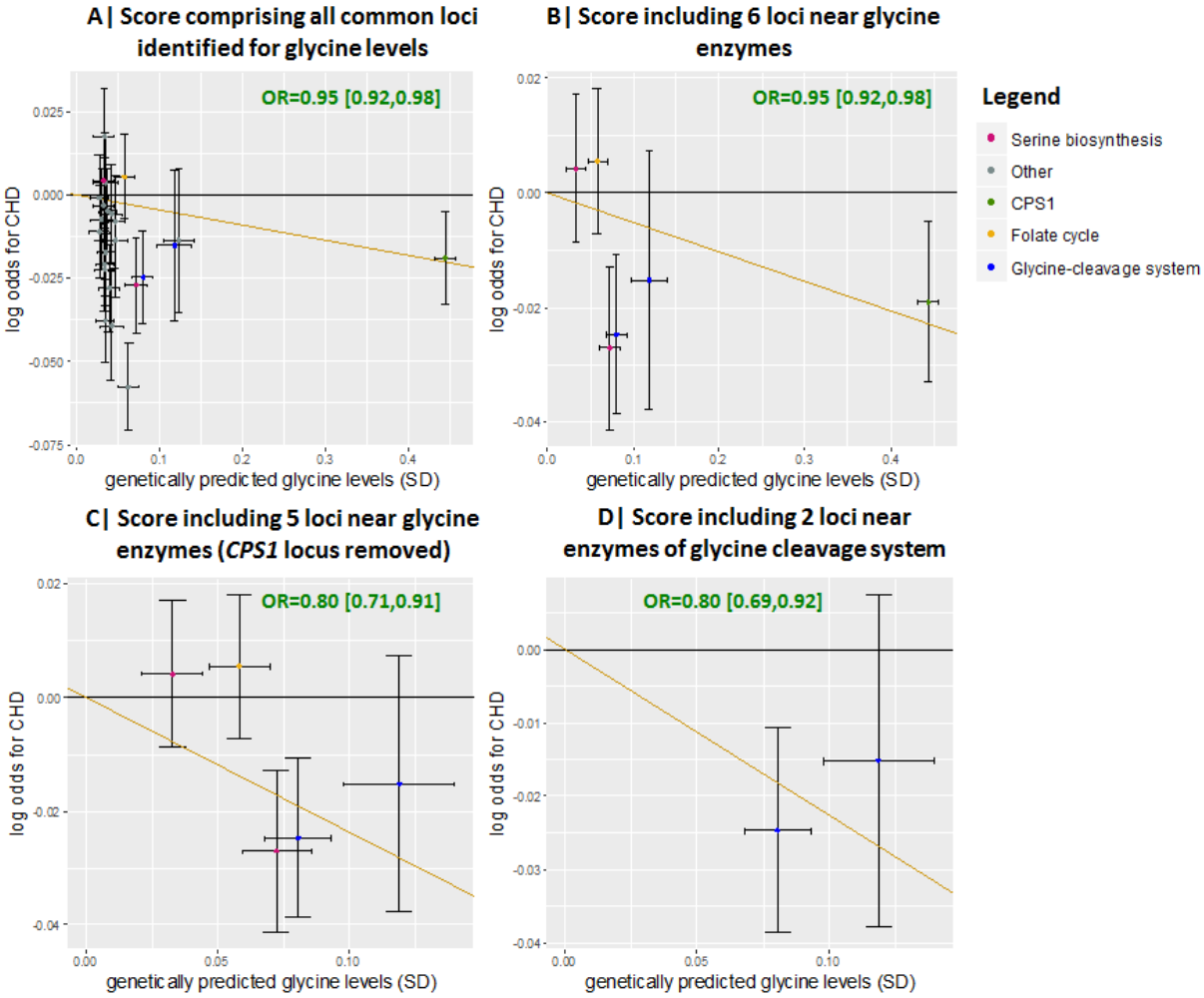


Figure 2.9: Dosage plots of the effect sizes of genetic variants for glycine on standard deviations of glycine levels versus the log odds for coronary heart disease. For the full glycine score (**A**), the score comprising 6 loci in/near enzymes related to glycine metabolism (**B**), a score comprising 5 glycine-related loci without the CPS1 locus (**C**) and a 2 SNP score including *GCSH* and *GLDC*- genes encoding enzymes part of the glycine cleavage system (**D**). The orange line represents the slope estimated using the weighted median method. SD: standard deviation, CHD: coronary heart disease.

In a non-genetic analysis of the EPIC-Norfolk study using Cox proportional hazards models, higher glycine levels were associated with lower incidence of CHD (hazard ratio [95% CI] for CHD per 1 SD higher glycine = 0.92 [0.87,0.96], p-value=4.7×10^{-4}). This observational association was similar to the genetically predicted association (Figure 2.8). Similar observational associations were found with myocardial infarction (HR=0.89 [0.82,0.97], p-value=5.5×10^{-3}), but glycine was

not associated with stroke (HR= 0.99 [0.93,1.05], p-value=0.70) or stroke sub-types (Supplementary Table 2.8). Associations of glycine with both CHD and myocardial infarction (MI) tended to be stronger in men than in women, but sex differences were not statistically significant (CHD: p-value for sex difference=0.22, MI: p-value=0.10) (Figure 2.8). Using log-transformed in place of within-sex standardised glycine measures as the exposure gave similar results (Supplementary Table 2.8).

Characterising the biological pathways mediating the protective effect of glycine on CHD

To explore the biological mechanisms through which glycine may influence risk of CHD, we investigated the downstream effects of genetic differences in glycine levels on CHD risk factors and assessed the extent to which the inverse genetic association of glycine levels with CHD risk may be driven by known aetiological mechanisms of CHD.

Using novel and published GWAS summary results, we assessed associations of genetically predicted glycine levels with systolic (SBP) and diastolic blood pressure (DBP), blood lipid fractions (triglycerides, HDL, LDL and total cholesterol) and 13 potentially relevant blood cell traits (Bonferroni-corrected significance threshold for 19 tests: p-value$<2.6\times10^{-3}$). Genetically predicted glycine levels based on the full genetic score were significantly associated with lower genetically predicted SBP (beta±SE per SD glycine on SD of SBP=-0.028±0.007, p-value=1.5×10^{-5}) and nominally with lower genetically predicted DBP (beta±SE=-0.019±0.009, p-value=0.039). Similar effect sizes were found for men and women separately (Figure 2.10, Supplementary Table 2.9). Effect estimates using the glycine-related score were similar as when using the full score. When using the glycine-related score without the *CPS1* locus or the 2 SNP score, effect sizes tended to increase (Figure 2.10, Supplementary Table 2.9).

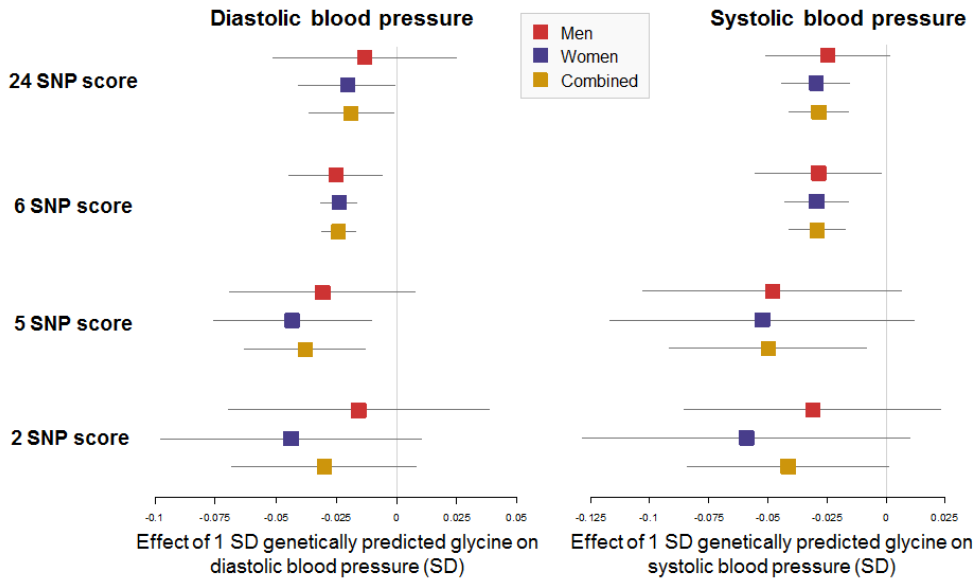


Figure 2.10: **Genetically predicted effect of glycine on diastolic and systolic blood pressure, in sex-combined and sex-specific analyses, using 4 different genetic scores for glycine.** The full score includes all 24 common loci associated with glycine, the 6 SNP score only includes loci in/near genes encoding enzymes related to glycine metabolism, the 5 SNP score excludes the CPS1 locus, and the 2 SNP score only includes loci near genes encoding enzymes that are part of the glycine cleavage system. Associations of the genetic scores for glycine with blood pressure were based on 203,943 male and 241,417 female white European participants of the UK Biobank.

Genetically predicted glycine levels were not associated with genetically predicted blood lipids and blood cell traits. Associations with lower HDL cholesterol (beta±SE=-0.056±0.010, p-value=7.4×10^{-9}) and 5 blood cell traits (lower platelet, neutrophil and lymphocyte count; higher platelet distribution width and mean corpuscular volume) reached significance when using the full genetic score, but the effect sizes drastically decreased when using the 5 or 2 SNP scores, suggesting that the associations were driven by *CPS1* only (Supplementary Table 2.9, Supplementary Figure 2.6).

To test if the genetic glycine-to-CHD association was mediated through lowering blood pressure, we adjusted the effect of genetically predicted glycine on CHD for genetically predicted SBP and DBP in a multivariable inverse variance-weighted Mendelian randomisation analysis. We used the 5 SNP score for glycine levels to reduce the likelihood of glycine-unrelated mechanisms. Adjusting for SBP and DBP separately and together progressively reduced the effect estimate of genetically predicted glycine on CHD risk (Figure 2.11). This suggests that the genetic association of glycine with CHD risk is mediated through blood pressure.

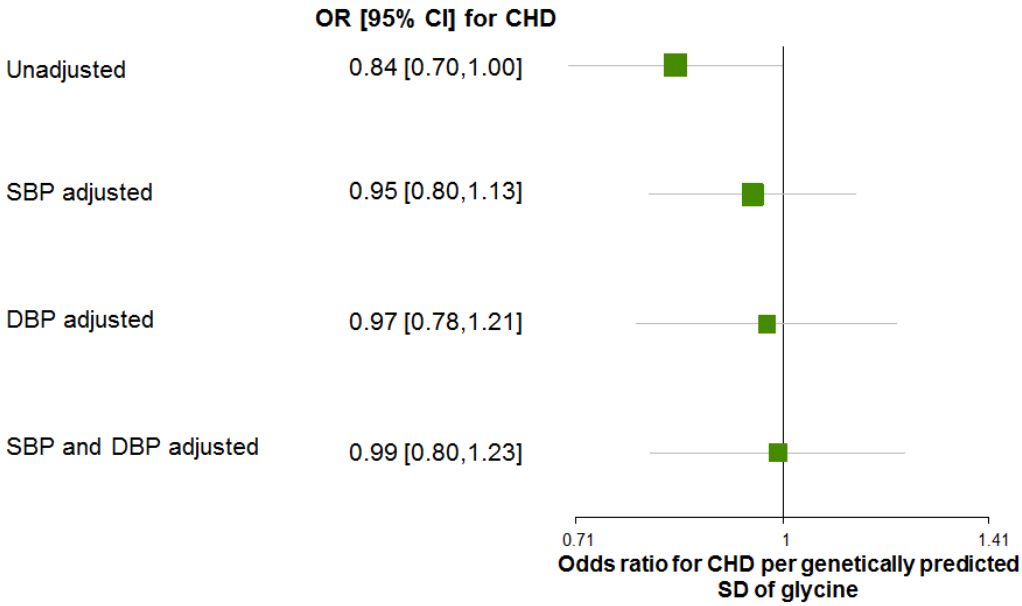


Figure 2.11: Forest plot of the odds ratios for coronary heart disease per genetically predicted standard deviation of glycine, unadjusted, adjusted for blood pressure. Analyses were based on the 5 SNP score for glycine levels, using summary-level inverse variance-weighted multivariable Mendelian randomisation. CHD: coronary heart disease, DBP: diastolic blood pressure, SBP: systolic blood pressure, SD: standard deviation, CI: confidence interval.

Discussion

Novel loci for glycine in genes involved in the *de novo* synthesis of serine

With the largest genetic discovery for glycine levels to date, I identified 22 loci not previously reported for glycine and replicated the five known loci. Two novel loci were located in the genes phosphoglycerate dehydrogenase (*PHGDH*) and phosphoserine phosphatase (*PSPH*), which encode the first and last enzymes catalysing the *de novo* biosynthesis of serine, a pathway which has not previously been genetically associated with glycine. Serine can be synthesised from and converted into glycine[227], and the associations of *PHGDH* and *PSPH* with glycine suggest that changes in the efficiency of the *de novo* biosynthesis of serine also affect glycine levels.

Eight and five of the novel loci coincide or are in high LD with loci reported for blood lipids (*COBLL1-GRB14*, *SLC22A3*, *LOC157273*, *TRIB1*, *ABO*, *REEP3* and *KPNB1*) and coagulation-related traits (*FGG, C5orf56*, *KLKB1*, *F12*, *ABO*), respectively. The direction of association with lipid and coagulation traits relative to the glycine-raising allele is not consistent between loci, which indicates that these pleiotropic loci do not reflect an overall shared genetic control of glycine, blood lipids and coagulation. Loci near tubulin gamma complex associated protein 4 (*TUBGCP4*, MAF=2.9%) and Kinesin Family Member 5B (*KIF5B*, MAF=0.3%), both involved in microtubule function, have to our knowledge not previously been reported for any trait.

The role of glycine in risk of T2D may be pathway-specific

Results of the Mendelian randomisation analyses on the causality of low glycine levels for T2D risk differed depending on the genetic instrument used for glycine. Analyses based on the two genetic scores which were most powered for but less specific to glycine indicated no causal effect of glycine on risk of T2D. This lack of association was mainly driven by the *CPS1* locus, which was the locus most strongly associated with glycine but did not alter T2D risk. After excluding the *CPS1* locus, a significant inverse association of genetically predicted glycine with lower risk of T2D was observed, and this association was driven by the two loci which were located near genes – glycine cleavage system protein H (*GCSH*) and glycine decarboxylase (*GLDC*) – encoding enzymes of the glycine cleavage system.

Our findings contradict the recently reported protective effect of glycine on T2D based on a Mendelian randomisation analysis using 5 loci for glycine, which included *CPS1*, *GLDC* and *GCSH*[147]. In this study, the significant inverse genetic association of glycine with T2D risk was largely driven by the *CPS1* locus, of which the glycine-raising allele was nominally associated with lower T2D risk in the 11,600 T2D cases and 33,000 controls of the GoT2D consortium. Our

analyses based on up to 63,767 T2D cases and 457,195 controls did not replicate a nominal association with T2D for *CPS1* nor for the genetic scores for glycine that included *CPS1*.

These findings suggest that overall levels of glycine may not be causal for T2D but that low glycine levels due to changes in the catalytic efficiency of the glycine cleavage system may cause higher risk of T2D. The glycine cleavage system is a protein complex comprised of four enzymes which catalyses the major catabolic pathway of glycine into ammonia and carbon dioxide[227]. Rare, severe mutations in genes of the glycine cleavage system have been identified as the cause of non-ketotic hyperglycinemia, a rare autosomal recessive inborn error of metabolism causing severe neurological symptoms, including seizures, apnea, hypotonia and spasms[228]. Due to the drastically shortened lifespan of most patients with non-ketotic hyperglycinemia and the rarity of this conditio[229,230], it is unknown if rare mutations in genes encoding enzymes of the glycine cleavage system affect risk of cardio-metabolic disease. Coding mutations in genes of the glycine cleavage system have also been linked to higher risk of neural tube defects[231]. To the best of our knowledge, this enzyme complex has not previously been linked to common metabolic diseases, such as T2D.

An alternative interpretation of our findings is that overall glycine levels are causally associated with T2D risk, but that this association was not observed when using the full or 6 SNP score due to the strong pleiotropic effects of *CPS1*. Carbamoyl-phosphate synthase 1 (*CPS1*) encodes the enzyme that catalyzes the first step of the urea cycle, which is responsible for the detoxification of ammonia, i.e., the breakdown product of glycine. The *CPS1* locus had widespread effects on the metabolome, which were not restricted to the glycine pathway, and of which some, including glycine, may protect against T2D while others could increase T2D risk. As *CPS1* had a more than 10 times stronger effect on glycine than most other glycine loci, the two genetic instruments for glycine which included *CPS1* in fact mostly represented the effect of *CPS1*. Therefore, the pleiotropic effects of *CPS1* could overshadow the genetic association of glycine levels with T2D risk.

In a few small-scale intervention studies, glycine has been suggested to increase insulin secretion after glucose intake[232,233]. *In vitro* experiments on islet β cells from donors with and without T2D indicated that glycine stimulated insulin secretion by binding with the glycine receptor expressed on the β cell membrane[234]. Islet β cells from donors with T2D had lower expression levels of the glycine receptor and responded more weakly to glycine[234]. Another small intervention study reported that glycine supplementation reduced HbA1C and pro-inflammatory cytokines in T2D patients[235] but these findings have not been replicated on a larger scale.

Insulin resistance as a driver of hypoglycinemia

We found strong genetic evidence that low glycine levels are a consequence of insulin resistance, which suggests that the inverse association of glycine levels with incidence of T2D is at least partly driven by this early disease process which is often already present several years before T2D can be diagnosed[236]. Observational epidemiological research has previously described inverse associations between measures of insulin resistance and glycine levels[179,180] and rodent experiments suggest that increased oxidative stress caused by insulin resistance leads to a higher rate of glutathione synthesis, for which glycine is one of the precursors, and thus lower glycine levels[176]. Another proposed mechanism that could link insulin resistance to lower levels of glycine is the increased demand for glycine to safely excrete fatty acyl-CoA esters which accumulate as the circulating levels of free fatty acids increase[237]. In an insulin resistant state, lipolysis in adipose tissue is incompletely suppressed by insulin, which leads to higher concentrations of free fatty acids in the circulation[238]. This leads to a build-up of beta oxidation intermediates, i.e., fatty acyl-CoA esters. Via a trans-conjugation reaction with glycine catalysed by acyl-CoA:glycine-N-acyltransferase, fatty acyl-CoA esters are converted to fatty acyl-glycine, which can be excreted via the urine[237,239]. As lipolysis and free fatty acids are elevated as a consequence of insulin resistance, the demand for glycine to safely excrete beta oxidation intermediates increases, which could ultimately lead to a depletion in glycine[237].

Glycine has previously been suggested as a biomarker for insulin resistance[180,240] and our findings suggesting that the inverse association of insulin resistance with glycine levels may represent a causal effect further strengthen this. Insulin resistance is not routinely screened for in primary health care settings, partly due to the high cost of measuring insulin. Using glycine as a biomarker for insulin resistance could be a financially more feasible approach to screening for insulin resistance, as glycine can be measured more cheaply than insulin, using fluorometric assays.

Glycine reduces risk of CHD equally in men and women

Based on genetic scores comprised of up to 24 SNPs, we demonstrate that low levels of glycine are genetically associated with higher risk of CHD, which supports a potential protective role of glycine against CHD. Hartiala *et al.* previously suggested that glycine metabolism may play a sex-specific role in the aetiology of CHD in women, based on the association of the glycine-raising allele of rs715 at the *CPS1* locus with lower risk of CHD in women only[188]. However, this sex-specific association with CHD did not take into account the known strong sex difference in the effect size on glycine levels[189], which we here estimated to be nearly three times stronger in women than in men. Besides its robust association with glycine levels, *CPS1*, which encodes the rate-limiting enzyme of the urea cycle, influences a range of other metabolites. Thus, it is possible that

causal inferences based on this non-specific locus were driven by associations with other metabolites unrelated to glycine metabolism. We demonstrate using four genetic scores with different degrees of specificity to glycine that the inverse genetic association of glycine levels with CHD risk is likely to be specific to glycine.

Based on sex-specific genetic scores for glycine that take into account sex differences in effect sizes, which we report for the first time for *GLDC* and *TRIB1*, we found no evidence that the effect on CHD per unit difference in glycine differed by sex. As the genetic score for glycine explains more than twice as much of the glycine variance in women than in men, the analyses in men were less powered, which caused the effect estimate of glycine on CHD risk for men to have wider confidence intervals.

Blood pressure as a potential mediating factor of the cardio-protective effect of glycine

We explored the underlying biological pathways through which glycine may protect against CHD by assessing the associations of genetically predicted glycine levels with CHD risk factors. Our findings suggested that glycine may reduce CHD risk by lowering blood pressure. High blood pressure is a major risk factor for cardiovascular diseases[241] and has also been linked to low levels of glycine[175,242]. Glycine supplementation has been shown to lower blood pressure in several studies on rodents[178,186,243] and in one small intervention study in participants with and without metabolic syndrome[244]. Our findings based on genetic associations with blood pressure in up to 445,360 individuals corroborate the blood pressure-lowering effect of glycine for the first time in a large-scale epidemiological setting. A potential mechanism through which glycine could lower blood pressure is through binding of glycine to the glycine-gated chloride channels expressed on the endothelium surface, which causes membrane hyperpolarisation. This results in the activation of the enzyme endothelial nitric oxide synthase which is responsible for the production of nitric oxide – a well-known vasodilator[187]. There was no evidence for a role of glycine in lipid metabolism or blood cell traits, despite previous studies suggesting that glycine may regulate lipid metabolism[175] and platelet[245] and immune cell activation[246].

The potential of glycine supplementation as a preventative strategy for cardiometabolic diseases

Our findings that higher glycine levels may be causally related to lower risk of cardiometabolic diseases raises the question if glycine supplementation should be evaluated as a preventative approach for CHD. The modest genetic effect size of glycine levels on CHD suggests that the potential benefit of glycine supplementation on cardiometabolic disease risk will be relatively minor. As a first step in identifying the preventative potential of glycine supplementation for

cardiometabolic disease, the safety of such an intervention should be assessed. To our knowledge, no long-term toxicological studies of glycine supplementation have been conducted on human participants. One study was set up to study the metabolic benefits of glycine supplementation in 60 participants, of which 30 received 15 mg of glycine per day for 3 months. This study did not report adverse effects associated with glycine supplementation and found that glycine supplementation reduced signs of oxidative stress, and lowered blood pressure in male participants only[244]. However, these findings have not been replicated in larger-scale studies with a longer intervention period. A potential concern of glycine supplementation is that this could lead to an increase in cancer risk. Research suggests that glycine and serine may promote oncogenesis by fuelling C1 and folate metabolism, thus increasing nucleotide biosynthesis and energy in the form of NADH and ATP[247]. This hypothesis is supported by findings from a systematic literature review on randomised controlled trials of folic acid supplements in human participants, which reported a suggestive increase in cancer incidence in individuals taking folic acid supplements[248]. Furthermore, in a recent study a glycine and serine-depleted diet was shown to improve survival in mouse models of colorectal cancer and lymphoma[249]. These findings indicate that a thorough assessment of the potential carcinogenicity of glycine should be conducted based on *in vitro* and *in vivo* animal-based experiments before further research on the cardiometabolic benefits of glycine supplementation is undertaken.

Strengths and limitations

Our investigation into the aetiological role of glycine metabolism on T2D on CHD has several strengths. The variance in glycine levels explained by the genetic score comprising 24 loci was more than 15%, which enabled us to conduct well-powered Mendelian randomisation experiments. The high proportion of explained variance can be attributed to the biological specificity of the exposure. As a comparison, the 751 loci identified for body mass index in the latest GWAS in up to 700,000 participants together explained approximately 6% of the variance in body mass index. Secondly, we conducted a thorough assessment of pleiotropic effects of the genetic score for glycine levels and generated a series of subsets of the full genetic score with increasing specificity to glycine metabolism, in order to reduce the likelihood of contributions of glycine-unrelated metabolic pathways to the causal effect estimate. We adopted multiple Mendelian randomisation approaches, including robust methods such as the weighted median method which are less sensitive to bias due to pleiotropic effects by a subset of genetic variants within the score, which further decreases the chance that the genetically determined effect estimate was driven by effects of the genetic scores of glycine on T2D and CHD through glycine-independent mechanisms.

We are aware of certain limitations to our study. First, some of the genetic loci we identified for glycine using a *p* value threshold of $<5\times10^{-8}$ may be false positives. It has recently been suggested that a lower *p* value cut-off, e.g., $<5\times10^{-9}$, is needed to account for the increased number of independent tests due to the higher density and increased coverage of rare variants by recent imputation reference panels[250]. Six of the 27 genetic loci for loci had $p>5\times10^{-9}$. Even if these 6 loci were not truly affecting glycine, this should not have had a substantial influence on the causal estimates from the MR analyses, as these loci were either not included in the genetic scores for glycine (*KIF5B* and *TUBGCP4*), or they had small effect sizes on glycine and were only included in the 24 SNP score (*ARL15*, *ADGRV1*, *GLS2* and *DAOA-AS1*).

Finally, despite undertaking sensitivity analyses, we cannot exclude the possibility that our findings may be driven by vertical pleiotropic effects of the genetic instruments. As the metabolome is comprised of thousands of metabolites connected through numerous reactions, some of which may be unknown, the distinction between horizontal and vertical pleiotropy is difficult to make in the context of metabolomics. The 2 SNP score, comprised of loci in genes of major enzymes of glycine catabolism, is the most specific score to the glycine pathway but still shows modest associations with some metabolites of which we cannot be sure that they are metabolically linked to glycine, e.g., the unknown metabolite X-16570, and with glycine-conjugated metabolites, which may have effects independent of glycine on cardio-metabolic disease risk. We can furthermore not exclude that the genetic scores for glycine were associated with metabolites not covered on the Metabolon platform or other aspects of the human "phenome" which were independently of glycine associated with cardio-metabolic disease risk. Therefore, we cannot fully exclude that glycine-independent mechanisms may have biased the genetically predicted association of glycine with disease risk. Moreover, as glycine is a metabolite on the intersection of many metabolic pathways, the genetic association of glycine with cardio-metabolic diseases may represent the causal effect of a metabolite to which glycine is metabolically close linked (e.g., tetrahydrofolate or serine).

Independent of potential pleiotropic effects of the glycine score, the genetically determined effect size of glycine on cardiometabolic disease may be up- or downwardly biased. The Mendelian randomisation approach estimates the effect size of an exposure on a certain outcome, assuming that exposure levels remain constant throughout the life course. As glycine levels may decrease, increase or fluctuate during a person's life, the estimated effect of glycine on CHD risk through genetics may be over- or underestimated. Furthermore, through a process called "canalization"[56], i.e., the process of developmental compensation that reduce or negate the phenotypic effect of

certain genetic mutations, the effect of genetically lower glycine levels on T2D and CHD risk may be weaker than the effect size of low glycine levels developed during the life course.

Finally, due to the relatively modest effect size of glycine levels on CHD risk, we had limited power to precisely quantify the extent to which the genetic association of glycine levels with CHD risk could be explained by blood pressure. There may also be other mechanisms underlying the genetic association between glycine levels and CHD risk, which we could not test here. For example, glycine may be linked to lower levels of oxidative stress, as it is a substrate for the biosynthesis of glutathione – a major antioxidant in human cells[251,252].

Conclusion

I show that low glycine is associated with higher incidence of CHD and that a genetic instrument for glycine is compatible with this relationship being causal. We also show that the association between genetically lower glycine and genetically higher blood pressure may at least in part explain the inverse relationship between glycine levels and CHD risk. No conclusive evidence for a causal, protective role of glycine on risk of T2D was found, but our findings suggest that changes in glycine catabolism by the glycine cleavage system may have a causal relationship with T2D. Finally, strong, genetic evidence was found that insulin resistance is causally associated with hypoglycinemia.

Chapter 3: The genetic determinants of total body fat percentage and fat-free mass index

Collaborations and contributions

This chapter focuses on the genetic determinants of total body fat percentage (BF%) and fat-free mass index (FFMI). Most of the work described in this chapter was done as part of a consortium project led by Prof Ruth Loos (Charles R. Bronfman Institute of Personalized Medicine, Icahn School of Medicine at Mount Sinai), in which I played a leading role as one of the main analysts. We started this project by conducting exome chip meta-analyses for BF% and FFMI. Study-level exome-wide association analyses were conducted by researchers at the research institutions running the studies. I ran exome chip analyses in the Ely and Fenland studies and conducted meta-analyses of all study-level results. It is standard practice in large consortia that all analyses are run in parallel by two separate analysts. Therefore, all exome chip meta-analyses were replicated by Dr Michael Preuss (Icahn School of Medicine at Mount Sinai). As a next stage in this project, genome-wide association studies for BF% and FFMI were conducted in UK Biobank. I conducted GWAS for BF% and FFMI on participants of European ancestry of UK Biobank. These analyses were done in parallel by Dr Hanieh Yaghootkar (University of Exeter). Meta-analyses of UK Biobank GWAS results with the exome chip results and published GWAS results were conducted by me for FFMI and Dr Kevin Lu (Vanderbilt University) for BF%. Significant independent variants for each trait were identified by me, Dr Hanieh Yaghootkar and Yingjie Ji (University of Exeter), and I conducted cross-trait LD clumping of the combined list of selected variants for BF% and FFMI. Collider bias tests for variants associated with FFMI were designed by Dr Zoltan Kutalik (University of Lausanne) and conducted by Dr Zoltan Kutalik and me. Annotations of the significant signals and LD score regression analyses were conducted by me and Dr Kevin Lu.

Based on the genetic loci identified for BF% and FFMI, I also conducted the following analyses: I assessed the observational and genetic correlations between BF%, FFMI and related anthropometric traits, conducted conditional analyses for the rare and low-frequency variants to identify the coding variants that are likely to be the causal signal in the region, investigated cross-trait associations of variants associated with BF% or FFMI, and identified BF%, FFMI and BMI loci with disproportionate effects on fat or lean mass.

Abstract

Background: The genetic determinants of overweight and obesity have thus far mostly been studied through large genetic discoveries for BMI, which has a low sensitivity to detect body fatness, as it does not distinguish fat from lean body mass. Evidence suggests that fat and lean mass may have discordant effects on metabolic health, but the distinct genetic determinants of fat and lean mass have not been studied at scale. Identifying genetic regions specific to body fat or lean mass can help us to assess their potential roles as aetiological factors for cardiometabolic diseases.

Aims: This chapter aims to (1) conduct genetic discoveries for total body fat percentage (BF%) and fat-free mass index (FFMI), (2) assess pleiotropic effects of loci for BF% and FFMI on related anthropometric phenotypes, (3) identify subsets of loci which disproportionately affect lean and/or fat mass and (4) assess association patterns of low-frequency coding variants for BF% and FFMI with cardiometabolic disease risk and related phenotypes.

Methods: (1) Genetic loci for BF% and FFMI were identified through exome chip meta-analyses including up to 112,443 participants and through GWAS on 443,391 UK Biobank participants of European ancestry. (2) Cross-anthropometric trait associations were assessed for the significant loci as well as across the entire genome using cross-trait LD score regression. (3) The proportionality of the effect sizes of loci for BF%, FFMI and BMI on fat versus lean mass was assessed based on genetic associations and observational correlations in UK Biobank. (4) For low-frequency and rare coding variants for BF% and FFMI, associations with phenotypes related to body size and cardiometabolic health were assessed using publicly available exome chip association data.

Results: (1) 673 novel genetic loci for BF% and 783 novel loci for FFMI were identified. Twenty low-frequency and rare coding variants for BF% and/or FFMI were identified as likely causal variants, including rare stop gained variants in *PDE3B* and *GPR151* with strong effects on BF%. (2) Of the 1,216 independent genetic loci for BF% and FFMI, 173 were associated with both traits in the same direction, and 29 loci had opposite effects on BF% and FFMI. 447 and 567 loci were associated only with BF% and FFMI, respectively. Subsets of BF% and FFMI loci were strongly associated with WHR, WHRadjBMI and height. (3) Despite very strong correlations of BF% and FFMI with BMI, 16 of the 1,216 genetic loci had disproportionate effects on BF% beyond BMI, of which 11 loci also had disproportionately large, inverse effects on FFMI. The vast majority of

the reported BMI loci affected BF% and FFMI proportionately, with the exception of 4 loci mainly driven by BF% and 3 other loci which were disproportionately associated with FFMI. (4) Heterogeneous association patterns of the low-frequency coding variants for BF% and FFMI with cardiometabolic diseases and traits were observed.

Conclusion: Through large-scale genetic discoveries for BF% and FFMI and systematic integration of genetic findings across anthropometric traits, genetic determinants of normal weight adiposity, i.e., high BF% in the absence of overweight, were identified. Furthermore, evidence was found that coding genetic factors associated with overall body fat and lean mass have a non-uniform relationship with cardiometabolic health. This relationship may be partly mediated by body fat distribution.

Introduction

The global health impact of overweight and obesity

Numerous studies have demonstrated that overweight and obesity are leading causes of morbidity and mortality worldwide[253]. According to estimates by the WHO's Global Health Observatory, 39% of the global adult population is overweight (BMI ≥25 kg/m^2) while 13% is obese (BMI ≥30 kg/m^2)[254]. In Europe alone, 59% and 23% of the population is overweight and obese, respectively[254]. 7% of all deaths worldwide in 2015 were attributed to overweight and obesity, of which 40% occurred among people with a BMI below the obese range[253]. The most common causes of death related to high BMI were cardiovascular disease (41% of all BMI-related deaths) and T2D (15%). Worldwide, nearly 5% of all disability-adjusted life years were lost due to high BMI[253], with cardiovascular disease (34%), chronic kidney disease (18%) and T2D (5%) as the leading underlying causes[253].

Fat and lean mass have discordant associations with metabolic phenotypes and outcomes

Most epidemiological and genetic research on adiposity thus far has focused on BMI[217,255,256], as it is an easy-to-obtain measure which only requires scales and a height measure. However, BMI does not distinguish fat from lean body mass. Consequently, BMI has a low sensitivity to detect excess adiposity, as it underestimates fat mass in individuals with a high fat mass but low lean mass[257,258]. Total body fat percentage (BF%), defined as the ratio of total mass of body fat to total body mass, has been shown to be a more sensitive measure of adiposity than BMI[257,258]. Available evidence suggests that 50-60% of individuals with a BMI within the normal weight range may be classified as obese when applying the WHO's cut-offs for BF% (men: BF% >25%, women: BF% >35%) to diagnose obesity[258,259]. In particular among men, elderly participants and among individuals with a BMI in the overweight range (25-29.9 kg/m^2), BMI was found to be a poor predictor of BF%-defined obesity[258]. Conversely, individuals with high lean but low fat mass, such as athletes, can, based on BMI, be misclassified as overweight[260].

Excess fat mass has been shown to be tightly associated with multiple pathophysiological processes, such as dyslipidaemia[261], insulin resistance[262,263] and hypertension[264,265] and has been causally linked to higher risk of cardiometabolic diseases in genetic[85,86] and intervention studies[19,25,26]. High BF% with a normal weight was found to be more strongly associated with abnormal blood glucose concentrations than overweight in the absence of elevated BF%[259], which emphasises the strong relevance of BF% to metabolic risk. Conversely, lean mass[266–268] and muscle density[269] and strength[270,271] have been associated with better cardiometabolic health. Muscles,

which make up around 50% of the total lean mass[272,273], are the most important depots for post-prandial glucose uptake[274] and are therefore considered to play an essential in glucose regulation[275]. Reduced sensitivity to insulin at the level of the muscles has been proposed as a major risk factor for developing T2D later in life[274], and maintenance of muscle mass and strength may play a role in the prevention of T2D[275] and improving glucose control in T2D patients[276]. These discordant associations of fat and lean mass with cardiometabolic diseases and risk factors highlight the importance of studying the molecular mechanisms associated specifically with fat and lean mass, instead of with overall body mass and size only.

Bio-impedance analysis is a cheap and non-invasive technique to estimate body fat and fat-free mass

Bio-impedance analysis (BIA) is an anthropometric analysis technique that estimates total body fat and fat-free mass[277], of which the latter includes the lean tissues, such as the muscles and internal organs, as well as total bone mass. Fat and fat-free mass are estimated by measuring the opposition, or electrical impedance, of the body to a small electrical current that is sent through the body, e.g. between the hands and the feet. Electrical impedance is inversely correlated with the total amount of water in the body and also with the total fat-free mass, as fat-free tissue contains a higher concentration of water than fat tissue. Using population and sex-specific equations which include age, height, sex and weight as covariates, fat and fat-free mass can be estimated from electrical impedance[277]. While BIA has most commonly been used to estimate total body fat and fat-free mass, so-called segmental BIA instruments can also measure impedance in the arms, legs and trunk separately and thus allow estimation of regional fat and fat-free mass. However, BIA has been shown to estimate overall and regional body fat and lean mass less accurately than imaging-based techniques such as dual-energy X-ray absorptiometry (DEXA) and magnetic resonance imaging (MRI). Despite relatively strong correlations between body fat measured by BIA and MRI (0.75 for women and 0.81 for men)[278], BIA tends to systematically underestimate total body fat mass by 2 to 6% when compared to DEXA[279] and by 5 kg when compared to MRI[278], although this may vary by adiposity status. However, as BIA is a cheap, non-invasive and quick technique, it has over the past decade been used in many epidemiological studies and has enabled the first genetic discoveries for fat and lean mass.

Genome-wide association studies on body mass index, BF% and lean mass

Numerous genetic studies seeking to identify the genetic variation driving overweight and obesity have been conducted. In particular, the Genetic Investigation of Anthropometric Traits (GIANT) consortium has led multiple large-scale meta-analyses of GWAS and exome chip meta-analyses on adiposity based on BMI[172,217,280]. In the most recent and thus far largest genetic discovery for BMI which included nearly 700,000 individuals, 751 independent genetic variants, together explaining 6% of the variance in BMI, reached genome-wide significance[172]. Through pathway enrichment analyses, the central nervous system has been identified as the most prominent site of action of genetic variation related to BMI[217]. Little evidence for sex-discordant effects of the identified genetic loci was found[217].

The strength of using BF%, measured through BIA or DEXA, as a proxy for adiposity was demonstrated in two genome-wide discovery efforts[88,89]. Twelve genetic loci in/near *FTO, IRS1, MC4R, TMEM18, COBLL1, SPRY2, TOMM40, TUFM, IGF2BP1, SEC16B, PLA2G6* and *CRTC1* for adiposity were identified, of which six loci (*IRS1, COBLL1, SPRY2, IGF2BP1, PLA2G6* and *CRTC1*) were robustly associated with BF% in sample sizes up to 100,716 individuals but were at the time not identified for BMI in sample sizes of up to 339,224 participants. Studies on the genetic determinants of lean mass are scarcer. In 2017, Zillikens *et al.* published the first GWAS of total lean mass and lean mass in the limbs, which is also called appendicular lean mass and serves as a proxy for total muscle mass. This study included up to 38,292 participants and identified the first 5 loci in/near *FTO, VCAN, IRS1, HSD17B11* and *ADAMTSL3* for lean mass. Other studies have focused on identifying genetic factors influencing muscle strength, by for example using hand grip strength as a proxy[281,282], but associations of the identified loci with lean mass are uncertain, because muscle strength and lean mass tend to be only weakly associated[283]. No studies to date have attempted to identify low-frequency coding variants for fat or lean mass through exome chip analyses, despite the availability of body composition data generated using BIA and, to a lesser extent, DEXA imaging, and exome-wide genotyping in several epidemiological cohorts. Coding variants identified through exome chip studies are more likely to be associated with biological phenotypes than variants in non-coding regions, and hypotheses on the biological mechanisms through which they influence the phenotype can be more easily formulated[284].

Opportunities and objectives

The implementation of genome or exome-wide genotyping and of BIA-based body composition analysis in the baseline health assessment of half a million UK Biobank participants and several tens of thousands of participants of other epidemiological cohorts, provides the opportunity to identify the genetic determinants of body fat and lean mass at an unprecedented scale. Genetic loci specifically associated with higher BF% or FFMI can be utilised to construct genetic scores and assess the causality of body fat and lean mass as aetiological risk factors for cardiometabolic risk factors and diseases. The large group of individuals with a high BF% in the absence of elevated BMI and observational evidence suggesting that BF% is more strongly associated with metabolic risk than BMI highlight the importance of understanding the causality of normal weight adiposity, i.e., high BF% in the absence of BMI-defined overweight. Furthermore, genetic discoveries for BF% and FFMI can contribute to a refined characterisation of loci identified for more crude measures of adiposity, such as BMI, and for cardiometabolic phenotypes.

This chapter aims to (1) identify the genetic determinants of BF% and FFMI through large-scale genome and exome-wide genetic discoveries, (2) assess the pleiotropy of loci identified for BF%, and FFMI across related anthropometric phenotypes, (3) identify subsets of genetic loci identified for BF% and FFMI and previously reported for BMI with disproportionately large effect sizes on BF% or FFMI and (4) assess associations of low-frequency coding variants affecting BF% or FFMI with cardiometabolic diseases and phenotypes.

Methods

Study design

BF%, i.e., the ratio of total fat mass (kg) to total body weight (kg), and FFMI, which is total fat-free mass (kg) divided by height squared (m^2), were used as phenotypes representing fat and lean mass, respectively. These phenotypes were chosen because they are only moderately correlated (Pearson's correlation coefficient based on BF% and FFMI in UK Biobank: R=0.47). Two strategies were used to identify the genetic determinants of BF% and FFMI (Figure 3.1). First, exome chip meta-analyses based on 28 studies were conducted to identify low-frequency coding mutations associated with BF% and FFMI. Exome chip meta-analyses were conducted for Europeans only and for all ancestries combined. Secondly, GWAS on UK Biobank participants of European ancestry were conducted to identify genome-wide genetic factors for BF% and FFMI.

To maximise sample sizes and thus power to identify low-frequency coding variants for BF% (Figure 3.1 A) and FFMI (Figure 3.1 B), summary-level results of the exome chip meta-analyses were meta-analysed with the GWAS results. For BF%, an additional meta-analysis of the UK Biobank GWAS results with the largest publicly available summary-level GWAS results for BF%[88] was conducted.

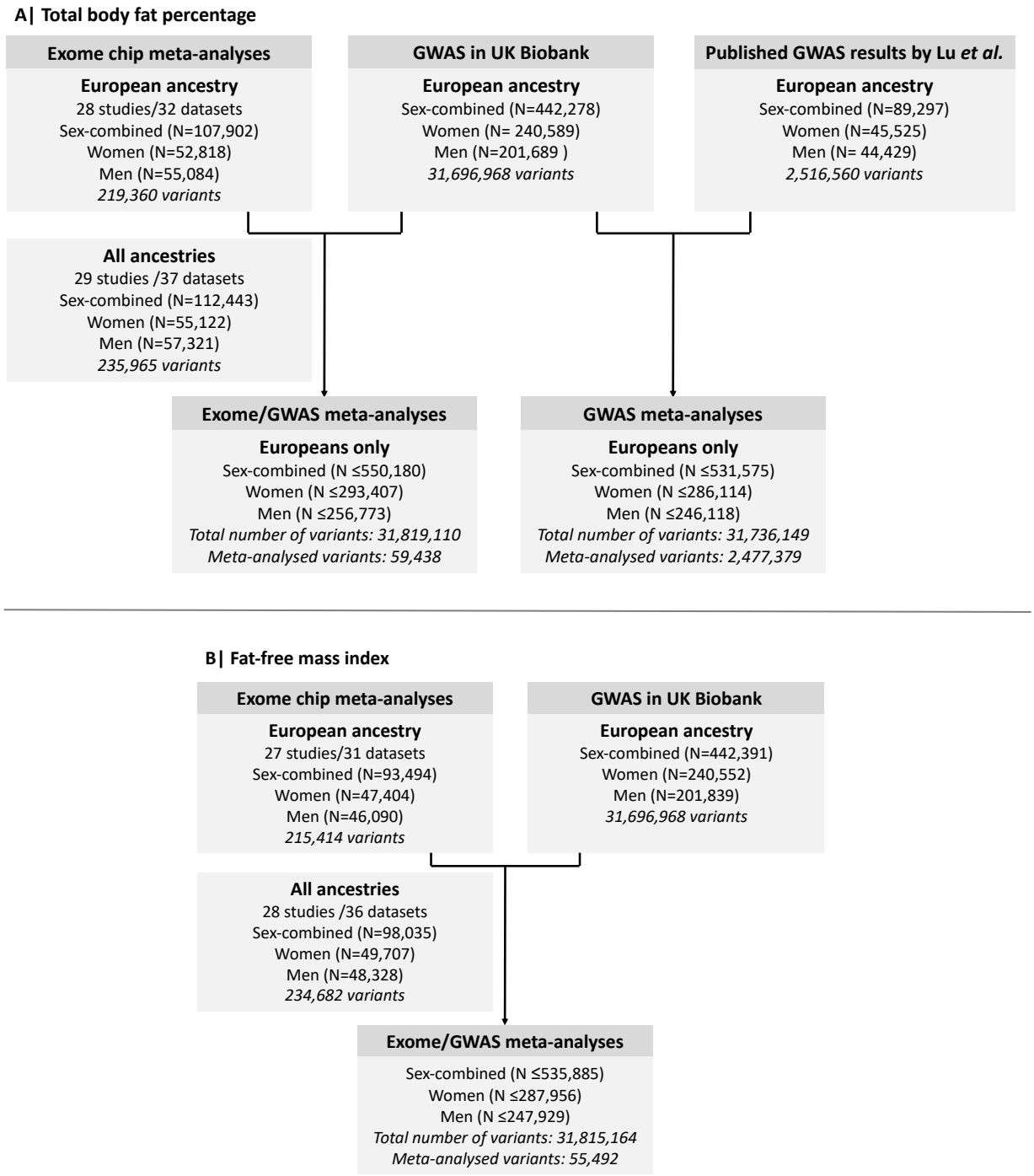


Figure 3.1: Flow chart of study design for genetic discoveries for BF% and FFMI. For both phenotypes, exome-chip meta-analyses for European ancestry and all ancestries combined, and GWAS on UK Biobank participants of European ancestry were conducted. Meta-analyses of the UK Biobank GWAS results and the exome chip meta-analyses were conducted for both FFMI and BF%. A separate meta-analysis of the UK Biobank GWAS results with published GWAS results was conducted for BF%.

Exome-wide association analyses

Study participants

We conducted exome chip meta-analyses for BF% and FFMI, for participants of European ancestry only and for all ancestries combined. The analyses for participants of European ancestry included 107,902 adult individuals from 28 studies (32 datasets) for BF% and 93,494 individuals from 27 studies (31 datasets) for FFMI. All ancestry analyses included 5 additional datasets, of which two datasets included participants of South-East Asian ancestry (N ≤2,330) and three datasets included participants of Afro-American ancestry (N ≤2,213), leading to sample sizes of up to 112,443 and 98,035 participants for BF% and FFMI, respectively (Supplementary Table 3.1). All participating institutions and coordinating centres approved this project and informed consent was obtained from all study participants.

Phenotypes

BF% and FFMI were derived in each cohort based on either BIA or DEXA (see Supplementary Table 3.1). Within each study, BF% and FFMI were adjusted for age, age^2, and genetic principal components (calculated from genome-wide genotyped data, all exome chip variants with minor allele frequency (MAF) >1%, or ancestry informative markers available on the exome chip) and additional study-specific covariates if applicable (e.g. recruitment centre) in linear regression models. The residuals of the linear regression model were then re-scaled based on the rank-based inverse normal transformation. For studies which included mostly unrelated participants, phenotypes were transformed for men and women separately, whereas for family-based studies sex was included as a covariate in the linear regression model. For case-control studies, phenotypes were transformed and analysed for cases and controls separately. The full analysis plan can be found in Supplementary Materials – Analysis plan 1.

Genotype calling

The majority of studies followed a standardized protocol and performed genotype calling using the designated manufacturer software, followed by zCall[285]. For studies participating in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, the raw intensity data for the samples were assembled into a single project for joint calling[286] (see Supplementary Materials – Analysis plan 1 for further information).

Study-level exome chip analyses

Study-level analyses were conducted for both sexes combined as well as by sex, were based on additive and recessive genetic models and were run using either RAREMETALWORKER[287] (https://genome.sph.umich.edu/wiki/RAREMETALWORKER) or RVTEST[288]

(https://genome.sph.umich.edu/wiki/Rvtests) (see Supplementary Table 3.1). Both software packages estimate genetic associations based on linear mixed models and account for potential cryptic relatedness by generating a genetic kinship matrix. Both software tools are designed to perform rare variant association analyses and can analyse data from studies of both unrelated and related individuals.

Centralised study-level quality control

A centralised quality control (QC) procedure, implemented in the R package 'EasyQC'[289], was applied to individual cohort association summary statistics to identify cohort-specific problems, such as errors in the phenotype transformation by plotting the median standard error versus the square root of the sample size for all studies, strand issues by testing the alignment of the observed allele frequencies against 1000 Genomes Project phase 1 reference data, and inflation of the test statistics due to e.g. population stratification, cryptic relatedness or genotype biases based quantile-quantile plots.

Exome chip meta-analyses, QC and annotation

Meta-analyses of the study-level exome-wide association results were carried out in parallel by two researchers using the R package 'RareMETALS'[287] (https://genome.sph.umich.edu/wiki/RareMETALS). Variants were omitted if call rate <95% or if p-value for Hardy-Weinberg equilibrium $<1\times10^{-7}$. Only coding variants which reached exome-wide significance ($p<2.0\times10^{-7}$) were considered. Variants were annotated based on the most severe annotation available from the Ensembl Variant Effect Predictor (VEP)[290]. Associations with height and BMI were obtained from publicly availably summary-level results from exome chip meta-analyses[280,291]. Associations with waist-to-hip ratio (WHR) and WHR adjusted for BMI (WHRadjBMI) from a recent exome chip analysis project[292] have only been released yet for variants reaching exome-wide significance for WHR or WHRadjBMI; therefore, associations of variants for BF% and FFMI with WHR and WHRadjBMI are only given for some of the identified variants.

Genome-wide association analyses in UK Biobank

Phenotypes

The UK Biobank study was introduced in the methods section of Chapter 2[202]. Participants underwent BIA using a Tanita BC418MA body composition analyser at baseline, from which BF% (UK Biobank data field 23099: http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=23099) and total fat-free mass (UK Biobank data field 23101: http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=23101) were derived. FFMI was calculated as total fat-free mass (kg)/standing height^2 (m^2). Standing height was measured using a SECA 240 height measure. QC of the phenotypes was performed centrally by UK Biobank. BF% and FFMI were adjusted for age, age^2, sex and assessment centre in linear regression models and the residuals of the linear regression models were rank-based inverse normally transformed. For the sex-specific analyses, phenotype transformations were conducted for men and women separately.

Genome-wide genetic data and identification of participants of white European ancestry

Genome-wide genotyping was conducted on 488,377 participants of UK Biobank using two different genotyping arrays, of which the probed genetic markers overlapped for 95%. 49,950 participants were genotyped on the Affymetrix UK BiLEVE Axiom array and 438,427 participants on the Affymetrix UK Biobank Axiom array. Quality control of the genome-wide genetic data and imputation using a combined linkage disequilibrium (LD) reference panel of the Haplotype Reference Consortium[192], the UK10K haplotype resource[193] and 1000 Genomes Phase 3[194] were performed centrally and have been described extensively elsewhere[213]. Genetic ancestry of the UK Biobank participants was determined using K-means clustering applied to the first four genetic principal components (PCs) anchored in the genetic PCs generated from the 1000 Genomes phase 3 reference panels for European (CEU), African (YRI) and Asian (CHB/JPT) ancestry populations[194]. Individuals who fell into the European cluster but who self-identified as being of any other ancestry than European were considered to be of non-European ancestry.

GWAS and QC

Phenotypes and genetic data were available for analyses on 442,278 (240,589 women and 201,689 men) and 442,391 (240,552 women and 201,839 men) UK Biobank participants of European ancestry for BF% and FFMI, respectively. We performed sex-combined and sex-specific genome-wide association analyses using BOLT-LMM[293] (v2.3) for inverse normally transformed BF% and FFMI under an additive genetic model and with genotyping array as a covariate. The BOLT-LMM (v2.3) software[203] was used as it fits linear mixed models, which can adjust for population structure and cryptic relatedness based on a genetic relatedness matrix. We used genotyped single nucleotide

polymorphisms (SNPs) present on both genotyping arrays, with MAF > 1% and which passed QC in all 106 genotyping batches, to estimate the parameters of the linear mixed model for the initial modelling step. The genome-wide association analyses were restricted to bi-allelic variants with INFO ≥0.4 if MAF ≥1%, and INFO ≥0.8 if MAF <1%. Variants in the human leukocyte antigen (HLA) region were excluded (variants on chromosome 6 with base-pair positions from 28,477,797 to 33,448,354).

Meta-analyses of the UK Biobank GWAS with exome chip meta-analyses and published GWAS data

Summary statistics of the GWAS in UK Biobank were meta-analysed with the European ancestry exome chip meta-analysis results to maximise power to identify coding variants for BF% and FFMI (Figure 3.1). A separate meta-analysis of the UK Biobank GWAS for BF% with published summary statistics of the European ancestry GWAS on BF% published by Lu *et al.*[88] was conducted (Figure 3.1). The published BF% GWAS included up to 89,297 individuals of European ancestry from 54 studies, of which 43 had genome-wide genetic data and 11 Metabochip data only[88]. Genotyped data in this previous study were imputed to the HapMap2 reference panel, which led to a total number of 2,516,560 testes variants, either directly genotyped or imputed. Meta-analyses were performed using the inverse variance-weighted fixed-effect method without adjustment for genomic control, as implemented in METAL[205]. Meta-analyses were conducted for the sex-combined and sex-specific analyses. The meta-analyses of the UK Biobank GWAS with the published GWAS for BF% and of the UK Biobank GWAS with the European ancestry exome chip data were used for all the following analyses, unless described otherwise.

Assessment of genomic inflation and SNP-based heritability

The linear mixed model-based GWAS method implemented in BOLT-LMM has been shown to sufficiently control for potential population stratification and cryptic relatedness in GWAS on UK Biobank participants[293]. To test if population stratification and cryptic relatedness caused inflation of the test statistics of our genetic discoveries for BF% and FFMI, we estimated genomic inflation using LD score regression (LDSC) based on the baseline LD model[294,295]. The LDSC intercept serves as an estimate for genomic inflation corrected for polygenicity, i.e., the inflation of the test statistics that is due to population stratification and cryptic relatedness. However, the LDSC intercept has been shown to increase with sample size and heritability due to attenuation bias[203]. The attenuation ratio (LDSC intercept - 1)/(mean $\chi 2$ - 1) has been proposed as an alternative to the LDSC intercept which is less biased by sample size and heritability[293]. We therefore made use of the attenuation ratio statistics to test for inflation of the test statistics due to residual population stratification and cryptic relatedness. LDSC was also used to estimate the SNP-based heritability

for BF% and FFMI. LDSC was conducted for the meta-GWAS results for BF% and the meta-analysis of the UK Biobank GWAS with the exome chip data for FFMI. Variants with MAF <0.1% or INFO <0.4, multi-allelic variants and variants in the HLA region were excluded from the LDSC analyses.

Identification of independent genetic loci for BF% and FFMI

We combined three approaches to identify the independent common and rare genetic loci associated with BF% and FFMI, based on the meta-GWAS for BF% (Figure 3.2), and on the meta-analysis of the UK Biobank GWAS and European exome chip data for FFMI (Figure 3.3). The genome-wide significance threshold was set at $p \leq 5 \times 10^{-9}$ to account for the increased number of tests that results from higher imputation density compared to earlier GWAS efforts[250].

For variants with MAF>1%, the independently associated lead variants that reached genome-wide significance were identified based on a combination of distance-based clumping and approximate conditional analyses conducted in GCTA-COJO[206]. Distance-based clumping of the common variants was conducted using 500 kb windows. In parallel, independent variant selection was conducted through a stepwise model selection procedure using the `--cojo-slct` function implemented in GCTA-COJO[206]. As LD reference panel for GCTA-COJO, we used SNPs with MAF >0.1% and INFO ≥0.3 from the approximately 343,000 UK Biobank participants who were unrelated and of British ancestry. Variants with a $p \leq 5 \times 10^{-9}$ in the joint model were considered to be conditionally independent. A combination of the variants selected through distance-based clumping (the "clumped variants") and approximate conditional analyses (the "GCTA variants") were used to generate the final list of common, independent loci. Starting from the list of GCTA variants, we replaced the GCTA variant by a clumped variant if the clumped variant was within 500 kb of the GCTA variant. This was done because GCTA selected in several cases not the most significant variant in the region as the sentinel variant in the region. Clumped variants which were not within 500 kb of a GCTA variant were added to the list of independently associated variants, unless they were located in previously reported regions of long-range LD[296] (see Supplementary Table 3.2). These clumped variants which were added in were mostly variants which had a marginal p-value near the significance threshold but for which the joint p-value dropped just below the significance threshold. Finally, GCTA variants were dropped if they were in LD (R^2>0.05) with any of the other GCTA variants. These steps were done for the European sex-combined and sex-specific analyses for BF% (Figure 3.2, Step A1 to 3) and FFMI (Figure 3.3, Step A1 to A3). Variants identified in the sex-specific analyses were only retained in the final list of independent variants if they were further than 500 kb from variants from selected in the sex-combined analysis (Figure 3.2 and 3.3, Step B).

Next, we identified the rare variants that reached genome-wide significance and were conditionally independent of common variants selected in step A within 1 Mb of the rare variant(s) (Figure 3.2 and 3.3, Step C). As GCTA-COJO is not optimised for identifying independently associated rare variants, we assessed the independence of the associations of rare variants using individual-level data-based conditional analyses. We first identified the "rare regions" by taking all genome-wide significant rare variants separated by less than 1 Mb, and added all common variants selected through the steps explained above. For each rare region, we fitted a joint model which included all rare and selected common variants in the subset of unrelated UK Biobank participants of British ancestry. Age, age^2, the first 10 genetic principal components, sex, genotyping chip and assessment centre were included as covariates. Rare variants with $p \leq 5\times10^{-4}$ in the conditional models were considered to be associated with the phenotype independently of the common variants in the region.

For lead variants identified in the sex-combined analyses, we tested for heterogeneity in the effect sizes for men and women based on the t statistic, which was calculated as follows:

$$t = (\beta_{women} - \beta_{men})/\sqrt{(SE^2_{women} + SE^2_{men}}$$

Loci for which the t statistic reached Bonferroni significance ($p \leq$ 0.05/total number of sex-combined lead variants) were considered to have sex-dimorphic effects.

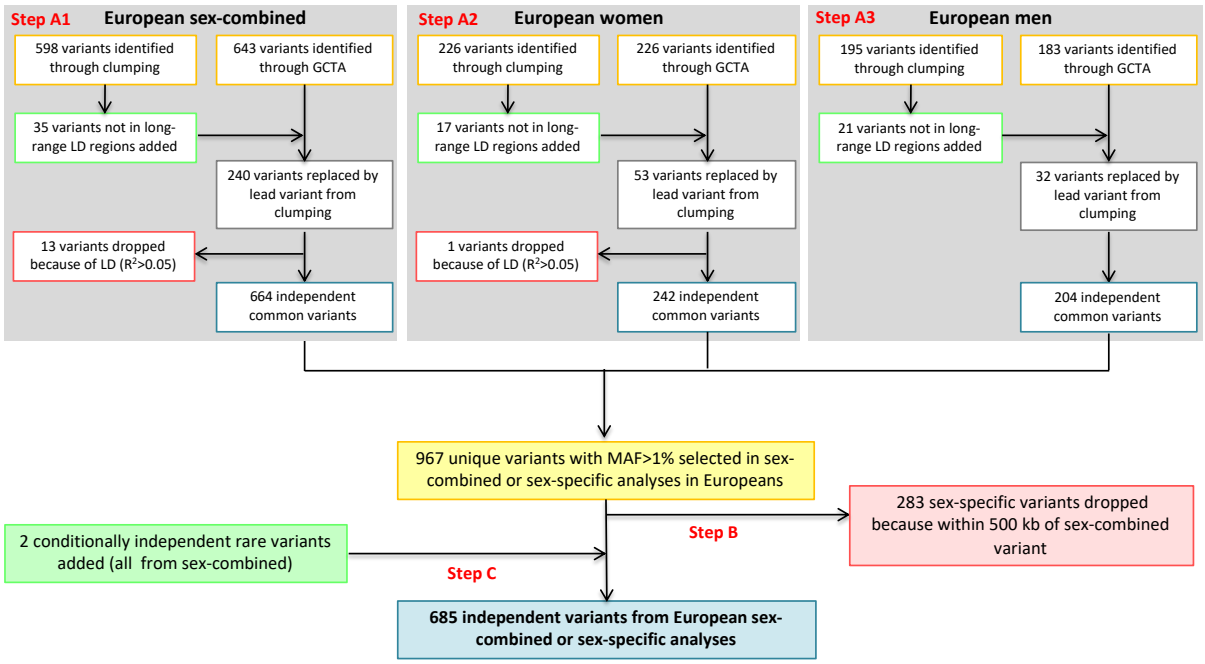


Figure 3.2: Overview of selection of common and rare independently associated lead variants for BF%, based on meta-analysis of GWAS results in UK Biobank and published summary-level GWAS results. In step A, the common, independent loci were identified based on distance-based clumping and approximate conditional analyses in GCTA-COJO, for the sex-combined and sex-specific analyses separately. In step B, sex-specific loci within 500 kb of a sex-combined locus were omitted, as they were likely to represent the same locus. In step C, rare variants which were associated independently of the common variants selected in step A, were identified based on individual-level data-based conditional analyses, and were added to the list of independently associated variants. This led to the identification of 685 independent loci for BF%.

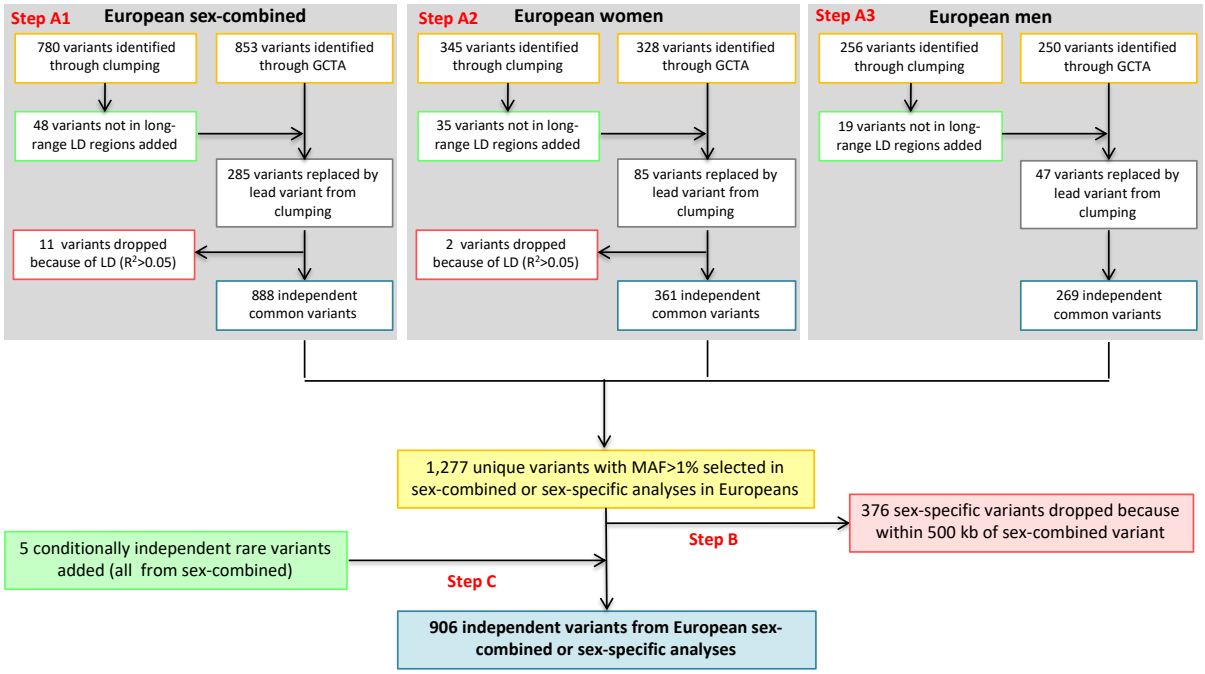


Figure 3.3: Overview of the selection of common and rare independent loci for FFMI, based on the meta-analyses of the GWAS in UK Biobank and the European exome chip meta-analyses. The same procedure as for BF% was followed, which led to the identification of 906 independent common and rare loci for FFMI.

Conditional analyses for rare and low-frequency coding variants based on individual-level data from the unrelated UK Biobank participants

We had a specific interest in rare and low-frequency coding variants that reached genome-wide significance for BF% and/or FFMI and asked if they were (one of) the causal variant(s) in the region or, alternatively, were located in a region driven by another, non-coding variant, without the low-frequency coding variant itself being the causal variant. Although the approximate conditional analyses were conducted on all variants with MAF >1% and thus also included the low-frequency (but not the rare) coding variants, we conducted a separate series of conditional analyses based on individual-level data. This separate series of conditional analyses for all low-frequency coding variants was conducted because, although GCTA-COJO has the advantage of identifying the statistically independent variants in an efficient manner and without the need of individual-level data, experience of the consortium members shows that GCTA-COJO may not always be reliable at identifying independently associated low-frequency and rare variants.

We first tested in UK Biobank if the genotyped coding variant, if available, showed the same direction and strength of association as the imputed variant which was analysed by BOLT-LMM for the GWAS. We then conducted exact conditional analyses using the same strategy as for the selection of the rare (non-coding and coding) variants, including the rare and low-frequency coding variant(s) and all selected variants for BF% (Figure 3.2) or FFMI (Figure 3.3) within 1 Mb of the coding variant(s) in the model. Coding variants which reached a significance threshold of $p \leq 5 \times 10^{-4}$ were considered to be independently associated with BF% or FFMI. Variants were annotated based on the most severe annotation found in VEP[290]. For the variants which were covered in previous exome chip projects, associations with related anthropometric phenotypes, including height[291] and BMI[280], and cardiometabolic risk factors and diseases, including blood lipids[297], T2D[40] and CHD[298], were extracted from publicly available summary-level results from European ancestry exome chip meta-analyses.

Assessment of collider bias in the GWAS for FFMI due to adjustment for height

Aschard *et al.* demonstrated that results of a GWAS for an outcome that is adjusted for a heritable covariate can be biased, and that this so-called collider bias increases with the strength of the correlation between the outcome and covariate[299]. In GWAS conducted on large sample sizes, collider bias can lead to significant associations that are solely driven by an association with the covariate without affecting the (unadjusted) outcome. As FFMI is calculated by the ratio of fat-free mass to height squared, of which the latter is both highly heritable and strongly correlated with lean mass, some of the variants reaching genome-wide significance for FFMI may be driven by an inverse association with height only. Collider bias was therefore assessed for each of the

genetic variants selected for FFMI, using the approach described by Aschard *et al.*[299]. The original approach to test for the presence of a collider bias was developed for linear covariates, while FFMI is adjusted for a quadratic covariate, i.e., height squared. Therefore, the original approach was adapted for quadratic covariates as follows:

Let us define fat-free mass (FFM) as y, height as x and a genetic marker as g, measured in a population of size n. We model FFM and height as linear functions of the genetic marker as follows:

$$\frac{x-\bar{x}}{\sigma_x} = \beta(g-\bar{g}) + \varepsilon_x \; with \; \varepsilon_x \sim \aleph(0,1)$$

$$\frac{y-\bar{y}}{\sigma_y} = \alpha(g-\bar{g}) + \varepsilon_y \; with \; \varepsilon_y \sim \aleph(0,1)$$

and where $\bar{x}$ and σ_x represent the sample mean and sample variance of variable x, respectively. Based on x and y, we define a new trait z (i.e., FFMI) as their ratio $z = y/x^2$. The new trait z can be written as

$$z = \frac{\alpha\sigma_y(g-\bar{g}) + \sigma_y\varepsilon_y + \bar{y}}{(\beta\sigma_x(g-\bar{g}) + \sigma_x\varepsilon_x + \bar{x})^2} = \frac{\alpha'(g-g') + \varepsilon'_y}{(\beta'(g-g') + \varepsilon'_x)^2} = f(g-g')$$

with $\alpha' = \alpha\sigma_y, \beta' = \beta\sigma_x, \varepsilon'_x = \bar{x} + \sigma_x\varepsilon_x, \varepsilon'_y = \bar{y} + \sigma_y\varepsilon_y$ and $f(t) = \frac{\alpha' t + \varepsilon'_y}{(\beta' t + \varepsilon'_x)^2}$.

We can approximate z using the first order Taylor series expansion as follows:

$$z = f(g-g') \approx f(0) + f'(0)\cdot t = \frac{\varepsilon'_y}{\varepsilon'^2_x} + \left(\frac{\alpha'}{\varepsilon'^2_x} - 2\frac{\varepsilon'_y\beta'}{\varepsilon'^3_x}\right)\cdot(g-\bar{g})$$

We can now estimate the effect of variant g on z as follows:

$$\hat{\gamma} = \frac{(g-\bar{g})'\cdot(z-\bar{z})}{(g-\bar{g})'\cdot(g-\bar{g})} = \frac{(g-\bar{g})'\cdot z}{(g-\bar{g})'\cdot(g-\bar{g})} \approx \frac{\alpha'}{\varepsilon'^2_x} - 2\frac{\varepsilon'_y\beta'}{\varepsilon'^3_x}$$

The estimator for γ can therefore be defined as

$$\hat{\gamma} = \widehat{\alpha'}\cdot E\left(\frac{1}{\varepsilon'^2_x}\right) - 2\cdot\widehat{\beta'}\cdot E\left(\frac{\varepsilon'_y}{\varepsilon'^3_x}\right) \; with \begin{pmatrix}\varepsilon'_x\\ \varepsilon'_y\end{pmatrix} \sim \aleph\left(\begin{pmatrix}\bar{x}\\ \bar{y}\end{pmatrix}, \begin{pmatrix}\sigma_x^2 & r(x,y)\sigma_x\sigma_y\\ r(x,y)\sigma_x\sigma_y & \sigma_y^2\end{pmatrix}\right)$$

with r(x,y) the correlation between x and y.

For simplicity we used 1,000 simulations to compute $E\left(\frac{1}{\varepsilon'^2_x}\right)$ and $E\left(\frac{\varepsilon'_y}{\varepsilon'^3_x}\right)$.

This enables us to estimate the effect of variant g on z, based on the effect estimates of g on x (α) and y (β). For variants which only have a non-zero effect on height but not on FFM, the estimated effect and variance of variant g on z become

$$\hat{\gamma} = -2 \cdot \widehat{\beta'} \cdot E\left(\frac{\varepsilon'_y}{\varepsilon'^3_x}\right)$$

$$Var(\hat{\gamma}) = 4 \cdot E^2\left(\frac{\varepsilon'_y}{\varepsilon'^3_x}\right) Var(\widehat{\beta'})$$

For variants with no apparent evidence for an association with FFM (p>0.05/number of independent FFMI variants) but with evidence for an association with height (p<0.05/number of independent FFMI variants), the observed effect size of the variant g on FFMI($\tilde{\gamma}$) was compared with the expected effect size of variant g on FFMI ($\hat{\gamma}$) if the association of variant g was solely driven by the association with height.

$$\begin{cases} H_0: \tilde{\gamma} = \hat{\gamma} \\ H_a: \tilde{\gamma} \neq \hat{\gamma} \end{cases}$$

Under the null hypothesis, the observed and expected effect sizes of variant g on FFMI are equal. The association of variant g is thus considered to be solely associated with FFMI through height. If the observed and expected effect sizes of variant g on FFMI are significantly different, then the null hypothesis is rejected and the association of variant g with FFMI is considered to be at least partially driven through an association with FFM. Variants for which the null hypothesis could not be rejected were omitted from the follow-up analyses.

The effect sizes of the FFMI variants on FFM and height were assessed in GWAS on the European participants of UK Biobank, using the same phenotype adjustments and transformations and analytical strategies as for the GWAS for FFMI. The mean value and variance of BF% and FFMI and their correlation were assessed in European ancestry participants of UK Biobank and were based on untransformed phenotypes.

Clumping of BF% and FFMI variants and associations with related anthropometric traits

To obtain a list of unique, LD-independent loci associated with BF% and/or FFMI, we conducted LD clumping of the lead variants selected for BF% and FFMI in the sex-combined analyses. Genetic loci for BF% and FFMI were considered to represent the same locus if they were within 500 kb and in LD (R^2 ≥0.2). LD clumping was conducted in PLINK 2.0 alpha[208] (https://www.cog-genomics.org/plink/2.0/) and using an LD reference panel based on 10,000

randomly selected, unrelated UK Biobank participants of British ancestry. This same strategy was used to identify the number of loci identified for BF% and FFMI which were also reported for BMI in the latest, thus far largest genetic discovery for BMI, which was based on a meta-analyses of a GWAS in UK Biobank with the published GWAS published by the GIANT consortium (total N= 681,275)[172].

Effect sizes of variants associated with BF% or FFMI with height, BMI, waist-to-hip ratio (WHR) and WHR adjusted for BMI (WHRadjBMI) were tested based on GWAS on UK Biobank participants of European ancestry, using the same trait transformations and analytical strategy as for the GWAS for BF% and FFMI.

Genetic and observational correlations with other anthropometric traits

We assessed the genetic correlations between BF%, FFMI, BMI, height, WHR, WHRadjBMI, waist circumference (WC), WC adjusted for BMI (WCadjBMI), hip circumference (HC) and hip circumference adjusted for BMI (WCadjBMI) using cross-trait LD score regression[294,300]. Briefly, LD score regression allows estimation of the genetic correlation between two traits based on summary-level GWAS results for the two traits and an LD reference panel. The genetic covariance between the two traits is estimated based on the slope from the regression of the product of the z scores for the two respective traits versus the LD score, for all genetic variants. Analyses were based on sex-combined GWAS on European UK Biobank participants for inverse normally transformed phenotypes adjusted for age, age^2 and sex. LD scores calculated for the 1000 Genomes phase 3 European population (downloaded from https://github.com/bulik/ldsc) were used. Variants with MAF<0.1% or INFO<0.4, multi-allelic variants and variants in the HLA region were omitted from the analyses. Observational correlations between the 10 anthropometric traits were assessed based on the Pearson's correlation coefficients and using the same phenotype transformations as for the GWAS.

Identification of BF%, FFMI and BMI loci with disproportionate effect sizes on fat and lean mass

Loci identified for higher BF% and FFMI could purely reflect higher BMI or, alternatively, could be more strongly associated with BF% or FFMI, respectively, than expected for a BMI locus, given the observational correlations of BF% and FFMI with BMI and the genetic effect size of the variant on BMI. To test if the variants identified for BF% and FFMI disproportionately affected the respective phenotypes, the method for testing collider bias proposed by Aschard *et al.*[299] was re-appropriated to estimate the expected effect size ($\beta_{pheno,exp}$) if the association with the phenotype (i.e., BF% or FFMI) was solely driven by BMI. The expected effect size and its variance were calculated as follows:

$$\beta_{pheno,exp} = \beta_{BMI} \cdot cor(pheno, BMI)$$
$$Var(\beta_{pheno,exp}) = SE_{BMI}^2 \cdot cor(pheno, BMI)^2$$

If the observed effect size ($\beta_{pheno,obs}$) did not differ from the expected effect size $\beta_{pheno,exp}$, then the association of the locus with the phenotype was only driven by BMI. If the observed effect size was significantly different from the expected effect size, then the association with the phenotype could not be fully explained by BMI.

$$\begin{cases} H_0: \beta_{pheno,obs} = \beta_{pheno,exp} \\ H_a: \beta_{pheno,obs} \neq \beta_{pheno,exp} \end{cases}$$

The t statistic for the difference between $\beta_{pheno,obs}$ and $\beta_{pheno,exp}$ was calculated as follows:

$$t = \frac{\beta_{pheno,obs} - \beta_{pheno,exp}}{\sqrt{SE_{pheno,obs}^2 + SE_{pheno,exp}^2 + 2 \cdot cor(\beta_{pheno,obs}, \beta_{pheno,exp}) \cdot SE_{pheno,obs} \cdot SE_{pheno,exp}}}$$

If the null hypothesis is rejected and $|\beta_{pheno,obs}| > |\beta_{pheno,exp}|$, then the locus is more strongly associated with the phenotype than a BMI locus would. If the null hypothesis is rejected and $|\beta_{pheno,obs}| < |\beta_{pheno,exp}|$, then the effect on the phenotype is significantly less than would be expected for a locus driven by BMI only.

The observed effect sizes on FFMI and BF% were taken from the meta-GWAS results and the effect on BMI was assessed in GWAS on the European participants of UK Biobank, using the same phenotype adjustments and transformations and analytical strategies as for the GWAS for FFMI and BMI%. The correlation between BF% and BMI and FFMI and BMI was assessed in European ancestry participants of UK Biobank and were based on the same phenotype transformations as the GWAS.

The same approach was applied to identify reported BMI loci with disproportionate effect sizes on FFMI and/or BF%, based on the observational correlations of BMI with BF% and FFMI and the effect sizes of the BMI loci on BF% and FFMI. This was done for loci reported in the latest genetic discovery for BMI by Yengo *et al.*[172], but analyses were restricted to loci for which the marginal $p \leq 5 \times 10^{-9}$.

Results

Body fat percentage and fat-free mass index have a curvilinear relationship

The observational relationships between BF%, FFMI, BMI and other anthropometric phenotypes were assessed in the UK Biobank population. Women had a higher BF% than men, for a given BMI and overall (mean BF% for women: mean±SE= 36.5±6.9%, for men: 25.3±5.8%, p for sex difference $<2.2\times10^{-16}$). A positive, curvilinear association of BF% with BMI was observed in the UK Biobank population, with a stronger increase in BF% per unit increase in BMI at the lower end of the observed BMI values (Figure 3.4 A). FFMI was higher in men than in women (women: mean±SE= 16.8±1.7 kg/m^2, men: 20.6±1.9 kg/m^2, p for sex difference $<2.2\times10^{-16}$) (Figure 3.4 B). The curvilinear and linear relationships of BF% and FFMI, respectively, with BMI, resulted in a positive, curvilinear relationship between BF% and FFMI, with a stronger increase in BF% per unit increase in FFMI in the lower FFMI range (Figure 3.4 C).

The plots in Figures 3.4 A-C indicate that for a small group of women with a BMI >40 kg/m^2, the relationship of BMI with BF% and FFMI seems to follow a steeper and more linear slope. The vast majority of these outlying values came from participants from the assessment centre in Sheffield (Supplementary Figure 3.1 A-B). For around 160 of the 15,789 women who visited the centre in Sheffield, BF% follows a steeper relationship with BMI than for all other female participants from Sheffield (Supplementary Figure 3.1 D). The vast majority of the 160 women were of British ancestry and were between 50 and 60 years old, the latter of which made pregnancy an unlikely explanation for the outlying values. As this sub-group of women only represented 0.0007% of the total number of women in the study, this small number of deviating values are unlikely to have any detectable influence on the results of the genetic analyses. Therefore, these values were not excluded from the analyses.

The observational relationships of BF% and FFMI with related anthropometric traits was assessed based on observational correlations between the inverse normally transformed phenotypes adjusted for age, age^2, sex and assessment centre (Figure 3.4 D). Strong correlations with BMI were observed for both BF% (Pearson's correlation coefficient R=0.85) and FFMI (R=0.84), while the correlation between BF% and FFMI was less pronounced (R=0.47). BF% and FFMI also showed strong correlations with waist circumference (WC) (FFMI: R=0.80, BF%: R=0.83), hip circumference (HC) (BF%: R=0.69, FFMI: R=0.84) and waist-to-hip ratio (WHR) (BF%: R=0.62, FFMI: R=0.45), but these correlations became much weaker after adjustment for BMI. Weak

inverse correlations with height were observed for FFMI (R=-0.14) and BF% (R=-0.01). Because of the large sample sizes which were used to assess these correlations, all correlations were highly significant ($p<2.2\times10^{-16}$), even though the strength of correlation between some pairs of traits was weak.

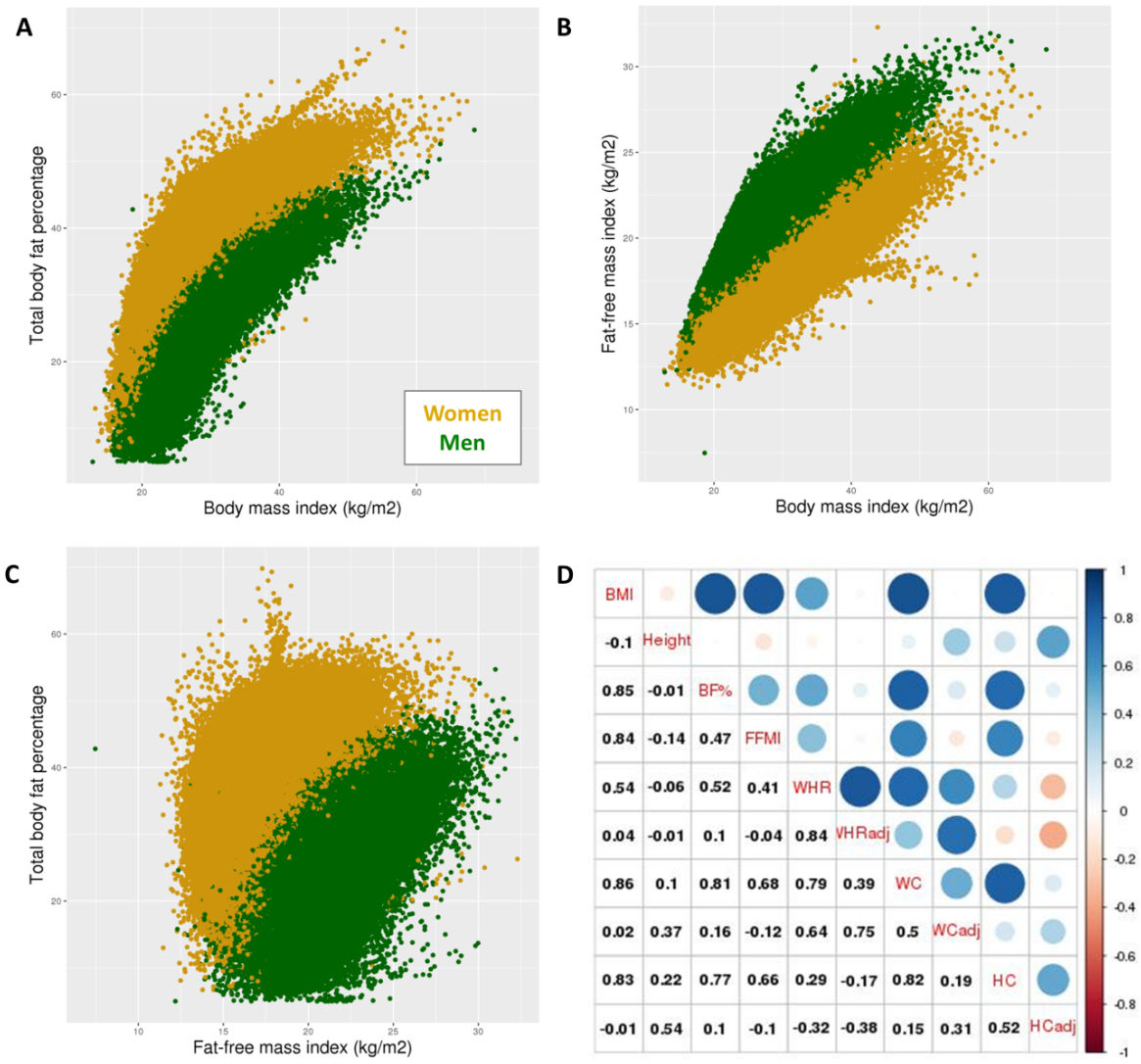


Figure 3.4: Scatter and correlation plots of total body fat percentage, fat-free mass index and related anthropometric phenotypes in UK Biobank. Plots are based on up to 207,585 male and 245,920 female participants of UK Biobank. **A|** Scatter plot of body mass index versus total body fat percentage (untransformed) **B|** Scatter plot of body mass index versus fat-free mass index (untransformed). **C|** Scatter plot of fat-free mass index versus total body fat percentage (untransformed). **D|** Plot of the Pearson's correlation coefficients between body fat percentage, fat-free mass index and related anthropometric traits. All phenotypes were adjusted for age, age^2, sex and assessment centre in linear regression models and the residuals were rank-based inverse normally transformed. BMI: body mass index, BF%: total body fat percentage, FFMI: fat-free mass index, WH: waist-to-hip ratio, WHadj: WH adjusted for BMI, WC: waist circumference, WCadj: WC adjusted for BMI, HC: hip circumference, HCadj: HC adjusted for BMI.

Exome chip meta-analyses reveal a low-frequency coding variant in *ZBTB7B* influencing FFMI

Five common coding variants associated with BF%

Based on exome chip meta-analyses for European ancestry only (N=107,902) and all ancestries combined (N=112,443), 40 genetic variants were associated with BF% at exome-wide significance level ($p \leq 2.0 \times 10^{-7}$) (Figure 3.5 upper panel and Supplementary Table 3.3). Five of these were coding, missense variants, which were all common (EAF>5%) (Table 3.1). rs11057401 in *CCDC92* was the only coding variant which only reached significance in the all ancestries meta-analysis. The *CCDC92* locus only reached significance for higher BF% in women and had a significantly lower effect on BF% in men (women: per-minor allele effect size on standard deviations (SDs) of BF%: beta=0.037, $p=1.0 \times 10^{-7}$; men: beta=0.005 , $p=0.45$, p for sex difference=9.8×10^{-4}). For 4 of the 5 significant coding variants, the association with BF% based on the additive model was most significant, while the association of rs2904880 in *CD19* only reached significance based on the recessive genetic model. All five coding variants had exome-wide or nominally (nominal significance threshold: $p \leq 1.25 \times 10^{-3}$, *CCDC92:* $p_{bmi}=0.001$) significant and directionally consistent effect sizes on BMI in published exome chip meta-analyses[280] which included 718,734 participants. The BF%-raising allele at rs11057401 in *CCDC92* is the only variant which has also been reported for body fat distribution in an exome chip study including 344,369 participants (sex-combined effect size on SDs on WHRadjBMI=-0.029, $p=7.3 \times 10^{-37}$)[292]. The sex-discordant effect of the *CCDC92* locus was also observed for WHRadjBMI (women: beta=-0.045, $p=7.3 \times 10^{-51}$; men: beta=-0.014, $p=4.7 \times 10^{-5}$, p for heterogeneity between sexes=7.2×10^{-11})[292]. The BF%-raising allele at rs11676272 in *ADCY3* was been reported for shorter height ($p=1.6 \times 10^{-32}$)[291] but none of the other BF% coding variants were associated with height.

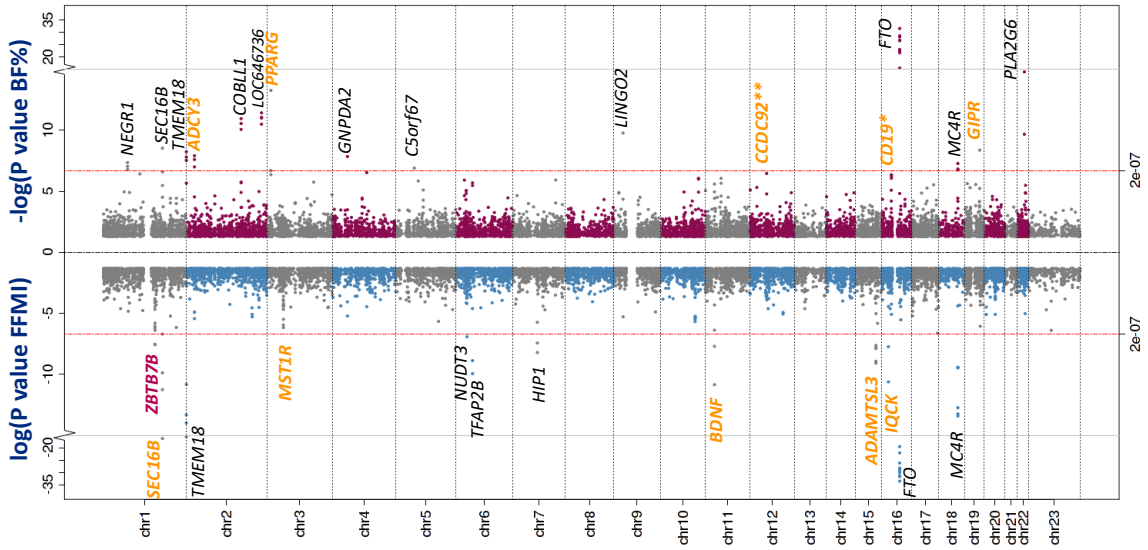


Figure 3.5: Manhattan plots for exome chip meta-analyses for BF% and FFMI. –10log(p-value) and log(p-values) are plotted for the meta-analyses of all ancestries based on the additive genetic model, for BF% (top) and FFMI (bottom), respectively. Loci in black only contain non-coding variants, loci in orange contain at least one common non-synonymous coding variant and the locus in red contains a low-frequency non-synonymous coding variant. * This locus reaches significance in the recessive sex-combined meta-analysis of the European ancestry only. ** This locus reaches significance in the meta-analysis of all ancestries for women, based on the additive model.

One low-frequency and 7 common coding variants for FFMI

Based on exome chip meta-analyses for FFMI including up to 98,035 participants, 40 genetic variants reached exome-wide significance (Figure 3.5, bottom panel and Supplementary Table 3.4). Eight of these were coding, of which all but one reached significance in the sex-combined, European ancestry meta-analysis based on the additive genetic model. Rs591120 in *SEC16B* only reached significance in the sex-combined meta-analysis of all ancestries (Table 3.1).

Rs141845046 in Zinc Finger and BTB Domain Containing 7B (*ZBTB7B*) was the only low-frequency coding variant that reached significance (c.670C>T, p.Pro224Ser, effect allele frequency (EAF)=3.1%, per-allele effect size on SDs of FFMI=0.074, $p=3\times10^{-8}$, p for heterogeneity between studies=0.70) (Table 3.1). The FFMI-raising allele at rs141845046 was not associated with BF% (per-allele effect size on SDs of BF%=0.018, $p=0.16$) and has previously been reported for higher BMI[280] (per-allele effect size on SDs of BMI=0.047, $p=4.2\times10^{-13}$), taller height[291] (per-allele effect size on SDs of height=0.058, $p=7.3\times10^{-17}$) and lower WHRadjBMI (per-allele effect size on SDs of WHRadjBMI=-0.037, $p=3.8\times10^{-8}$)[292]. This indicates that *ZBTB7B* increases BMI through lean mass without increasing overall or central body fat.

Seven common, coding variants in *SEC16B*, *MST1R*, *BDNF*, *ADAMTSL3* (3 variants) and *IQCK* were also significantly associated with FFMI. Apart from the three coding variants in *ADAMTSL3* (p_{bmi}>0.05), the common coding variants associated with FFMI also reached exome-wide significance for BMI[280], and effect sizes were directionally consistent between BMI and FFMI. The FFMI-raising alleles at the three coding variants in *ADAMTSL3* have previously been reported for shorter height (p-values ranging from 2.3×10^{-25} to 1.8×10^{-79}), while the FFMI-raising allele in *IQCK* was nominally associated with shorter height ($p=1.3\times10^{-4}$), which indicates that the associations of these variants with FFMI is at least partly driven by the inverse associations with height.

Table 3.2: Coding variants associated at exome-wide significance with BF% or FFMI. Chr; chromosome, pos: base-pair position, EA: effect allele, OA: other allele, EAF: effect allele frequency, sex-comb: sex-combined analyses, add: additive model, rec: recessive model.

rsid	chr	pos	Gene	Amino acid change	Analysis	EA	OA	EAF	Beta	p-value
BF%										
European ancestry only										
rs11676272	2	25,141,538	*ADCY3*	p.Ser107Pro	sex-comb add	G	A	0.455	0.025	4.0E-08
rs1801282	3	12,393,125	*PPARG*	p.Pro12Ala	sex-comb add	G	C	0.130	0.050	1.3E-13
rs2904880	16	28,944,396	*CD19*	p.Leu174Val	sex-comb rec	G	C	0.715	-0.053	1.3E-07
rs1800437	19	46,181,392	*GIPR*	p.Glu318Gln	sex-comb add	C	G	0.210	-0.030	5.7E-08
All ancestries combined										
rs11057401	12	124,427,306	*CCDC92*	p.Ser70Cys	women add	A	T	0.314	0.037	1.0E-07
FFMI										
European ancestry only										
rs141845046	1	154,987,704	*ZBTB7B*	p.Pro224Ser	sex-comb add	T	C	0.033	0.074	3.9E-08
rs1062633	3	49,924,940	*MST1R*	p.Arg1286Gly	sex-comb add	C	T	0.484	0.026	1.9E-07
rs6265	11	27,679,916	*BDNF*	p.Val66Met	sex-comb add	T	C	0.186	-0.041	1.1E-10
rs4483821	15	84,488,636	*ADAMTSL3*	p.His146Arg	sex-comb add	G	A	0.447	-0.028	7.2E-09
rs4842838	15	84,582,124	*ADAMTSL3*	p.Val661Leu	sex-comb add	T	G	0.516	-0.029	1.4E-08
rs34047645	15	84,611,367	*ADAMTSL3*	p.Gly713Arg	sex-comb add	C	G	0.183	0.041	7.0E-09
rs7191155	16	19,800,213	*IQCK*	p.Leu220Pro	sex-comb add	C	T	0.162	-0.039	5.1E-09
All ancestries combined										
rs591120	1	177,902,753	*SEC16B*	p.Pro864Ala	sex-comb add	C	G	0.435	0.025	2.0E-07

Genome-wide association analyses for BF% and FFMI

685 conditionally independent genetic variants identified for BF%

Based on meta-analyses of the UK Biobank GWAS with the published GWAS results, leading to a total sample size of up to 531,566 participants (286,110 women) of European ancestry, 685 conditionally independent variants reached genome-wide significance for BF% ($p \leq 5\times10^{-9}$) (Figure 3.6, bottom; Supplementary Table 3.5). The 5 most significant loci for BF% were located in *FTO* ($p=9.9\times10^{-189}$), *SEC16B* ($p=2.8\times10^{-59}$), *ADCY3* ($p=8.6\times10^{-59}$), *TUFM* ($p=3.3\times10^{-56}$) and *MC4R* ($p=1.0\times10^{-54}$), and are all known as major loci for BMI. The variant with the largest effect size on BF% was the rare, stop gained variant (rs150090666, c.2347C>T, p.Arg783Ter) in Phosphodiesterase 3B (*PDE3B*), of which each minor allele (EAF=0.1%) was associated with 0.150 SDs higher BF% ($p=2.6\times10^{-9}$). This association was based on the imputed variant in UK Biobank only, but the association of the genotyped variant in UK Biobank was very similar (beta=0.199, $p=2.6\times10^{-9}$). This variant had a scaled CADD score of 35, which indicates that it is predicted to be among the 0.03% most deleterious variants in the human genome[301].

Limited evidence was found for sex-discordant effects of the genetic loci identified for BF%. Of the 685 genetic loci identified for BF%, 10 and 14 loci only reached significance in the women-only and men-only analyses, respectively (Supplementary Table 3.5). Of the 661 variants which reached significance in the sex-combined analyses, a significant sex difference (Bonferroni corrected p for 661 tests $\leq7.6\times10^{-5}$) in the effect sizes on BF% in men versus women was identified for three variants, which all had stronger effects in men than women and did not reach significance in women alone (Supplementary Table 3.5). The first variant was located near *LOC646736* (rs952227, women: beta±SE on SDs of BF% =0.011±0.003, men: beta±SE =-0.038±0.003, p for sex difference = 2.0×10-10). The two other variants were located in the non-pseudo-autosomal region of the X chromosome, near *FAM9B* (rs5934505, women: beta±SE on SDs of BF% =0.003±0.004, men: beta±SE =-0.024±0.003, p for sex difference = 2.7×10^{-5}) and near *EDA2R* (X:66222744_CAA_C, women: beta±SE on SDs of BF% =0.003±0.003, men: beta±SE =-0.023±0.002, p for sex difference = 2.9×10^{-8}). I also found a strong correlation between the summary-level GWAS results for BF% in men and women based on LD score regression (R=0.89±0.01), which confirms that genetic effects on BF% are largely shared between men and women.

The 12 previously reported loci for BF% also reached significance in the present study (p of lead variant at locus $\leq5\times10^{-9}$, lead variant within 500 kb of and in high LD ($R^2>0.5$) with the published lead variant). Nearly a third (215 of the 685 independent variants, or 31.4%) of the genetic loci

which reached genome-wide significance for BF% have also been reported for BMI in up 681,275 participants by Yengo *et al.*[172].

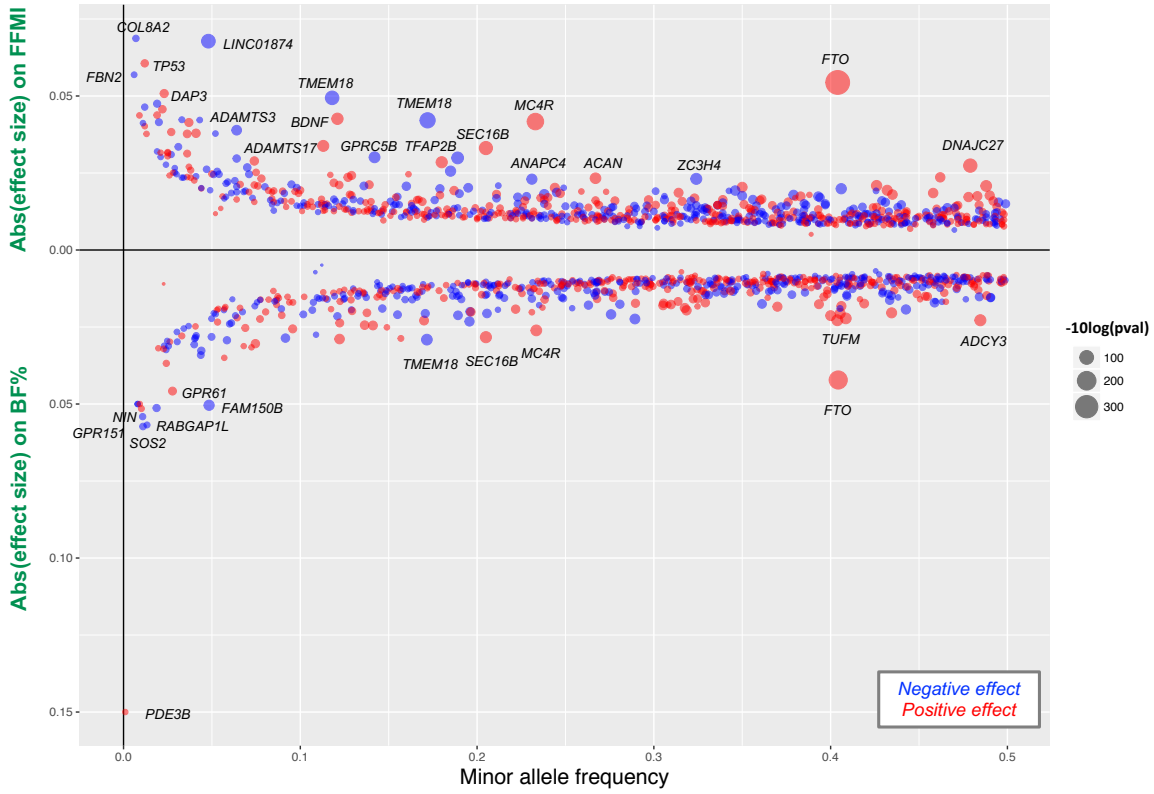


Figure 3.6: scatter plot of the minor alleles frequencies versus the effect sizes of the independent variants selected for FFMI and BF%. 685 independent variants reaching genome-wide significance in the meta-analysis of the GWAS in UK Biobank and the published GWAS results by Lu *et al.* were plotted for BF%. For FFMI, the 787 independent variants which reached genome-wide significance in the meta-analysis of the GWAS in UK Biobank and the exome chip data and for which the association was not solely driven by height were plotted. The y axis represents the absolute effect sizes on standard deviations of FFMI (top) and BF% (bottom). Red dots indicate positive and blue dots negative effect sizes of the minor alleles. The diameter of the dots are scaled to the -10log(p-value).

906 independent variants identified for FFMI

Based on meta-analyses of the UK Biobank GWAS and the exome chip data, leading to a total sample size of up to 535,885 participants (287,956 women), 906 conditionally independent variants reaching genome-wide significance ($p \leq 5 \times 10^{-9}$) were identified for FFMI (Supplementary Table 3.6). The five loci most significantly associated were located in/near *FTO* ($p=2.1 \times 10^{-346}$), *MC4R* ($p=1.0 \times 10^{-151}$), *TMEM18* ($p=2.2 \times 10^{-126}$), *LINC01874* ($p=3.0 \times 10^{-98}$) and *ADAMTSL3* ($p=1.1 \times 10^{-92}$), of which the first three are major BMI and the latter is a well-known locus for height.

As for BF%, little evidence for sex-dimorphic genetic loci for FFMI was found. Of the 906 lead variants, 8 only reached significance in women and 6 were specific to men (Supplementary Table

3.6). Of the 892 variants identified in the sex-combined analyses, only 3 variants had significant sex differences in their effect sizes on FFMI (Bonferroni corrected p-value for 892 tests $\leq 5.6\times10^{-5}$). Two variants in close proximity had stronger effects in women (rs713586 near *ADCY3* and *DNAJC27* and 2:25365642_TAATA_T in *ERF3B*) while rs76895963 in *CCND2-AS1* had a stronger effect on FFMI in men (Supplementary Table 3.6). A strong, positive correlation was found between the summary-level GWAS results for BF% in men and women based on LD score regression ($R=0.90\pm0.01$), which confirms the shared genetic architecture of FFMI between men and women.

Four of the 5 genetic loci (*FTO*, *IRS1*, *VCAN* and *ADAMTSL3*) which were reported for total or appendicular lean mass, adjusted for age, age^2, height and fat mass in the published GWAS by Zillikens *et al.*[302] were replicated in the present GWAS for FFMI (p of lead variant at locus $\leq 5\times10^{-9}$, lead variant within 500 kb of and in high LD ($R^2>0.5$) with the published lead variant). The reported locus in *HSD17B11* (sentinel variant rs9991501) was not in LD with any of the loci in the region that reached genome-wide significance in the present analysis, nor was the sentinel variant itself associated with FFMI (rs9991501: $p=0.21$). The adjustment for different covariates, in particular total fat mass, or a false positive finding in the study by Zillikens *et al.*[302] are possible reasons why this locus was not associated with FFMI in the present analysis. More than a third of the identified loci for FFMI (312 of the 906 or 34.5%) were also reported for BMI by Yengo *et al.*[172].

120 of the 906 genetic loci identified for FFMI are driven by collider bias due to height squared

To test if the association of any of the genetic variants identified for FFMI were the result of collider bias due to the adjustment for height squared, we assessed the association of all 906 FFMI variants with height and fat-free mass (FFM). For the variants associated with height ($p_{height}\leq 5.5\times10^{-5}$) but not with FFM ($p_{FFM}>5.5\times10^{-5}$), the observed effect size of the variant on FFMI was compared with the expected effect size if the association would by driven by height squared only. If the observed and expected effect size were not significantly different then the variant was associated solely through height and was omitted from further analyses.

601 of the 906 genetic loci identified for FFMI were significantly associated with FFM (p for 906 tests $\leq 5.5\times10^{-5}$) (Figure 3.7, blue dots). 156 other variants were not associated with FFM nor with height (Figure 3.7, green dots). These variants were weakly associated with FFMI and may therefore not have reached significance for FFM. The remaining 228 loci did not reach significance for FFM ($p>5.5\times10^{-5}$) but were associated with height ($p\leq 5.5\times10^{-5}$, Figure 3.7, red and yellow

dots). For 108 of these 228 variants, the observed effect size on FFMI was significantly different from the predicted effect size if the association would only be driven by height squared (Figure 3.7, yellow dots), which indicates that their associations with FFMI could not be fully explained by height. For the other 120 variants which reached significance for height only, the observed effect size on FFMI did not differ significantly from the expected effect size (Figure 3.7, red dots). The associations of these 120 variants with FFMI were therefore likely to be solely driven by the adjustment for height squared and they were therefore not considered for further analyses (Supplementary Table 3.7). In conclusion, we identified 787 independent genetic loci for FFMI which were not solely driven by height. 773 loci reached significance in the sex-combined analyses, while 6 were only associated in men and 8 only in women (Figure 3.6, top). This subset of FFMI variants were taken forward for further analyses.

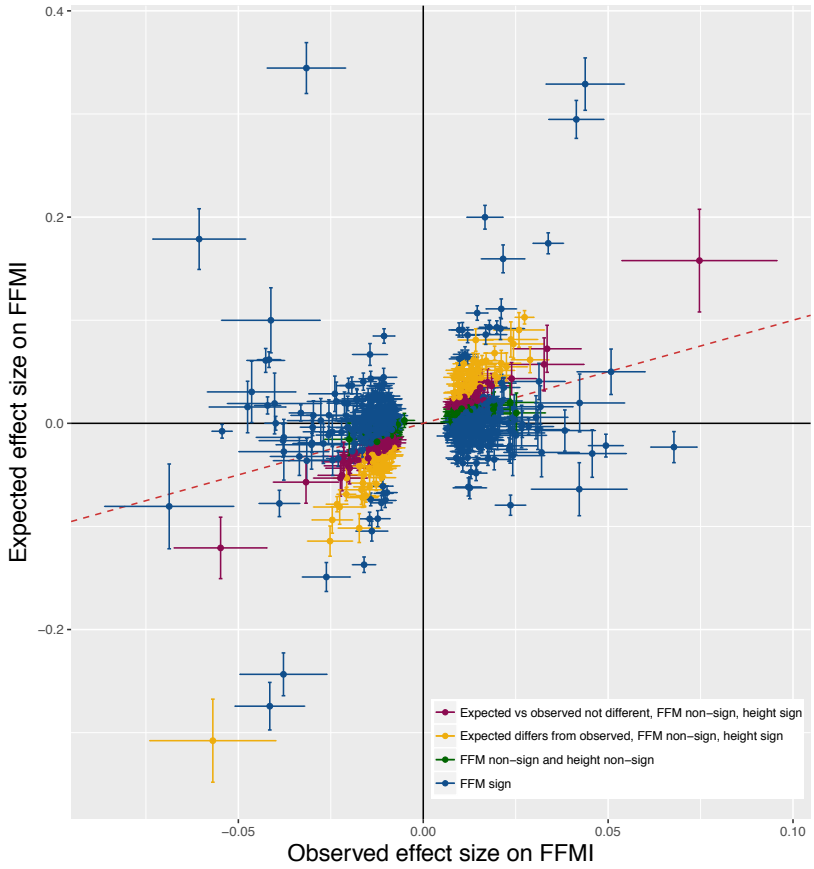


Figure 3.7: Scatter plot of the observed effect size on FFMI versus the expected effect size on FFMI if the association is entirely driven by the adjustment for height squared. The plot includes the 906 variants identified for FFMI in sex-combined and sex-specific analyses. Variants in blue reached Bonferroni-corrected significance for FFM ($p \leq 5.5\times10^{-5}$), variants in green were not significant for FFM, nor for height ($p > 5.5\times10^{-5}$). Variants in yellow and red reached significance for height but not for FFM. The expected and observed effect sizes on FFMI were significantly different ($p \leq 5.5\times10^{-5}$) for the yellow variants but not for the red variants. The red dashed line represents equal observed and expected effect sizes on FFMI. FFM: fat-free mass; FFMI: fat-free mass index.

Genomic inflation, polygenicity and SNP-based heritability

Substantial genomic inflation was observed in the meta-GWAS of both BF% and FFMI (BF%: $\lambda_{GC} = 2.456$; FFMI: $\lambda_{GC} = 2.6745$). LDSC was conducted to assess to what extent this inflation was due to polygeneticy and population stratification or cryptic relatedness. The LDSC intercepts and attenuation ratios based on the Baseline LD model indicated that this inflation was due to polygenicity (BF%: LDSC intercept ±SE = 1.124±0.019, attenuation ratio ± SE =0.054±0.008; FFMI: LDSC intercept ±SE = 1.160±0.025, attenuation ratio ± SE =0.058±0.009). Similar and slightly higher attenuation ratios have been reported for GWAS for other phenotypes in UK Biobank[293]. The SNP-based heritability for BF% estimated using LDSC was h^2±SE =23.3±0.6%, while for FFMI the SNP-based heritability was h^2±SE =30.2±0.9%.

Associations of BF% and FFMI loci with related anthropometric traits

Substantial overlap in genetic loci identified for BF% and FFMI

The genetic loci identified for BF% and FFMI overlapped substantially, with the majority of shared loci affecting both traits in the same direction. Of the 1,216 LD-independent loci associated with BF% (661 loci) and/or FFMI (787 loci) in the sex-combined analyses, 202 (16.6%) were genome-wide significantly associated with both BF% and FFMI. 173 (86.5%) of the 202 shared loci had directionally consistent effects on BF% and FFMI (Figure 3.8, green dots), while 29 (14.3%) were associated with both phenotypes in opposite direction (Figure 3.8, blue dots). 447 of the 1,216 loci (36.8%) only reached genome-wide significance for BF% (Figures 3.8, pink dots) and 567 (46.6%) only for FFMI (Figure 3.8, yellow dots). However, of the variants only selected for BF%, 140 (32.5%) also reached nominal significance (Bonferroni-corrected significance threshold for 1,216 tests: $p \leq 4.1 \times 10^{-5}$) for FFMI, of which 14 had an opposite direction of effect on FFMI. 151 (26.6%) of the variants only selected for FFMI also reached nominal significance for BF%, of which 19 had opposite effects on FFMI and BF%. Overall, around 60% of the 1,216 identified loci were specific to one of the two traits, with 307 only affecting BF% and 416 only affecting FFMI. The remaining 40% were at least nominally significant for both phenotypes, of which approximately half reached genome-wide significance for both traits.

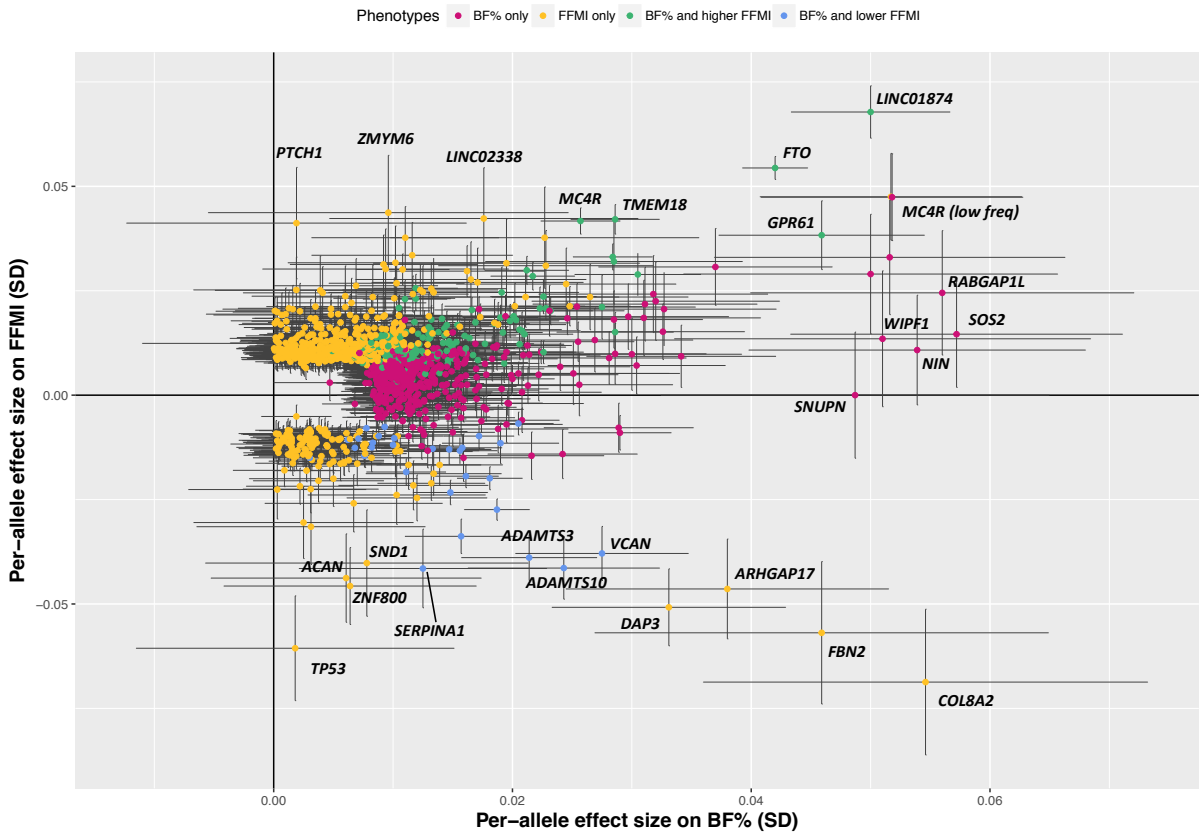


Figure 3.8: Scatter plot of the effect sizes of the 1,216 LD-independent genetic loci on BF% versus FFMI. Effect sizes are aligned to the BF%-raising allele. Variants in red reached genome-wide significance for BF% only, variants in yellow for FFMI only, and variants in green and blue for both BF% and FFMI with the same (green) and opposite (blue) directions of effect on BF% and FFMI. Because of its large effect size on BF%, the locus in *PDE3B* was omitted from this plot. The effect size of *PDE3B* on FFMI was non-significant: beta=0.030, p=0.21.

Association of BF% and FFMI loci with related anthropometric traits

As evidence was found that a substantial proportion of the genetic determinants of BF%, FFMI and BMI are shared and that BF% and FFMI are observationally correlated with other anthropometric phenotypes, I next asked if genetic loci associated with BF% and FFMI also affect other phenotypes reflecting body size and composition. The genetic associations of the 1,216 loci for BF% or FFMI with related anthropometric traits, i.e., BMI, height, WHR and WHRadjBMI were assessed based on GWAS on the UK Biobank participants of European ancestry. For the vast majority of 649 BF%-raising loci, of which 202 were common with FFMI, a similar, positive association with BMI was found (Figure 3.9, upper left). 526 of the 649 BF% loci reached at least nominal significance for BMI (Bonferroni-corrected significance threshold for 1,216 tests: $p \leq 4.1 \times 10^{-5}$), of which all but two had the same directions of effect on both BMI and BF% (Figure 3.9, upper left). The BF%-raising alleles at the common missense variant rs72755233 in *ADAMTS17* and intronic variant rs61327788 in *COQ10B* were significantly associated with *lower*

BMI (rs72755233: per-BF%-raising allele effect size on SDs of BMI = -0.016, p=5.7×10^{-8}; rs61327788: beta= -0.010, p=9.3×10^{-8}). Although no consistent pattern of association of the BF% loci with height was observed, 189 BF%-raising loci were also associated with height, of which 107 loci were positively and 82 were negatively associated with height, including several well-established height loci, such as *ADAMTS10*, *ADAMTS17* and *MAP3K3* (Figure 3.9, upper right). More than a third (255 of the 649 loci) of the BF%-raising loci also reached nominal significance for higher WHR, which may be due to the observational correlation of BF% with WHR (Figure 3.9, bottom left). However, 27 of the BF%-raising genetic regions were associated with *lower* WHR. After adjusting WHR for BMI, only 27 BF%-raising loci remained associated with higher central-to-lower body fat distribution; while 46 BF% loci were associated with lower WHRadjBMI (Figure 3.9, bottom right). Among the genetic loci associated with higher BF% but lower WHR or WHRadjBMI were well-known loci for body fat distribution, including *DNAH10*, *VEGFA* and in particular the low-frequency variant rs72959041 in *RSPO3* which has a particularly strong effect size on lower WHR (beta=-0.130, p=1.9×10^{-169}) and WHRadjBMI (beta=-0.165, p=1.3×10^{-167}) This indicates that these loci increase overall body fat mainly through allowing more fat accumulation in the lower body instead of in the waist area.

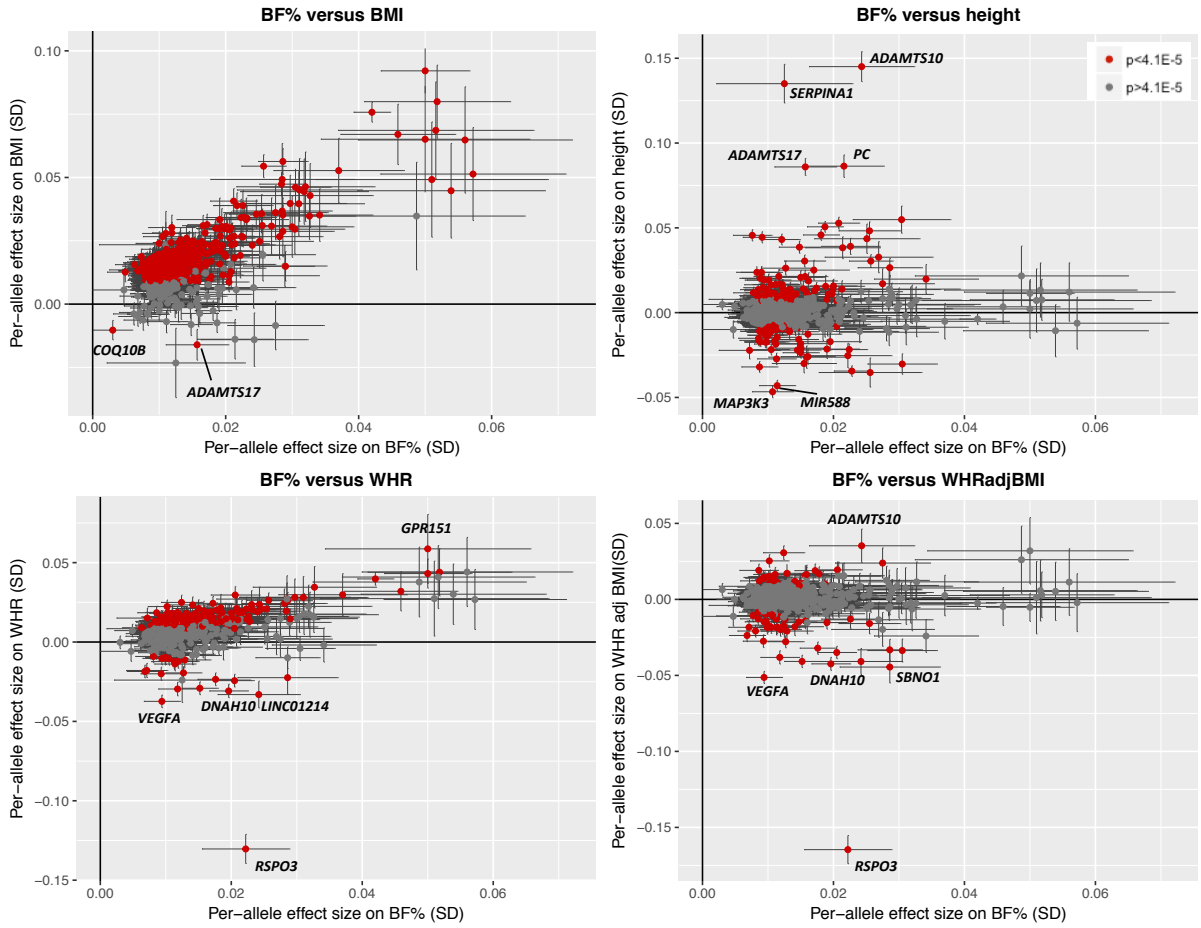


Figure 3.9: Scatter plots of the effect sizes of loci selected for BF% on BF% versus related anthropometric traits. Effect sizes are plotted for the 658 variants selected for BF% of which 202 were also selected for FFMI. All effect sizes are aligned to the BF% raising allele. Variants in red are significantly (Bonferroni corrected p-value $\leq 4.1\times10^{-5}$) associated with the trait on the y axis. Effect sizes on BMI, height, WHR and WHRadjBMI were estimated through GWAS on inverse-normally transformed traits on the UK Biobank participants of European ancestry.

The rare, stop gained variant rs150090666 in *PDE3B* with a large effect size on BF% (beta=0.150, $p=2.6\times10^{-9}$) but not on FFMI (beta=0.030, $p=0.21$) was strongly associated with taller height (beta=0.228, $p=1.4\times10^{-16}$) and lower WHRadjBMI (beta=-0.241, $p=1.3\times10^{-11}$). The variant also reached nominal significance for higher BMI (beta=0.132, $p=2.0\times10^{-5}$) but not for lower WHR (beta=-0.131, $p=2.0\times10^{-4}$) (Figure 3.10). Similar to *RSPO3*, this locus thus increases body fat by disproportionately increasing fat storage in the hip region.

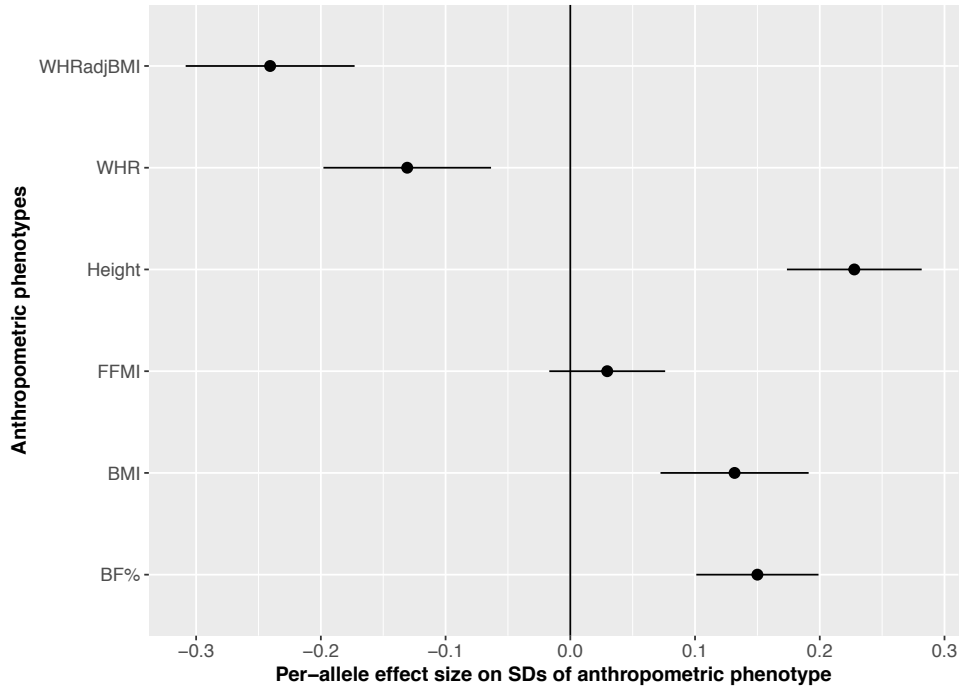


Figure 3.10: Forest plot of the effect sizes of the stop-gained variant in *PDE3B* on BF%, FFMI and related anthropometric phenotypes. All effect sizes were from GWAS in UK Biobank only and are expressed as per-allele effect size on standard deviations of the phenotype.

As for the loci identified for BF%, the vast majority of loci identified for higher FFMI showed similar and positive effect sizes on BMI (Figure 3.11, upper left). 611 out of the 769 loci associated with FFMI reached at least nominal significance ($p \leq 4.1 \times 10^{-5}$) for BMI. Rs11118310 near *LYPLAL1* was the only FFMI-raising locus which was associated with *lower* BMI (beta for BMI per FFMI-raising allele=-0.009, $p=1.6 \times 10^{-5}$). Although these analyses excluded the FFMI loci for which the associations were entirely driven by inverse effects on height, 88 (11%) FFMI-raising variants were associated with shorter height, while 76 reached nominal significance for taller height (Figure 3.11, upper right). As expected given the positive correlation between FFMI and WHR, 172 of the FFMI-raising loci were also associated with higher WHR, while 17 loci were associated with lower WHR (Figure 3.11, bottom left). Conversely, of the 100 FFMI-raising loci which reached at least nominal significance for WHRadjBMI, 92 were associated with *lower* WHRadjBMI (Figure 3.11, bottom right). The locus near *LYPLAL1* was one of the exceptions, as the FFMI-raising and BMI-lowering allele was associated with *higher* WHRadjBMI.

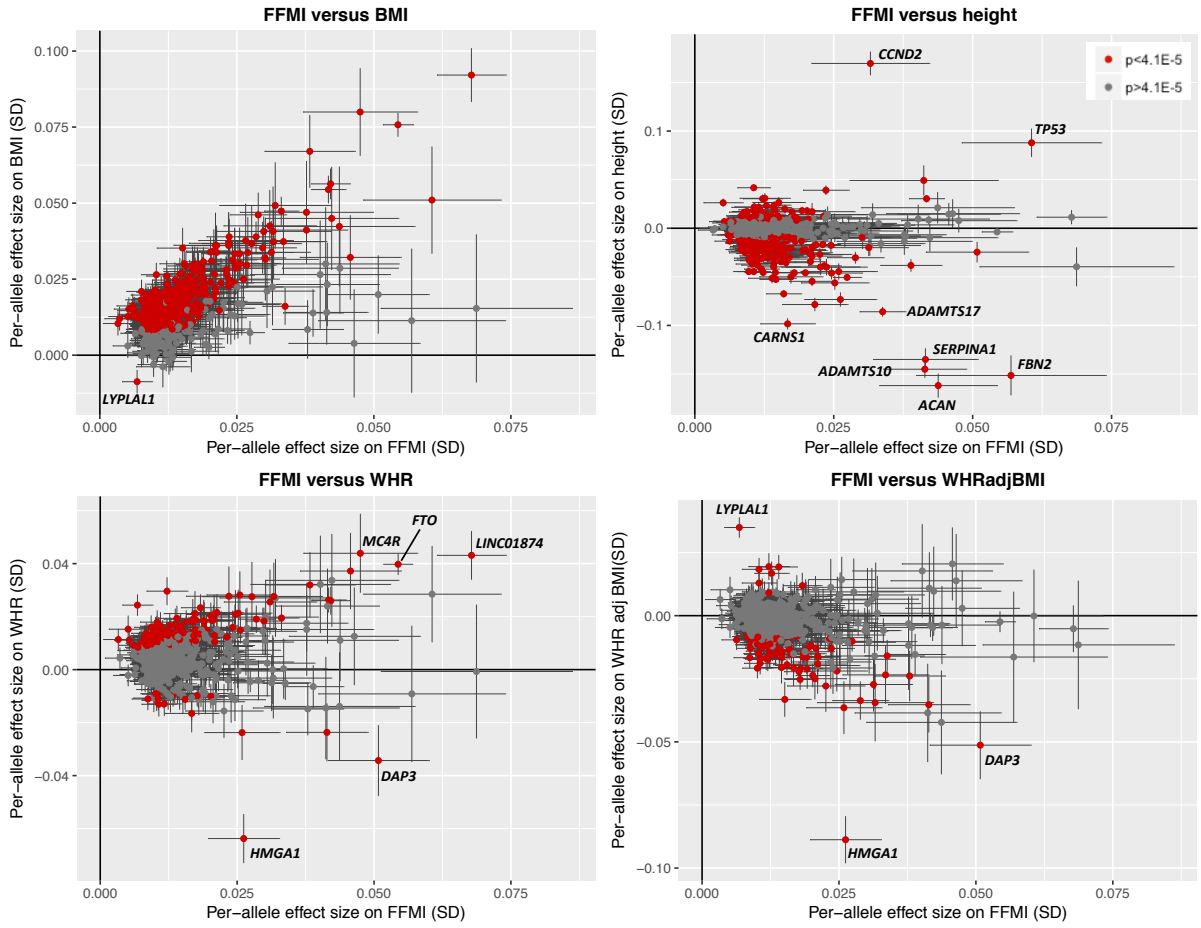


Figure 3.11: Scatter plots of effect sizes on FFMI versus related anthropometric traits for the loci selected for FFMI. Effect sizes are plotted for the 769 variants selected for FFMI of which 202 were also selected for BF%. All effect sizes are aligned to the FFMI raising allele. Variants in red are significantly (Bonferroni corrected p-value $\leq 4.1\times10^{-5}$) associated with the trait on the y axis. Effect sizes on BMI, height, WHR and WHRadjBMI were estimated through GWAS on inverse normally transformed traits in UK Biobank participants of European ancestry.

Genetic correlations of FFMI and BF% with other anthropometric traits mirror observational correlations

As strong association patterns of the genetic loci reaching significance for BF% and FFMI with related anthropometric phenotypes were found, I next extended this exercise to the entire genome using cross-trait LD score regression analyses. Based on GWAS in UK Biobank, I assessed the genome-wide genetic correlations between BF%, FFMI and eight related anthropometric traits, i.e., BMI, height, WHR, WHRadjBMI, waist circumference (WC), WC adjusted for BMI (WCadjBMI), hip circumference (HC) and hip circumference adjusted for BMI (HCadjBMI) (Figure 3.12). BF% and FFMI were genetically moderately correlated (R=0.49, $p=8\times10^{-154}$) (Bonferroni corrected p-value threshold for 17 tests <0.003), and the strength of the genetic correlation was very similar to the observational correlation of BF% and FFMI (observational

correlation coefficient: R=0.47). The strongest genome-wide genetic correlations of BF% and FFMI were found with BMI (BF%: R±SE=0.849±0.006; FFMI: R±SE=0.879±0.005), WC (BF%: R±SE=0.878±0.006; FFMI: R±SE=0.68±0.01) and HC (BF%: R±SE=0.813±0.008; FFMI: R±SE=0.66±0.02). WHR was genetically more strongly associated with BF% (R±SE=0.62±0.02) than with FFMI (R±SE=0.44±0.02). However, after adjusting the fat distribution traits for BMI, their genetic correlations with BF% became weak (WCadjBMI: R±SE=0.14±0.02; HCadjBMI: R±SE=0.10±0.02, WHRadjBMI: R±SE=0.08±0.02). FFMI was weakly correlated with the BMI-adjusted fat distribution traits (WCadjBMI: R±SE=-0.36±0.02; HCadjBMI: R±SE=-0.24±0.02, WHRadjBMI: R±SE=-0.18±0.02), which is likely to be caused by the strong genetic correlation of BMI with FFMI. Height was inversely correlated with FFMI (R±SE=-0.22±0.02) but was not associated with BF% (p=0.35). The genome-wide genetic correlations of BF% and FFMI with related anthropometric traits were very similar the observational correlations (Figure 3.12).

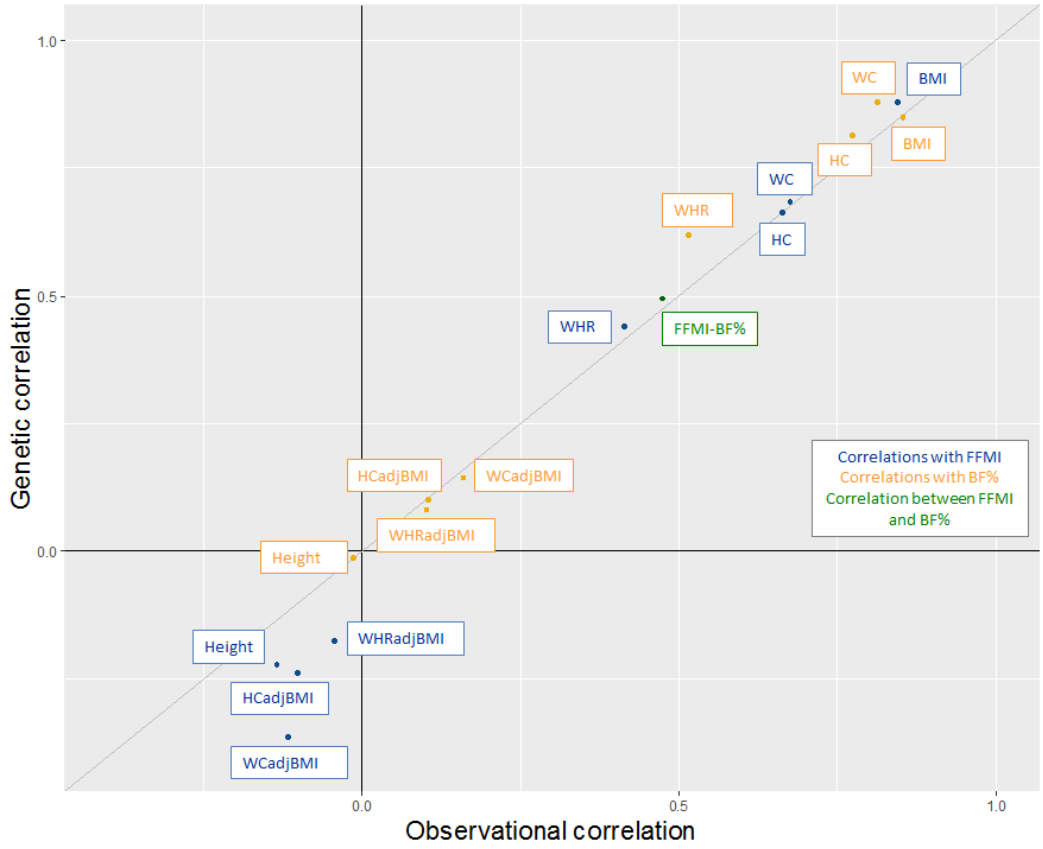


Figure 3.12: Genetic and observational correlations of total body fat percentage and fat-free mass index with related anthropometric traits in UK Biobank. All traits were adjusted for age, age^2 and assessment centre and residuals were inverse normally transformed. Correlations with body fat percentage are shown in yellow and with fat-free mass index in blue. Correlations between body fat percentage and fat-free mass index are shown in green. The confidence intervals of the genetic and observational correlations were too narrow to visualize on the plot. HC: hip circumference, HCadjBMI: HC adjusted for BMI, WC: waist circumference, WCadjBMI: WC adjusted for BMI.

16 loci act on BF% or FFMI beyond BMI

Given the strong observational and genetic correlations of both BF% and FFMI with BMI, several of the identified loci associated with higher BF% or FFMI may purely reflect higher BMI. To identify the genetic loci which affect BF% or FFMI more strongly than a BMI locus would, given the correlations of both phenotypes with BMI, I calculated for each of the 1,216 BF% and FFMI loci the predicted effect size if the effect of that variant on BF% or FFMI were only due to an effect on BMI (Figure 3.13). Of 649 loci associated with higher BF%, 16 loci (*ADAMTS3*, *LYPLAL1*, *MTMR11*, *ADAMTSL3*, *RIN3*, *UBE2Q2P1*, *ADAMTS14*, *LINC01184*, *PLCE1*, *TNP1*, *LIN28B*, *ACAN*, *VCAN*, *ADAMTS10*, *ADAMTS17* and *LOC646736*) had an effect size on BF% that was significantly larger (Bonferroni-corrected p for 1,216 tests $\leq 4.1\times10^{-5}$) than the predicted effect size if the association with BF% had been due to BMI only. For 62 additional loci, the effect size on BF% was higher than the predicted effect size at nominal significance ($p<0.05$). Four BF% loci had a significantly *lower* effect size on BF% than expected if its association only reflected BMI (*FTO*, *MC4R*, *SEMA3F-AS1*, *TMEM18* and *LINC01874*), while 36 other loci had a nominally lower than predicted effect on BF% (Figure 3.13, left panel).

Of the 769 FFMI-raising loci, 11 had a significantly ($p \leq 4.1\times10^{-5}$) stronger effect on FFMI than expected if the variant was associated through BMI only (*ADAMTS3, LYPLAL1, MTMR11, ADAMTSL3, UBE2Q2P1, ADAMTS14, PLCE1, ACAN, VCAN, ADAMTS17* and *LOC646736*) and 71 loci had an observed effect size that was higher than the expected effect size at nominal significance. *TUFM* was the only FFMI-raising locus which had a significantly *lower* observed compared to predicted effect size on FFMI. 35 loci had a nominally lower observed effect size on FFMI than expected if was driven by BMI only (Figure 3.13, right panel).

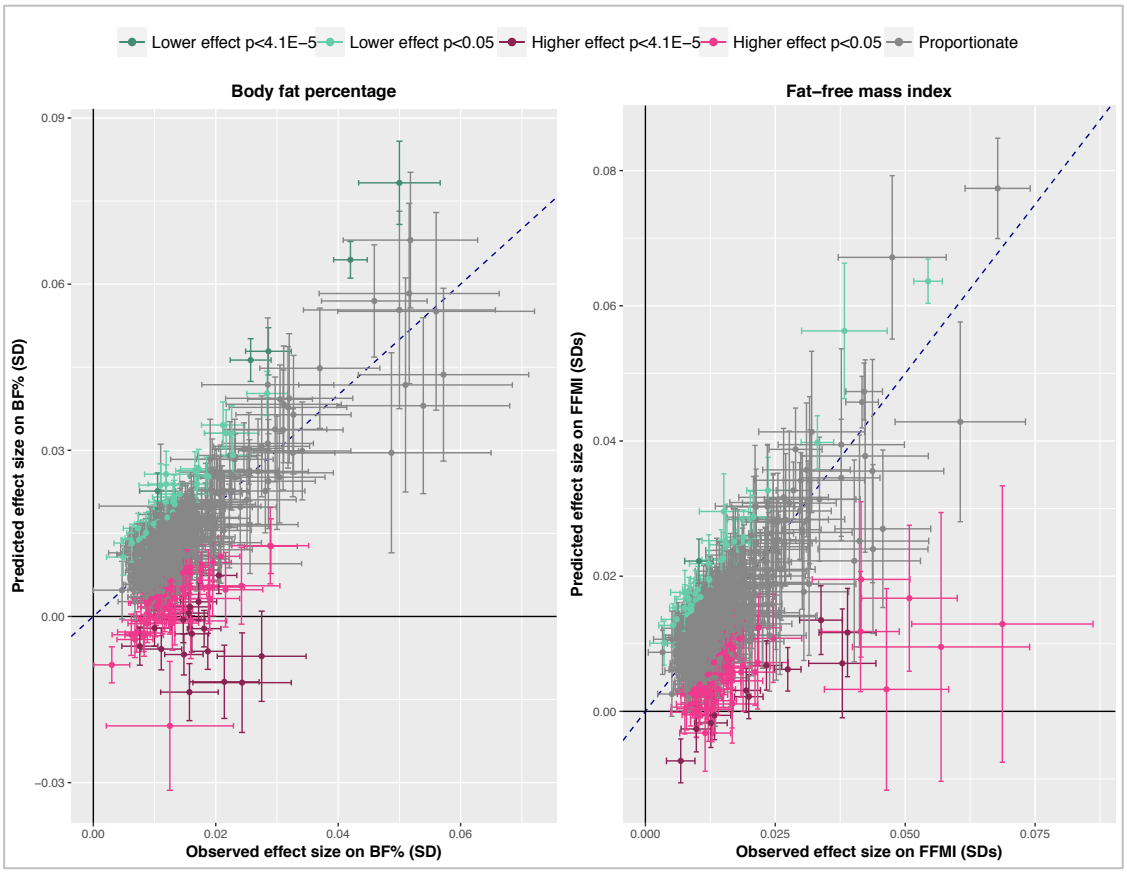


Figure 3.13: Scatterplots of observed versus predicted effect sizes on FFMI and BF% of the genetic loci associated with BF% and/or FFMI. The predicted effect size on the y axis reflects the expected effect size on BF% (left) or FFMI (right) if the association is entirely due to the association of the variant with BMI and were calculated based on a GWAS for BMI in the European ancestry participants of UK Biobank and the correlation of BF% and FFMI with BMI. The observed effect sizes on BF% and FFMI come from the meta-GWAS results. For the variants in pink, the observed effect size was stronger than the predicted effect size (dark pink: Bonferroni-corrected significance: $\leq 4.1\times10^{-5}$, light pink: nominal significance: $p<0.05$). The variants in green had a lower observed than predicted effect size (dark green: Bonferroni-corrected significance: $\leq 4.1\times10^{-5}$, light green: nominal significance: $p<0.05$). For variants in grey, the predicted effect size did not differ from the observed effect size ($p>0.05$) The dashed blue line indicates equal observed and predicted effects.

Overall, 16 loci had stronger effects on both BF% and FFMI than would be expected if the associations had only been driven by BMI (*ADAMTS3, LYPLAL1, MTMR11, ADAMTSL3, UBE2Q2P1, ADAMTS14, PLCE1, ACAN, VCAN, LOC646736* and *ADAMTS17*) or on BF% only (*RIN3, LINC01184, TNP1, LIN28B* and *ADAMTS10*). After aligning the alleles to the BF%-increasing allele, all 11 loci that were more strongly associated with both BF% and FFMI had an inverse effect on FFMI (Figure 3.14A, yellow variants). The 5 variants with disproportionate effects on higher BF% only also all had negative effects on FFMI (Figure 3.14A, red variants). None of the 16 variants reached genome-wide significance for BMI ($p>5\times10^{-9}$), but the BF%-raising alleles at 6 loci reached nominal significance for *lower* BMI (rs10518106 near *ADAMTS3*, beta for BMI=-0.014, $p=8.4\times10^{-4}$; rs11259936 in *ADAMTSL3*, beta=-0.007, $p=1.7\times10^{-4}$;

rs2250127 in *LINC01184*, beta=-0.007, p=2.3×10^{-3}; rs314274 in *LIN28B*, beta=-0.006, p=1.9×10^{-3}; rs3817428 in *ACAN*, beta=-0.008, p=1.7×10^{-4} and rs72755233 in *ADAMTS17*, beta=-0.016, p=5.7×10^{-8}), while the BF%-raising allele near *LYPLAL1* was nominally associated with higher BMI (rs11118310, beta for BMI =0.009, p=1.6×10^{-5}) (Figure 3.14 B and C).

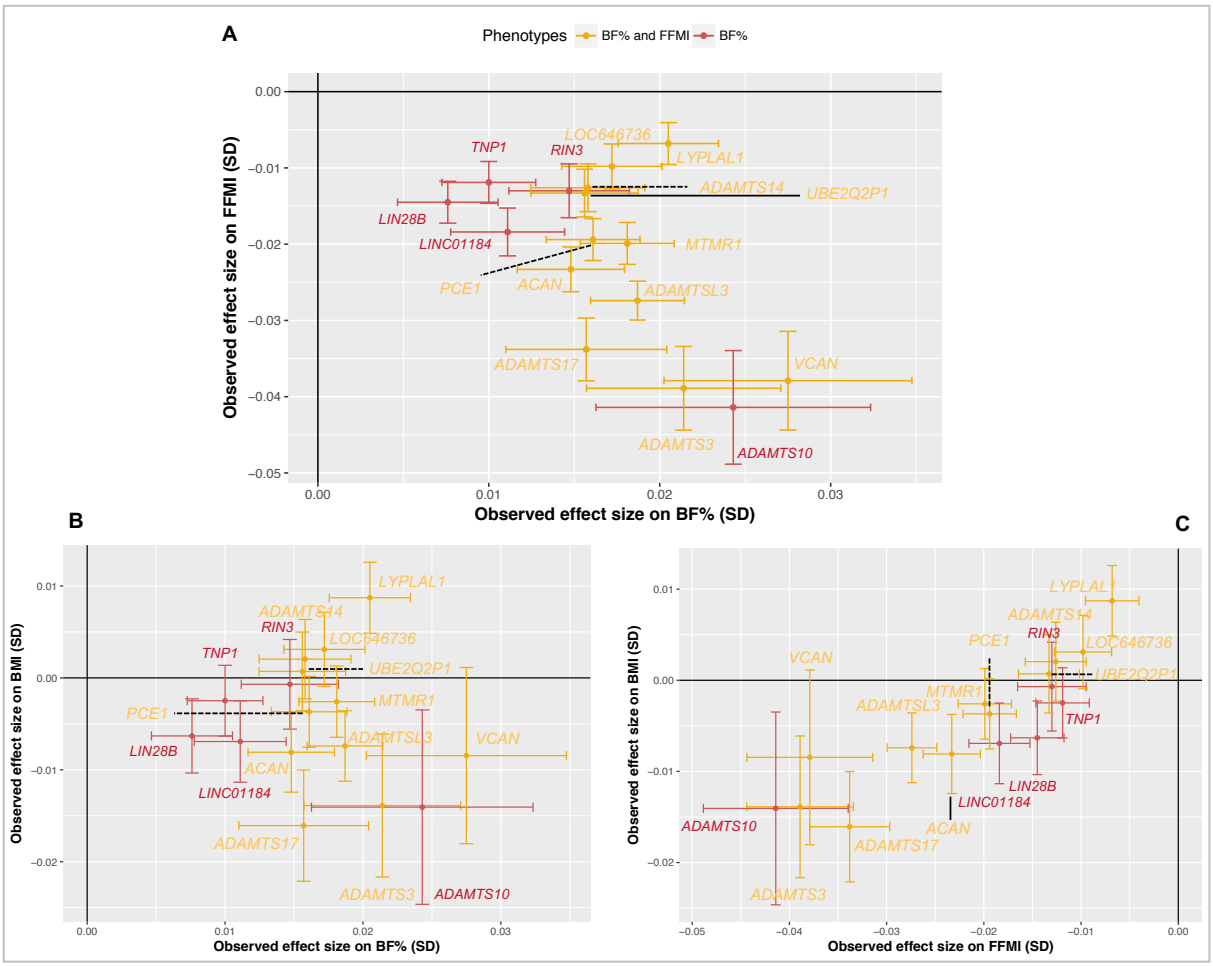


Figure 3.14: scatter plots of effect sizes of the 16 genetic loci associated with BF% and FFMI beyond BMI, on BF%, FFMI and BMI. A| Effect sizes on BF% versus FFMI, **B|** Effect sizes on BF% versus BMI, **C|** Effect sizes on FFMI versus BMI. All effect sizes are aligned to the BF%-raising allele. Effect sizes on BMI were estimated in a GWAS including the European ancestry participants of UK Biobank. For variants in red, the observed effect was significantly higher than the predicted effect for BF% only. For variants in yellow, the observed effect was significantly higher than the observed effect on both BF% and FFMI.

The majority of genetic loci for BMI have proportionate effects on fat and lean mass

Given that a subset of the BF% and FFMI loci were associated with fat or lean mass beyond BMI, I next assessed if there are loci which have been reported with BMI with disproportionately strong effect sizes on either fat or lean body mass. Based on the GWAS results, I assessed the effect sizes of the 620 genetic loci reported for BMI with $p \leq 5 \times 10^{-9}$ reported by Yengo *et al.*[172] on BF% and BMI. Of the 620 loci, 130 reached genome-wide significance for both BF% and FFMI, while 106 and 193 reach genome-wide significance only for BF% and FFMI, respectively (Figure 3.15). The vast majority of the BMI loci had directionally consistent, nominally significant effects on BF% and FFMI (Bonferroni corrected p for 620 tests $\leq 8.1 \times 10^{-5}$). However, 23 BMI variants were associated with FFMI (p for FFMI $\leq 8.1 \times 10^{-5}$) but had no effect on BF% (p for BF% >0.05): *JAZF1, BC021693, GAB2, DOCK3, MMP16, NEK6, L3MBTL3, ATP2B1, EIF4G1, LINC01286, TBX15, SMG6, ITIH1, MSL2, CEP120, CAST, NADK, ANAPC4, ZC3HAV1, KCNH2, CCNL1, SLC9B1* and *VIL1*. Seven other BMI loci were only associated with higher BF% (p for BF% $\leq 8.1 \times 10^{-5}$) but not with FFMI (p for FFMI >0.05): *LYPLAL1, DNAH10, ZNF664, PEPD, GFPT1, CDK5RAP3* and *HOXA13*. *PPARG* (sentinel variant rs1899951) was the only BMI locus which had significant effects on BF% and FFMI in opposite directions. The BMI raising allele at the *PPARG* locus was associated with higher BF% (beta±SE on SDs of BF% =0.028±0.002, $p=1.2 \times 10^{-36}$) but lower FFMI (beta±SE on SDs of FFMI =-0.009±0.002, $p=3.1 \times 10^{-5}$), which means that the BMI raising effect of this locus (beta±SE on SDs of BMI =0.017±0.002, $p=3.1 \times 10^{-12}$) is entirely driven by fat mass.

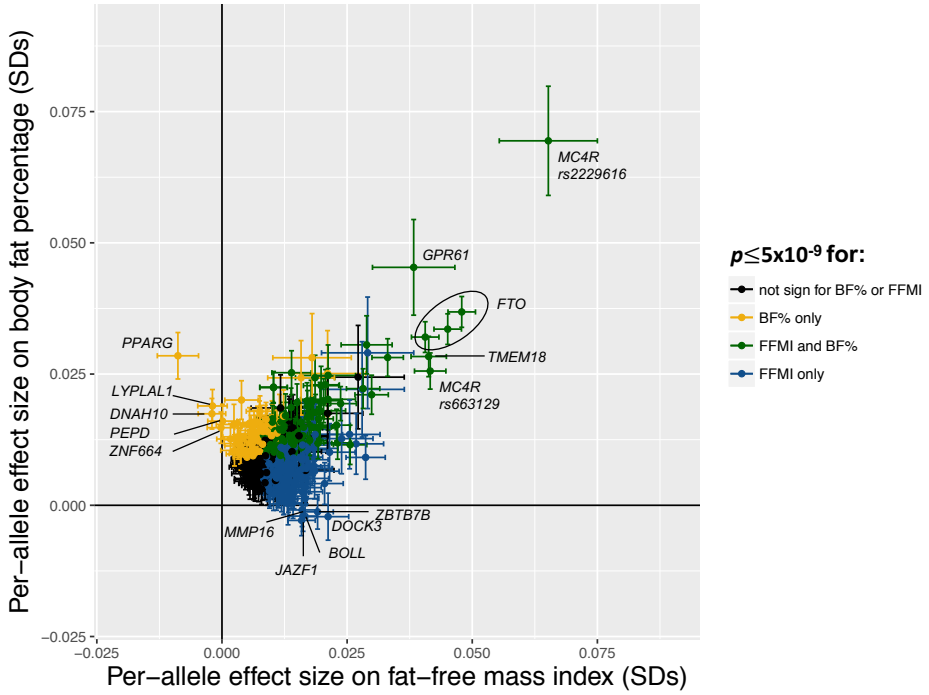


Figure 3.15: Effect sizes of 620 genetic loci reported for BMI on fat-free mass index and body fat percentage. Per-allele effect sizes on standard deviations of BF% or FFMI are aligned to the BMI-raising alleles. Variants in green reached genome-wide significance ($p \leq 5\times10^{-9}$) for both BF% and FFMI, variants in yellow only reached genome-wide significance for BF%, variants in blue only reached genome-wide significance for FFMI and variants in black did not reach significance for either phenotypes.

As a next step, I identified the BMI variants which disproportionately affected BF% and/or FFMI by comparing the observed effect size on BF% and FFMI with the predicted effect size based on the correlation of BMI with BF% and FFMI (Figure 3.16). For 10 of the 620 genetic variants for BMI, the observed effect size on BF% was significantly ($p \leq 8.1\times10^{-5}$) lower than the predicted effect size (rs9816226 in *JAZF1*, rs10497807 near *BOLL*, rs905938 in *ZBTB7B*, rs13021737 near *TMEM18*, rs12446632 in *LOC105371116*, rs3751813 in *FTO*, rs8047395 in *FTO*, rs663129 near *MC4R* and rs9816226 near *ETV5* and rs9922708 in *FTO*), which indicates that these BMI loci more strongly influence lean than fat mass (Figure 3.16, left panel). For 101 other BMI loci, the observed effect on BF% was weaker than the predicted effect at nominally significance ($p_{diff}<0.05$). No BMI variants had disproportionately large effect sizes on BF% at Bonferroni-corrected significance, but for 4 loci the observed effect size on BF% was disproportionately large at nominal significance ($p_{diff}<0.05$): rs10408013 in *PEPD* ($p_{diff}=0.030$), rs11118308 near *LYPLAL1* ($p_{diff}=0.0011$), rs1899951 in *PPARG* ($p_{diff}=6.5\times10^{-4}$) and rs7133378 in *DNAH10* ($p_{diff}=0.0056$).
5 of the 620 BMI loci had a significantly lower observed effect size on FFMI than predicted based on the observational correlation between BMI and FFMI and the effect size of the BMI variants on FFMI: rs1899951 in *PPARG*, rs7133378 in *DNAH10*, rs7124681 in *CELF1*, rs7187776 near

TUFM and rs9922708 in *FTO* (Figure 3.16 right panel). These 5 loci affect fat mass more strongly than lean mass, given the observational correlation between BMI and FFMI. For 54 of the 620 BMI loci, the observed effect size on FFMI was lower than the expected effect size at nominal significance. None of the BMI loci were more strongly associated with FFMI than predicted at Bonferroni-corrected significance, but for 3 loci (rs9816226 in *JAZF1*, p_{diff}=0.013; rs905938 in *ZBTB7B,* p_{diff}=0.039; and rs10497807 near *BOLL*, p_{diff}=0.011) the observed effect size was larger than the expected effect size at nominal significance. Overall, evidence was found that 4 BMI loci were disproportionately associated with BMI through BF% (*PPARG*, *DNAH10*, *CELF1* and *TUFM*) and that 8 BMI loci were disproportionately associated through FFMI (*JAZF1*, *BOLL*, *ZBTB7B*, *TMEM18*, *LOC105371116*, *MC4R*, *ETV5* and *FTO*), while no strong evidence was found for disproportionate effects on fat or lean mass for all other 148 BMI loci.

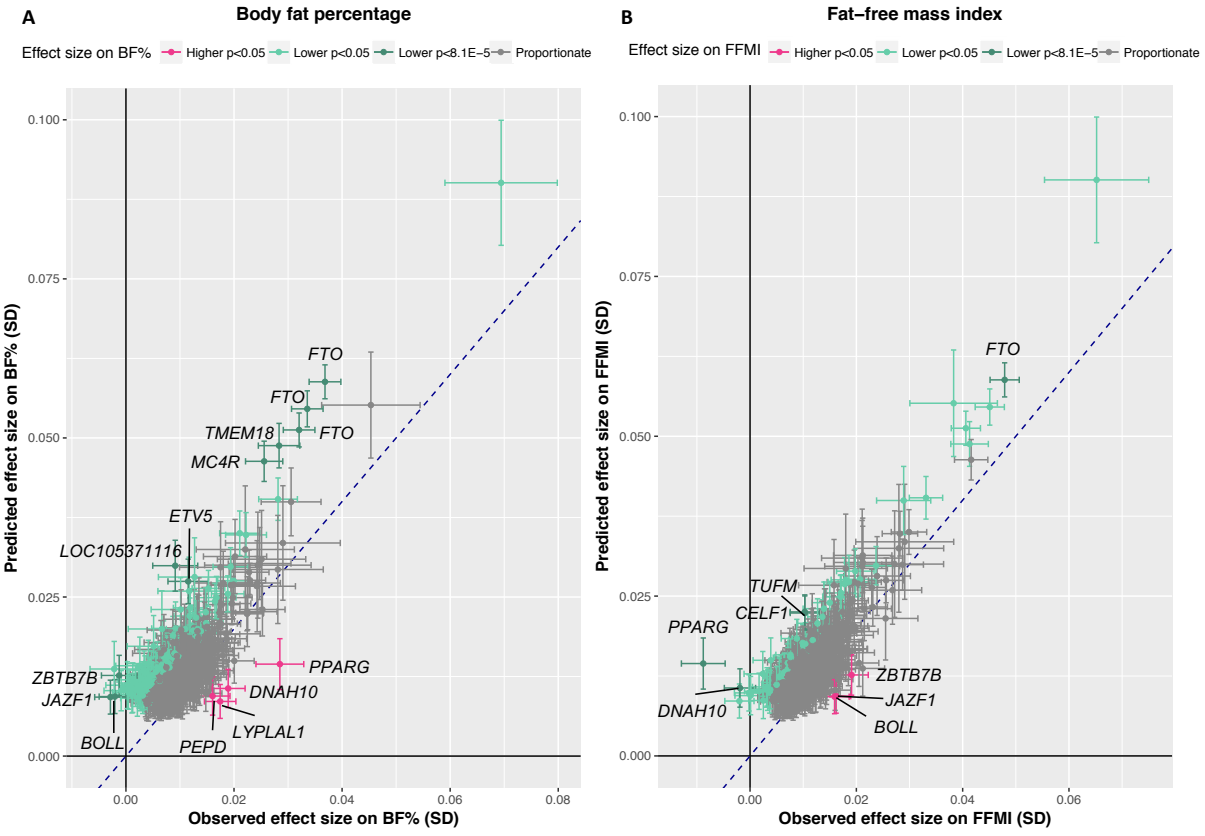


Figure 3.16: Scatter plots of observed versus expected effect sizes of the loci for BMI on BF% and FFMI. All effect sizes are aligned to the BMI-raising allele. The predicted effect size reflects the expected effect size on **A|** BF% or **B|** FFMI if the association with BF% or FFMI is proportionate to the effect size on BMI, given the observational correlations of BMI with BF% and FFMI, respectively. The observed effect sizes on BF% and FFMI come from the meta-GWAS results. For variants in dark and green, the observed effect size was smaller than the predicted effect size on Bonferroni-corrected significance ($p \leq 8.1 \times 10^{-5}$) and nominal significance, respectively. For variants in pink, the observed effect size was larger than the predicted effect size at nominal significance. The dashed blue line indicates equal observed and predicted effects.

20 low-frequency coding variants identified as likely causal variant for association with BF% or FFMI

Low-frequency coding variants associated with BF%

31 low-frequency coding variants (MAF <5%) reached genome-wide significance for BF%, based on meta-analyses of the UK Biobank GWAS with the European exome chip data or the published GWAS results by Lu *et al.*[88] (Supplementary Table 3.8). Three of the 31 coding variants were rare (MAF <1%) and had strong effect sizes on BF%: The missense variant rs148330006 in *CYR61* (MAF=0.8%, per-rare allele beta=0.055, $p=4.1\times10^{-12}$), the stop gained variant rs114285050 in *GPR151* (MAF=0.8%, beta=-0.048, $p=5.5\times10^{-10}$) and the stop gained variant rs150090666 in *PDE3B* (MAF=0.1%, beta=0.151, $p=2.6\times10^{-9}$). The low-frequency missense variant rs74580294 in *MLXIP* only reached significance in the women-only analyses (MAF=4.96%, per-rare allele effect on BF%=-0.039, $p=5.4\times10^{-11}$). 23 of the 31 coding variants were also genotyped in UK Biobank, and effect sizes of the genotyped variants correlated perfectly (Pearson's correlation coefficient =0.9997) with those of the imputed variants (Supplementary Figure 3.2). For 8 of the 31 coding variants, evidence was found that they were associated with BF% independently of common variants in the region, based on individual-level data-based conditional analyses (Table 3.2). Among these 8 variants were the 2 rare stop gained variants in *PDE3B* and *GPR151*. Six of the 8 independently associated coding variants were themselves sentinel variants for BF%. Coding variants in *ARHGAP17* and *MC4R* were not selected as sentinel variants in the GWAS but were independently associated with BF% in the exact conditional analyses.

Low-frequency coding variants associated with FFMI

44 low-frequency coding variants reached genome-wide significance for FFMI in the meta-analyses of the GWAS in Europeans of UK Biobank and the European ancestry exome chip meta-analyses (Supplementary Table 3.9). One missense variant only reached significance in the analyses in men (rs77169818 in *GALR1*, MAF=4.6%, beta per rare allele=-0.042, $p=1.7\times10^{-10}$) and 5 missense variants were rare (rs141374503 in *ADAMTS3*, MAF=0.4%, beta per rare allele=0.075, $p=2.6\times10^{-12}$; rs61744853 in *ZMYM6*, MAF=0.9%, beta=0.044, $p=4.1\times10^{-10}$; rs78727187 in *FBN2*, MAF=0.6%, beta=-0.057, $p=5.8\times10^{-11}$; rs41290587 in *SPARC*, MAF=0.7%, beta=0.051, $p=3.0\times10^{-10}$; rs149615348 in *CRISPLD2*, MAF=0.7%, beta=0.049, $p=5.6\times10^{-10}$). For the 33 of the 44 variants that were also genotyped in UK Biobank, the associations of the genotyped variants with FFMI correlated perfectly (Pearson's correlation coefficient = 0.9999) with the associations of the imputed variants (Supplementary Figure 3.3). 14 low-frequency and 3 rare coding variants remained with FFMI associated after conditioning on the common sentinel variants within 1 Mb of the coding variant, of which one was specific for men (rs77169818 in *GALR1*) and 3 were rare

(rs61744853 in *ZMYM6*, rs141374503 in *ADAMTS3* and rs78727187 in *FBN2*) (Table 3.2). 4 of the 17 independently associated coding variants were not among the selected sentinel variants for FFMI, i.e., the coding variants in *ZBTB7B*, *NSD1*, *R3HDM2* and *MC4R*. Collider bias analyses showed that the associations of rs141374503 in *ADAMTS3* and rs145878042 in *RAPGEF3* with FFMI were solely driven by their associations with height and were therefore omitted from further analyses.

Table 3.2: 20 low-frequency and rare coding variants independently associated with BF% and FFMI

rsid	Chr:pos	EA/ OA	EAF	Gene	Trait	BF%		FFMI	
						Beta±-SE	*p*	Beta±-SE	*p*
Sex-combined analyses									
rs61744853	1:35,453,651	C/A	0.009	*ZMYM6*	FFMI	0.007±0.337	0.34	0.044±0.007	4.1E-10
rs141845046	1:154,987,704	T/C	0.023	*ZBTB7B*	FFMI	-0.004±0.005	0.35	0.034±0.005	8.8E-14
rs11722554	4:5,016,883	A/G	0.038	*CYTL1*	FFMI	-0.009±0.004	0.012	0.026±0.004	1.4E-13
rs1229984	4:100,239,319	T/C	0.025	*ADH1B*	BF%	-0.032±0.005	4.6E-11	0.021±0.005	4.5E-06
rs61749613	5:82,815,170	G/A	0.041	*VCAN*	BF%, FFMI	-0.030±0.004	5.5E-17	0.038±0.003	8.4E-30
rs78727187	5:127,668,685	T/G	0.006	*FBN2*	FFMI	0.043±0.009	3.3E-06	-0.057±0.009	5.8E-11
rs114285050	5:145,895,394	A/G	0.008	*GPR151*	BF%	-0.048±0.008	5.5E-10	-0.029±0.007	7.3E-05
rs78247455	5:176,722,005	A/G	0.026	*NSD1*	FFMI	0.005±0.005	0.29	0.033±0.004	7.7E-15
rs41271629	6:86,257,229	G/T	0.029	*SNX14*	FFMI	-0.003±0.004	0.47	0.025±0.004	3.0E-10
rs62621812	7:127,015,083	A/G	0.022	*ZNF800*	FFMI	-0.008±0.005	0.13	0.046±0.005	5.5E-22
rs150090666	11:14,865,399	T/C	0.001	*PDE3B*	BF%	0.151±0.025	2.6E-09	0.030±0.024	0.21
rs1126930	12:49,399,132	C/G	0.036	*PRKAG1*	FFMI	0.009±0.004	0.019	0.031±0.004	4.7E-18
rs78607331	12:57,648,644	T/C	0.045	*R3HDM2*	FFMI	0.015±0.004	1.2E-05	0.020±0.003	7.5E-10
rs61754230	12:72,179,446	T/C	0.02	*RAB21*	BF%	0.032±0.005	1.9E-10	0.023±0.005	2.9E-06
rs28929474	14:94,844,947	T/C	0.02	*SERPINA1*	FFMI	0.015±0.005	0.0039	-0.042±0.005	3.9E-18
rs12595158	15:62,316,035	T/C	0.025	*VPS13C*	BF%, FFMI	-0.030±0.005	3.8E-09	0.029±0.005	7.5E-10
rs78457529	16:24,950,880	T/C	0.012	*ARHGAP17*	BF%, FFMI	0.039±0.006	1.5E-09	-0.046±0.006	3.3E-14
rs2229616	18:58,039,276	T/C	0.02	*MC4R*	BF%, FFMI	-0.068±0.005	1.0E-40	-0.065±0.005	3.1E-39
rs62621197	19:8,670,147	T/C	0.037	*ADAMTS10*	BF%, FFMI	-0.025±0.004	1.7E-09	0.041±0.004	5.7E-27
Men-only analyses									
rs77169818	18:74,980,601	T/A	0.046	*GALR1*	FFMI	-0.005±0.007	0.55	-0.042±0.007	1.7E-10

Associations of coding variants for BF% and FFMI with related anthropometric and metabolic phenotypes.

Evidence was found that 20 low-frequency and rare coding variants were the likely causal signals in the genomic region for BF% and/or FFMI. Of the 20 variants, 5 reached genome-wide significance for both BF% and FFMI: rs12595158 in *VPS13C,* rs2229616 in *MC4R,* rs78457529 in *ARHGAP17*, rs62621197 in *ADAMTS10,* and rs61749613 in *VCAN*. The former two (*VPS13C* and *MC4R*) were associated with BF% and FFMI in the same direction, while the latter 3 (*ARHGAP17*, *ADAMTS10* and *VCAN*) affected the phenotypes in opposite directions. However, some uncertainty remains if the variants in *VPS13C* and *VCAN* are the causal variants in the region because conditional analyses for BF% and FFMI gave contradictory results. For *VPS13C*, conditional analyses for FFMI indicated that this association was not independent of common variants in the region, despite the independent association found for BF%. Similarly, the coding variant in *VCAN* was found to be independently associated with FFMI but not with BF%. 4 of the 20 likely causal coding variants only reached genome-wide significance for BF%, including the rare stop gained variant in *PDE3B*, which was the only coding variant for BF% not also nominally associated with FFMI (p for BF%>2.1×10^{-3}). Two of the coding variants for BF% were nominally associated with FFMI in the same direction (*RAB21* and *GRP151*), while *ADH1B* had an opposite effect on FFMI. Of the 11 coding variants reaching genome-wide significance for FFMI only, 9 did not reach nominal significance for BF% (p>2.1×10^{-3}) (*ZMYM6, ZBTB7B, CYTL1, NSD1, SNX14, ZNF800, PRKAG1, SERPINA1* and *GALR1*), while *R3HDM2* had a directionally consistent effect on BF% and *FBN2* affected BF% in the opposite direction.

To characterise the broader anthropometric and cardiometabolic implications of the coding variants for BF% and FFMI, cross-trait associations were assessed for the coding variants which were covered in previous exome chip projects on anthropometric and cardiometabolic risk factors and diseases. Look-ups were available for all but 3 (*PDE3B*, *MC4R*, *ADAMTS10*) of the 19 coding variants reaching significance in the sex-combined analyses (Figure 3.17).

Cross-trait look-ups of the coding variants revealed heterogeneous association patterns of the BF% and FFM-altering variants with cardiometabolic risk factors and diseases (Figure 3.17). For a subset of the 16 loci, including *ZBTB7B*, *VCAN*, *CYTL1*, *RAB21* and *GPR151*, the alleles associated with higher BF% and/or lower FFMI were associated with higher cardiometabolic risk factors or disease risk, or vice versa, which is in line with evidence from observational epidemiological research. The FFMI-raising, low-frequency variant in *ZBTB7B* (rs141845046, c.670C>T, p.Pro224Ser, beta=0.034, p=8.8×10^{-14}) has previously been reported for higher BMI

(beta=0.047, p=4.2×10^{-13}) and taller height (beta=0.058, p=7.3×10^{-17}). However, the variant did not affect BF% (beta=-0.004, p=0.35), which indicates that the association of *ZBTB7B* with BMI is only due to higher lean mass. Suggestive, non-significant associations with lower LDL (beta=-0.025, p=0.0026) and total cholesterol (beta=-0.020, p=0.013) suggest that the high lean mass associated with *ZBTB7B* may have a metabolically favourable effect. *VCAN* (rs61749613, c.1045A>G, p.Lys349Glu) was associated with lower BF% (beta=-0.030, p=5.5×10^{-17}) and higher FFMI (beta=0.038, p=8.4×10^{-30}) but not with BMI (beta=0.012, p=0.03), and was also nominally associated with a favourable lipid profile: higher HDL cholesterol (beta=0.030, p=7.1×10^{-6}) and lower total triglycerides (beta=-0.028, p=3.4×10^{-5}). The rare, stop gained variant in *GPR151* (rs114285050, c.283C>T, p.Arg95Ter) was more strongly associated with lower BF% than with lower FFMI and has also been reported for lower BMI. While no associations with cardiometabolic disease risk were identified in published exome chip studies, a recent study by Emdin *et al.* on associations of rare loss-of-function mutations with cardiometabolic phenotypes reported suggestive associations of the rare nonsense mutation in *GPR151* with lower risk of T2D (per-allele OR for T2D=0.86, p=0.006) and CHD (per-allele OR for CHD=0.91, p=0.01)[303]. Other examples include *CYTL1*, of which the FFMI-increasing variant was associated with lower total cholesterol, and *RAB21*, which increases BF% more strongly than FFMI and is associated with higher total and LDL cholesterol.

However, for other coding variants, including *ZNF800*, *PRKAG1*, *SERPINA1* and *PDE3B*, the BF%-increasing and/or FFMI-decreasing allele was inversely associated with cardiometabolic phenotypes or disease risk, or vice-versa. *ZNF800* (rs62621812, c.307C>T, p.Pro103Ser) was associated with higher FFMI (beta=0.046, p=5.5×10^{-22}) and was associated with BMI (beta=0.021 , p=0.0012) through lean mass only (BF%: beta=-0.006, p=0.20). Positive associations with LDL (beta=0.032, p=3.3×10^{-4}) and total cholesterol (beta=0.028, p=9.5×10^{-4}) indicated that this lean mass locus may have a metabolically adverse effect. Similarly, *PRKAG1* (rs1126930, c.293C>G, p.Thr98Ser) was associated with BMI (beta=0.039, p=2.6×10^{-11}) through lean mass (FFMI: beta=0.031, p=5.5×10^{-22}) but not through body fat (BF%: beta=0.009, p=0.016), and was associated with an metabolically unfavourable blood lipid profile, i.e., with higher triglycerides (beta=0.026 , p=7.7×10^{-4}) and lower HDL cholesterol (beta=-0.038 , p=7.0×10^{-7}). The rare, stop gained variant in *PDE3B* was strongly associated with higher BF% but not with FFMI. This variant was not covered in published exome chip data, but in the study by Emdin *et al.,* a suggested *inverse* association with CHD risk was found (per-allele OR for CHD=0.65, p=0.03)[303]. The CHD-protective effect of the BF%-raising, rare allele may be due to increased body fat storage in the hip region, as indicated by the strong, inverse association with WHR and WHRadjBMI (Figure 3.10).

Finally, the missense mutation in *SERPINA1* (c.1096C>T, p.Glu366Lys), which was associated with lower FFMI (beta=-0.042, p=3.9×10^{-18}), has been described as a causal mutation for a severe form of alpha-1-antitrypsin deficiency[304] – a recessive, monogenetic condition which leads to high risk of lung emphysema and chronic liver disease[305] and, in the more severe cases, shortened lifespan[306]. The minor, FFMI-lowering allele was also associated with taller height (beta=0.120, p=1.4×10^{-45}) but not with BMI (beta=-0.014, p=0.09) or BF% (beta=0.013, p=0.023). Although the rare allele was associated with higher LDL (beta=0.081, p=4.3×10^{-14}) and total cholesterol (beta=0.078, p=5.5×10^{-14}), a nominal, inverse association with *lower* CHD risk (OR=0.889, p=0.0014) and a suggestive association with *lower* risk of T2D have been reported (OR=0.922, p=0.0026). This at first sight contradictory association pattern can be a consequence of the shortened lifespan or more intense clinical monitoring of the risk allele carriers.

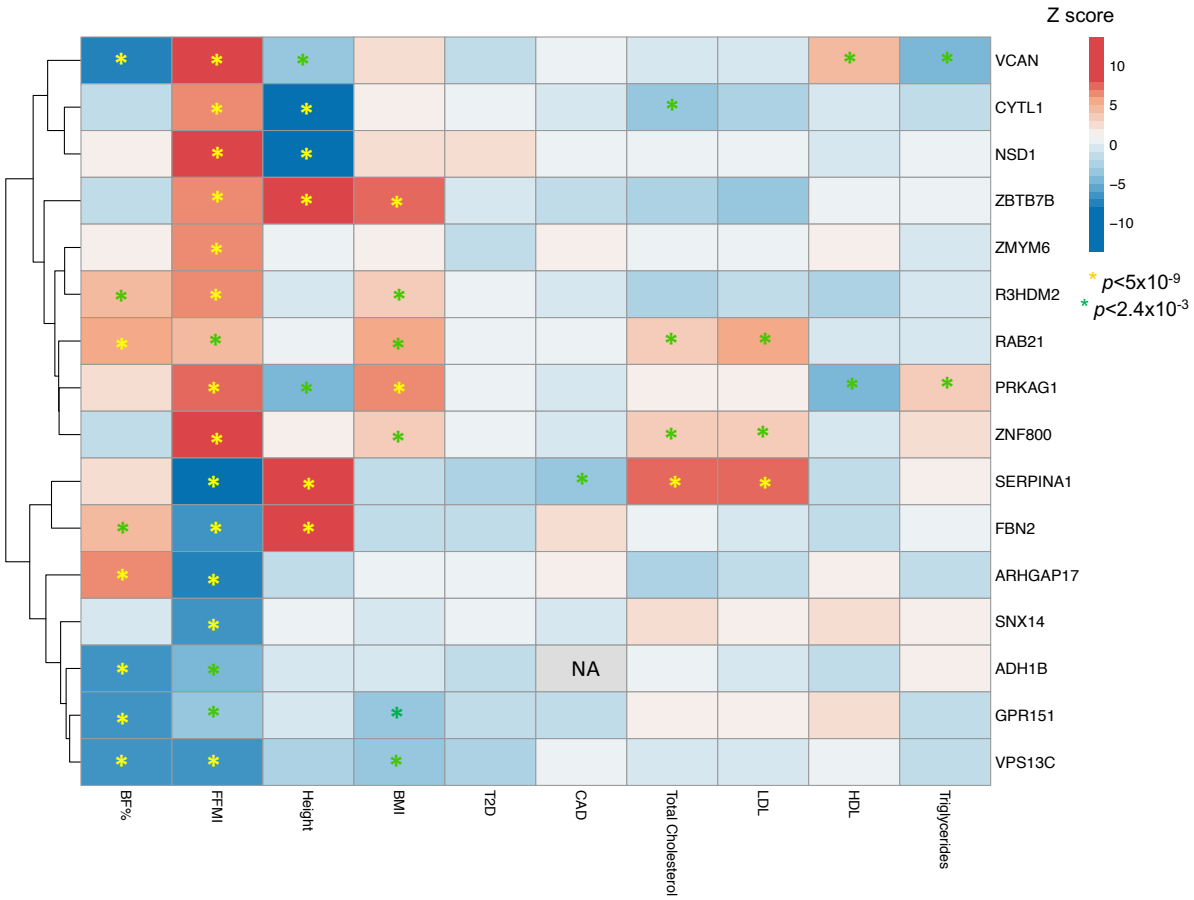


Figure 3.17: Heat plot of reported associations of independent low-frequency and rare coding variants for FFMI and BF% with related anthropometric and metabolic traits and diseases. Z scores are aligned to the rare allele. Associations highlighted with a yellow asterisk reach genome-wide significance (p<5×10^{-9}) and associations highlighted with a green asterisk reach nominal significance (Bonferroni corrected p for 20 tests <0.0025).

Discussion

In this chapter, I aimed to identify genetic factors specifically associated with overall body fatness and leanness, using BF% and FFMI as the respective phenotypes, in order to enable genetics-based causal assessments of total body composition phenotypes as aetiological risk factors for cardiometabolic diseases. To this end, large-scale genetic discoveries through exome chip analyses and GWAS were conducted, and loci for BF%, FFMI or BMI with disproportionately strong effect sizes on total body fat or lean mass were identified through an approach that integrated genetic associations with BF%, FFMI and BMI and observational relationships between the 3 traits. I furthermore aimed to characterise association patterns of the low-frequency coding variants that were identified as the likely causal signals for FFMI and BF%, with related anthropometric and cardiometabolic phenotypes.

BF% and FFMI are highly heritable traits with no strong evidence for sex-dimorphic genetic control

In the thus far largest genetic discoveries for BF% and FFMI, 673 novel genetic loci for BF% and 783 loci for FFMI not previously reported for total lean body mass were identified. Despite the strong observational and genetic correlations of both BF% and FFMI with BMI, only around 30 to 35% of the identified loci for BF% and FFMI were reported in the latest meta-GWAS for BMI in up to 700,000 participants[172].

Although we confirmed the salient sex differences in overall body composition, no strong evidence was found that BF% and FFMI are controlled by distinct genetic factors in men and women. This is in line with findings from GWAS on BMI, in which the vast majority of loci had similar effect sizes in women and men and very few sex-specific loci have been identified[172,217]. In contrast, GWAS on WHR and WHRadjBMI found that the difference in fat distribution between men and women[307] was also reflected by the identified genetic loci, of which nearly half were found to have sex-dimorphic effects[308,309]. Sherman *et al.*[310] recently identified 300 million base pairs, accounting for 10% of the estimated total size of the human genome, which do not map to the current reference sequence of the human genome. It is therefore possible that genetic signals driving the sex dimorphism in BF% and FFMI may be located on parts of the genome which are currently not yet queried because they are missing from the human genome reference sequence upon which array-based genome-wide genotyping and imputation are based.

The estimated SNP-based heritability for BF% was very similar to the heritability for BMI, i.e., approximately 23% for BF% and 22% for BMI, while the heritability estimated for FFMI was higher (30%). However, the estimated heritability for FFMI was likely to be substantially increased

by loci which were only associated with FFMI through height, of which the latter is highly heritable, with a SNP-based heritability of around 48%[172]. Therefore, the higher estimated SNP-based heritability for FFMI than for BF% is unlikely to reflect a true, higher heritability of lean than fat mass. Zillikens *et al.* reported a SNP-based heritability for total lean mass of around 43%, although the inclusion of different covariates and a different analytical approach[302] does not allow direct comparison between the published and our heritability estimates.

The importance of interpreting loci for BF% and FFMI in the context of associations with BMI and other standard anthropometric risk factors

BF% and FFMI were chosen as the phenotypes representing fat and lean mass, respectively, because they are only moderately correlated, despite each having a strong but distinct relationship with BMI. While FFMI is linearly associated with BMI, the relationship of BF% with BMI was curvilinear, with stronger variation and increases per BMI unit in BF% at the lower end of the BMI scale than at the higher end. Individuals with a high BF% typically have a higher BMI and therefore also higher absolute levels of lean mass[311], as higher total body weight due to higher BMI and/or taller height, acts as a chronic overload stimulus on the muscles, which leads to higher muscle mass and strength[312]. Overweight and obesity are therefore typically positively associated with both BF% and FFMI. However, the observational and genetic relationships between BF% and FFMI are weaker than their respective correlations with BMI are, because BF% represents total fat adjusted for total body weight while FFMI represents lean mass adjusted for height. As total body weight and height are weakly correlated at a population level, this results in a modest but curvilinear relationship between BF% and FFMI. However, for a given BF%, a large variation in FFMI was observed, especially among individuals with a high BF%.

Because BF% and FFMI are strongly correlated with BMI and moderately with each other and other standard anthropometric risk factors, genetic loci identified for BF% or FFMI need to be interpreted in the context of their associations with the other trait as well as with standard anthropometric risk factors, in order to obtain a full overview of their effects on body size and composition. Despite the positive observational and genetic correlations between BF% and FFMI, only a third of the 1,216 loci identified for BF% and/or FFMI in the sex-combined analyses had directionally consistent effects on BF% and FFMI, while 307 and 416 genetic loci were specific to BF% and FFMI, respectively, and 92 loci had opposite effects on BF% and FFMI. In line with the strong observational and genetic correlations with BMI, 4 out of 5 of all loci identified for higher BF% or FFMI also reached at least nominal significance for BMI with concordant effect sizes,

with the exception of two BF%-raising loci in/near *COQ10B* and *ADAMTS17* and the FFMI-raising locus near *LYPLAL1,* which were nominally associated with *lower* BMI.

A third of the BF% loci was also positively associated with WHR but most of these associations disappeared after adjusting for BMI, which mirrors the strong observational and genetic correlations of BF% with WHR but not with WHRadjBMI. However, loci in/near *RSPO3*, *VEGFA* and *DNAH10* were associated with higher BF% and have been reported for *lower* WHR with and without adjustment for BMI[308,309]. This indicates that these loci drive higher body fat by an absolute increase in fat mass accumulation in the hip region. Although all three loci are associated with higher overall adiposity, each of these has been linked to *lower* metabolic risk factor levels, including lower fasting insulin[94] and total triglycerides and higher HDL cholesterol[76,214] and/or lower risk of T2D[16,37,46]. The associations of the BF%-increasing loci with higher distribution of body fat to the hip region and lower cardiometabolic risk provide evidence that genetic loci for adiposity do not have uniformly detrimental effects on cardiometabolic health and that the inverse associations of adiposity-raising loci with cardiometabolic risk factors and diseases may be at least partly mediated by positive associations with fat accumulation in the hip region but not in the waist. These findings corroborate the hypothesis that having a high capacity to store fat in the gluteo-femoral region has a protective effect against cardiometabolic diseases, even in the context of high overall adiposity, as has been suggested in the literature[79].

The genetic determinants of "normal weight adiposity"

The strong correlations of BF% and FFMI with BMI suggested that many loci reaching genome-wide significance for BF% and FFMI may simply represent BMI, without disproportionately affecting fat or lean mass. In this chapter, I adopted a quantitative approach to identify genetic loci which affect fat or lean mass disproportionately, by re-appropriating the method proposed by Aschard *et al.* to test for collider bias[299]. The vast majority of the genetic loci identified for BF% or FFMI had proportionate effect sizes on fat and lean mass and thus represented BMI. However, strong evidence was found that 16 loci had disproportionately strong effects on BF%, of which 11 were also disproportionately strong and inversely associated with FFMI. Only one of the 16 loci (*LYPLAL1*) reached nominal significance for higher BMI and 6 others reached nominal significance for lower BMI; this indicates that the identified loci may represent the molecular determinants of a body composition pattern characterised by higher relative adiposity and lower relative lean mass in the absence of an elevated BMI[259]. To the best of our knowledge, these 16 loci represent the first reported genetic determinants of "normal weight adiposity", which has in

observational epidemiological research been proposed as an aetiological pathway to cardiometabolic diseases. Variants in lysophospholipase-like protein 1 gene (*LYPLAL1*) have previously been reported for higher insulin resistance and waist-to-hip ratio, but it is the first time that this locus has been reported for disproportionately high body fatness and low leanness, while only modestly affecting BMI.

5 of the 16 loci with disproportionate effect on BF% were loci located near genes encoding enzymes of the A Disintegrin and Metalloproteinase with Thrombospondin motifs (ADAMTS) family: *ADAMTS3*, *ADAMTS10*, *ADAMTS14*, *ADAMTS17* and *ADAMTSL3*. These secreted, matrix-associated metallo-endopeptidases play multiple roles in tissue morphogenesis and all exert their function by regulating the structure and function of the extracellular matrix[313]. Several members of the ADAMTS family have been linked to fat distribution[308,309] (*ADAMTS3, ADAMTS9, ADAMTSL3*) and height[172,314] (*ADAMTS3, ADAMTS10*, ADAMTS17, *ADAMTSL3*), but also to metabolic risk factors (*ADAMTS3, ADAMTS10*), cardiovascular disease (*ADAMTS1, ADAMTS4, ADAMTS5*) and cancers[313].

Further research is required to identify the molecular mechanisms through which loci in genes of the ADAMTS family and in other identified regions increase relative fatness and/or reduce relative lean body mass without increasing BMI. Conducting phenome-wide associations and targeted, more detailed cross-trait associations on refined anthropometric and cardiometabolic phenotypes will be the first step to characterise the identified loci. Through fine-mapping of the identified loci, hypotheses about the causal variant(s) in the regions can be generated, which can enable functional *in vitro* and *in vivo* studies, to elucidate the physiological and molecular mechanisms driving normal weight adiposity.

BMI loci driven by fat but not lean mass have been reported for *lower* cardiometabolic risk

The same strategy as for the BF% and FFMI was applied to investigate if there are loci reported for BMI for which the association is predominantly driven by either fat or lean mass. The vast majority of loci reported for BMI had proportionate effects on fat and lean mass, with the exception of 7 loci, which all had modest effect sizes on BMI. Compelling evidence was found that *ZBTB7B*, *JAZF1* and *BOLL* were associated with higher BMI only through affecting lean mass, while loci in *PPARG*, *DNAH10*, *PEPD* and *LYPLAL1* affected BMI only through fat mass. Despite their disproportionate effects on higher adiposity, *PPARG, LYPLAL1, PEPD* and *DNAH10* have all been associated with *lower* insulin resistance[79,94] and *PPARG* and *PEPD* have been reported for *lower* risk of T2D[16,37,40]. However, the BF%-increasing alleles at all 4 loci have

also been reported for higher fat distribution to the lower body, which has been proposed to have protective effects against cardiometabolic disease. Our findings corroborate this hypothesis that higher lower body fat could be a protective factor against cardiometabolic conditions, even in the context of overall higher adiposity, and emphasise the importance interpreting overall higher body fat in combination with body fat distribution, both in the context of clinical assessment and of genetic research on anthropometric risk factors.

The major BMI loci in *FTO*, *MC4R* and *TMEM18* were also identified as loci for both BF% and FFMI in the present study. Our analyses suggested that these BMI loci may have larger effects on lean than on fat percentage, although this disproportionately is minor in magnitude but did still reach statistical significance due to the large effect sizes of the three loci on BMI, BF% and FFMI.

Implications of the collider bias affecting the genetic discovery for FFMI

In a recent study, Aschard *et al.* provided evidence that collider bias may occur in a GWAS for an outcome adjusted for a heritable covariate which is correlated with the outcome[299]. Concretely, this means that false positive genetic associations for the outcome may be identified if the genetic variant is only associated with the covariate but not with the unadjusted outcome. As FFM and height are strongly correlated and both heritable[172,302], we tested for collider bias in the GWAS results for FFMI. Using a novel adaptation for quadratic covariates to the approach proposed by Aschard *et al.*, we found strong evidence for collider bias in the GWAS on FFMI, with associations of more than 13% of the FFMI loci that reached genome-wide significance being solely driven by associations with height. We have however not tested collider bias for variants which did not reach genome-wide significance. Testing for collider bias for larger groups of variants which met a lower significance threshold would have increased the multiple testing burden and would thus have further decreased the power of a statistical test that is already limited in its power due to the uncertainty on 3 phenotypes that needs to be taken into account. Therefore, look-ups of genetic variants in our summary-level GWAS results for FFMI, which we intend to make publicly available, always need to be carefully interpreted and in the context of associations with FFM and height.

This finding also raises questions about the role of collider bias in other GWAS adjusted for height or height squared, in particular for BMI, in which collider bias has thus far not been systematically investigated. If collider bias affects GWAS on BMI similar to GWAS on FFMI, then several of the currently known BMI loci may not be truly associated with overweight or obesity.

Low-frequency coding mutations for BF% or FFMI

In this chapter, I presented the first exome chip meta-analyses for BF% and FFMI, of which the results were meta-analysed with the GWAS in UK Biobank. This resulted in the identification of 34 and 44 low-frequency coding variants reaching genome-wide significance for BF% and FFMI, respectively. Through exact conditional analyses conducted in UK Biobank, I found evidence that only 20 of all 67 low-frequency coding variants were the likely causal variant in the region. This emphasises the importance of interpreting low-frequency coding variants in the context of GWAS results, as has previously also been demonstrated for coding variants associated with T2D[40]. 5 of the 20 causal low-frequency variants were associated with both BF% and FFMI, 14 with FFMI only and one with BF% only. Heterogeneous association patterns of the BF%-increasing and/or FFMI-decreasing alleles with cardiometabolic risk factors were identified, with some showing associations with higher cardiometabolic risk factors (*ZBTB7B*, *VCAN*, *CYTL1*, *RAB21* and *GPR151*), while the for other variant the fat-increasing or lean mass-decreasing alleles were associated with lower cardiometabolic risk (*ZNF800*, *PRKAG1*, *SERPINA1* and *PDE3B*). Particularly interesting findings were the missense variant in *ZBTB7* associated with FFMI only, the nonsense mutation in *PDE3B* which strongly affected BF% and the pathogenic mutation in *SERPINA1*.

The low-frequency coding variant in Zinc finger and BTB domain-containing protein 7B (*ZBTB7B*) was identified as a likely causal variant for higher FFMI. Although this variant has recently been reported for BMI[280], our findings indicate that the association with BMI was only driven by higher lean mass. *ZBTB7B* encodes a zinc finger-containing transcription factor, which was recently identified as one of the driving factors of the development of brown adipose tissue and cold-induced beige fat in murine models[315]. In humans, brown adipose tissue has been linked to lower overall fat mass[316], but its association with lean mass is unknown. However, brown adipocytes have been shown to originate from the same cellular lineage as skeletal myocytes[317], but it is unclear if ZBTB7 also promotes myocyte development. ZBTB7 has also been reported to inhibit type 1 collagen gene expression in the skin[318] and regulates T cell fate in the thymus[319]. The suggestive associations of the lean mass-increasing allele with lower blood lipids suggest that this lean mass-increasing locus may have a cardioprotective effect, although no associations with T2D or CHD were found. Further analyses are needed to investigate if *ZBTB7B* could be a potential drug target for treating sarcopenia and reducing cardio-metabolic disease risk. First, it needs to be determined which components of fat-free mass – skeletal muscle, bone and non-muscle lean mass – are directly affected by *ZBTB7B*. This can be done by assessing the genetic associations with the 3 distinct components of fat-free mass, and through tissue-specific gene expression and eQTL

analyses. If these analyses suggest that *ZBTB7B* influences skeletal muscle mass, *in vitro* and *in vivo* experiments will be needed to investigate if *ZBTB7B* plays a role in skeletal muscle development, maintenance and/or function and how increasing the activity or expression of *ZBTB7B* alters cardio-metabolic phenotypes.

The rare stop gained mutation in phosphodiesterase 3B (*PDE3B*) which was predicted to have deleterious effects on protein function and structure was recently described by Emdin *et al.* in a study of loss-of-function variants in UK Biobank[303]. Based on the identified inverse associations with CHD risk, physician-prescribed lipid-lowering medication and WHR, Emdin *et al.* proposed PDE3B as a novel potential target for the treatment of CHD. In this chapter, we link *PDE3B* for the first time to higher overall body fatness and BMI, but not to lean mass. Our findings in combination with previously published results indicate that loss of PDE3B leads to higher overall BMI and body fat but may decrease cardiovascular risk by promoting distribution of fat to the hip region over the central region. *PDE3B* encodes a phosphodiesterase that is expressed in several tissues involved in energy homeostasis, such as adipose tissue, the pancreas and liver[320], and regulates metabolic processes by breaking down the second messengers cyclic AMP and cyclic GMP. PDE3B has been identified as a key intermediate through which insulin exerts its anti-lipolytic action in adipocytes[321] and was found to enhance glucose-induced insulin secretion by pancreatic β cells[322]. Furthermore, *PDE3B* knock-out mice were found to gain less weight on a high-fat diet, carry less fat tissue and developed more brown adipose tissue[323]. Although these *in vitro* and *in vivo* experiments clearly indicate a role for *PDE3B* in adiposity, our finding that the nonsense mutation in *PDE3B* causing a truncated, likely non-functional gene product is strongly associated with *higher* BF% seems to be opposite to the experimental findings which suggest that non-functional PDE3B increases lipolysis and reduce fat mass. Further research on the potential of PDE3B as a novel target of cardiometabolic diseases is needed and should seek to clarify the seemingly opposite findings from observational genetic research on humans and functional studies on cell and rodent models.

The low-frequency missense mutation in serpin family A member 1 (*SERPINA1*) reached genome-wide significance for lower FFMI, taller height and higher LDL and total cholesterol. *SERPINA1* encodes alpha-1-antitrypsin, which is a member of the serine protease inhibitors, a class of proteins responsible for inhibiting the activity of certain enzymes. Alpha-1-antitrypsin inhibits the activity of elastases secreted by neutrophils in lungs, which prevents damage to the connective tissue in the lungs[305]. The mutation that reached significance for lower FFMI is known to be a pathogenic mutation that causes a severe form of alpha-1-antitrypsin deficiency, a recessive condition characterised by lung emphysema and, in some cases, liver cirrhosis. Duckers *et al.*

reported lower lean mass in patients with alpha-1-antitrypsin deficiency compared to unaffected individuals[324]. Sarcopenia is likely to be a mere consequence of the reduced ability to engage in physical activity due to the restricted lung capacity. However, findings by Heinzelmann *et al.* showed that patients with alpha-1-antritrypsin deficiency showed less skeletal muscle adaptation to exercise as compared to patients with chronic obstructive pulmonary disease due to other causes[325]. Further research is needed to determine if *SERPINA1* has a primary function in body composition and to clarify the role of sarcopenia in the pathophysiology of alpha-1-antitrypsin effect of pathogenic mutations in *SERPINA1* on reduced lean mass.

Strengths and limitations

With a sample size of up to more than half a million participants, the unprecedented scale at which the genetic discoveries for BF% and FFMI were conducted and thus the power to identify genetic associations is one of the major strengths of the study presented in this chapter. The 5- and 10-times larger sample sizes compared to the previous GWAS for BF% and lean mass, respectively, led to the identification of 673 not previously reported for BF% and 783 not previously reported for FFMI. Secondly, I set out a framework for a systematic and quantitative assessment of the extent to which loci identified for fat and lean mass merely reflect BMI or, alternatively, disproportionately affect fat and lean mass, beyond BMI. Finally, exact conditional analyses were systematically conducted for all low-frequency coding variants reaching genome-wide significance, which allowed us to distinguish coding variants which were the likely causal variant in the region from coding variants hitchhiking on a GWAS locus without themselves being causal.

While conducting a genetic study in one large biobank has the advantage that all data is collected and processed in a more uniform manner than is typically the case when data from multiple cohorts are meta-analysed, this also has the disadvantage that findings may be population or even study-specific. As UK Biobank is unique in its scale and availability of genome-wide data imputed to three LD reference panels, we could not replicate our findings in an independent study. Secondly, BF% and FFMI were derived from a BIA-based body composition assessment. While BIA has the advantages of being a cheap, quick and non-invasive technique, which are attractive characteristics for any type of measurement that needs to be conducted on half a million participants, BIA is a less accurate technique for analysing body composition than certain imaging-based techniques such as dual-energy X-ray absorptiometry (DEXA) or magnetic resonance imaging (MRI), as it has been shown to systematically underestimate body fat. However, the systematic underestimation of body fat by BIA is unlikely to have influenced the findings, as BF% and FFMI were rank-based inverse normally transformed. A final limitation is that BIA does not allow to

distinguish fat-free muscle mass from the fat-free mass made up by the internal organs and bones. As muscle mass only represents about half of the total lean mass[326], genetic loci identified for FFMI do not only reflect muscle mass but also represent variation in the mass of the internal organs and, to a much lesser extent, bone mass. FFMI loci of specific interest may therefore need to be replicated using gold-standard methods, i.e., MRI or computerised tomography (CT).

Conclusion and future research directions

The genetic determinants of BF% and FFMI were found to largely coincide with the genetic factors driving BMI. However, by systematically integrating genetic findings of multiple anthropometric phenotypes, subsets of genetic variants with disproportionately strong effects on fat and/or lean mass were identified, including 15 genetic loci associated with disproportionately high fat mass in the absence of overweight. Furthermore, evidence was found that the associations loci for high BF% with cardiometabolic risk factors is highly dependent on the pattern of fat distribution associated with the locus, as evidenced by the identification of multiple loci associated with higher BF% but with *lower* central-to-lower body fat distribution and *lower* cardiometabolic risk factors and/or disease risk.

Chapter 4: The genetic determinants of specific fat and lean mass compartments

Collaborations and contributions

In this chapter, I describe an ongoing project that focuses on the genetic determinants of fat and lean mass compartments based on DEXA imaging. GWAS for 9 fat compartments and 7 lean mass were conducted in 5 cohorts, including the Fenland, EPIC-Norfolk, ORCADES, Oxford Biobank and UK Biobank studies. I conducted all analyses described in this chapter myself, with the exception of the study-level GWAS in the Oxford Biobank, which were conducted by Dr Matthew Neville (Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford) and the study-level GWAS in ORCADES, which were conducted by Katherine Kentistou (Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh).

Publications related to this chapter

- M. J. Neville, L. B.L. Wittemans (co-first author), K. E. Pinnick, M. Todorčević, R. Kaksonen, K. H. Pietiläinen, J. Luan, R. A. Scott, N. J. Wareham, C. Langenberg, F. Karpe**. Regional fat depot masses are influenced by protein-coding gene variants.** *Under review at PLoS ONE*

  This project was conducted in collaboration with Prof Fredrik Karpe and Dr Matthew Neville (Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford) and focused on the identification and functional characterisation of protein-coding variants associated with 6 fat mass compartments adjusted for BF%. I designed the genetic analyses and conducted exome chip analyses in two cohorts, conducted exome chip meta-analyses, annotated the identified coding variants and interpreted the findings in the context of the existing literature on the genetic determinants of anthropometric and cardiometabolic phenotypes. As different decisions were made in terms of the included covariates, the findings from this paper cannot be directly compared to the GWAS results presented in this chapter. Therefore, the decision was made to exclude the exome chip from the thesis.

- L. A. Lotta, L. B. L. Wittemans, [15 authors], C. Langenberg. **Specific genetic determinants of gluteo-femoral and abdominal fat distribution and risk of type 2 diabetes and cardiovascular disease.** *JAMA* 320(24), 2553-2563 (2018).

  This project sought to identify genetic factors specifically associated with abdominal and gluteo-femoral fat mass and to assess the independent unconfounded effect sizes of central and lower body fat on risk of cardiometabolic disease, based on the identified genetic instruments. I contributed to the genetic discovery for WHR and waist and hip circumference in UK Biobank by preparing the phenotypes for analysis, identifying the independent, significant loci and generating plots of the GWAS results. I furthermore assessed the associations of the loci for WHR, waist and hip with compartmental fat phenotypes based on DEXA imaging and contributed to the manuscript.

- C. Hilton, M. J. Neville, L. B. L. Wittemans, [12 authors], F. Karpe. **MicroRNA-196a links human body fat distribution to adipose tissue extracellular matrix composition.** *Under review at EBioMedicine.*

  This study sought to characterise the function of microRNA-196a in body fat distribution through a combination of *in vitro* functional studies and *in silico* genetic analyses. I contributed to this study by assessing the associations of eQTL signals for microRNA-196a with body fat compartments in the Fenland, EPIC-Norfolk and UK Biobank studies.

Abstract

Background: Waist-to-hip ratio (WHR) is a prominent cardiometabolic risk factor that increases disease risk independently of BMI. However, the role of other patterns of fat distribution and of regional lean mass as cardiometabolic risk factors is largely unknown. Identifying the genetic determinants of regional fat and lean mass forms a first important step towards assessing the causality of refined patterns of fat and lean mass distribution on cardiometabolic disease risk.

Aims: This chapter aims to (1) identify the genetic determinants of absolute and relative levels of fat and lean mass in different compartments of the body and (2) to assess fat and lean mass distribution patterns associated with the identified genetic loci.

Methods: (1) Meta-analyses of GWAS were conducted for 9 fat and 7 lean mass compartments and included up to 25,452 participants from 5 cohorts. Analyses for absolute fat and lean mass compartments were adjusted for age, sex and genetic principal components only, while analyses for relative fat and lean mass compartments were additionally adjusted for total fat mass and height squared, respectively. (2) The heterogeneity of the effect sizes of genetic loci associated with absolute fat or lean mass compartments across all compartments was assessed based on the Cochran's Q statistic. Fat and lean mass distribution patterns of the identified loci for relative and absolute fat and lean compartments were assessed based on hierarchical clustering applied to the associations across all compartments.

Findings: (1) 34 and 13 independent genetic loci reached genome-wide significance for absolute and relative lean mass compartments, respectively, while 11 and 51 genetic loci were identified for absolute and relative fat compartments. *FTO* was the only locus associated with both fat and lean mass compartments. (2) 38 of the 51 loci reaching genome-wide significance for relative fat mass compartments were associated with a higher central-to-lower body fat distribution pattern that is highly similar to WHR. The 38 loci had consistent positive effects on all central fat compartments with the exception of subcutaneous android fat mass, and consistent, negative effects on the leg and gynoid fat compartments but not on arm fat mass. An inverse correlation between the effect sizes of the central fat-increasing alleles on visceral versus subcutaneous android fat adjusted for total fat mass was found. 7 loci were identified for higher arm fat adjusted for total fat, of which two loci (*PPARG* and *SPATA20*) were also associated with absolute fat mass in the arms but not in any of the other compartments. For lean mass, no heterogeneity in the effect sizes on the 7

different lean mass compartments was found for any of the 34 and 13 genetic loci reaching genome-wide significance for absolute or relative lean mass compartments, respectively.

Conclusion: Although the majority of loci associated with relative regional fat mass represented WHR, a subset of the identified loci was associated with other patterns of fat distribution. Loci with strong effect sizes on arm but not on leg fat mass were identified, which indicates that peripheral fat mass in the upper and lower limbs is controlled by distinct genetic factors. Genetic loci identified for absolute levels of regional fat mass tend to have heterogeneous effects across different body fat compartments. Conversely, all loci identified for lean mass compartments had similar effects across all lean mass compartments, which suggests that lean mass in the different regions of the body is controlled by shared genetic factors.

Introduction

Fat distribution is a major independent risk factor for cardiometabolic disease

Despite the well-established role of overall adiposity as an aetiological risk factor for T2D and cardiovascular diseases[21,253], strong differences in cardiometabolic risk have been observed between individuals who are overweight or obese[327,328]. Fat distribution, and central fat accumulation in particular, has been proposed as an important driver of differences in cardiometabolic risk among individuals with generalised adiposity[327,329,330]. Central fat accumulation, usually measured by waist-to-hip ratio (WHR) or WHR adjusted for BMI (WHRadjBMI), has been shown to be a prominent risk factor of cardiometabolic diseases independently of overall adiposity[85,331,332]. Central fat is distributed over two compartments, i.e., so-called "visceral fat" that is stored in the abdominal cavity, and subcutaneous fat. The distribution over the visceral and subcutaneous compartments is highly variable between individuals and between men and women. Research suggests that it is mostly through the former that central fat accumulation deteriorates insulin sensitivity and increases cardiometabolic risk factor levels[30,330,332], but simple indices of fat distribution, such as WHR or waist circumference, cannot distinguish visceral from subcutaneous fat.

Gluteo-femoral fat as an independent, protective factor against T2D and CHD

Studies on the cardiometabolic sequelae of high WHR have predominantly focused on the numerator of the ratio and considered the denominator, i.e., hip circumference, merely as a correction factor for overall body size. More recently, interest in the role of hip circumference independent of waist as an aetiological factor has grown, after observational epidemiological research associated higher hip circumference with *lower* risk of cardiometabolic disease[30,327,333,334]. In a few genetics-based studies, higher hip circumference was suggested to be causally related to lower risk of cardiometabolic diseases[79,335]. Our recent causal investigation based on genetic instruments specific for waist and hip, corroborated this hypothesis and found a strong, protective effect of higher hip circumference and gluteo-femoral fat against CHD and T2D, independent of overall BMI and waist circumference[336]. These findings emphasise that both central and lower body fat play an aetiological role in cardiometabolic disease risk and raise the question about the metabolic relevance of fat accumulation in other regions of the body that are not represented by WHR.

Lean mass distribution patterns and their metabolic sequelae are largely unknown

The extensive research on fat distribution and its relationship with cardiometabolic disease risk has thus far not been mirrored by a similar research interest in the metabolic role of lean mass distribution patterns. Studies on lean mass distribution have mostly focused on sarcopenia in the

context of aging[337,338], bone mineral density[339] and rare neurological[340] and musculoskeletal conditions[341]. Differences in absolute and relative lean mass and lean mass distribution have been described between men and women, with the strongest relative difference being found for muscle mass in the arms[342]. Evidence for pronounced differences in lean mass distribution between women of different ethnicities has been reported[339]. Few studies to date have attempted to identify metabolic risk profiles associated with different patterns of lean mass distribution. Hamasaki *et al.* reported that higher lean mass in the lower extremities relative to total body weight or lean mass in the upper extremities was associated with a better blood lipid profile in patients with T2D[338]. Overall, the variation in lean mass distribution and its role as a metabolic risk factor in the general population remains unclear.

Imaging techniques for regional body composition assessment

Imaging techniques such as dual-energy X-ray absorptiometry (DEXA), computerised tomography (CT) and magnetic resonance imaging (MRI), allow for accurate measurement of fat, lean and bone mass in specific regions of the body. MRI and CT are considered the "gold-standard" techniques for measuring fat and lean mass compartments but are not routinely used for whole-body scans in large-scale epidemiological studies because of the lack of software tools to efficiently process whole-body images, the high cost and time investments and, for CT, concerns about radiation exposure.

DEXA is a non-invasive, two-dimensional imaging technique which is considered the gold-standard method for bone densitometry but has more recently also been used to measure overall and regional fat and lean mass. DEXA technology uses the differences in attenuation of X rays at two different energies as they pass through bone or lean versus fat tissue to estimate bone mineral, fat and lean mass, for the whole body and by region. Although DEXA does not allow direct measurement of visceral fat mass, software packages have been developed which allow estimation of visceral fat mass and volume, based on measured total android fat and estimated subcutaneous fat in the android region. In epidemiological studies, DEXA-based measures of body composition have been shown to correlate more strongly with metabolic risk factors than traditional anthropometric risk factors, such as BMI and waist and hip circumference[30].

In a recent validation study based on 4,753 participants of UK Biobank[343], total fat and lean mass and visceral fat mass measured by DEXA were compared with estimates based on MRI. DEXA and MRI measurements correlated very well for total body fat (R=0.99, prediction error standard deviation relative to the mean (CV) = 4.5%) and lean tissue (R=0.97, CV=4.6%), but visceral fat measured by MRI diverged more substantially from visceral fat predicted by DEXA (R=0.94, CV=21%)[343]. The main advantages of DEXA over MRI and CT is that it is a relatively low-cost

and high-throughput technique which causes a much lower radiation exposure to the participants than CT imaging. Furthermore, image processing for DEXA has become efficient and straightforward with the software packages developed by the manufacturers of DEXA instruments, while constraints on processing the large numbers of images generated by MRI imaging have limited the use of MRI for whole-body composition analysis. An important limitation of DEXA imaging is that it does not have sufficient spatial and contrast resolution to detect ectopic fact, such as fat infiltration in the liver and muscles.

Genome and exome-wide association studies on WHR and refined anthropometric phenotypes

The genetic determinants of fat distribution have been extensively studied through GWAS[308,309] and exome chip analyses[292] for WHR and WHRadjBMI. GWAS in up to 695,000 individuals, mostly of European ancestry, have led to the identification of more than 460 loci associated with WHR[308,309]. Strong evidence for distinct genetic factors controlling fat distribution in men and women was found, with nearly half of the loci identified for WHR having sex-discordant effects[308,309], of which the majority had stronger effect sizes or only reached significance in women. WHR was also found to be more heritable in women (SNP-based heritability estimate of 26%) than in men (17%)[309]. Adipogenesis, angiogenesis and insulin resistance have been identified through genetic enrichment analyses as the prominent physiological processes linked to WHR[308]. To our knowledge, no previous studies aiming to systematically characterise the genetic determinants of all fat and lean mass compartments have been conducted. However, a series of genetic discoveries on highly refined adiposity phenotypes based on CT imaging have been conducted in up to 18,300 participants of multiple ancestries. The studied phenotypes included ectopic fat, such as pericardial fat, visceral and subcutaneous fat volume and fat attenuation – an index for adipocyte size, lipid content and maturity – have been conducted[344–346]. Through GWAS, multiple genetic loci, including regions in/near *ATXN*, *TRIB2*, *UBE2E2*, *EBF1*, *RREB1*, *GSDMB*, *GRAMD3* and *ENSA*, were identified for ectopic fat traits. Interestingly, none of these loci reached significance in the at the time largest GWAS for BMI[217] or WHR[308], which indicates that ectopic fat may be controlled by genetic factors that are distinct from those controlling overall adiposity and fat distribution in the adipose tissues.

Zillikens *et al.* conducted genome-wide association analyses for appendicular lean mass, i.e., the combined lean mass in the legs and arms, which serves as a proxy for total muscle mass[302]. Loci in *IRS1*, *ADAMTSL3* and *VCAN* reached genome-wide significance for appendicular lean mass. However, all three loci had nearly identical effect sizes on total lean mass, for which a GWAS was

conducted in the same study. This suggests that the three identified loci were not specific to appendicular lean mass but instead act on overall lean mass.

Opportunities, objectives and outline

In recent years, DEXA-based body composition analysis has been adopted in several epidemiological cohort studies, and is currently being undertaken in 100,000 UK Biobank participants, as part of the UK Biobank imaging study[343]. However, large-scale genetic studies of regional body composition traits measured by DEXA have not yet been conducted. Through identifying the genetic factors that influence specific body fat and lean mass compartments, the molecular mechanisms driving fat and lean mass in specific regions of the body can be identified, genetic loci known for standard anthropometric phenotypes can be characterised in terms of the body compartments they affect the most, and genetic instruments for specific body fat and lean mass compartments can be constructed, which, in turn, can enable causal investigations of patterns of fat and lean distributions as aetiological factors for cardiometabolic diseases.

In the final research chapter of this thesis, I aim to identify the genetic determinants of absolute and relative fat and lean mass in distinct compartments of the body and to assess fat and lean mass distribution patterns associated with the identified genetic loci. As an introduction, I investigate relationships between the fat and lean mass compartments and in relation to standard anthropometric phenotypes in a healthy middle-aged population of more than 10,000 individuals. I then describe the results of the first genome-wide genetic discoveries for 16 fat and lean mass compartments measured by DEXA, which were in the initial stage of this ongoing project conducted for the sexes combined only, as has traditionally been done in the initial genetic discoveries for novel phenotypes.

Methods

DEXA-based assessment of whole-body and regional body composition

All cohorts included in the genome-wide association studies made use of DEXA technology to assess overall and regional body composition. DEXA-based body composition analysis is based on whole-body X ray images of the bone and soft tissues. Seven body regions (android region, gynoid region, legs, arm , trunk, head and pelvis), some of which partially overlap, were automatically demarcated by the software developed by the manufacturers of the DEXA instruments[79,347] and were manually checked and corrected if necessary. The head region is defined by the area above the horizontal line that touches the base of the skull. A line through the gleno-humeral joint and the crease of the axilla demarcates the arm region. The pelvic region is defined by a horizontal line that touches the upper edge of the iliac crest and diagonal lines that cut through the acetabulofemoral joints. The trunk region includes the entire upper body and pelvic regions without the head and arms regions. The leg region includes the entire area below the pelvic region. The android region is defined as the area between horizontal line on top of the iliac crest and a second line at a distance equal to 20% of the total distance between the iliac crest and the base of the skull. The gynoid region partly overlaps with both the legs and trunk regions. Its upper horizontal edge is placed at a distance of 1.5 times the height of the android region from the iliac crest and its lower edge is at a distance of 2 times the height of the android region from the upper edge. Visceral fat mass was estimated by measuring the thickness of the subcutaneous fat in the android region and subtracting the estimated subcutaneous android fat mass from the total android fat mass.

GWAS were conducted for fat mass in the arms, legs, trunk, gynoid region, android region, subcutaneous android region, visceral region and for total body and peripheral fat mass, and for lean mass in the arms, legs, trunk, gynoid region, android region and total body and appendicular lean mass. Peripheral fat and appendicular lean mass were calculated by adding up the arm and leg fat mass or arm and leg lean mass, respectively.

Studies

Fenland

The participants and the design of the Fenland study and genome-wide genotyping and imputation procedures have been described in the methods section of Chapter 2. DEXA imaging was performed on 11,869 participants (6,396 women) of the Fenland study, using the Lunar Prodigy advanced fan beam scanner (GE Healthcare, Madison, Wisconsin, USA). The manufacturer's enCORE software (version 14.10.022, GE Healthcare) was used to identify the regions and estimate overall and regional fat and lean masses. Scanning was performed by trained operators and the images were manually reviewed by a researcher, who discarded images of low quality due to movement artefacts or positioning errors and manually adjusted the boundaries defining the regions placed by the software where necessary. DEXA variables on 11,741 participants (6,334 women) passed QC. Related individuals were identified on the basis of the identity-by-descent (IBD) statistic computed in PLINK[208,349]. Individuals with IBD > 10% were excluded from genome-wide analyses. In total, DEXA and genome-wide genetic data on 10,309 participants (5,507 women) were available for analyses.

EPIC-Norfolk

The study design, participants and genotyping and imputation procedures for the EPIC-Norfolk study have been described in Chapter 1. Whole-body DEXA scanning and processing and quality control of the images was conducted using the same instruments, software and protocols as for the Fenland study. DEXA imaging was conducted on 5,568 participants, of which 5,547 participants (3,098 women) passed QC. Related individuals were excluded if IBD> 10%, which resulted in 4,440 participants (2,441 women) with DEXA and genome-wide data included in the genome-wide association analyses.

Oxford Biobank

The Oxford Biobank includes more than 8,000 participants living in Oxfordshire (United Kingdom)[30]. Participants were randomly invited to participate in the study based on all individuals on the UK National Health Service population registry who were between 29 and 55 years of age and free of any known chronic diseases. Whole-body DEXA imaging using the GE Lunar iDXA machine was performed as part of a health assessment, and images were processed using the enCORE software (version 14.0; GE Healthcare, Bucks, UK). Genome-wide genotyping was performed using the Affymetrix UK Biobank Axiom array and genome-wide genotyped data were

imputed to the HRC LD reference panel. Participants of non-European descent were identified based on a scatter plot of the first two genetic principal components of the Oxford Biobank participants, calculated in PLINK, versus those of the CEU, YRI and CHB/JPT reference panels from 1000 Genomes phase 3, and were omitted from the analyses. DEXA and genome-wide data were available for 4,572 participants, of which 2,603 were women.

Orkney Complex Disease Study

The Orkney Complex Disease Study (ORCADES) is set in the Orkney Isles of northern Scotland and is a cross-sectional, family-based epidemiological study focusing on the genetic determinants of complex diseases[350]. Ethical approval for this study was given by the local research ethics committee and all participants provided informed consent. The participants underwent whole-body DEXA scans using the Hologic QDR4500 densitometer machine and images were processed using the manufacturer's APEX4 software. During the image processing, images were excluded if the regions could not be correctly identified due to low quality of the scans. Fat and lean mass measures >5 SDs from the mean were omitted. Genome-wide genotyping was done in two phases. Panel A consisted of 890 participants and was genotyped using the Illumina HumanHap300v2. The Illumina OmniX or Omni1 were used to genotype panel B which was comprised of 1,300 participants. Only variants represented on both the OmniX and Omni1 array were retained for analyses. SNPs were omitted if the call rate <98%, MAF <0.01 or Hardy-Weinberg equilibrium (HWE) p-value $<10^{-6}$. Duplicate samples were removed, and individuals of non-European ancestry were identified and excluded based on a scatter plot of the first two genetic principal components anchored in the 1000G phase 3 CEU, YRI and CHB/JPT reference panels. Imputation to the HRC LD reference panel was done for panels A and B separately. Up to 1,194 women and 785 men participating in ORCADES were included in the analyses.

UK Biobank

As part of the ongoing UK Biobank Imaging Study[351], which has the aim to conduct detailed imaging on 100,000 participants of the UK Biobank study, whole-body DEXA scans using the GE-Lunar iDXA (GE Healthcare, Madison, Wisconsin, USA) are conducted[343]. DEXA images are processed using the GE enCORE software by the DEXA operator during or shortly after the imaging is completed. DEXA data on 5,170 participants (2,713 women), which passed QC conducted by UK Biobank, were available at the time the analyses were conducted. As an additional QC step, DEXA data were omitted if the total weight estimated by DEXA deviated more than 2.5% from the weight measured using scales on the day that DEXA imaging was conducted, as this suggests that the participant's body was not entirely within the scanning area.

This led to a total number of 4,995 participants (2,604 women) who passed DEXA QC. Details on genotyping, imputation and quality control have been described in Chapter 2. Analyses in UK Biobank were restricted to participants of European ancestry, as current sample sizes for non-European ancestries with available DEXA data are very low. DEXA and genome-wide data were available for 4,707 participants (2,463 women) of European ancestry. Genotyping, imputation, quality control and identification of participants of European descent have been described in Chapter 2.

Genetic discovery for fat and lean mass compartments

Phenotype transformations

All DEXA traits were analysed with minimal adjustments and with adjustment for total fat mass for the fat compartments, and for height squared for the lean compartments. For the minimally adjusted analyses, the traits were first natural log-transformed and then adjusted for age and study-specific covariates. Additional adjustment for total fat mass or height for the fat and lean mass compartments, respectively, were done for the fully adjusted models. The residuals of these linear regression models were rank-based inverse normally transformed. Transformations were done for men and women separately.

Study-level genome-wide association analyses

In the Fenland and EPIC-Norfolk studies, GWAS adjusted for the first four genetic principal components were conducted using BGENIE v1.2[213]. For the Fenland study, separate GWAS for the Genome-Wide Human SNP Array, the UK Biobank axiom array and the CoreExome array were run. In the Oxford Biobank, GWAS adjusted for the first four genetic principal components were conducted using SNPTEST[352]. As ORCADES is a family-based study, a genetic relatedness matrix based on identity-by-state inferred from the genotyping array was calculated and phenotypic residuals were calculated using GenABEL[353]. Genome-wide association analyses adjusted for genotyping array were run using RegScan[354]. GWAS in UK Biobank were adjusted for the first 10 genetic principal components and genotyping chip and were conducted using BOLT-LMM[203].

Quality control of GWAS

Variants were omitted if the minor allele count was less than 10, if the variant represented an insertion or deletion, if more than two alleles were present for the same chromosome number and

base pair position, if the p-value for HWE $<10^{-8}$ or if the absolute value of the beta or standard error was larger than 5. Next, a second-stage quality control was conducted using the R package 'EasyQC'[289], in which scatter plots of the observed allele frequencies versus the allele frequencies from CEU 1000G, scatter plots of the p-values versus the Z scores, quantile-quantile plots and λ_{GC} plots were generated to identify analytical errors.

Meta-analyses of the genome-wide association studies

Fixed-effect meta-analyses of the study-level GWAS results based on the effect sizes and standard errors were conducted using METAL[205]. Variants were dropped from the meta-analyses results if they were not available in at least 50% of the maximum sample size and at least half of the included datasets, if the maximum difference in effect allele frequency between the studies was larger than 0.2, if the minor allele frequency <1% or if the absolute value of the effect size or standard error was larger than 5.

Identification and clumping of genetic loci

Loci and lead variants were first identified within traits, using distance-based clumping. Variants with p-value$<5\times10^{-8}$ more than 1 Mb apart were considered to represent independent, genome-wide significant loci. Based on the significant lead variants for all traits, a second distance-based clumping was applied across traits and analyses, to identify the overall number of independent loci across traits. Variants which were less than 1 Mb apart were considered to represent the same locus. At each locus, the variant selected for the highest number of traits was selected as the sentinel variant at the locus. In case any of the other variants was in low linkage disequilibrium ($R^2<0.6$ and $D'<0.6$) then this variant was retained as a secondary signal at the locus.

Annotations, reported associations and cross-trait associations for loci reaching genome-significance for fat and lean mass compartments

Sentinel variants were annotated based on the most severe annotation of the variant in the Ensembl Variant Effect Predictor (VEP)[290]. Reported associations for the variants reaching genome-wide significance were identified through searches using the PhenoScanner[355] and loci reported in the largest GWAS for standard anthropometric traits, including BMI[172], height[172], WHR[309] and WHRadjBMI[309]. Associations with BF% and FFMI were based on the GWAS described in Chapter 3. Associations of variants not reported for BMI, height, WHR and WHRadjBMI but reaching nominal significance were based on GWAS conducted on the UK Biobank participants of European ancestry for the respective traits, as described in Chapter 3.

Fat and lean mass distribution patterns associated with the loci reaching genome-wide significance for fat and lean mass compartments were assessed by visualising and clustering the effect sizes of all significant loci across all fat and lean mass compartments using the R package 'pheatmap'[356] v1.0.10. Hierarchical clustering of both the variants and the body compartments was based on the Euclidean distances between the variants and phenotypes.

Results

Body composition by body region in a healthy middle-aged population

The distribution of fat and lean mass across the different regions of the body was assessed among 6,316 female and 5,295 male middle-aged, generally healthy participants of the Fenland study. Overall and regional fat and lean mass differed significantly between men and women, both on the absolute (kilograms (kg) fat, lean or bone mass) and relative scales (percentage of total body weight) (Figure 4.2, Table 4.1). In women, fat and lean body mass represent 37.4±6.9% (27.2±9.8 kg) and 59.3±6.6% (41.3±5.3 kg) of the total body weight, respectively, while in men around 28.9±5.9% (25.2±8.2 kg) of the total body weight is accounted for by body fat mass and 67.4±6.9% (57.2±6.8 kg) by lean mass (Table 4.1). While women store half of their total body fat in the trunk (18.1±5.1%) and the other half in the peripheral compartments (18.1±2.9%), men tend to store a higher proportion in the trunk (16.5±4.4%) than peripherally (11.3±2.1%). The distribution of peripheral fat over the legs and arms was similar in men and women, with around 77% of the total peripheral fat being stored in the legs and the remaining 23% in the arms. Despite men and women having similar relative amounts of android fat (men: 2.9±1.0% of total body weight, women: 2.8±1.1%), a larger relative proportion of women's android fat is located subcutaneously (2.1±0.6%) than viscerally (0.78±0.60%), while men store approximately equal amounts of android fat in the subcutaneous (1.4±0.5%) and visceral (1.5±0.8%) compartments. Although men had higher lean mass than women on the absolute and relative scale, the distribution of lean mass over the central and appendicular compartments was similar in men and women, with appendicular lean mass representing about 45% of the total lean mass.

Table 4.1: Overall and regional fat, bone and lean mass for 5,395 men and 6,316 women of the Fenland study, expressed in kilograms and as percentages of total body weight.

Trait	Men (mean±SD)	Women (mean±SD)	*p* for sex difference	Men % of total weight±SD	Women % of total weight±SD	*p* for sex difference
Age (years)	48.64±7.61	48.71±7.43	0.6	-	-	-
Weight (kg)	85.54±13.24	70.75±13.76	<2.2E-16	-	-	-
Height (cm)	177.43±6.62	163.99±6.31	<2.2E-16	-	-	-
Total fat mass (kg)	25.22±8.16	27.16±9.76	<2.2E-16	28.9±5.9%	37.4±6.9%	<2.2E-16
Legs fat mass (kg)	7.45±2.30	10.16±3.39	<2.2E-16	8.6±1.7%	14.1±2.6%	<2.2E-16
Arms fat mass (kg)	2.36±0.71	2.90±0.93	<2.2E-16	2.7±0.5%	4.0±0.07%	<2.2E-16
Peripheral fat mass (kg)	9.81±2.92	13.05±4.17	<2.2E-16	11.3±2.1%	18.1±2.9%	<2.2E-16
Gynoid fat mass (kg)	3.71±1.19	4.96±1.60	<2.2E-16	4.3±0.9%	6.9±1.2%	<2.2E-16
Android fat mass (kg)	2.54±1.15	2.13±1.16	<2.2E-16	2.9±1.0%	2.8±1.1%	0.39
Visceral fat mass (kg)	1.35±0.82	0.60±0.54	<2.2E-16	1.5±0.8%	0.78±0.60%	<2.2E-16
Subcutaneous fat mass (kg)	1.19±0.54	1.52±0.73	<2.2E-16	1.4±0.5%	2.1±0.6%	<2.2E-16
Trunk fat mass (kg)	14.51±5.50	13.32±5.98	<2.2E-16	16.5±4.4%	18.1±5.1%	<2.2E-16
Total lean mass (kg)	57.17±6.79	41.29±5.28	<2.2E-16	67.4±6.9%	59.3±6.6%	<2.2E-16
Legs lean mass (kg)	20.13±2.78	14.19±2.24	<2.2E-16	23.7±2.2%	20.3±2.2%	<2.2E-16
Arms lean mass (kg)	6.80±1.10	3.93±0.70	<2.2E-16	8.0±0.9%	5.6±0.7%	<2.2E-16
Appendicular lean mass (kg)	26.93±3.71	18.12±2.85	<2.2E-16	31.7±2.8%	25.9±2.7%	<2.2E-16
Gynoid lean mass (kg)	9.12±1.19	6.54±0.90	<2.2E-16	10.7±1.1%	9.4±1.1%	<2.2E-16
Android lean mass (kg)	4.25±0.55	3.14±0.44	<2.2E-16	5.0±0.5%	4.5±0.6%	<2.2E-16
Trunk lean mass (kg)	26.96±3.19	20.45±2.51	<2.2E-16	31.7±3.1%	29.4±3.8%	<2.2E-16
Total bone mass (kg)	3.15±0.38	2.39±0.31	<2.2E-16	3.7±0.4%	3.4±0.5%	<2.2E-16
Legs bone mass (kg)	1.21±0.17	0.85±0.12	<2.2E-16	1.42±0.20%	1.23±0.20%	<2.2E-16
Arms bone mass (kg)	0.45±0.06	0.30±0.039	<2.2E-16	0.53±0.08%	0.43±0.08%	<2.2E-16
Gynoid bone mass (kg)	0.32±0.05	0.23±0.037	<2.2E-16	0.38±0.06%	0.34±0.05%	<2.2E-16
Trunk bone mass (kg)	0.95±0.15	0.72±0.12	<2.2E-16	1.11±0.14%	1.03±0.16%	<2.2E-16

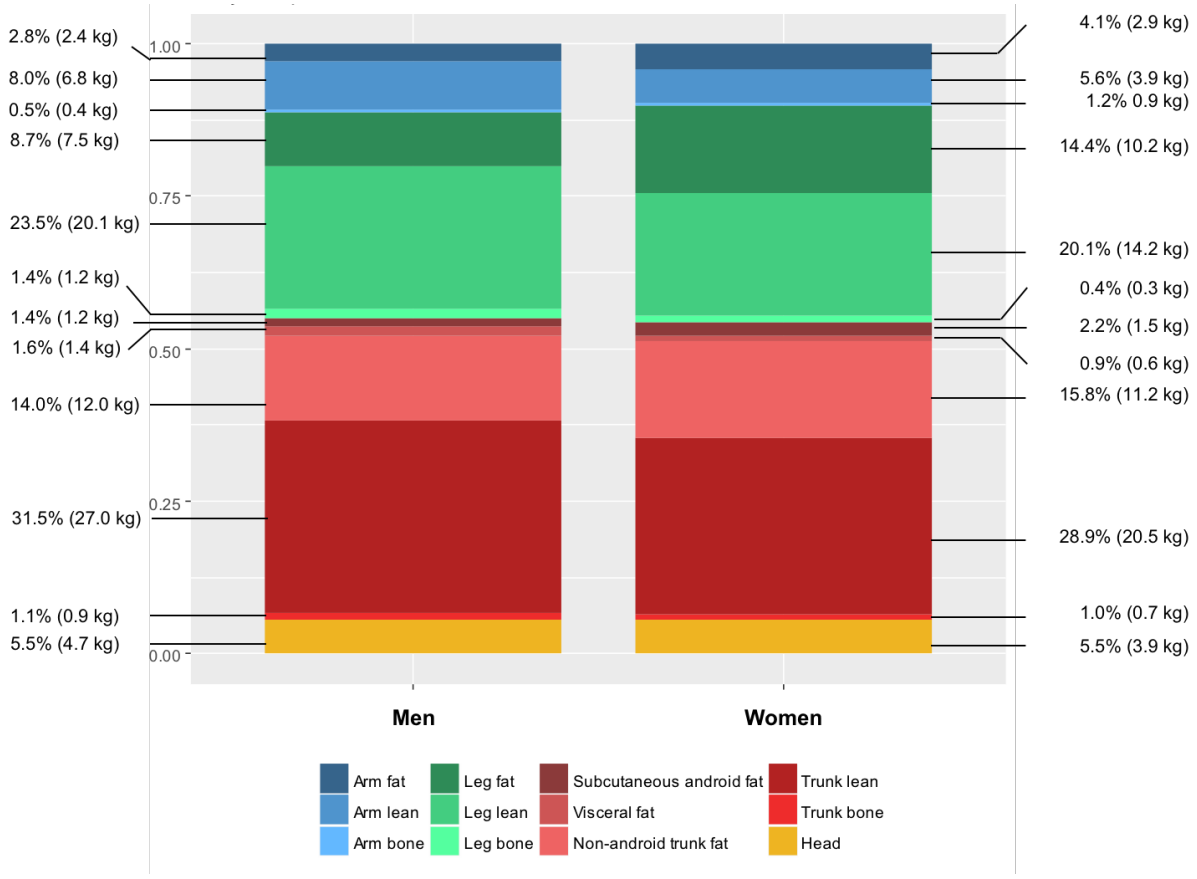


Figure 4.2: Histograms of the body composition by anatomical region in men and women of the Fenland study. Based on DEXA measures of regional fat, lean and bone mass on 11,711 participants (6,316 women).

To assess the extent to which fat and lean mass compartments measured by DEXA are captured by standard anthropometric traits, correlations of 6 standard anthropometric traits (weight, height, BMI, waist and hip circumference and WHR) with fat and lean mass compartments measured by DEXA were assessed in the Fenland study (Figure 4.3). Weight, BMI and waist and hip circumference were strongly correlated with all fat compartments (range of Pearson's correlation coefficients for the correlations of total body weight with fat compartments: R = 0.82-0.92; BMI: R = 0.74-0.92; waist: R = 0.72-0.90; hip: R = 0.61-0.91), while correlations of WHR with fat compartments were moderate (R = 0.27-0.68). Height correlated very weakly with regional fat mass (R = 0.01-0.21). Strong positive correlations between all fat compartments were observed (R = 0.53-0.98) (Figure 4.3 A).

The standard anthropometric trait most strongly correlated with lean mass compartments was total body weight (range of Pearson's correlation coefficients of height with lean mass compartments: R = 0.74-0.82), and substantial correlations with lean mass were also found for height (R = 0.37-0.56), BMI (R = 0.54-0.62), and waist (R = 0.51-0.60) and hip circumference (R = 0.55-0.67). Total

and regional lean mass were very strongly inter-correlated (R = 0.73-0.99) and also showed substantial correlations with all fat compartments (R = 0.38-0.56) (Figure 4.3 A).

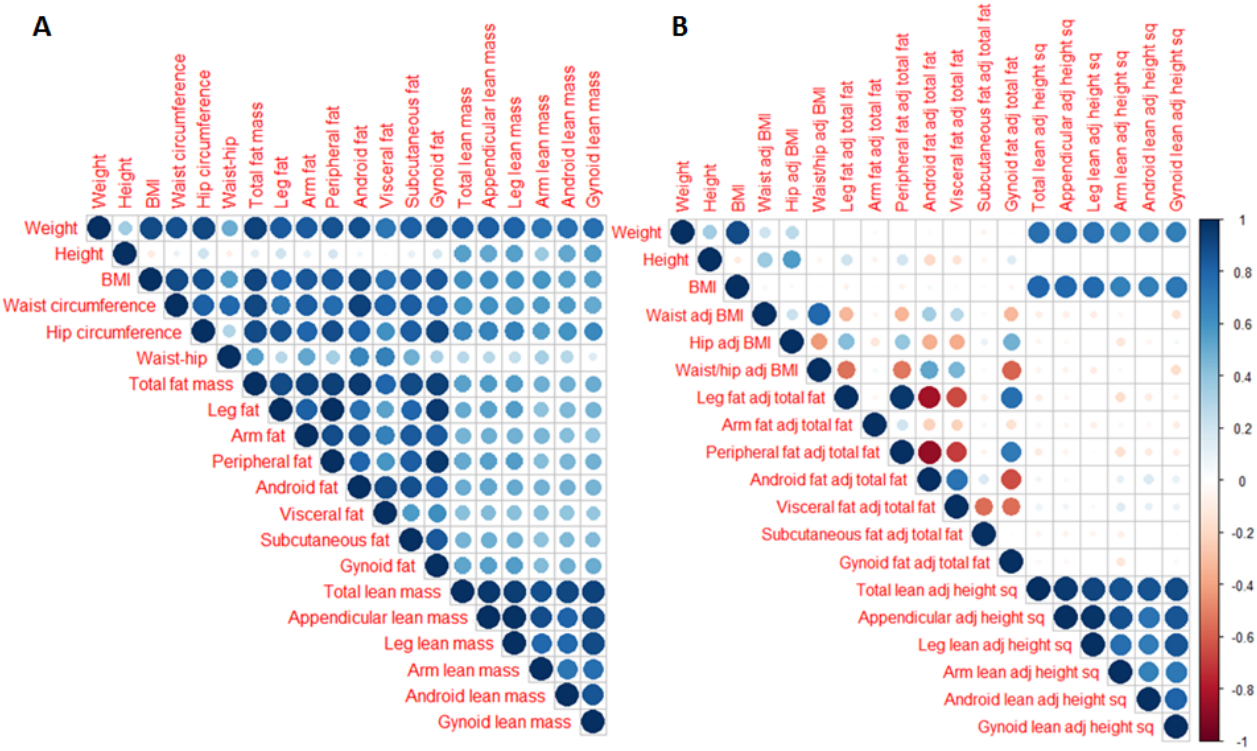


Figure 4.3: Plots of the correlations between standard and DEXA-based anthropometric traits. The Pearson's correlation coefficients were calculated based on data from 11,711 participants (6,316 women) from the Fenland study for which measurements of all included anthropometric traits were available. **A|** Correlations between anthropometric traits which were only adjusted for age and sex. **B|** Correlations between anthropometric traits which were all adjusted for age and sex, and additionally for BMI (waist, hip and waist-to-hip ratio), total body fat mass (fat compartments) or height squared (lean compartments).

After adjusting compartmental fat mass for total fat mass, no substantial correlations of weight (absolute value of R <0.02), height (absolute value of R <0.21) and BMI (absolute value of R <0.07) with adjusted fat compartments were found (Figure 4.3 B). Waist and WHR adjusted for BMI were positively correlated with relative fat mass in the android (waist adjusted for BMI: R =0.35; WHRadjBMI: R =0.52) and visceral compartments (waist adjusted for BMI: R =0.28; WHRadjBMI: R =0.46) and inversely with relative fat mass in leg (waist adjusted for BMI: R =-0.33; WHRadjBMI: R =-0.55) and gynoid compartments (waist adjusted for BMI: R =-0.32; WHRadjBMI: R =-0.58), but no correlations of BMI-adjusted waist, hip or WHR with arm (waist: R =-0.05, hip: R =-0.13, WHR: R =0.04, p for correlations ranging between 0.56-0.88) and subcutaneous android fat adjusted for total fat mass (waist: R =0.02, hip: R =0.09, WHR: R =-0.02, p for all 3 correlations >0.05) were found. Relative arm and subcutaneous android fat did

also not correlate with any of the other fat compartments (p for correlation between 0.62 and 0.95), with the exception of a suggestive inverse correlation of subcutaneous android fat with visceral android fat (R =-0.43, p=0.06). Total and visceral android fat were positively correlated (R =0.91), and were both inversely correlated with lower body fat compartments and total peripheral fat (Figure 4.3 B).

Total body weight and BMI – but not waist, hip or WHR adjusted for BMI – were substantially correlated with lean mass compartments adjusted for height squared (total body weight: R =0.65-76; BMI: R =0.67-0.79). Adjusted lean mass compartments remained strongly inter-correlated (R = 0.63-0.98) but did not correlate with adjusted fat compartments (p for correlations between adjusted fat and lean mass compartments >0.05) (Figure 4.3 B).

Genome-wide association analyses for absolute and relative fat and lean mass compartments

46 independent genetic variants associated with absolute fat or lean mass compartments

75 variants and 150 associations reaching genome-wide significance ($p \leq 5 \times 10^{-8}$) were identified for 9 fat (legs, arms, trunk, gynoid, total android, subcutaneous android, visceral android, total peripheral and total fat) and 7 lean mass compartments (arms, legs, android, gynoid, trunk, appendicular and total lean mass) through meta-analyses of GWAS adjusted for age and genetic principal components in up to 25,430 participants from 5 cohorts. 45 loci and 46 independent variants were identified through cross-trait LD clumping of the 75 lead variants (Figure 4.4, Supplementary Table 4.1). 11 of the 46 independent variants were only associated with fat mass compartment and 34 were only associated with one or more lean mass compartments. The locus near *FTO* was the only locus associated with both fat and lean mass compartments. *FTO* reached genome-wide significance for higher fat mass in all 9 regions, and with higher total, leg and appendicular lean mass and nominal significance for higher lean mass in the other 4 regions ($p \leq 1.1 \times 10^{-3}$). No heterogeneity in the effect sizes of the *FTO* locus across the 16 fat and lean mass compartments was identified (p for Cochran's Q statistic = 0.37).

The genetic effect sizes of the 46 independent variants on lean mass compartments clustered separately from the effect sizes on fat compartments when hierarchical clustering was applied to the effect sizes of the 46 independent variants on all 16 compartments (Figure 4.4) Overall, effect sizes across the lean mass compartments were more similar than effects on the fat compartments. Total android, subcutaneous android, trunk and total fat mass clustered separately from leg, gynoid and peripheral fat mass. Visceral android and, to a lesser extent, arm fat, clustered less closely with the other fat compartments (Figure 4.4).

64 independent variants identified for relative regional fat and lean mass

64 independent variants across 61 loci and 157 associations reached genome-wide significance in meta-analyses of GWAS for 8 fat mass compartments adjusted for total fat mass and 7 lean mass compartments adjusted for height squared in up to 25,260 participants from 5 cohorts (Figure 4.5, Table 4.2). 13 of the 64 independent variants reached genome-wide significance for one of more lean mass compartments adjusted for height squared, while 51 variants were associated with at least one the 8 fat compartments adjusted for total fat mass. No variants reached significance for both relative fat and lean mass compartments. Hierarchical clustering applied to the effect sizes of the 64 variants on all lean and fat mass compartments revealed three main clusters: one for central fat compartments, including trunk and total and visceral android fat, one for lower body fat

compartments, including leg, gynoid and peripheral fat, and one large cluster which was comprised of all lean mass compartments. Subcutaneous android fat clustered more closely with lean compartments than with central fat compartments, and relative arm fat clustered separate from gynoid, leg and total peripheral fat (Figure 4.5). Of the 51 loci associated with relative regional fat, 7 loci were only associated with arm fat mass (loci in/near *RDH14, PPARG, HOXA10-HOXA9, FGFR2, HOXC-AS2, TBX5* and *SPATA20*), 38 signals showed opposite associations with central versus lower body fat, 3 loci reached significance for subcutaneous android fat only (*LMX1A*, *WT1* and *PIWIL3*) and 3 other loci for gynoid fat only (*CALCRL*, *TMCC1* and *LY86*).

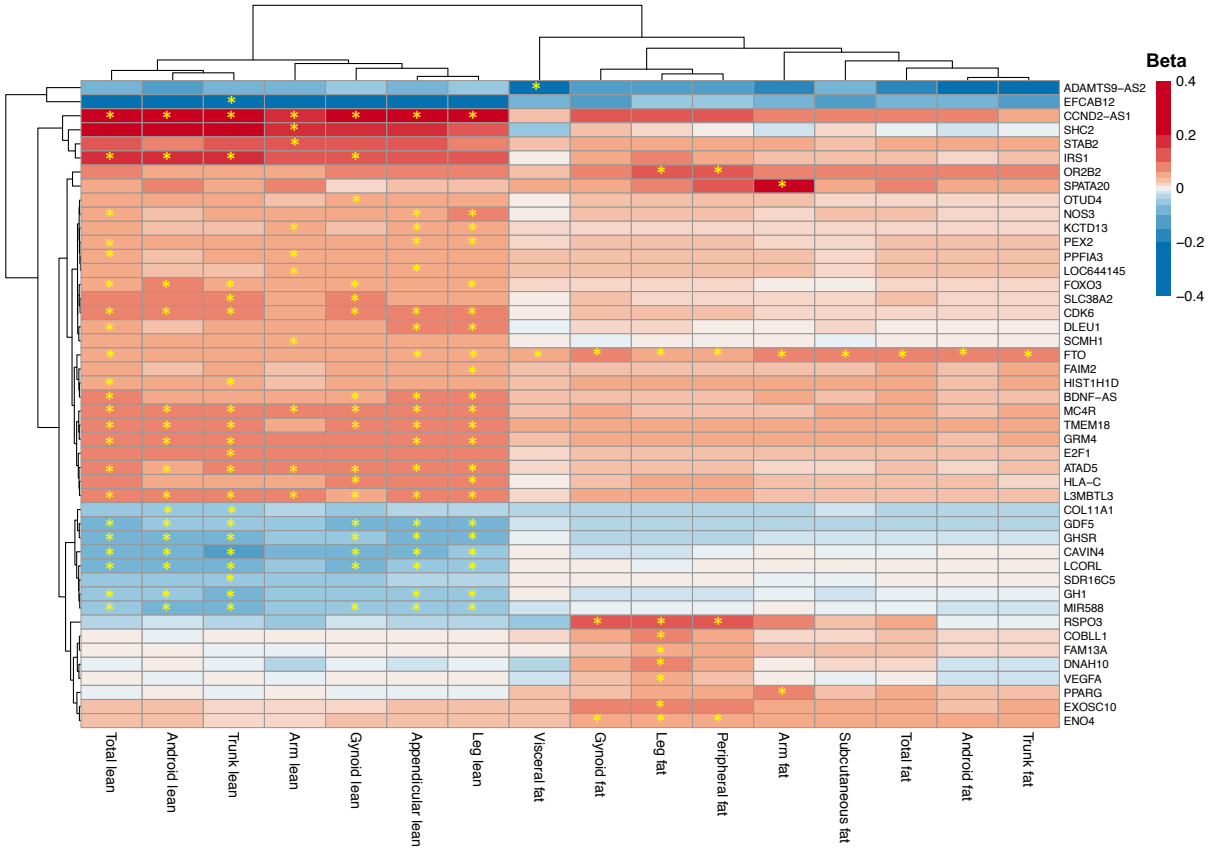


Figure 4.4: Heat plot and hierarchical clustering of the loci reaching genome-wide significance for absolute fat and lean mass compartments measured by DEXA. Associations of 46 independent variants at 45 loci reaching genome-wide significance ($p \leq 5 \times 10^{-8}$) for at least one of the 7 lean mass and/or 9 fat mass compartments are represented in this plot. The yellow asterisks indicate associations that reach genome-wide significance and colour coding indicates the per-allele effect sizes on SDs of the DEXA phenotype. Effect sizes are aligned to the allele with a positive effect size on BMI in the UK Biobank. Phenotypes were adjusted for age, genetic principal components and study-specific covariates only.

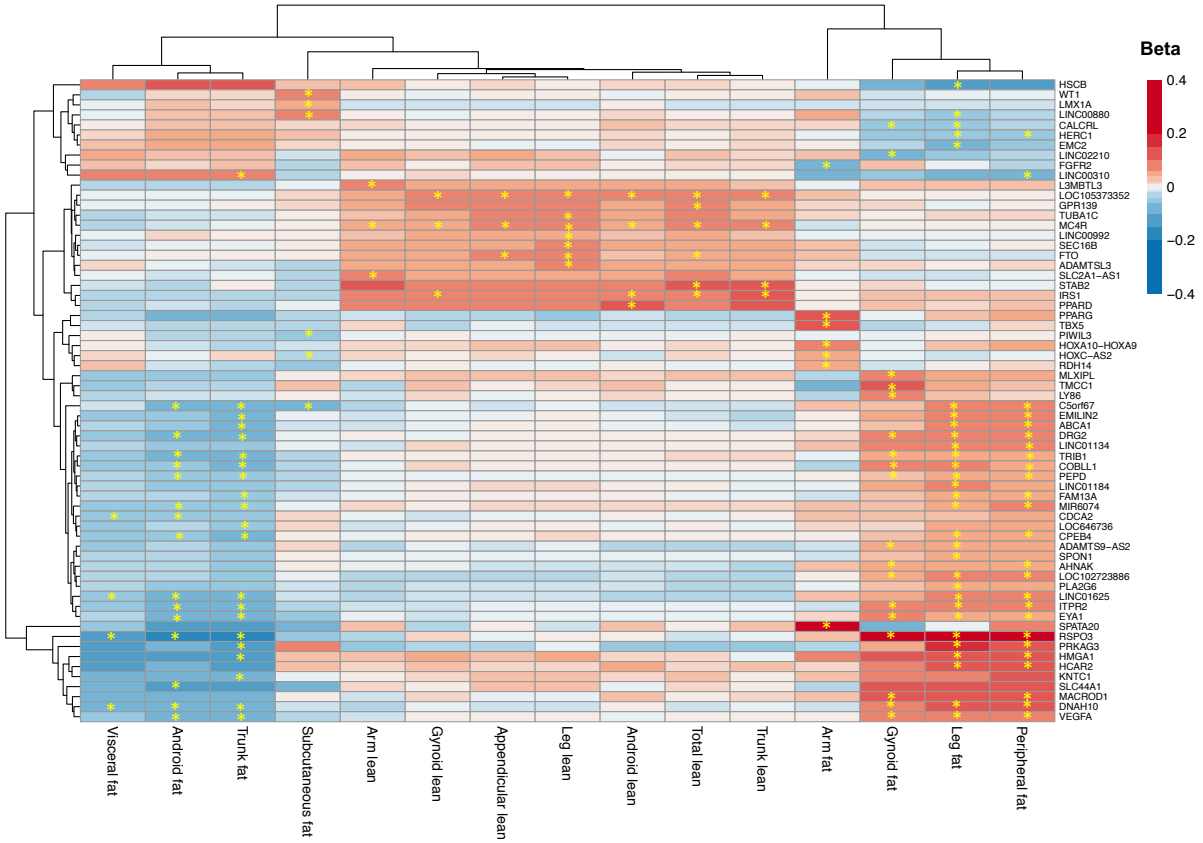


Figure 4.5: Heat plot of 64 independent variants associated with fat or lean mass distribution at genome-wide significance. Models were adjusted for age, genetic principal components and study-specific covariates, and additionally for total fat mass for the fat compartments and for height squared for the lean mass compartments. The colour coding indicates the per-allele effect size on standard deviations of the phenotypes. Effect sizes are aligned to the allele with a positive effect size on BMI in the UK Biobank. Associations reaching genome-wide significance are indicated by yellow asterisks.

Seven genetic loci reach genome-wide significance for arm fat

Seven genetic loci in/near *RDH14, PPARG, HOXA10-HOXA9, FGFR2, HOXC-AS2, TBX5* and *SPATA20* reached genome-wide significance for arm fat mass adjusted for total fat, of which the loci in *PPARG* and *SPATA20* also reached significance for absolute arm fat. None of the arm fat loci reached genome-wide significance for any other fat or lean mass compartments (Figures 4.4 and 4.5).

Three loci in *PPARG*, *SPATA20* and *TBX5* for arm fat adjusted for total fat mass were associated with a pattern of fat distribution characterised by higher relative fat mass in the arms and less relative fat mass in the central compartments, without affecting relative leg fat mass (Figure 4.6). The intronic variant rs13064760 (EAF=11.8%) in *PPARG* was strongly associated with higher relative arm fat mass (per-allele effect size on SDs of arm fat mass adjusted for total fat mass=0.117, p=2.2×10^{-17}), and reached nominal significance for lower relative android (beta=-0.066, p=1.7×10^{-6}) and trunk fat (beta=-0.062, p=7.8×10^{-6}) but was not associated with relative leg

(beta=0.020, p=0.14) or gynoid fat mass (beta=-0.001, p=0.94). The arm fat-raising allele showed suggestive associations with lower android visceral (beta=-0.039, $p=5.5\times10^{-3}$) and subcutaneous fat mass (beta=-0.030, p=0.032). The *PPARG* locus also reached significance for higher absolute arm fat mass (sentinel variant rs13083375, EAF=11.8%, in perfect LD with the sentinel variant rs13064760 for relative arm fat mass), with a per-allele effect size of 0.086 SDs on arm fat. Heterogeneity among the effect sizes of the *PPARG* locus across all absolute fat compartments was identified (p for Cochran's Q statistic=0.035), which corroborates the effect of *PPARG* on fat distribution. Nominally significant but weaker associations with absolute leg (beta=0.046, $p=8.1\times10^{-4}$) and peripheral fat mass (beta=0.059, $p=2.1\times10^{-5}$) were also found, while effect sizes on absolute central fat compartments did not reach nominal significance (trunk fat: beta=0.031, p=0.025, total android fat: beta=0.027, p=0.054, subcutaneous android fat: beta=0.035, p=0.012, visceral android fat: beta=0.020, p=0.14). The arm fat-raising allele at *PPARG* has been reported for higher BMI[172] (beta=0.017, $p=1.3\times10^{-11}$) and is known to be associated with *lower* T2D risk[46] (BMI unadjusted: OR=0.878, $p=1.4\times10^{-12}$; BMI adjusted: OR=0.861, $p=1.1\times10^{-11}$). Variants in/near *PPARG* have been reported for WHRadjBMI[308] but these were LD-independent (R^2=0.03) of the sentinel variant for arm fat mass, which reached nominal significance for lower WHR ($p=7.4\times10^{-6}$) but not for WHRadjBMI (p=0.025) in UK Biobank, indicating that the association with WHR was mostly driven by its effect on BMI.

The low-frequency, missense sentinel variant in *SPATA20* (rs62621401, p.Lys406Arg, EAF=98.5%) had a per-allele effect size of 0.391 SDs on arm fat adjusted for total fat mass ($p=1.4\times10^{-26}$). Weaker effects on lower relative android (beta=-0.110, p=0.0029), trunk (beta=-0.119, p=0.0012) and gynoid fat mass (beta=-0.0927, p=0.012) but not on relative leg fat (beta=-0.007, p=0.85) were observed (Figure 4.6). The missense variant was also associated with higher absolute arm fat mass (per-allele effect size on SDs of arm fat mass=0.222, $p=1.0\times10^{-9}$), but did not reach significance for other compartments. Weaker, suggestive associations (p<0.05) were found with higher total fat (beta=0.076, p=0.036), leg fat (beta=0.073, p=0.045) and peripheral fat (beta=0.107, p=0.003), but evidence for heterogeneity in the effect sizes of rs62621401 across the absolute fat compartments was found (p for Cochran's Q statistic=0.012). The variant was not associated with BMI (effect size on SDs of BMI per arm fat-raising allele=0.012, p=0.12), height (beta=0.010, p=0.24) in published exome chip meta-analyses[280,291], nor did it reach significance in recently published GWAS[309] and exome chip analyses[292] for WHR or WHRadjBMI. The *SPATA20* locus was also not associated with BF% in the analyses presented in Chapter 3 of this thesis (beta=-0.012, p=0.47). A suggestive (p<0.05) inverse association of the arm fat-increasing allele with

higher T2D risk (BMI unadjusted: OR=1.03, p=0.031; BMI adjusted: OR=1.02, p=0.14)[40] but not with CHD risk (OR=1.05, p=0.12)[298] has been reported.

The intronic variant rs2551377 in *TBX5* (EAF=16.8%) was associated with higher relative arm fat mass (per-allele effect size on SDs of arm fat adjusted for total fat mass=0.132, $p=3.4\times10^{-28}$), had suggestive associations with lower relative android (beta=-0.039, p=0.0011), trunk (beta=-0.029, p=0.017) and gynoid fat (beta=-0.035, p=0.0039) but was not associated with relative leg (beta=-0.017, p=0.16), visceral (beta=-0.012, p=0.33) or subcutaneous android fat (beta=-0.021, p=0.085) (Figure 4.6). The *TBX5* locus has not been reported for BMI, WHR or WHRadjBMI.

HOXA10-HOXA9, *RDH14*, *HOXC-AS2* and *FGFR2* were also significantly associated with relative arm fat mass but had no effect on total android fat mass (p>0.05) (Figures 4.5 and 4.6). None of these 4 loci have been reported for BMI, WHR or WHRadjBMI. *HOXA10-HOXA9* (rs1859164, EAF=53.8%) was associated with higher arm fat mass (beta=0.080, $p=7.1\times10^{-19}$), showed weak associations with lower trunk fat (beta=-0.028, p=0.0016) and higher leg fat (beta=0.021, p=0.019) but did not affect android (p=0.084), gynoid (p=0.66), visceral (p=0.99) and subcutaneous fat (p=0.05). *RDH14* (rs11096542, EAF=59.2%) was associated with higher arm fat mass (beta=0.050, $p=4.0\times10^{-8}$), reached nominal significance for lower subcutaneous android fat (beta=-0.042, $p=5.3\times10^{-6}$) and showed suggestive associations with higher visceral android fat (beta=0.023, p=0.011) and lower trunk fat (beta=-0.026, p=0.0049), but not with leg (p=0.59), gynoid (p=0.12) or total android fat (p=0.22). *HOXC-AS2* (rs1109391, EAF=56.6%) increased arm fat (beta=0.050, $p=2.5\times10^{-8}$) while suggestively decreasing subcutaneous android fat (beta=-0.030, p=0.0010) and increasing visceral android fat (beta=0.018, p=0.042), but not android (p=0.30), trunk (p=0.21), leg (p=0.05), or gynoid fat (p=0.33). Rs11199849 at *FGFR2* (EAF=74.8%) reached genome-wide significance for higher arm fat (beta=0.073, $p=1.2\times10^{-12}$), nominal significance for lower visceral fat (beta=-0.037, $p=3.7\times10^{-4}$) and showed suggestive associations with lower trunk (beta=-0.028, p=0.0063) and gynoid fat (beta=-0.028, p=0.0062) and higher subcutaneous android fat (beta=0.022, p=0.033). *FGFR2* was not associated with leg (p=0.61) or total android fat mass (p=0.062).

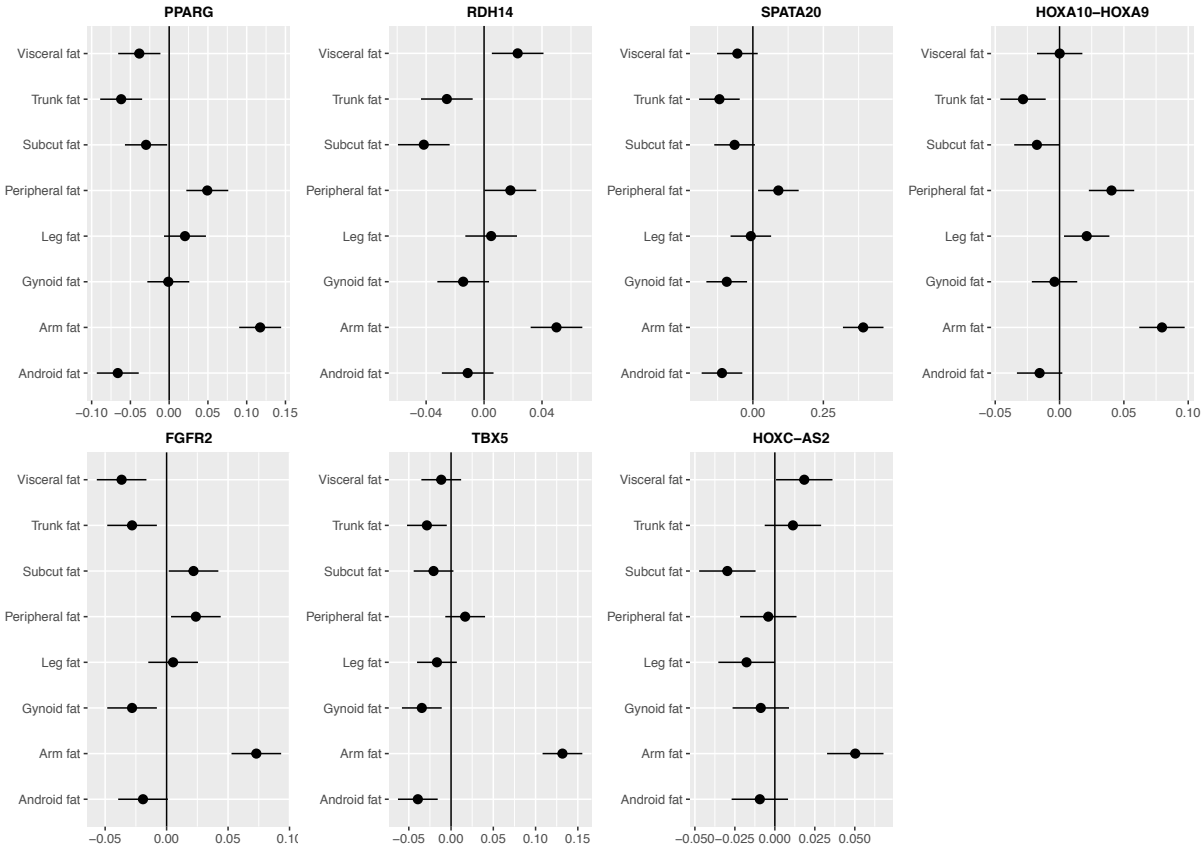


Figure 4.6: Forest plots of associations of 7 arm fat loci with all 8 fat compartments adjusted for total fat mass. The 7 presented loci reached genome-wide significance for arm fat only ($p \leq 5\times10^{-8}$) in GWAS adjusted for age, genetic principal components, study-specific covariates and total fat mass. Effect sizes ± 95% confidence intervals are expressed as per-allele betas on SDs of fat compartments.

No uniform association pattern of the 7 relative arm fat-raising loci with T2D risk was found, based on published T2D GWAS[38,219] and exome chip[40] summary-level results (Figure 4.7). The arm fat-raising allele at rs13064760 (*PPARG*) reached genome-wide significance for lower T2D risk (OR=0.90, $p=5.9\times10^{-23}$), while rs11096542 at *RDH14* reached nominal significance (OR=0.97, $p=3.6\times10^{-4}$). Suggestive associations of the arm fat-raising alleles at *SPATA20* with lower T2D risk (OR=0.97, $p=0.031$) and at *HOXC-AS2* (OR=1.02, $p=0.040$) with higher T2D risk were found. *TBX5* ($p=1.00$) and *FGFR2* ($p=0.74$) were not associated with T2D risk.

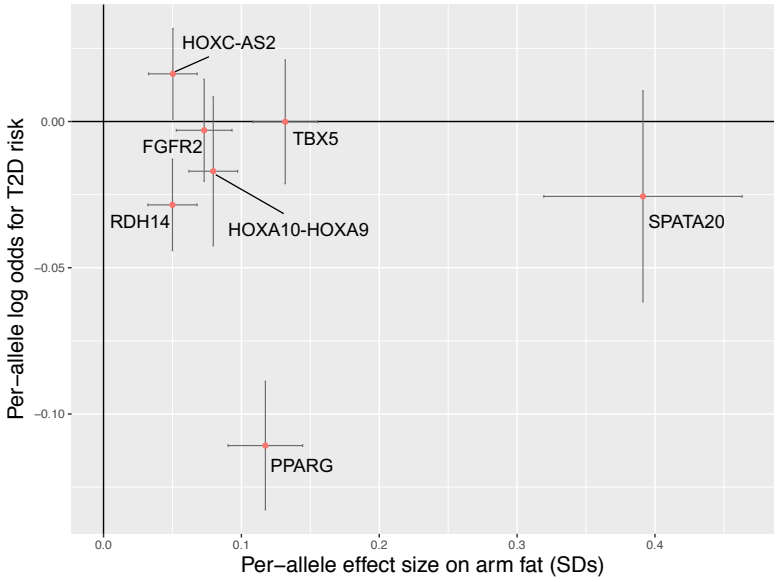


Figure 4.7: Scatter plot of effect sizes of 7 arm fat loci on SDs of arm fat adjusted for total fat mass versus log odds for T2D. Effect sizes on arm fat mass were adjusted for total fat mass, age, genetic PCs and study-specific covariates. Look-ups for T2D came from the summary-level GWAS results by Xue *et al.*[38], except for *HOXC-AS2* (Scott *et al.*[46]) and *SPATA20* (Mahajan *et al.*[40]). Alleles were aligned to the arm fat-raising allele.

Majority of loci for relative regional fat mass associated with central-to-lower body fat distribution

Loci associated with central-to-lower body fat distribution are captured by WHR

38 of the 51 loci reaching genome-wide significance for fat compartments adjusted for total fat mass were associated with a pattern of higher central-to-lower body fat distribution (Figures 4.5 and 4.8). Loci were considered to be associated with higher central-to-lower body fat distribution if they were associated with at least one relative central fat mass compartment (android or trunk) or one relative lower body fat compartment (legs or gynoid region) at genome-wide significance, and showed nominally significant associations (Bonferroni corrected significance threshold for 64 loci associated with relative fat or lean mass compartments: $p \leq 7.8 \times 10^{-4}$) in the opposite direction with lower or central body fat, respectively. Central-to-lower body fat distribution is also captured by WHR and WHRadjBMI, as evidenced by the fact that all but 3 (*LINC01134*, *EMC2* and *PLA2G6*) of the 38 loci have previously been reported for WHR and/or WHRadjBMI[308,309]. The central fat-increasing allele at *LINC01134* (rs4073647) reached nominal significance for WHRadjBMI (beta=0.012, $p=1.7 \times 10^{-4}$) but not for WHR (beta=0.0084, $p=0.0049$) ($p \leq 7.8 \times 10^{-4}$ in GWAS for WHR and WHRadjBMI in UK Biobank) and was not associated with BF% (beta=-0.0030, $p=0.22$) or BMI (beta=-0.0022, $p=0.36$). The central fat-increasing allele at rs13264315 (*EMC2*) was not associated with BF% (beta=0.0030, $p=0.17$), BMI (beta=0.0030, $p=0.14$), WHR

(beta=0.0027, p=0.23) or WHRadjBMI (beta=0.0016, p=0.43). The leg fat-lowering allele at *PLA2G6* (rs4380, EAF=52.8%) reached genome-wide significance for lower BF% (beta=-0.019, $p=2.0\times10^{-22}$) and nominal significance for lower BMI (beta=-0.0066, $p=6.5\times10^{-4}$), but was not associated with WHR (beta=-0.0001, p=0.98) or WHRadjBMI (beta=0.0044, p=0.033).

Five loci for central-to-lower body fat distribution affect absolute leg fat mass only

5 of the 38 loci (*DNAH10*, *RSPO3*, *FAM13A*, *VEGFA* and *COBLL1*) associated with central-to-lower body fat distribution also reached genome-wide significance for absolute leg fat mass, but none of these loci affected central fat mass (p>0.05) (*DNAH10*: p_{trunk}=0.13, $p_{android}$=0.24; *RSPO3*: p_{trunk}=0.97, $p_{android}$=0.64; *COBLL1*: p_{trunk}=0.13, $p_{android}$=0.23; *FAM13A*: p_{trunk}=0.15, $p_{android}$=0.11; *VEGFA*: p_{trunk}=0.14, $p_{android}$=0.06). Significant heterogeneity between the effect sizes on the 9 fat compartments was observed for all five loci (*DNAH10*: p for Cochran's Q statistic=5.9×10^{-15}, *RSPO3*: $p=3.1\times10^{-14}$, *FAM13A*: $p=4.9\times10^{-3}$, *COBLL1*: $p=1.1\times10^{-3}$, *VEGFA*: $p=9.6\times10^{-11}$). These findings indicate that the associations of *DNAH10*, *RSPO3*, *FAM13A*, *VEGFA* and *COBLL1* with relative central-to-lower body fat distribution and WHR is solely driven by leg fat mass, without substantial effects on central fat mass.

The effects of WHR loci on visceral versus subcutaneous android fat are inversely correlated

The effect sizes of the 38 loci associated with central-to-lower body fat across all 8 fat compartments were assessed and hierarchical clustering was applied (Figure 4.8 A). As expected, the effect sizes on lower relative gynoid, leg and peripheral fat mass and the effect sizes on higher relative trunk and total and visceral android fat mass clustered together. However, associations with relative arm and subcutaneous android fat were generally weaker, followed different patterns and formed a separate, third cluster (Figure 4.8 A).

To further investigate the effects of loci for higher central-to-lower body fat distribution on relative visceral and subcutaneous fat mass, I compared the effect sizes on the two separate android compartments (Figure 4.8 B). 36 of the 38 loci associated with higher central-to-lower body fat distribution were at least suggestively associated with higher relative visceral fat mass (p values ranging from 0.037 for *HERC1* to 1.4×10^{-16} for *DNAH10*). Conversely, 14 of the 38 loci, which were all positively associated with visceral fat (p-values between 1.2×10^{-3} and 2.2×10^{-10}), had no effect (*LOC646736, ADAMTS9-AS2, CPEB4, HMGA1, MLXIPL, CDCA2, SPON1, AHNAK, MACROD1, HCAR2, LINC02210, LINC00310*) (p>0.05) or a negative effect (*PRKAG3*: beta for relative subcutaneous fat=-0.064 , p=0.005) on relative subcutaneous android fat. *LINC00880* and

C5orf67 were the only two loci which did not affect relative visceral fat mass (*LINC00880*: beta=-0.0019, p=0.84; *C5orf67*: beta=0.0102, p=0.36). However, both loci were significantly associated with higher relative subcutaneous android fat (*LINC00880*: beta=0.068, $p=6.6\times10^{-13}$; *C5orf67*: beta=0.094, $p=3.5\times10^{-15}$), which suggests that these two known loci for WHR increase waist circumference by acting on the subcutaneous android compartment but not on the visceral compartment.

A weak inverse correlation between the effect sizes of the central-to-lower fat distribution-increasing alleles on relative visceral and subcutaneous android fat was observed (R = -0.359, p=0.027) (Figure 4.8 B). This suggests that WHR loci do not tend to affect visceral and subcutaneous android fat similarly. An important exception to this was the intronic variant rs577721086 in *RSPO3*, which had strong positive effect sizes on both relative visceral and subcutaneous android fat (visceral: beta=0.147, $p=9.9\times10^{-13}$; subcutaneous: beta=0.059, p=0.005).

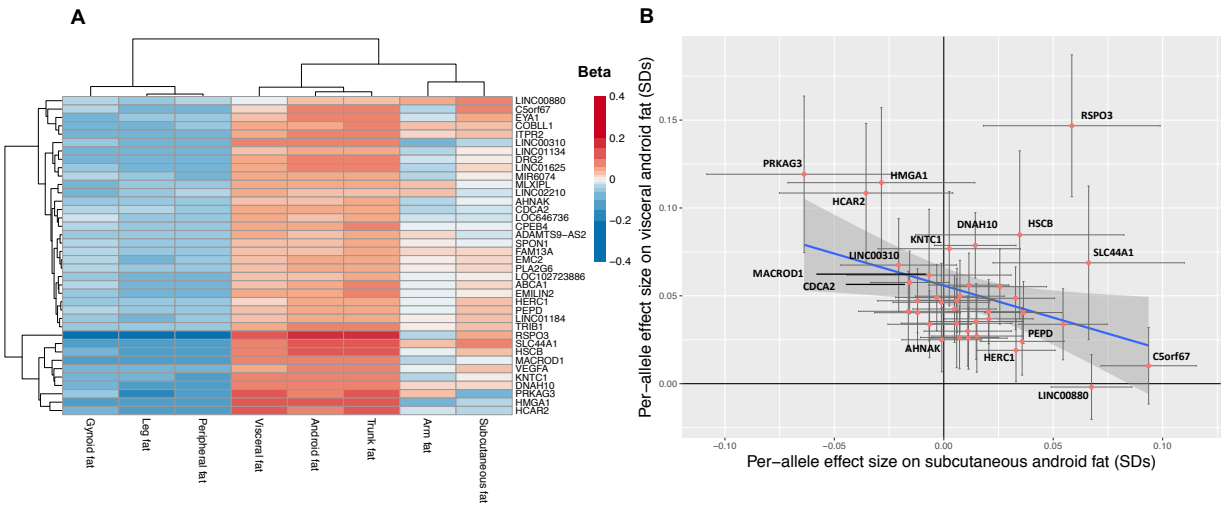


Figure 4.8: Cross-compartment associations for the 38 loci associated with central-to-lower body fat distribution. A| Heat plot and hierarchical clustering of the effect sizes on 8 body fat compartments adjusted for total fat mass for the 38 loci associated with higher central-to-lower body fat distribution. The colour coding represents the per-allele effect size on SDs of fat compartments adjusted for total fat mass, age, genetic PCs and study-specific covariates. **B|** Scatter plot of the effect sizes (±95% confidence intervals) of the 38 central-to-lower body fat-increasing alleles on subcutaneous versus visceral android fat adjusted for total fat mass. The blue line and grey shadow represent the best line of fit ± standard error.

Central-to-lower body fat distribution and risk of type 2 diabetes

10 of the 38 loci for higher central-to-lower body fat distribution were associated with higher T2D risk[38,46] ($p\leq7.8\times10^{-4}$). All 10 loci were positively associated with relative visceral fat mass (p<0.05), of which 5 reached nominal significance (*ITPR2*, *COBLL1*, *LOC646736*, *ADAMTS9-AS2*, *LINC01184* and *DNAH10*). Conversely, only 5 of the 10 loci that were associated with T2D were

also associated with higher subcutaneous fat (*C5orf67*, *HERC1*, *COBLL1*, *ITPR2* and *LINC01184*) while the five other loci (*LOC102723886*, *MACROD1*, *DNAH10*, *LOC646736* and *ADAMTS9-AS2*) were not associated with subcutaneous android fat (p>0.05). Positive but non-significant correlations were observed between the effect sizes of the 38 variants on total android fat and T2D risk (R = 0.198, p=0.24) and on visceral android fat and T2D risk (R=0.281, p=0.092), while the effect sizes on subcutaneous android fat and T2D risk were negatively but non-significantly correlated (R=-0.111, p=0.51) (Figure 4.9). Overall, these findings indicate that loci for central-to-lower body fat distribution are more likely to increase T2D risk if they are associated with visceral fat mass.

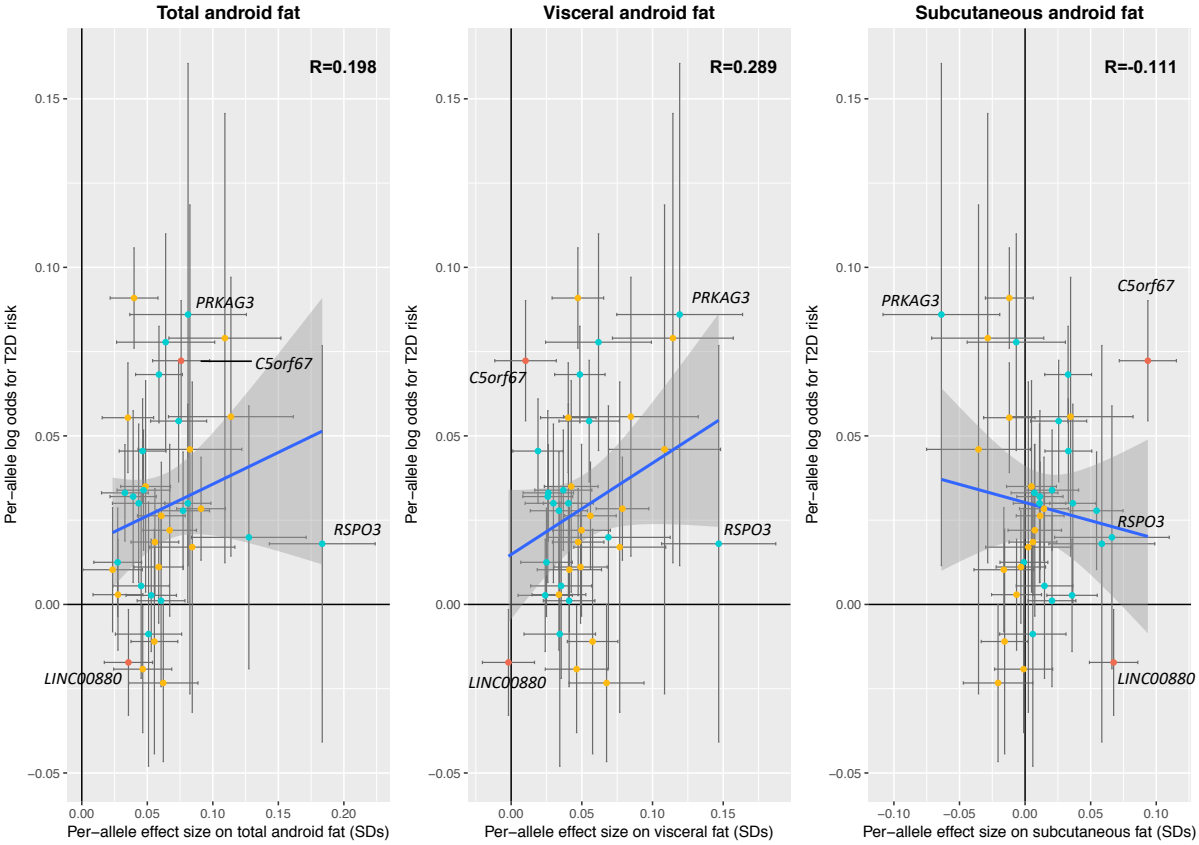


Figure 4.9: Scatter plots of effect sizes of 38 loci for higher central-to-lower body fat distribution on total, visceral and subcutaneous android fat versus T2D risk. All alleles are aligned to the android fat-increasing allele and effect sizes on fat compartments are adjusted for total fat mass, age, genetic PCs and study-specific covariates. Associations with T2D risk came from Xue *et al.*[38], if available, and from Scott *et al.*[46] when not available in the results by Xue *et al.* The blue line and grey shadow represent the best line of fit ± standard error. Loci in red reached nominal significance for subcutaneous fat only but were not associated with visceral fat, loci in yellow reached nominal significance for visceral fat but were not associated with subcutaneous fat and loci in blue reach nominal or suggestive associations for both visceral and subcutaneous fat.

6 loci associated with gynoid or subcutaneous android fat

Finally, 3 loci only reached genome-wide significance for relative gynoid fat mass (rs7586854 near *CALCRL*, rs72628505 in *TMCC1* and rs744056 in *LY86*) and three genetic signals were identified for relative android subcutaneous fat only (rs1553607 in *LMX1A*, rs11031749 near *WT1* and rs9612738 near *PIWIL3*). Despite the fact that all 3 loci associated with higher gynoid fat have been reported for lower WHR and/or WHRadjBMI, they did not reach nominal significance for lower relative android or trunk fat ($p>7.8\times10^{-4}$) (Figure 4.10 upper row). However, all 3 loci did reach nominal significance for higher leg fat mass ($p\leq7.8\times10^{-4}$), and *TMCC1* and *LY86* also reached nominal significance for *lower* arm fat mass.

The subcutaneous android fat-raising loci at *LMX1A* and *WT1* have been reported for WHRadjBMI[309] but not for BMI (p for BMI in UK Biobank GWAS =0.46 and 0.003, respectively) and were also not associated with BF% (p for BF% in UK Biobank GWAS=0.78 and 0.67, respectively). The nominally significant effect of *LMX1A* on total android fat (beta=0.037, $p=4.0\times10^{-5}$) was only due to subcutaneous fat, as no association with visceral fat mass (beta=-0.002, $p=0.82$) was found. The subcutaneous android fat-raising allele at *WT1* had a suggestive inverse association with visceral fat mass (beta=-0.028, $p=0.010$) and a positive association with arm fat mass (beta=0.027, $p=0.010$). Finally, *PIWIL3* has not been reported for any of the traditional anthropometric traits ($p_{bmi}=0.66$, $p_{BF\%}=0.76$, $p_{WHR}=0.65$, $p_{WHRadjBMI}=0.85$, based on GWAS in UK Biobank) and the allele associated with higher subcutaneous android fat did not affect visceral fat mass (beta=-0.001, $p=0.94$) (Figure 4.10, bottom row).

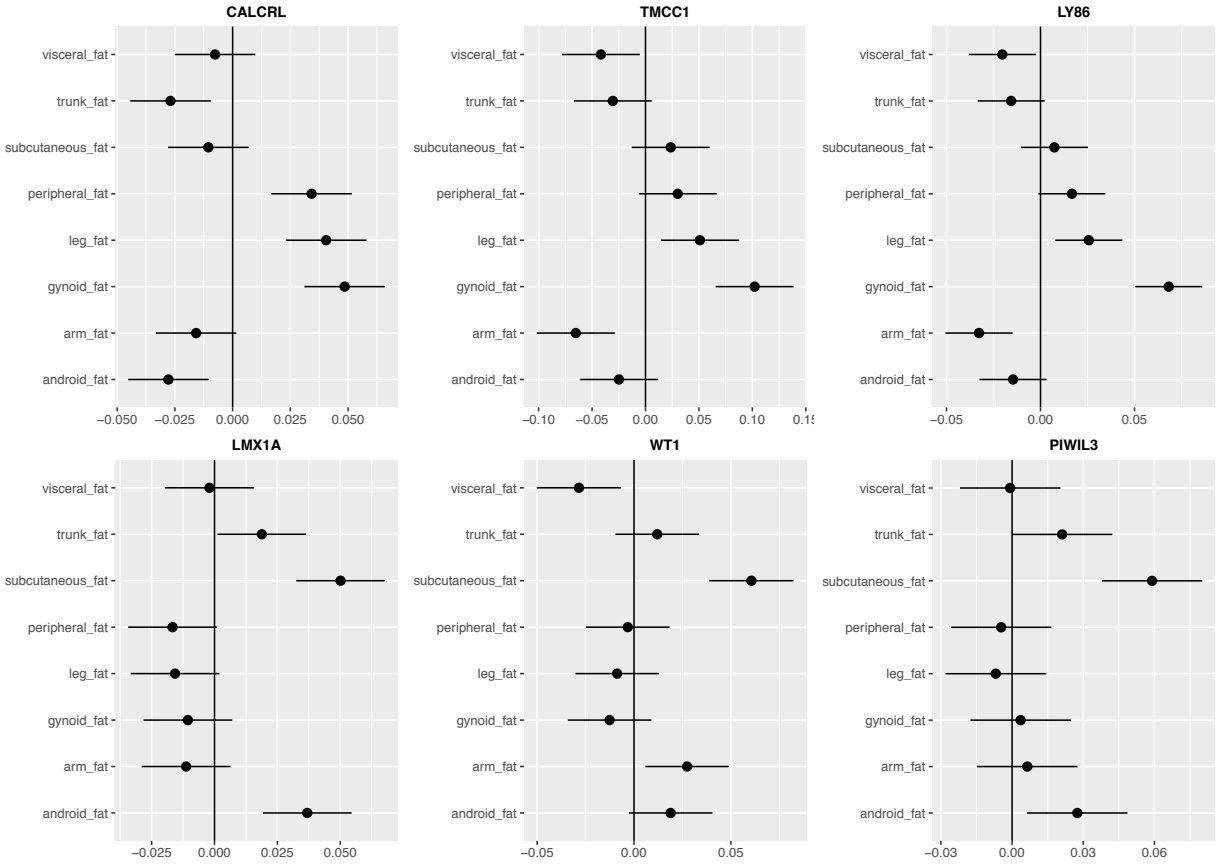


Figure 4.10: Forest plots of effect sizes 3 loci for gynoid fat and 3 loci for subcutaneous android fat on fat compartments adjusted for total fat mass. The 3 loci on the first row reached genome-wide significance for gynoid fat only and the three loci on the second row reached genome-wide significance for subcutaneous android fat only. Results are based on GWAS adjusted for age, genetic principal components, study-specific covariates and total fat mass. Effect sizes ± 95% confidence intervals are expressed as per-allele effect sizes on SDs of fat compartments.

No evidence for distinct genetic factors driving lean mass in different body compartments

No heterogeneity between the effect sizes on the 7 lean mass compartments was observed for any of the 34 loci identified for absolute lean mass compartments (p for Cochran's Q statistic ranging from p=0.18 for *GDF5* to p=0.99 for *BDNF*). 4 of the 34 loci were associated with all 7 lean mass compartments (*MC4R, ATAD5, CCND2, L3MBTL3*) and 7 loci were associated with all compartments except lean mass in the arms (*GDF5, GH1, LCORL1, TMEM18, CDK6, GHSR, MIR588*). 5 of the 34 lean mass loci were only significantly associated with trunk lean mass (*CAVIN4, SDR16C5, EFCAB12, SHC2, E2F1*), and *OTUD4, LOC644145, FAIM2* and *SCMH1* only reached significance for gynoid, appendicular, leg and arm lean mass, respectively. The remaining 14 loci were associated with between 2 and 5 lean mass compartments. Of the 34 lean mass loci, 16 have been reported for height[172,314] and 8 loci have been reported for BMI[172,217], all with directionally consistent effects on lean mass and height or BMI. The remaining 5 lean mass loci (*STAB2, CAVIN4, EFCAB12, PPFIA3* and *LOC644145*) have not been reported for height or BMI. 11 and 7 of the 34 lean mass loci reached genome-wide and nominal significance, respectively, for FFMI in the GWAS presented in Chapter 3. 16 loci were not associated with FFMI ($p>1.1\times10^{-3}$), of which all but two (*CAVIN4*: p_{height} in GWAS UK Biobank=0.67, *LOC644145*: $p_{height}=1.1\times10^{-3}$) reached genome-wide significance for height in the same direction as with lean mass, which suggests that their association with absolute regional lean mass could have been driven mostly be height.

Of the 13 loci identified for lean mass compartments adjusted for height squared (*SLC2A1, SEC16B, LOC105373352, IRS1, LINC00992, PPARD, L3MBTL3, TUBA1C, STAB2, ADAMTSL3, GPR139, FTO* and *MC4R*), 6 loci also reached genome-wide significance for unadjusted lean mass compartments (*TMEM18, IRS1, L3MBTL3, STAB2, FTO* and *MC4R*). Although none of the 13 loci reached genome-wide significance for all lean mass compartments, all lean loci had directionally consistent effects and reached nominal significance (Bonferroni corrected p-value for 64 tests $\leq7.8\times10^{-4}$) across all lean mass compartments, with the exception of *TUBA1C* for arm lean mass (p=0.002) and *PPARD* for leg lean mass (p=0.0009) (Figure 4.11). For none of the 13 lean mass loci heterogeneity was found among the effect sizes on the 7 lean mass compartments (Cochran's Q statistic's p-value>0.05) (Figure 4.11).

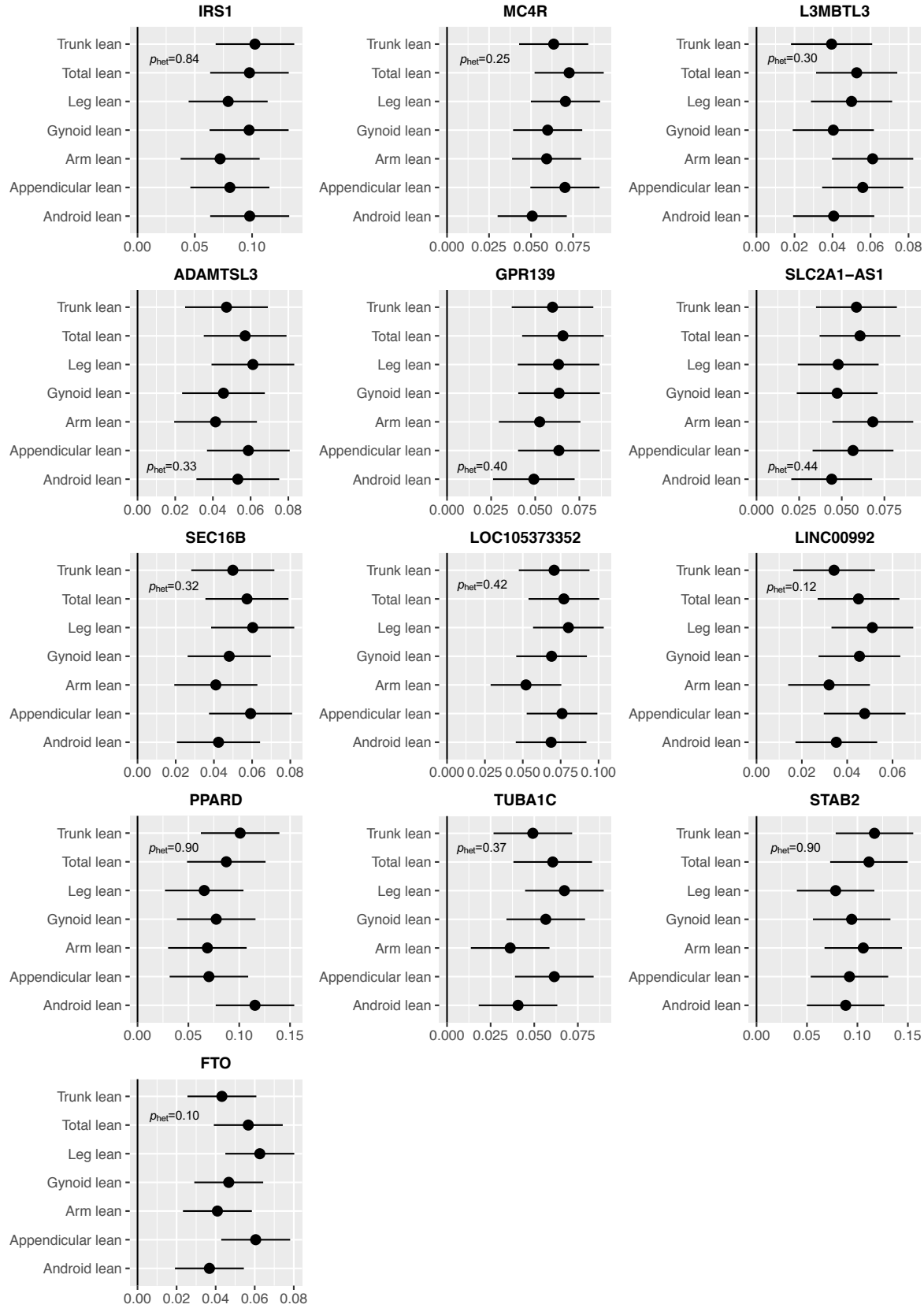


Figure 4.11: Forest plots of 13 loci reaching genome-wide significance for at least one lean mass compartment adjusted for height squared. Effect sizes are expressed as per-allele beta ± upper and lower 95% confidence intervals on standard deviations of lean mass compartments adjusted for height squared, age, genetic principal components and study-specific covariates.

8 of the 13 loci for relative lean mass (*SEC16B, LOC105373352, LINC00992, PPARD, L3MBTL3, ADAMTSL3, GPR139, FTO* and *MC4R*) also reached genome-wide significance for FFMI ($p \leq 5\times10^{-9}$) in the GWAS in UK Biobank presented in Chapter 3, with the same directions of effect for lean mass compartments and FFMI. Loci at *IRS1* ($p_{FFMI}=3.3\times10^{-6}$) and *TUBA1C* ($p_{FFMI}=6.4\times10^{-5}$) reached nominally significant ($p \leq 7.8\times10^{-4}$), directionally consistent associations with FFMI, but *LINC00992* ($p_{FFMI}=0.0011$), *STAB2* ($p_{FFMI}=0.017$) and *SLC2A1* ($p_{FFMI}=0.22$) were not associated with FFMI. 5 of the 13 relative lean mass loci (*IRS1, MC4R*, *L3MBTL3, ADAMTSL3* and *GPR139*) have been reported for height[172,314]. The lean mass-increasing alleles at *IRS1* (rs4675093, EAF=93.0%) and *MC4R* (rs538656, EAF=23.4%) have been reported for taller height, while the other three loci had opposite effects on lean mass and height (*L3MBTL3, ADAMTSL3, GPR139*), which suggests their associations with lean mass compartments may have partly been driven by the adjustment for height squared. However, collider bias analyses in Chapter 3 indicated that associations of these 3 loci with FFMI were not solely driven by the adjustment for height. 8 of the 13 lean mass loci have been reported for BMI[172,217] with directionally consistent effect sizes on lean mass and BMI (*SEC16B, LOC105373352, PPARD, L3MBTL3, ADAMTSL3, GPR139, FTO* and *MC4R*). The lean mass-raising alleles at *IRS1* ($p_{bmi}=4.5\times10^{-5}$) and *LINC00992* ($p_{bmi}=7.7\times10^{-4}$) reached nominal significance for higher BMI (p in GWAS for BMI in UK Biobank $\leq 7.7\times10^{-4}$), while loci at *SLC2A1* ($p_{bmi}=0.19$), *TUBA1C* ($p_{bmi}=0.27$) and *STAB2* ($p_{bmi}=0.049$) were not associated with BMI (p in GWAS for BMI in UK Biobank $>7.7\times10^{-4}$).

Table 3.2: 61 loci and 64 variants reaching genome-wide significance for DEXA fat and lean mass compartments adjusted for total fat and height squared, respectively.

Locus	Chr	Position	Sentinel variant	EA	OA	EAF	Nearest genes	Phenotype	Beta±SE	p-value	hetpval	N	Lead variant for trait (if other than sentinel variant)
1	1	3,830,619	rs4073647	c	a	0.860	*LINC01134*	Leg fat	0.0729±0.0129	1.6E-08	7.4E-01	25,259	
								Peripheral fat	0.0746±0.0129	7.3E-09	8.8E-01	25,260	
2	1	43,443,005	rs7534178	g	a	0.836	*SLC2A1-AS1*	Arm lean	0.0682±0.0121	1.9E-08	2.0E-01	25,256	
3	1	165,221,696	rs1553607	a	c	0.528	*LMX1A*	Subcutaneous android fat	0.0501±0.009	2.5E-08	8.7E-01	25,177	
4	1	177,843,479	rs2094510	t	c	0.208	*LINC01741, SEC16B*	Leg lean	0.0603±0.0111	4.9E-08	6.9E-01	25,252	
5	1	219,721,283	rs11118336	c	t	0.419	*LOC102723886*	Trunk fat	-0.05±0.009	3.3E-08	5.7E-01	25,256	rs77962041
								Leg fat	0.0631±0.009	2.8E-12	9.1E-01	25,259	
								Peripheral fat	0.0625±0.009	4.7E-12	8.6E-01	25,260	
								Gynoid fat	0.0547±0.009	1.5E-09	7.9E-01	25,252	rs4846565
6	2	610,603	rs2867131	c	t	0.832	*LOC105373352, TMEM18*	Android lean	0.0687±0.0119	7.6E-09	2.0E-02	25,244	
								Appendicular lean	0.0759±0.0119	1.7E-10	2.2E-04	25,252	
								Leg lean	0.0801±0.0119	1.5E-11	5.4E-05	25,252	
								Total lean	0.0771±0.0119	8.9E-11	8.1E-04	25,254	
								Trunk lean	0.0707±0.0119	2.8E-09	2.6E-02	25,252	
								Gynoid lean	0.069±0.0119	6.5E-09	1.7E-04	25,248	rs2867115
7	2	18,707,873	rs11096542	a	g	0.592	*KCNS3, RDH14*	Arm fat	-0.05±0.0091	4.0E-08	3.5E-01	25,254	
8	2	165,539,661	rs6717858	c	t	0.406	*COBLL1*	Gynoid fat	0.0619±0.0091	8.3E-12	5.4E-01	25,252	rs10184004
								Android fat	-0.0589±0.0091	7.7E-11	8.7E-01	25,247	rs10195252
								Leg fat	0.063±0.0091	3.5E-12	9.1E-01	25,259	
								Peripheral fat	0.0558±0.0091	7.4E-10	8.9E-01	25,260	
								Trunk fat	-0.0622±0.0091	6.4E-12	7.8E-01	25,256	

Locus	Chr	Position	Sentinel variant	EA	OA	EAF	Nearest genes	Phenotype	Beta±SE	p-value	hetpval	N	Lead variant for trait (if other than sentinel variant)
9	2	188,075,837	rs7586854	c	t	0.500	*ZSWIM2, CALCRL*	Gynoid fat	-0.0485±0.0089	4.9E-08	9.7E-01	25,252	
								Leg fat	-0.0405±0.0089	5.4E-06	7.0E-01	25,259	rs7592188
10	2	219,699,999	rs78058190	g	a	0.947	*PRKAG3, WNT6*	Leg fat	0.155±0.0227	8.0E-12	6.5E-01	25,259	
								Peripheral fat	0.1377±0.0227	1.3E-09	9.2E-01	25,260	
								Trunk fat	-0.1271±0.0227	2.1E-08	8.6E-01	25,256	
11	2	227,100,698	rs2972146	g	t	0.356	*LOC646736, MIR5702*	Trunk fat	-0.0515±0.0092	2.4E-08	2.9E-01	25,256	
		227,637,763	rs4675093	a	t	0.930	*IRS1*	Android lean	0.0979±0.0176	2.5E-08	7.7E-01	25,244	
								Total lean	0.0977±0.0175	2.6E-08	7.1E-01	25,254	
								Trunk lean	0.1025±0.0175	5.1E-09	8.1E-01	25,252	
								Gynoid lean	0.0974±0.0176	2.9E-08	5.4E-01	25,248	rs111851950
12	3	12,369,401	rs13064760	t	c	0.118	*PPARG*	Arm fat	0.1174±0.0138	2.2E-17	5.8E-01	25,254	
13	3	64,718,258	rs2371767	c	g	0.275	*ADAMTS9-AS2*	Gynoid fat	0.0477±0.0099	1.6E-06	1.1E-01	25,252	rs574435140
								Leg fat	0.0583±0.0099	4.5E-09	1.9E-01	25,259	
14	3	129,398,165	rs72628505	t	g	0.935	*TMCC1*	Gynoid fat	0.1021±0.0186	3.8E-08	2.5E-01	25,252	
15	3	156,795,408	rs56406311	c	t	0.609	*LEKR1, LINC00880*	Leg fat	-0.0532±0.0094	1.3E-08	2.2E-01	25,259	
								Subcutaneous android fat	0.0675±0.0094	6.6E-13	5.7E-02	25,177	rs900400
16	4	89,741,269	rs3822072	g	a	0.547	*FAM13A*	Peripheral fat	0.0481±0.0089	7.4E-08	3.0E-01	25,260	rs4544678
								Leg fat	0.0552±0.0089	6.8E-10	4.6E-01	25,259	
								Trunk fat	-0.0536±0.0089	2.1E-09	6.8E-01	25,256	
17	5	55,860,866	rs3936510	g	t	0.802	*C5orf67*	Android fat	-0.0758±0.0111	8.1E-12	8.5E-01	25,247	
								Leg fat	0.0822±0.0111	1.2E-13	9.2E-01	25,259	
								Peripheral fat	0.0868±0.0111	4.8E-15	7.4E-01	25,260	
								Subcutaneous android fat	-0.0936±0.0111	3.5E-17	1.5E-01	25,177	
								Trunk fat	-0.0862±0.0111	7.2E-15	6.4E-01	25,256	

Locus	Chr	Position	Sentinel variant	EA	OA	EAF	Nearest genes	Phenotype	Beta±SE	p-value	hetpval	N	Lead variant for trait (if other than sentinel variant)
18	5	116,895,323	rs187859	t	c	0.641	*LINC00992*	Leg lean	0.0511±0.0092	3.2E-08	4.6E-01	25,252	
19	5	127,357,526	rs17764730	t	c	0.249	*LINC01184*	Leg fat	0.0617±0.0103	2.1E-09	8.3E-02	25,259	
20	5	173,334,219	rs75049939	t	g	0.693	*CPEB4*	Leg fat	0.048±0.0096	5.7E-07	1.2E-01	25,259	rs56364564
								Peripheral fat	0.0498±0.0096	2.2E-07	8.1E-02	25,260	rs56364564
								Android fat	-0.0587±0.0096	1.0E-09	4.2E-01	25,247	
								Trunk fat	-0.0633±0.0096	4.6E-11	5.6E-01	25,256	
21	6	6,733,540	rs744056	a	g	0.413	*LY86, RREB1*	Gynoid fat	0.0682±0.0091	6.6E-14	5.1E-01	25,252	
22	6	34,190,449	rs10214450	c	t	0.955	*GRM4, HMGA1*	Leg fat	0.1307±0.0217	1.6E-09	8.3E-01	25,259	
								Peripheral fat	0.1467±0.0217	1.4E-11	6.4E-01	25,260	
								Trunk fat	-0.1338±0.0217	6.8E-10	4.3E-01	25,256	
23	6	35,373,877	rs6942357	a	g	0.055	*PPARD*	Android lean	0.1155±0.0197	4.6E-09	9.7E-01	25,244	
24	6	43,757,896	rs998584	c	a	0.517	*VEGFA, LINC01512*	Android fat	-0.0809±0.009	2.1E-19	1.6E-01	25,247	
								Gynoid fat	0.075±0.009	6.6E-17	9.7E-01	25,252	
								Leg fat	0.0983±0.009	6.0E-28	6.5E-01	25,259	
								Peripheral fat	0.0945±0.009	6.5E-26	5.2E-01	25,260	
								Trunk fat	-0.0857±0.009	1.3E-21	8.1E-02	25,256	
25	6	127,440,047	rs577721086	t	c	0.948	*RSPO3*	Android fat	-0.1835±0.0205	4.0E-19	6.5E-01	25,247	
								Gynoid fat	0.201±0.0205	1.1E-22	4.2E-01	25,252	
								Leg fat	0.211±0.0205	7.9E-25	9.5E-01	25,259	
								Trunk fat	-0.1947±0.0205	2.4E-21	8.8E-01	25,256	
								Visceral android fat	-0.1468±0.0206	9.9E-13	2.3E-01	25,180	rs1936807
								Peripheral fat	0.2081±0.0205	3.7E-24	8.3E-01	25,260	rs72959041
26	6	130,379,160	rs12661232	t	c	0.311	*L3MBTL3*	Arm lean	0.0611±0.0109	2.4E-08	7.0E-01	19,438	
27	6	139,831,180	rs679582	a	g	0.628		Android fat	-0.0607±0.0092	5.0E-11	7.1E-01	25,247	

Locus	Chr	Position	Sentinel variant	EA	OA	EAF	Nearest genes	Phenotype	Beta±SE	p-value	hetpval	N	Lead variant for trait (if other than sentinel variant)
								Leg fat	0.0634±0.0092	6.1E-12	5.6E-01	25,259	
								Peripheral fat	0.0681±0.0092	1.6E-13	6.4E-01	25,260	
							LINC01625, LOC100132735	Visceral android fat	-0.0561±0.0093	1.4E-09	5.8E-01	25,180	
								Trunk fat	-0.0741±0.0092	9.4E-16	8.1E-01	25,256	rs632057
28	7	27,218,419	rs1859164	t	c	0.538	*HOXA10-HOXA9*	Arm fat	0.0796±0.009	7.1E-19	3.7E-01	25,254	
29	7	73,042,085	rs62466318	t	c	0.198	*MLXIPL, VPS37D*	Gynoid fat	0.073±0.0112	7.9E-11	7.2E-01	25,252	
30	8	25,464,690	rs11992444	g	t	0.488	*CDCA2, EBF2*	Visceral android fat	-0.0576±0.0091	2.2E-10	8.3E-01	25,180	rs73221948
								Android fat	-0.0554±0.0091	9.7E-10	2.7E-01	25,247	
31	8	72,469,742	rs4738141	g	a	0.251	*EYA1, MSC*	Trunk fat	0.061±0.0102	2.5E-09	1.5E-01	25,256	rs13260920
								Android fat	0.0773±0.0102	4.2E-14	3.2E-01	25,247	
								Gynoid fat	-0.0652±0.0102	1.9E-10	5.8E-01	25,252	
								Leg fat	-0.0583±0.0102	1.2E-08	3.3E-01	25,259	
								Peripheral fat	-0.0589±0.0102	8.6E-09	1.7E-01	25,260	
32	8	109,432,178	rs13264315	t	a	0.341	*EIF3E, EMC2*	Leg fat	-0.0607±0.011	3.2E-08	5.1E-01	19,438	
33	8	126,506,694	rs112875651	g	a	0.605	*TRIB1, LINC00861*	Android fat	0.0605±0.0092	5.6E-11	5.8E-01	25,247	
								Gynoid fat	-0.052±0.0092	1.7E-08	1.3E-01	25,252	
								Leg fat	-0.0585±0.0092	2.2E-10	7.4E-01	25,259	
								Trunk fat	0.0611±0.0092	3.5E-11	5.8E-01	25,256	
								Peripheral fat	-0.0556±0.0092	1.7E-09	7.5E-01	25,260	rs6470361
34	9	107,722,705	rs1962883	t	c	0.459	*ABCA1, SLC44A1*	Leg fat	-0.0684±0.0104	5.5E-11	3.1E-02	19,438	
								Peripheral fat	-0.0684±0.0104	5.7E-11	3.9E-02	19,438	
								Trunk fat	0.0628±0.0104	1.8E-09	4.8E-02	19,438	

Locus	Chr	Position	Sentinel variant	EA	OA	EAF	Nearest genes	Phenotype	Beta±SE	p-value	hetpval	N	Lead variant for trait (if other than sentinel variant)
		107,869,756	rs41396353	t	g	0.042	*ABCA1, SLC44A1*	Android fat	-0.1275±0.0223	1.1E-08	5.7E-01	25,247	
35	10	122,995,312	rs11199849	c	t	0.252	*WDR11, FGFR2*	Arm fat	-0.073±0.0103	1.2E-12	2.7E-01	25,254	
36	11	14,258,655	rs4757244	g	a	0.694	*SPON1*	Leg fat	0.0538±0.0097	2.5E-08	8.2E-02	25,259	
37	11	32,387,562	rs11031749	t	g	0.205	*SNORA88, WT1*	Subcutaneous android fat	0.0606±0.0111	4.3E-08	9.0E-01	25,177	
								Visceral android fat	-0.0284±0.0111	1.0E-02	7.6E-01	25,180	rs11031796
38	11	62,316,195	rs11231138	g	c	0.372	*AHNAK, EEF1G*	Peripheral fat	-0.0525±0.0092	1.4E-08	7.1E-01	25,260	
								Gynoid fat	-0.0534±0.0092	7.5E-09	9.0E-01	25,252	rs11231156
39	11	63,929,215	rs11231711	a	g	0.059	*MACROD1*	Peripheral fat	-0.1051±0.019	3.1E-08	7.6E-01	25,260	
								Gynoid fat	-0.1161±0.019	9.9E-10	9.4E-01	25,252	rs56271783
40	12	26,457,650	rs2129869	t	a	0.217	*SSPN, ITPR2*	Gynoid fat	-0.0668±0.0108	7.2E-10	3.0E-02	25,252	rs10842703
								Android fat	0.0739±0.0108	9.1E-12	8.2E-01	25,247	
								Peripheral fat	-0.0625±0.0108	8.0E-09	2.9E-01	25,260	
								Trunk fat	0.0739±0.0108	9.0E-12	9.1E-01	25,256	
								Leg fat	-0.0682±0.0108	3.0E-10	2.4E-01	25,259	rs2175723
41	12	49,609,808	rs11168921	a	g	0.185	*TUBA1A, TUBA1C*	Leg lean	0.0673±0.0115	4.8E-09	7.9E-01	25,252	
42	12	54,387,377	rs1109391	g	t	0.434	*MIR196A2, HOXC-AS2*	Subcutaneous android fat	-0.0297±0.009	1.0E-03	6.3E-01	25,177	rs11614913
								Arm fat	0.0503±0.009	2.5E-08	9.6E-01	25,254	
43	12	66,411,692	rs1871675	g	a	0.366	*HMGA2, MIR6074*	Leg fat	0.0597±0.0093	1.3E-10	8.3E-01	25,259	rs7979129
								Peripheral fat	0.0657±0.0093	1.4E-12	7.6E-01	25,260	rs7979129
								Android fat	-0.0558±0.0093	2.0E-09	6.5E-01	25,247	
								Trunk fat	-0.0582±0.0093	3.9E-10	5.7E-01	25,256	

Locus	Chr	Position	Sentinel variant	EA	OA	EAF	Nearest genes	Phenotype	Beta±SE	p-value	hetpval	N	Lead variant for trait (if other than sentinel variant)
44	12	104,001,177	rs11111676	a	t	0.868	*STAB2*	Total lean	0.1114±0.0196	1.3E-08	5.5E-01	19,438	
								Trunk lean	0.1169±0.0196	2.5E-09	6.5E-01	19,438	
45	12	114,877,400	rs2551377	t	g	0.168	*TBX5-AS1, TBX3*	Arm fat	0.1318±0.012	3.4E-28	2.5E-02	25,254	
46	12	123,024,476	rs147730268	g	t	0.914	*KNTC1*	Trunk fat	-0.0991±0.0166	2.1E-09	9.2E-01	25,256	
		123,192,599	rs11059476	c	t	0.942	*HCAR2, HCAR3*	Leg fat	0.1257±0.0201	4.1E-10	3.7E-01	25,259	
								Peripheral fat	0.1234±0.0201	8.4E-10	5.3E-01	25,260	
47	12	124,409,502	rs7133378	g	a	0.672	*DNAH10*	Android fat	0.091±0.0095	9.3E-22	8.6E-01	25,247	
								Gynoid fat	-0.0845±0.0095	5.6E-19	6.1E-01	25,252	
								Leg fat	-0.1132±0.0095	7.6E-33	9.2E-01	25,259	
								Peripheral fat	-0.105±0.0095	1.9E-28	8.4E-01	25,260	
								Trunk fat	0.0996±0.0095	9.2E-26	5.7E-01	25,256	
								Visceral android fat	0.0787±0.0095	1.4E-16	9.7E-01	25,180	
48	15	63,922,474	rs11071759	t	c	0.397	*HERC1*	Leg fat	-0.0537±0.0091	3.4E-09	8.9E-01	25,259	
								Peripheral fat	-0.0509±0.0091	2.2E-08	8.6E-01	25,260	
49	15	84,542,945	rs56100529	t	a	0.282	*ADAMTSL3*	Leg lean	0.0612±0.0112	4.3E-08	7.9E-01	19,438	
50	16	19,980,931	rs7199285	c	t	0.819	*GPRC5B, GPR139*	Total lean	0.0657±0.0118	2.5E-08	7.5E-01	25,254	
51	16	53,809,123	rs55872725	t	c	0.405	*FTO*	Appendicular lean	0.0605±0.009	2.1E-11	2.2E-01	25,252	
								Leg lean	0.0626±0.009	4.2E-12	2.5E-01	25,252	
								Total lean	0.0567±0.009	3.4E-10	2.3E-01	25,254	rs11642841
52	17	17,988,591	rs12952818	a	g	0.606	*GID4, DRG2*	Gynoid fat	0.062±0.0104	2.3E-09	8.5E-01	19,438	rs2955382
								Leg fat	0.0649±0.0104	3.9E-10	9.5E-01	19,438	rs2955382
								Android fat	-0.0671±0.0104	1.0E-10	6.0E-01	19,438	
								Peripheral fat	0.0652±0.0104	3.3E-10	9.3E-01	19,438	
								Trunk fat	-0.0658±0.0104	2.3E-10	9.4E-01	19,438	

Locus	Chr	Position	Sentinel variant	EA	OA	EAF	Nearest genes	Phenotype	Beta±SE	p-value	hetpval	N	Lead variant for trait (if other than sentinel variant)
53	17	43,691,377	rs382362	c	t	0.206	*MAPK8IP1P2, LINC02210*	Gynoid fat	-0.0638±0.0116	3.7E-08	9.1E-01	25,252	
54	17	48,628,160	rs62621401	a	g	0.985	*SPATA20*	Arm fat	0.3913±0.0367	1.4E-26	3.5E-01	25,254	
55	18	2,846,812	rs11664106	t	a	0.371	*EMILIN2*	Leg fat	0.0655±0.0096	1.0E-11	3.4E-01	25,259	
								Peripheral fat	0.0625±0.0096	8.7E-11	3.2E-01	25,260	
								Trunk fat	-0.0659±0.0096	8.4E-12	6.1E-01	25,256	
56	18	57,850,422	rs538656	t	g	0.234	*PMAIP1, MC4R*	Android lean	0.0507±0.0105	1.3E-06	9.0E-01	25,244	rs8084834
								Trunk lean	0.0635±0.0105	1.3E-09	9.6E-01	25,252	rs7235626
								Gynoid lean	0.0608±0.0106	9.4E-09	7.2E-01	25,248	rs7240682
								Appendicular lean	0.0702±0.0105	2.0E-11	7.7E-01	25,252	
								Leg lean	0.0705±0.0105	1.7E-11	6.4E-01	25,252	
								Total lean	0.0727±0.0105	3.8E-12	8.6E-01	25,254	
								Arm lean	0.0593±0.0105	1.5E-08	8.4E-01	25,256	rs35693910
57	19	33,925,588	rs11084734	t	c	0.305	*PEPD*	Leg fat	0.054±0.0097	2.8E-08	1.2E-01	25,259	
								Trunk fat	-0.0558±0.0097	9.9E-09	2.6E-01	25,256	
								Gynoid fat	0.0363±0.0097	1.9E-04	4.1E-02	25,252	rs60497719
								Android fat	-0.053±0.0097	5.3E-08	4.6E-01	25,247	rs35968262
								Peripheral fat	0.0527±0.0097	6.1E-08	8.3E-02	25,260	rs182863881
58	21	35,593,827	rs28451064	a	g	0.132	*LINC00310, KCNE2*	Peripheral fat	-0.0788±0.0135	4.9E-09	6.4E-02	25,260	
59	22	25,093,121	rs9612738	c	t	0.773	*POM121L10P, PIWIL3*	Subcutaneous android fat	-0.059±0.0108	4.8E-08	9.7E-01	25,177	
60	22	29,142,286	rs6005856	a	g	0.035	*HSCB*	Leg fat	-0.1328±0.0243	4.4E-08	5.2E-01	25,259	
61	22	38,569,529	rs4380	c	t	0.472	*PLA2G6*	Leg fat	0.0577±0.0101	1.3E-08	9.3E-02	19,438	

Discussion

In this chapter, I have sought to fill a gap in current knowledge of genetic factors driving regional adiposity between large-scale GWAS on WHR (a crude measure of one pattern of fat distribution) and smaller-scale genetic discoveries of highly refined but selective adiposity traits based on CT imaging. I furthermore aimed to conduct the first systematic genetic investigation of regional absolute and relative lean mass.

GWAS of DEXA phenotypes allows refined characterisation of WHR loci

I aimed to identify genetic loci associated with patterns of fat distribution by conducting GWAS for absolute fat mass in different regions of the body and regional fat mass adjusted for total fat mass. 38 of the 51 loci identified for relative regional fat mass were associated with a WHR-like fat distribution pattern, which suggests that differences in WHR are the most common source of variation in body fat distribution. However, it cannot be excluded that this finding may be due to greater power to identify loci for central and leg fat, which are by far the largest fat depots in the human body and are, after transformation to the standard distribution, measured with greater precision than arm fat mass, which among the Fenland participants only account for approximately 9 to 10% of the total body fat. Variations in central-to-lower body fat distribution are largely captured by WHR based on simple tape measures, as evidenced by the at least nominally significant associations of all but two of the 38 loci with WHRadjBMI.

However, genetic associations with refined absolute and relative fat compartments allowed us to conduct a detailed characterisation of the effects of the WHR-like loci across all fat compartments, and to distinguish their effects on visceral and subcutaneous central fat mass. I provide evidence that loci increasing central-to-lower body fat distribution do not have consistent effects on arm fat nor on subcutaneous android fat. Neither of these two relative fat compartments showed substantial observational correlations with WHRadjBMI, which further corroborates the independence of relative arm and subcutaneous android fat from WHR. However, similar to the moderate inverse observational association between visceral and subcutaneous fat mass adjusted for total fat, a weak inverse correlation between the genetic effect sizes of the 38 WHR-like loci on visceral and subcutaneous android fat was observed. This finding suggests that visceral and subcutaneous central fat are to some extent regulated by distinct genetic factors.

An exception to this observation was the locus in R-spondin 3 (*RSPO3)*, which encodes the activator of the canonical Wnt signalling pathway and had strong positive effects on both visceral and subcutaneous android fat mass adjusted for total fat mass. Variants in *RSPO3* have been reported for higher WHRadjBMI[308], higher fasting insulin[94], lower HDL cholesterol and higher

triglycerides[214]. The allele for higher central-to-lower fat distribution was also identified for lower BF% in Chapter 3 and without adjustment for total fat mass, *RSPO3* was only associated with lower leg and gynoid fat mass. Altogether, this indicates that the *RSPO3* locus decreases overall BF% by decreasing only absolute fat mass in the legs and hip and that the relative increase in central-to-lower body fat distribution affects both the visceral and the subcutaneous android compartment in the same direction.

The upper peripheral fat compartments are observationally and genetically distinct from the lower peripheral compartments and standard anthropometric traits

In the extant literature, peripheral fat mass is a frequently used term that refers to the combined fat mass in the legs and arms. While multiple studies have focused on the genetic determinants and cardiometabolic correlates of lower body fat, based on indirect measures such as hip circumference, very few studies have explicitly focused on arm fat mass. The results presented in this chapter indicate that arm fat may not simply be the "upper body equivalent" of lower body peripheral fat mass. Firstly, after adjusting for total fat mass, arm fat mass did not correlate with leg fat mass or any other fat compartments, nor with standard anthropometric phenotypes, including WHRadjBMI and BMI. The latter observation indicates that relative fat distribution to the arms is not captured by the commonly used anthropometric risk factors. Furthermore, evidence was found that arm fat may be controlled by genetic factors that are distinct from those controlling leg fat mass, as indicated by the two genetic loci in *PPARG* and *SPATA20* reaching genome-wide significance for absolute arm fat mass but not (*SPATA20*) or only weakly (*PPARG*) affecting leg fat mass, and the 7 loci associated for proportionately higher arm fat mass but not (*PPARG*, *SPATA20*, *TBX5, RDH14, FGFR2, HOXC-AS2*) or to a lesser extent (*HOXA10-HOXA9*) with relative leg fat mass.

These findings raise the question of whether the peripheral fat compartments in the upper versus the lower extremities may also differ in terms of their relationships to cardiometabolic disease risk. Although the trunk and legs provide a larger storage capacity for adipose tissue than any of the other compartments, this does not preclude arm fat from having a metabolically relevant function. For instance, only a small fraction of the total fat mass is stored in the visceral compartment, and yet visceral fat is highly relevant in the aetiology of cardiometabolic diseases[330,332]. As a first step to identifying the aetiological role of arm fat in T2D, genetic associations of the arm fat loci with T2D risk were investigated. The absolute arm fat-raising alleles at *PPARG* and *SPATA20* were genome-wide significantly and suggestively associated with *lower* risk of T2D, which suggests that

having a larger capacity to store fat in the arms may have a similar protective effect against cardiometabolic disease risk as does leg fat mass. However, associations of the other 5 loci for relative arm fat mass with T2D risk did not follow a clear trend. Further investigations of observational associations of absolute and relative arm fat mass with detailed cardiometabolic risk factors and disease outcomes, as well as a further expansion of the genetic discovery for arm fat mass, which will allow us to generate larger and more powered genetic instruments for arm fat mass, are needed to understand the metabolic relevance of the arm fat compartment.

The T2D risk-reducing locus at *PPARG* is most strongly associated with absolute and relative higher arm fat mass

The locus in *PPARG* was associated with higher absolute levels of arm fat mass and with arm fat distribution. The peroxisome proliferator-activated receptor-γ gene (*PPARG*) encodes a nuclear fatty acid receptor that is highly expressed in adipose tissue and which, upon binding of a fatty acid, regulates gene expression by binding to the peroxisome proliferator-response element. PPAR-γ has been identified as a key regulator of adipocyte differentiation, lipid metabolism and uptake of fatty acids and glucose in adipose and muscle tissue[357]. Coding mutations in *PPARG* have been identified as a cause of familial partial lipodystrophy[358] while the arm fat-increasing allele of the common locus in *PPARG* has been reported for lower risk of T2D[16,359], lower fasting insulin levels[94], higher WHR[308] and more recently with higher BMI[172]. In the previous chapter of this thesis, I found that the association of *PPARG* with BMI was entirely driven by fat mass and that a significantly inverse association with lean mass was observed. In this chapter, I show that *PPARG* has heterogeneous effects on fat mass in different areas of the body, with an effect nearly twice as strong on SDs of arm fat mass than on leg fat mass and only weak effects on central fat mass. After adjustment for total fat mass, *PPARG* remained most strongly associated with higher arm fat mass and was nominally associated with lower central fat mass but not with leg fat mass. These findings suggest that *PPARG* is more strongly associated with a higher arm-to-central fat distribution than with central-to-lower fat distribution, as suggested by the previously reported association with WHR.

No evidence for distinct genetic determinants for different lean mass compartments

In contrast to the analyses of regional fat mass, GWAS for lean mass compartments with or without adjustment for height squared did not lead to the identification of any loci which affected lean mass more strongly in some compartments than in others. For none of the identified lean mass loci was evidence found for heterogeneity in the effect sizes across the absolute lean mass compartments, while for the majority of loci identified for absolute fat compartments at least suggestive evidence for heterogeneity was found. These findings are in line with the strong observational correlations between the lean mass compartments, with and without adjustment for height squared. Larger sample sizes and analyses adjusted for total lean mass may be required to identify loci associated with lean mass distribution patterns, and epidemiological research to identify lean mass compartments and distribution patterns associated with cardiometabolic risk factors and outcomes are needed.

Strengths and limitations

I conducted the first large-scale genome-wide association analyses of fat and lean mass distribution based on direct, DEXA-based measures in up to 25,000 participants. The analyses were not restricted to certain fat compartments for which there was a prior interest, as was the case for the previously published CT-based GWAS of specific fat traits[330,344]. Instead, I analysed fat and lean mass in all regions of the body and was therefore able to characterise for all the identified loci the patterns of association with all fat and lean mass compartments in the body. This enabled a complete overview of the anthropometric effects of the identified loci. Finally, the analyses were based on meta-analyses of 5 cohorts and heterogeneity of the effect sizes between the studies was assessed. Therefore, the findings are unlikely to be study-specific.

However, the presented analyses also suffer from certain limitations. Firstly, this ongoing project started off by conducting analyses for the sexes combined and may therefore not have been able to identify loci that are specific to men or women and were not able to assess sex differences in the effect sizes of the identified loci. The high number of sex-specific loci and loci with sex-discordant effects for WHR and WHRadjBMI[309] suggest that also for other patterns of fat distribution sex-dimorphic associations may be identified. Secondly, no conditional analyses to identify the independent variants at each locus were performed. The reasons for this are that our main focus was on refined characterisation of known loci rather than on identifying novel loci, and because current tools designed to conduct such analyses based on summary-level data cannot identify independent variants across multiple traits. Instead, a pragmatic, distance-based clumping

approach was adopted to identify the cross-trait lead variant at each locus. The vast majority of the loci identified in this chapter have previously been reported in large discovery projects for BMI, WHR, BF% and FFMI led by GIANT and others, and the refined structure of these loci has in most cases already been described. Furthermore, collider bias due to the adjustment for height, which is heritable and correlated with regional and overall lean mass, was not systematically assessed. Therefore, some of the loci identified in the adjusted analyses, in particular for lean mass compartments adjusted for height squared, may be false positives, as was the case for the GWAS findings for FFMI in Chapter 3. Moreover, the analyses are restricted to common and low-frequency variants, as the current sample size was relatively modest and does not allow me to reliably assess the effects of rare loci on regional fat and lean mass. Finally, certain limitations are associated with DEXA-based body composition analysis[279,343]. DEXA is known to modestly underestimate fat mass in very large individuals[279]. Furthermore, visceral fat mass estimates based on DEXA imaging have been shown to be less accurate than for the other compartments, when compared to MRI, which is the "gold standard"[343]. These DEXA-related limitations may have led to inaccurate measurements for some participants and thus lower power.

Conclusion and future directions

In this chapter, loci associated with patterns of fat distribution not captured by standard anthropometric risk factors, such as relative arm and subcutaneous android fat, were identified based on GWAS for refined fat and lean traits measured by DEXA imaging. The majority of loci associated with regional fat mass represented variation in WHR, but by assessing the cross-compartment associations of these loci, a refined characterisation of their effects on body size and composition was obtained. While strong evidence was found for genetic factors influencing fat distribution, the presented findings did not provide evidence for the existence of genetic factors which affect lean mass in distinct compartments differently.

The results presented in this chapter are preliminary findings of an ongoing project. We are currently in the process of expanding this discovery further with approximately 3,000 individuals from the Avon Longitudinal Study of Parents and Children (ALSPAC) study (GWAS ongoing) and around 10,000 participants from the Canadian Longitudinal Study on Aging (phenotype QC ongoing). Furthermore, DEXA variables for at least 5,000 additional UK Biobank participants are expected within the next 6 months, which we intend to include in the following round of GWAS and meta-analyses. Sex-specific GWAS analyses are also currently being conducted.

Concluding discussion

Summary of the findings

This PhD thesis set out to identify aetiological pathways to cardiometabolic diseases by integrating genome-wide data with refined metabolic traits, based on high-throughput metabolic profiling (Chapter 2) and body composition measurement techniques (Chapters 3 and 4) in large-scale population-based and disease-specific cohort studies.

While metabolic disruption lies at the heart of the aetiology of cardiometabolic diseases, the role of metabolic pathways has not been studied systematically. In this thesis, I developed a strategy based on the Mendelian randomisation framework to assess the aetiological role of metabolic pathways on cardiometabolic diseases and to characterise the underlying mechanisms through which a metabolic pathway acts on disease risk. By integrating high-throughput untargeted metabolomics with large-scale genome-wide data, I identified glycine metabolism as a potential aetiological pathway to CHD and found evidence that low glycine levels may increase the risk of CHD by elevating blood pressure. Evidence for glycine as a potential causal factor for T2D risk was weaker and may depend on the biological mechanism through which glycine levels are reduced. The strong association of genetically predicted insulin resistance with lower genetically predicted glycine suggests that the observational, inverse glycine-to-T2D association may at least partly be due to confounding by insulin resistance. My findings indicate that glycine supplementation may be beneficial for individuals diagnosed with, or at high risk of developing cardiometabolic diseases but potential negative side effects, in particular on cancer risk, must be assessed prior to setting up trials.

Secondly, this thesis focused on the genetic characterisation of overall and regional body size and composition, with the aim to generate genetic risk scores for specific patterns of overall body composition and fat and lean mass distribution that are not, or not fully, captured by standard indices of obesity, such as BMI and WHR. Through genetic discoveries for BF% and FFMI, I identified genetic loci associated with high relative body fat in the absence of overweight or obesity. Observational epidemiological research has proposed normal weight adiposity as a risk factor for cardiometabolic and other complex diseases. The genetic loci identified in this thesis for normal weight adiposity allow for causal assessment and molecular characterisation of this emerging anthropometric risk factor, which may lead to the identification of novel targets for treatment. Furthermore, I identified genetic loci associated with fat compartments, such as arm and subcutaneous android fat mass, that are largely independent of BMI and WHR, based on GWAS for regional body composition phenotypes measured by DEXA imaging. Genetic loci associated

with arm fat mass did not or only weakly affect leg fat mass, which indicates that fat mass in the upper and lower peripheral compartments are controlled by distinct genetic factors. Whether arm fat has a similar, protective role on cardiometabolic disease risk as leg fat mass remains unclear, but the identified genetic loci for arm fat mass can enable future investigations that seek to clarify the metabolic function of arm fat mass. While the majority of the loci identified for regional fat had heterogeneous effect sizes across the different compartments, genetic evidence suggested that lean mass in different regions of the body is controlled by the same genetic factors.

In this concluding chapter, I will first discuss the limitations of the research described in this thesis and discuss how these may have had an impact on the findings. I will then continue with a discussion of the implications of the presented findings and an outline of new research avenues to which the work of the thesis leads.

Limitations

Limitations of using genetics to identify causal risk factors

The presented research into the aetiological factors of cardiometabolic diseases relied on the assumption that genetic effects represent unconfounded estimates for risk factor-to-disease associations. This assumption is made most explicitly in the Mendelian randomisation analyses on glycine presented in Chapter 2, but it is also relevant to the genetic loci identified for refined anthropometric phenotypes (Chapters 3 and 4), if these were used as genetic instruments to study causal associations of body size and composition with cardiometabolic diseases and traits in the future. The 3 main reasons why a genetically predicted effect size of a risk factor with a disease outcome may not reflect a causal effect are the following: (1) Some of the genetic loci reaching statistical significance may not be truly associated with the risk factor, (2) genetic loci often violate the instrumental variable assumptions and may thus lead to biased causal estimates and (3) even genetic associations based on valid genetic instruments may under- or overestimate the causal effect size. I will outline these three reasons and discuss how each of them may have affected the research presented in the thesis.

Statistical associations of genetic loci with the risk factor may be false positive, study-specific or biased

Genetic loci that reach genome-wide significance in a GWAS for a risk factor may not be truly associated with the risk factor, for reasons of chance, study-specific effects and collider bias. For many years, the GWAS community has adopted a p-value threshold of 5×10^{-8}, based on the assumption that conducting genome-wide tests for associations with a phenotype represents 1 million independent statistical tests and that therefore, to limit the chance of false positive findings to 5%, the significance threshold for GWAS should be set at $p<5\times10^{-8}$. However, with the emergence of new and denser LD reference panels, some argue that the number of independent statistical tests has increased and that therefore stricter p-value thresholds should be used, in particular when studying rare variants[360]. Despite the stringent Bonferroni-corrected significance threshold used throughout this thesis to limit type I errors, 5% of the loci may have reached significance due to chance, without being truly associated with the phenotype. In the meta-analyses of GWAS presented in Chapter 2 (glycine) and Chapter 4 (regional fat and lean mass) in which the effects were replicated across multiple studies, the likelihood of such false positive associations was less likely than for loci identified for BF% and FFMI (Chapter 3), which were mostly based on the GWAS in UK Biobank only. As no studies with a sufficiently large sample size were

immediately available to us in which we could attempt to replicate loci identified for BF% and FFMI in UK Biobank, it is likely that a subset of the identified loci were statistically associated with the phenotypes by chance only.

The external validity of genetic loci identified in UK Biobank only was not tested. Therefore, some of the identified loci in UK Biobank may be study-specific, e.g. due to technical issues related to the genotyping array, genotype calling or imputation. Furthermore, participants of UK Biobank do not represent a cross-section of the British population, as the cohort is strongly enriched for individuals of European ancestry who are in good health, follow lifestyle recommendations and have a higher than average socio-economic status[361]. Therefore, some of the genetic findings based on UK Biobank only may not be generalisable to the entire UK and other populations.

Finally, work by Aschard *et al.*[299] and Day *et al.*[362] highlighted that genetic loci reaching genome-wide significance for a phenotype adjusted for another heritable and correlated trait, such as WHR adjusted for BMI, may be associated through the covariate (i.e., BMI) only, without having any effect on the phenotype (WHR) itself. Loci for which the association in the GWAS is entirely driven by this so-called collider bias have no effect on the intended phenotype. In this thesis I undertook genetic discoveries for several adjusted anthropometric phenotypes, either by taking the ratio of two traits (e.g., FFM to height squared) or by adjusting the anthropometric phenotype (e.g., arm lean mass adjusted for height squared). Even if collider bias is formally assessed for each identified locus, as was done for all loci reaching genome-wide significance for FFMI, false positive findings due to collider bias can still remain due to the limited power of such assessments. Collider bias was not tested for in the other genetic discoveries presented in this thesis. Therefore, the identified loci presented in this thesis may require replication in an independent study.

Pleiotropic genetic instruments

Pleiotropy is the phenomenon in which a gene or genetic locus is associated with two or more apparently unrelated traits[363]. One of the conditions which must be met by a genetic locus for it to be a valid genetic instrument is that it cannot influence the outcome through any other pathway than through the risk factor for which it was identified. The challenge of this no-pleiotropy condition is that it is not fully empirically testable. Potential bias of the causal estimates due to pleiotropic genetic instruments is therefore one of the most commonly raised limitations of the Mendelian randomisation framework[363–366]. It is generally assumed that the chance of bias due to pleiotropy decreases with increasing numbers of genetic instruments included in the MR analysis[363],

as the biases introduced by the individual loci, which each have a unique set of pleiotropic effects, may cancel out each other[363]. However, in a recent study by Verbanck *et al.*[60], 48% of the significant causal estimates based on the MR framework were estimated to be biased due to pleiotropy of the genetic instruments. So-called robust MR methods which are less sensitive to bias caused by pleiotropic instruments have been developed[58–60]. The MR analyses presented in Chapter 2 of the thesis were based on such robust methods, but bias due to known and unknown pleiotropic effects can never be fully excluded. Therefore, a significant MR-based association of a risk factor with an outcome based on multiple, robust MR methods is highly indicative but not final, conclusive evidence for a causal relationship between a risk factor and a disease.

The Mendelian randomisation framework is based on a simple model of causality

Even if a significant genetic risk factor-to-disease association is identified based on genetic variants which meet the instrumental variable assumptions, the genetic effect size may differ from the causal effect size[366]. A significant MR-based genetic association of the risk factor with disease risk represents a linear causal effect of lifelong higher (or lower) exposure levels to the risk factor on disease risk. In reality, however, causal relationships may be non-linear and could vary over time, or so-called "legacy effects"[367] of the risk factor on disease risk may exist. Furthermore, there may be non-random time trends in exposure levels of certain risk factors. The current MR framework is limited in its ability to model these more complex and time-varying causal relationships. Applying MR in situations where the causal relationship between a risk factor and outcome is non-linear and/or differs over time can lead to inaccurate conclusions. For example, evidence suggests that the association of BMI with breast cancer risk may be non-linear and could be different for pre and post-menopausal breast cancer and disease sub-types[368]. The possible non-linear and time-varying causal relationship of BMI with breast cancer could be the reason why MR studies on BMI and breast cancer have thus far led to contradicting findings[369,370].

Finally, the genetic effect may in some cases underestimate the causal effect due to "canalisation". This developmental phenomenon occurs if the phenotypic consequences of low or high risk factor levels driven by a certain genetic mutation are weakened by compensatory mechanisms that act to negate part of the potentially detrimental effects caused by the mutation and its direct phenotypic consequences. A genetic association of the risk factor with the exposure based on such a genetic locus will underestimate the effect of the risk factor.

These general limitations of causal inference based on the MR framework indicate that evidence from observational epidemiological research on the nature of the relationship between a risk factor

and disease should be taken into account and that findings from MR should be primarily interpreted as the likelihood for causality, without over-interpretation of the estimated effect size.

Eurocentric bias

A strong Eurocentric bias dominates the field of human genomic research and of biomedical research in general[371,372]. The systematic under-representation of participants of non-European ancestries in genetic research is not only problematic for ethical reasons, but also forms an important limitation to scientific progress[373]. In the context of genetic research, a higher availability of genome-wide data from individuals of non-European ancestry would increase the overall sample size and thus the power to conduct genetic discoveries. Moreover, certain genetic variants for which the allele frequencies are very low in Europeans, may only be discovered in non-European ancestries. Trans-ethnic genetic discoveries are also crucial for the fine-mapping of known loci and may thus facilitate the identification of the causal variants and genes[373].

Eurocentric bias also forms an important limitation in research on the aetiological risk factors of cardiometabolic diseases. While the major risk factors of cardiometabolic diseases are well understood for individuals of European ancestry living in high-income countries, this is less the case for individuals of e.g. South-Asian and African ancestries living in low or middle-income settings, despite the fact that the latter groups carry a disproportionately share of the T2D burden[1]. Differences in body composition have been suggested as one of the factors that could explain ethnic differences in susceptibility to cardiometabolic illnesses at a given BMI[98,99,374]. Expanding the work in this thesis on body composition patterns and their genetic determinants to participants of African, Latin-American and south and east Asian ancestries may therefore be in important step towards understanding ethnic differences in susceptibility to cardiometabolic diseases and improving risk stratification among individuals of non-European ancestries.

General limitations of using the biomedical paradigm to study the aetiological pathways to cardiometabolic diseases

This thesis focuses on identifying aetiological pathways to cardiometabolic diseases through the lens of biomedical science. The biomedical paradigm of studying the aetiology of cardiometabolic diseases mostly focuses on the proximal causes of disease that are located at the level of the individual[375]. However, a vast body of evidence indicates that these proximal causes of disease are under the influence of powerful, upstream determinants of health that lie outside the individual, including the build environment, socio-economic deprivation and other macro-institutional factors[375–377]. While biomedical research is essential for the development of treatments for

individuals who are already sick or are at high risk of developing a disease, the vast scale of the current metabolic disease epidemic also requires a societal approach to population health. In the words of Sir Michael Marmot: "Why treat people and send them back to the conditions that made them sick?"[378]. The need for such an approach is increasingly being recognised in the medical community, as evidenced by a recent editorial in *The Lancet* calling for a more holistic approach to public health[379].

Implications and directions for future research

Assessing metabolic pathways as aetiological factors for complex diseases

Expanding the approach to other metabolic pathways and "non-metabolic" complex diseases

The approach developed in this thesis to assess the role of metabolic pathways as aetiological risk factors for disease can be expanded to a broader systematic causal investigation of associations between the metabolome and a range of disease outcomes. Conducting such genetics-based causal investigations of metabolic pathways is being increasingly facilitated for research teams worldwide by the publicly available summary-level results of several published metabolome-wide GWAS efforts[169–171,184], with results of several larger-scale ongoing efforts expected to become publicly available in the near future.

The most obvious diseases to which metabolic pathways could be causally related are cardiometabolic diseases, as dysregulation of metabolic processes are key elements in their aetiology. Further research on the aetiological role of the many other metabolic pathways which have been associated with T2D and CHD, but which were not addressed in this thesis will therefore likely lead to the identification of other causal, metabolic pathways to cardiometabolic diseases, of which some may be good candidates for pharmacological intervention.

In recent years, metabolic risk factors have been linked to higher incidence of complex diseases which are not traditionally considered to have a metabolic aetiology, such as cancers[368,380], autoimmune conditions[381] and neurodegenerative diseases[382,383]. Furthermore, associations between the metabolome and disease incidence have been identified for several of these "non-metabolic" complex diseases[384–386]. These findings highlight that a causal role of metabolic pathways may not be restricted to the traditional metabolic diseases, and that the approach adopted in this thesis to study the causality of metabolic pathways may also lead to the identification of novel aetiological pathways to other complex diseases and thus novel treatment strategies.

Attention also needs to be given to the aetiological role of metabolic pathways across diseases, as this may not only reveal causal effects of a potential modifiable pathway on multiple disease outcomes, but also predict potential adverse effects. This was evidenced by the work in this thesis which indicated a causal link between low glycine levels and high cardiometabolic disease risk, while glycine may potentially increase cancer risk. Another recent example is the subtle increase in T2D risk by statins – the major class of LDL cholesterol-lowering drugs that are commonly prescribed for primary and secondary CHD prevention[63,75]. These findings emphasise the

importance of a cross-disease approach, which can be conducted efficiently based on genetics, provided that disease-specific summary-level GWAS results are made publicly available.

Two strategies for prioritising metabolic pathways for causal assessment

Two main strategies can be adopted to prioritise metabolic pathways for causal assessment. Firstly, metabolic pathways can be prioritised based on those most strongly and consistently associated with disease risk in published and novel observational association analyses. This strategy was used in this thesis and led to prioritising the glycine pathway, which, together with the branched-chain amino acid pathway, is one of the most frequently reported pathways for T2D in longitudinal observational studies. With large-scale metabolite profiling being increasingly adopted in epidemiological cohorts and plans to conduct metabolite profiling for very large groups of participants, such as those of UK Biobank, the scale and thus power of such observational studies will drastically increase over the next decade, which will facilitate observational studies on links between the metabolome and diseases. However, as shown for the glycine pathway, strong and consistent observational evidence will not necessarily translate into strong causal evidence. Conversely, a metabolic pathway that is causal for disease risk is highly likely to be also observationally associated with disease incidence.

An alternative strategy is to prioritise metabolic pathways for causal assessment based on the availability of good candidate genetic instruments for the pathway. Genetic loci are strong candidate instruments for a metabolic pathway if the locus is specifically associated with the metabolites of the pathway and if there is a plausible biological mechanism through which the genetic region influences metabolite levels. Loci which are most likely to fulfil these criteria are located in genes encoding enzymes that catalyse a reaction in the metabolic pathway of interest. Based on the identified loci for metabolites through metabolome-wide GWAS, such instruments can be found and tested for cross-metabolome associations, while phenome-wide association studies for such instruments in e.g. UK Biobank can give a first indication of potential causal effects on disease outcomes.

In reality, a combination of both strategies will be most efficient, because specific genetic instruments may not be available for certain metabolic pathways prioritised through observational research, which may be particularly the case for lipid pathways. Furthermore, even if a powerful and specific instrument is found for a particular metabolic pathway, causal investigations may not lead to relevant findings if the pathway does not show any observational associations with disease outcomes.

The genetic determinants of refined anthropometric phenotypes: implications and future research avenues

The metabolic sequelae of overall body fatness may be highly dependent on fat distribution

This thesis aimed to identify the genetic determinants of refined anthropometric phenotypes which have a less heterogeneous relationship with cardiometabolic risk than the currently used anthropometric risk factors, such as BMI. The large heterogeneity in cardiometabolic risk among people with a similar BMI, especially among those with a BMI in the overweight but not obese range, has partly been attributed to differences in overall body composition, and partly to differences in fat distribution.

This PhD thesis provides several lines of evidence that fat distribution is a more important factor for cardiometabolic risk than BF%. First, several of the loci which were identified for higher BF% have elsewhere been reported for *lower* cardiometabolic risk. The protective effects of these loci on cardiometabolic risk are likely to be driven by the positive effects of these loci on body fat in the hip or leg region but not on central body fat, as evidenced by the positive associations of these loci with higher leg fat mass only and the reported inverse associations with WHR. Secondly, known BMI loci for which findings from this thesis indicate that they only increase BMI through fat but not through lean mass, including *PPARG*, *PEPD* and *DNAH10* have previously been reported for *lower* cardiometabolic risk, which may again be explained by the associations of these loci with fat distribution to the lower body.

Loci associated for higher body fat and lower cardiometabolic risk were located in/near *PDE3B*, *RSPO3*, *VEGFA*, *DNAH10*, *PPARG*, *ZNF664* and *LYPLAL1*, for the majority of which a functional role in adipogenesis has been described[79,357,387,388]. These findings highlight that the capacity to generate and expand subcutaneous adipose tissue plays a key role in protection against cardiometabolic conditions[100].

The use of pharmacological agonists of PPARγ, the so-called thiazolidinediones (TZD), for the treatment of T2D is based on the principle that insulin resistance and other pathophysiological processes driven by a long-term positive energy balance can be ameliorated by increasing the capacity to safely store excess energy in the form of lipids in the peripheral fat compartments[357,389]. Rosiglitazone, a thiazolidinedione (TZD) developed for T2D treatment, was found to improve glycaemic control and increase fat accumulation in individuals with T2D, but, due to concerns about side effects, including oedema, congestive heart failure and increased risk for bladder cancer, and fractures in women[390], its use was discontinued in most countries. However, other loci which were here reported for higher BF% but lower cardiometabolic risk may point to novel potential

targets through which the capacity to safely store excess energy can be increased. Functional characterisation of these loci and experiments based on animal models of cardiometabolic diseases are needed to identify the potential of these gene products as targets for pharmacological intervention. Increased risk of congestive heart failure, potentially driven by water retention, was one of the major reasons the use of rosiglitazone was discontinued, but it is unclear if this side effect is a general consequence of increasing fat depots beyond their "natural" limits, or a TZD-specific effect. Effects on fluid balance, which can be measured using BIA, should therefore be a primary focus in the early phases of novel studies on identifying targets to reduce cardio-metabolic risk by increasing peripheral fat storage.

A shift in focus from expanding discoveries to in-depth characterisation of genetic loci

In the past decade of genome-wide studies on body size and other complex traits, the main focus has been on accumulating data on ever increasing numbers of individuals in order to enhance power and thus increase the number of associated loci that reach genome-wide significance. The number of identified genetic loci for each standard anthropometric trait has now reached a thousand or more but for most of these loci the biological mechanisms through which they influence the anthropometric risk factor are entirely unknown. Therefore, the point has been reached at which further efforts into expanding discoveries for simple measures of obesity will only lead to limited further gains in biological understanding. Instead, a shift in focus to in-depth characterisation of the already known loci is more likely to lead to significant improvements in understanding of the pathophysiology of cardiometabolic diseases and to the identification of novel targets for prevention or treatment. The systematic integration of genetic, refined phenotypic and clinical data to create detailed biological pictures of already known loci may become the dominant approach in the next decade of genetic research on body composition and on complex traits in general.

The findings presented in Chapters 3 and 4 regarding the genetic characterisation of overall and regional body fat and lean mass highlight that testable hypotheses on the biological functions of genetic loci for an anthropometric trait can be formulated by integrating knowledge on observational and genetic associations with a range of other anthropometric and metabolic phenotypes. Further enrichment of this integrative approach with molecular phenotypes based on -omics techniques and even more detailed imaging-based anthropometric data, which is facilitated by the increasing scale and the granularity at which phenotypic and clinical data are becoming available for research, is expected to accelerate the discovery of novel molecular targets for treatment. For instance, data generated by the UK Biobank Imaging Study, which aims to obtain

MRI imaging of the brain, heart and abdomen, ultrasound imaging of the carotid arteries and DEXA imaging for body composition and bone density analysis on 100,000 participants[202,351], is of particular relevance to the anthropometric and metabolic research community and will help to inform research on the metabolic implications of genetic loci associated with body size and composition.

Clinical implications

While BMI predicts cardiometabolic risk well for those at the lower and higher ends of the BMI scale, there is a large variation in incidence of cardiometabolic disease risk among the large group of individuals who are mildly overweight. Based on the findings from this thesis which suggest that fat distribution is more relevant than overall body composition for cardiometabolic risk, risk stratification for those in the overweight category is more likely to improve by measuring fat distribution than by focussing on overall body composition. Therefore, central and hip fat accumulation should be routinely assessed by waist and hip circumference measurement and should be interpreted separately instead of as their ratio. Secondly, an estimation of the proportion of total central fat accounted for by subcutaneous fat, based on skinfold thicknesses in the waist region, could further improve risk stratification. Finally, once the metabolic role of the arm fat compartment is corroborated, this could be assessed by circumference and skinfold thickness measures. These simple and low-cost measures can easily be done in general practices, pharmacies or even at home, and an easy-to-use scoring system that also takes overall BMI and potentially total fat and lean mass based on BIA into account, could help to identify those individuals who are at highest risk and should undergo tests to measure a panel of blood-based cardiometabolic risk factors.

Conclusion

This PhD thesis develops and demonstrates an integrative approach, drawing on genome-wide genetic data and detailed metabolic phenotypes based on metabolomics and refined anthropometric assessment techniques, to identifying and characterising novel aetiological pathways to cardiometabolic diseases. By combining large-scale genetic, metabolomic and clinical data, I develop a strategy to assess metabolic pathways as aetiological factors of cardiometabolic diseases. Based on this strategy, I identify a novel causal metabolic pathway to cardiometabolic diseases and formulate testable hypotheses about its mechanisms of action. Furthermore, I set out a new framework for genetics-based causal investigations into the role of refined anthropometric

traits, such as normal weight adiposity and emerging patterns of body fat distribution, on cardiometabolic disease risk.

Bibliography

1. International Diabetes Federation. *IDF Diabetes Atlas Eighth Edition 2017*. *International Diabetes Federation* (2017).
2. DeFronzo, R. A., Ferrannini, E., Zimmet, P. & Alberti, G. *International Textbook of Diabetes Mellitus*. (Wiley, 2015).
3. World Health Organization. *Global Report on Diabetes*. **978,** (2016).
4. Daneman, D. Type 1 diabetes. *Lancet* **367,** 847–858 (2006).
5. American Diabetes Association, (ADA). Gestational diabetes mellitus. *Diabetes Care* **Suppl 1,** S88–S90 (2004).
6. Shah, A. D. *et al.* Type 2 diabetes and incidence of cardiovascular diseases: A cohort study in 1·9 million people. *Lancet Diabetes Endocrinol.* **3,** 105–114 (2015).
7. Stratton, I. M. *et al.* Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. *BMJ* **321,** 405–412 (2000).
8. Yau, J. W. Y. *et al.* Global Prevalence and Major Risk Factors of Diabetic Retinopathy. *Diabetes Care* **35,** 556–564 (2012).
9. Zhuo, X. *et al.* The lifetime cost of diabetes and its implications for diabetes prevention. *Diabetes Care* **37,** 2557–2564 (2014).
10. Hex, N., Bartlett, C., Wright, D., Taylor, M. & Varley, D. Estimating the current and future costs of type 1 and type 2 diabetes in the UK, including direct health costs and indirect societal and productivity costs. *Diabet. Med.* **29,** 855–862 (2012).
11. Caro, J. J., Ward, A. J. & O'Brien, J. A. Lifetime costs of complications resulting from type 2 diabetes in the US. *Diabetes Care* **25,** 476–481 (2001).
12. Nolan, C. J., Damm, P. & Prentki, M. Type 2 diabetes across generations: From pathophysiology to prevention and management. *Lancet* **378,** 169–181 (2011).
13. Savage, D. B., Falk Petersen, K. & Schulman, G. I. Disordered lipid metabolism and the pathogenesis of insulin resistance. *Physiol. Rev.* **87,** 507–520 (2007).
14. Thanabalasingham, G. & Owen, K. R. Diagnosis and management of maturity onset diabetes of the young (MODY). *BMJ* **343,** d6044–d6044 (2011).
15. Semple, R. K., Savage, D. B., Cochran, E. K., Gorden, P. & O'Rahilly, S. Genetic syndromes of severe insulin resistance. *Endocr. Rev.* **32,** 498–514 (2011).
16. Mahajan, A. *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* (2018). doi:10.1038/s41588-018-0241-6
17. Scott, R. A. *et al.* The link between family history and risk of type 2 diabetes is not explained by anthropometric, lifestyle or genetic risk factors : the EPIC-InterAct study. *Diabetologia* **56,** 60–69 (2013).
18. Shai, I. *et al.* Ethnicity, obesity, and risk of type 2 diabetes in women: a 20-year follow-up study. *Diabetes Care* **29,** 1585–1590 (2006).
19. Diabetes Prevention Program Research Group. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N. Engl. J. Med.* **346,** 393–403 (2002).
20. Wang, Y., Rimm, E. B., Stampfer, M. J., Willett, W. C. & Hu, F. B. Comparison of abdominal adiposity and overall obesity in predicting risk of type 2 diabetes among men. *Am. J. Clin. Nutr.* **81,** 555–563 (2005).
21. Hu, F. B. *et al.* Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. *N. Engl. J. Med.* **345,** 790–797

(2001).

22. Helmerhorst, H. J. F., Wijndaele, K., Brage, S., Wareham, N. J. & Ekelund, U. Objectively measured sedentary time predicts insulin resistance, independent of moderate and vigorous physical activity. *Diabetes* **58,** 1776–1779 (2009).
23. Hu, F. B., Li, T. Y., Colditz, G. a, Willett, W. C. & Manson, J. E. Television watching and other sedentary behaviors in relation to risk of obesity and type 2 diabetes mellitus in women. *J. Am. Med. Assoc.* **289,** 1785–1791 (2003).
24. O'Connor, L. *et al.* Prospective associations and population impact of sweet beverage intake and type 2 diabetes, and effects of substitutions with alternative beverages. *Diabetologia* **58,** 1474–1483 (2015).
25. Lindström, J. *et al.* The Finnish Diabetes Prevention Study (DPS). *Diabetes Care* **26,** 3230–3236 (2003).
26. Ramachandran, A. *et al.* The Indian Diabetes Prevention Programme shows that lifestyle modification and metformin prevent type 2 diabetes in Asian Indian subjects with impaired glucose tolerance (IDPP-1). *Diabetologia* **49,** 289–297 (2006).
27. Dunbar, J. A. *et al.* Challenges of diabetes prevention in the real world: results and lessons from the Melbourne Diabetes Prevention Study. *BMJ Open Diabetes Res. Care* **3,** e000131 (2015).
28. Li, G. *et al.* The long-term effect of lifestyle interventions to prevent diabetes in the China Da Qing Diabetes Prevention Study: a 20-year follow-up study. *Lancet* **371,** 1783–1789 (2008).
29. Floegel, A. *et al.* Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. *Diabetes* **62,** 639–648 (2013).
30. Vasan, S. K. *et al.* Comparison of regional fat measurements by dual-energy X-ray absorptiometry and conventional anthropometry and their association with markers of diabetes and cardiovascular disease risk. *Int. J. Obes.* **42,** 850–857 (2018).
31. Mora, S. *et al.* Lipoprotein particle size and concentration by nuclear magnetic resonance and incident type 2 diabetes in women. *Diabetes* **59,** 1153–1160 (2010).
32. Hulley, S. *et al.* Randomized Trial of Estrogen Plus Progestin for Secondary Prevention of Coronary Heart Disease in Postmenopausal Women. *JAMA* **280,** 605–613 (1998).
33. Stampfer, M. J. & Colditz, G. A. Estrogen replacement therapy and coronary heart disease: a quantitative assessment of the epidemiologic evidence. *Prev Med* **20,** 47–63 (1991).
34. Nelson, M. R. *et al.* The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47,** 856–860 (2015).
35. Hurle, M. R., Nelson, M. R., Agarwal, P. & Cardon, L. R. Trial watch: Impact of genetically supported target selection on R&D productivity. *Nat. Rev. Drug Discov.* **15,** 596–597 (2016).
36. Altshuler, D. in *Chief Medical Officer annual report 2016: Generation Genome* Chapter 4 (2017).
37. Morris, A. P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44,** 981–990 (2012).
38. Xue, A. *et al.* Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat. Commun.* **9,** 2941 (2018).
39. Ng, M. C. Y. *et al.* Meta-Analysis of genome-wide association studies in African Americans provides insights into the genetic architecture of type 2 diabetes. *PLoS Genet.* **10,** e1004517 (2014).
40. Mahajan, A. *et al.* Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. *Nat. Genet.* **50,** 559–572 (2018).

41. Fuchsberger, C. *et al.* The genetic architecture of type 2 diabetes. *Nature* **536,** 41–47 (2016).
42. Mahajan, A. *et al.* Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.* **46,** 234–244 (2014).
43. Zhang, Y. *et al.* Association of PPP1R3B polymorphisms with blood lipid and C-reactive protein levels in a Chinese population. *J. Diabetes* **5,** 275–281 (2013).
44. Imamura, M. *et al.* Genome-wide association studies in the Japanese population identify seven novel loci for type 2 diabetes. *Nat. Commun.* **7,** 10531 (2016).
45. Tabassum, R. *et al.* Genome-wide association study for type 2 diabetes in indians identifies a new susceptibility locus at 2q21. *Diabetes* **62,** 977–986 (2013).
46. Scott, R. A. *et al.* An expanded genome-wide association study of type 2 diabetes in Europeans. *Diabetes* **66,** 2888–2902 (2017).
47. Benner, C. *et al.* FINEMAP: Efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32,** 1493–1501 (2016).
48. Malik, R. *et al.* Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* **50,** 524–537 (2018).
49. Tachmazidou, I. *et al.* Identification of new therapeutic targets for osteoarthritis through genome-wide analyses of UK Biobank data. *Nat. Genet.* **51,** 230–236 (2019).
50. Hormozdiari, F. *et al.* Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am. J. Hum. Genet.* **99,** 1245–1260 (2016).
51. Sun, B. B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558,** 73–79 (2018).
52. Kemp, J. P. *et al.* Identification of 153 new loci associated with heel bone mineral density and functional involvement of GPC6 in osteoporosis. *Nat. Genet.* **49,** 1468–1475 (2017).
53. Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6,** 5890 (2015).
54. Barbeira, A. N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.* **9,** 1825 (2018).
55. Keavney, B. Genetic epidemiological studies of coronary heart disease. *Int. J. Epidemiol.* **31,** 730–736 (2002).
56. Ebrahim, S. & Davey Smith, G. Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? *Hum. Genet.* **123,** 15–33 (2008).
57. Burgess, S., Bowden, J., Fall, T., Ingelsson, E. & Thompson, S. G. Sensitivity analyses for robust causal inference from Mendelian randomization analyses with multiple genetic variants. *Epidemiology* **28,** 30–42 (2017).
58. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44,** 512–525 (2015).
59. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40,** 304–314 (2016).
60. Verbanck, M., Chen, C. Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **50,** 693–698 (2018).
61. Flannick, J. & Florez, J. C. Type 2 diabetes : genetic data sharing to advance complex disease research. *Nat. Rev. Genet.* **17,** 535–549 (2016).

62. Wittemans, L. B. L., Lotta, L. A. & Langenberg, C. Prioritising Risk Factors for Type 2 Diabetes: Causal Inference through Genetic Approaches. *Curr. Diab. Rep.* **18,** 40 (2018).

63. Sattar, N. *et al.* Statins and risk of incident diabetes: a collaborative meta-analysis of randomised statin trials. *Lancet* **375,** 735–742 (2010).

64. Haase, C. L., Tybjaerg-Hansen, A., Nordestgaard, B. G. & Frikke-Schmidt, R. HDL Cholesterol and Risk of Type 2 Diabetes: A Mendelian Randomization Study. *Diabetes* **64,** 3328–3333 (2015).

65. Qi, Q., Liang, L., Doria, A., Hu, F. B. & Qi, L. Genetic Predisposition to Dyslipidemia and Type 2 Diabetes Risk in Two Prospective Cohorts. *Diabetes* **61,** 745–752 (2012).

66. Maneka, N. *et al.* Mendelian randomization studies do not support a role for raised circulating triglyceride levels influencing type 2 diabetes, glucose levels, or insulin resistance. *Diabetes* **60,** 1008–1018 (2011).

67. Burgess, S. & Thompson, S. G. Multivariable Mendelian Randomization: The Use of Pleiotropic Genetic Variants to Estimate Causal Effects. *Am. J. Epidemiol.* **181,** 251–60 (2015).

68. Mora, S. *et al.* Lipoprotein particle profiles by nuclear magnetic resonance compared with standard lipids and apolipoproteins in predicting incident cardiovascular disease in women. *Circulation* **119,** 931–939 (2009).

69. Wilson, P. W. F. *et al.* Prediction of Incident Diabetes Mellitus in Middle-aged Adults: The Framingham Offspring Study. *Arch. Intern. Med.* **167,** 1068–1074 (2007).

70. Gupta, A. K. *et al.* Determinants of New-Onset Diabetes Randomized in the Anglo-Scandinavian Cardiac Outcomes Trial – Blood Pressure Lowering Arm and the Relative Influence of Antihypertensive Medication. *Diabetes Care* **31,** 982–988 (2008).

71. White, J. *et al.* Association of Lipid Fractions With Risks for Coronary Artery Disease and Diabetes. *JAMA Cardiol.* **1,** 692–699 (2016).

72. Fall, T. *et al.* Using Genetic Variants to Assess the Relationship Between Circulating Lipids and Type 2 Diabetes. *Diabetes* **64,** 2676–84 (2015).

73. Besseling, J., Kastelein, J. J. P., Defesche, J. C., Hutten, B. A. & Hovingh, G. K. Association Between Familial Hypercholesterolemia and Prevalence of Type 2 Diabetes Mellitus. *JAMA* **313,** 1029–1036 (2015).

74. Xu, H. *et al.* Familial hypercholesterolemia and type 2 diabetes in the old order amish. *Diabetes* **66,** 2054–2058 (2017).

75. Lotta, L. A. *et al.* Association Between Low-Density Lipoprotein Cholesterol–Lowering Genetic Variants and Risk of Type 2 Diabetes. *JAMA* **316,** 1383–1391 (2016).

76. Liu, D. J. *et al.* Exome-wide association study of plasma lipids in >300,000 individuals. *Nat. Genet.* **49,** 1758–1766 (2017).

77. Burgess, S., Freitag, D. F., Khan, H., Gorman, D. N. & Thompson, S. G. Using multivariable Mendelian randomization to disentangle the causal effects of lipid fractions. *PLoS One* **9,** e108891 (2014).

78. Silverman, M. G. *et al.* Association Between Lowering LDL-C and Cardiovascular Risk Reduction Among Different Therapeutic Interventions. *JAMA* **316,** 1289–1297 (2016).

79. Lotta, L. A. *et al.* Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. *Nat. Genet.* **49,** 17–26 (2017).

80. Gusarova, V. *et al.* Genetic inactivation of ANGPTL4 improves glucose homeostasis and is associated with reduced risk of diabetes. *Nat. Commun.* **9,** 2252 (2018).

81. Do, R. *et al.* Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nat. Genet.* **45,** 1345–1352 (2013).

82. Helgadottir, A. *et al.* Variants with large effects on blood lipids and the role of cholesterol and triglycerides in coronary disease. *Nat. Genet.* **48,** 634–639 (2016).

83. Lean, M. E. J. *et al.* Primary care-led weight management for remission of type 2 diabetes (DiRECT): An open-label, cluster-randomised trial. *Lancet* **391,** 541–551 (2017).

84. Frayling, T. M. *et al.* A Common Variant in the FTO Gene Is Associated with Body Mass Index and Predisposes to Childhood and Adult Obesity. *Science (80-.).* **316,** 889–894 (2007).

85. Dale, C. E. *et al.* Causal Associations of Adiposity and Body Fat Distribution With Coronary Heart Disease, Stroke Subtypes, and Type 2 Diabetes Mellitus: A Mendelian Randomization Analysis. *Circulation* **135,** 2373–2388 (2017).

86. Corbin, L. J. *et al.* Body mass index as a modifiable risk factor for type 2 diabetes: Refining and understanding causal estimates using Mendelian randomisation. *Diabetes* **65,** 3002–3007 (2016).

87. Emdin, C. A. *et al.* Genetic association of waist-to-hip ratio with cardiometabolic traits, type 2 diabetes, and coronary heart disease. *JAMA* **317,** 626–634 (2017).

88. Lu, Y. *et al.* New loci for body fat percentage reveal link between adiposity and cardiometabolic disease risk. *Nat. Commun.* **7,** 10495 (2016).

89. Kilpeläinen, T. O. *et al.* Genetic variation near IRS1 associates with reduced adiposity and an impaired metabolic profile. *Nat. Genet.* **43,** 753–760 (2011).

90. Minster, R. L. *et al.* A thrifty variant in CREBRF strongly influences body mass index in Samoans. *Nat. Genet.* **48,** 1049–1054 (2016).

91. Barroso, I. *et al.* Dominant negative mutations in human PPARγ associated with severe insulin resistance, diabetes mellitus and hypertension. *Nature* **402,** 880–883 (1999).

92. Yaghootkar, H. *et al.* Genetic evidence for a link between favorable adiposity and lower risk of type 2 diabetes, hypertension, and heart disease. *Diabetes* **65,** 2448–2460 (2016).

93. Manning, A. . *et al.* A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat. Genet.* **44,** 659–669 (2012).

94. Scott, R. A. *et al.* Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat. Genet.* **44,** 991–1005 (2012).

95. Scott, R. A. *et al.* Common genetic variants highlight the role of insulin resistance and body fat distribution in type 2 diabetes, independent of obesity. *Diabetes* **63,** 4378–4387 (2014).

96. Yaghootkar, H. *et al.* Genetic Evidence for a Normal-Weight " Metabolically Obese " Phenotype Linking Insulin Resistance , Hypertension , Coronary Artery Disease , and Type 2 Diabetes. *Diabetes* **63,** 4369–4377 (2014).

97. Ramachandran, A., Wan Ma, R. C. & Snehalatha, C. Diabetes in Asia. *Lancet* **375,** 408–418 (2010).

98. Babai, M. A. *et al.* Defining a BMI cut-off point for the iranian population: The shiraz heart study. *PLoS One* **11,** e0160639 (2016).

99. Frank, L. K. *et al.* Measures of general and central obesity and risk of type 2 diabetes in a Ghanaian population. *Trop. Med. Int. Heal.* **18,** 141–151 (2013).

100. Virtue, S. & Vidal-Puig, A. Adipose tissue expandability, lipotoxicity and the Metabolic Syndrome - An allostatic perspective. *Biochim. Biophys. Acta* **1801,** 338–349 (2010).

101. Danforth, E. Failure of adipocyte differentiation causes type II diabetes mellitus? *Nat. Genet.* **26,** 13 (2000).

102. Abbasi, A. *et al.* A Systematic Review of Biomarkers and Risk of Incident Type 2 Diabetes: An Overview of

Epidemiological, Prediction and Aetiological Research Literature. *PLoS One* **11,** e0163721 (2016).

103. Ye, Z. *et al.* Association between circulating 25-hydroxyvitamin D and incident type 2 diabetes: a mendelian randomisation study. *Lancet Diabetes Endocrinol.* **3,** 35–42 (2015).
104. Mente, A. *et al.* Causal Relationship between Adiponectin and Metabolic Traits: A Mendelian Randomization Study in a Multiethnic Population. *PLoS One* **8,** e66808 (2013).
105. Yaghootkar, H. *et al.* Mendelian randomization studies do not support a causal role for reduced circulating adiponectin levels in insulin resistance and type 2 diabetes. *Diabetes* **62,** 3589–3598 (2013).
106. Sluijs, I. *et al.* A Mendelian Randomization Study of Circulating Uric Acid and Type 2 Diabetes. *Diabetes* **64,** 3028–36 (2015).
107. Prins, B. P. *et al.* Investigating the Causal Relationship of C-Reactive Protein with 32 Complex Somatic and Psychiatric Outcomes: A Large-Scale Cross-Consortium Mendelian Randomization Study. *PLoS Med.* **13,** e1001976 (2016).
108. Noordam, R., Smit, R. A. J., Postmus, I., Trompet, S. & van Heemst, D. Assessment of causality between serum gamma-glutamyltransferase and type 2 diabetes mellitus using publicly available data: A mendelian randomization study. *Int. J. Epidemiol.* **45,** 1953–1960 (2016).
109. Lee, Y. S. *et al.* Serum gamma-glutamyl transferase and risk of type 2 diabetes in the general Korean population: a Mendelian randomization study. *Hum. Mol. Genet.* **25,** 3877–3886 (2016).
110. Liu, J., Au Yeung, S. L., Lin, S. L., Leung, G. M. & Schooling, C. M. Liver Enzymes and Risk of Ischemic Heart Disease and Type 2 Diabetes Mellitus: A Mendelian Randomization Study. *Sci. Rep.* **6,** 38813 (2016).
111. Wang, Q. *et al.* Sex hormone-binding globulin associations with circulating lipids and metabolites and the risk for type 2 diabetes: Observational and causal effect estimates. *Int. J. Epidemiol.* **44,** 623–637 (2015).
112. Ding, E. L. *et al.* Sex Hormone–Binding Globulin and Risk of Type 2 Diabetes in Women and Men. *N. Engl. J. Med.* **361,** 1152–1163 (2009).
113. Jujić, A. *et al.* Atrial natriuretic peptide and type 2 diabetes development - Biomarker and genotype association study. *PLoS One* **9,** e89201 (2014).
114. Pfister, R. *et al.* Mendelian Randomization Study of B-Type Natriuretic Peptide and Type 2 Diabetes: Evidence of Causal Association from Population Studies. *PLoS Med.* **8,** e1001112 (2011).
115. Abbasi, A. *et al.* Bilirubin as a potential causal factor in type 2 diabetes risk: A mendelian randomization study. *Diabetes* **64,** 1459–1469 (2015).
116. Walford, G. A. *et al.* Metabolite profiles of diabetes incidence and intervention response in the diabetes prevention program. *Diabetes* **65,** 1424–1433 (2016).
117. Shi, L. *et al.* Plasma metabolites associated with type 2 diabetes in a Swedish population: a case–control study nested in a prospective cohort. *Diabetologia* **61,** 849–861 (2018).
118. Peddinti, G. *et al.* Early metabolic markers identify potential targets for the prevention of type 2 diabetes. *Diabetologia* **60,** 1740–1750 (2017).
119. Lu, Y. *et al.* Metabolic signatures and risk of type 2 diabetes in a Chinese population: an untargeted metabolomics study using both LC-MS and GC-MS. *Diabetologia* **59,** 2349–2359 (2016).
120. Forouhi, N. G. *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 Endocrinol.* **2,** 810–818 (2014).
121. Suvitaival, T. *et al.* Lipidome as a predictive tool in progression to type 2 diabetes in Finnish men. *Metabolism.*

78, 1–12 (2018).

122. Rhee, E. P. *et al.* Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. *J. Clin. Invest.* **121,** 1402–1411 (2011).
123. Lu, Y. *et al.* Serum lipids in association with type 2 diabetes risk and prevalence in a Chinese population. *J. Clin. Endocrinol. Metab.* **103,** 671–680 (2018).
124. Belongie, K. J. *et al.* Identification of novel biomarkers to monitor β-cell function and enable early detection of type 2 diabetes risk. *PLoS One* **12,** e0182932 (2017).
125. Chambers, J. C. *et al.* Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. *Lancet Diabetes Endocrinol.* **3,** 526–534 (2015).
126. Patti, G. J., Yanes, O. & Siuzdak, G. Metabolomics: The apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* **13,** 263–269 (2012).
127. Newgard, C. B. Metabolomics and metabolic diseases: Where do we stand? *Cell Metab.* **25,** 43–56 (2017).
128. Fall, T. *et al.* Non-targeted metabolomics combined with genetic analyses identifies bile acid synthesis and phospholipid metabolism as being associated with incident type 2 diabetes. *Diabetologia* **59,** 2114–2124 (2016).
129. Palmer, N. D. *et al.* Metabolomic profile associated with insulin resistance and conversion to diabetes in the Insulin Resistance Atherosclerosis Study. *J. Clin. Endocrinol. Metab.* **100,** 463–468 (2015).
130. Nowak, C. *et al.* Effect of insulin resistance on monounsaturated fatty acid levels: A multi-cohort non-targeted metabolomics and Mendelian randomization study. *PLoS Genet.* **12,** e1006379 (2016).
131. Yamada, C. *et al.* Association between insulin resistance and plasma amino acid profile in non-diabetic Japanese subjects. *J. Diabetes Investig.* **6,** 408–415 (2015).
132. Würtz, P. *et al.* Branched-chain and aromatic amino acids are predictors of insulin resistance in young adults. *Diabetes Care* **36,** 648–655 (2013).
133. Cirulli, E. T., Guo, L., Swisher, C. L., Shah, N. & Huang, L. Profound perturbation of the human metabolome by obesity. *bioRxiv* 298224 (2018). doi:https://doi.org/10.1101/298224
134. Würtz, P. *et al.* Metabolic signatures of adiposity in young adults: Mendelian randomization analysis and effects of weight change. *PLoS Med.* **11,** e1001765 (2014).
135. Menni, C. *et al.* Metabolomic Profiling of Long-Term Weight Change: Role of Oxidative Stress and Urate Levels in Weight Gain. *Obesity* **25,** 1618–1624 (2017).
136. Molnos, S. *et al.* Metabolite ratios as potential biomarkers for type 2 diabetes: a DIRECT study. *Diabetologia* **61,** 117–129 (2018).
137. Wang-Sattler, R. *et al.* Novel biomarkers for pre-diabetes identified by metabolomics. *Mol. Syst. Biol.* **8,** 615 (2012).
138. Rebholz, C. M. *et al.* Serum metabolomic profile of incident diabetes. *Diabetologia* **61,** 1046–1054 (2018).
139. Yengo, L. *et al.* Impact of statistical models on the prediction of type 2 diabetes using non-targeted metabolomics profiling. *Mol. Metab.* **5,** 918–925 (2016).
140. Yu, D. *et al.* Plasma metabolomic profiles in association with type 2 diabetes risk and prevalence in Chinese adults. *Metabolomics* **12,** 10.1007/s11306-015-0890-8. (2016).
141. Drogan, D. *et al.* Untargeted metabolic profiling identifies altered serum metabolites of type 2 diabetes mellitus in a prospective, nested case-control study. *Clin. Chem.* **61,** 487–497 (2015).
142. Padberg, I. *et al.* A new metabolomic signature in type-2 diabetes mellitus and its pathophysiology. *PLoS One*

9, e85082 (2014).

143. Zhao, J. *et al.* Novel Metabolic Markers for the Risk of Diabetes Development in American Indians. *Diabetes Care* **38,** 220–227 (2015).
144. Liu, J. *et al.* Metabolomics based markers predict type 2 diabetes in a 14-year follow-up study. *Metabolomics* **13,** 104 (2017).
145. Qiu, G. *et al.* Plasma metabolomics identified novel metabolites associated with risk of type 2 diabetes in two prospective cohorts of Chinese adults. *Int. J. Epidemiol.* **45,** 1507–1516 (2016).
146. Chen, T. *et al.* Branched-chain and aromatic amino acid profiles and diabetes risk in Chinese populations. *Sci. Rep.* **6,** 20594 (2016).
147. Merino, J. *et al.* Metabolomics insights into early type 2 diabetes pathogenesis and detection in individuals with normal fasting glucose. *Diabetologia* **61,** 1315–1324 (2018).
148. Stancáková, A. *et al.* Hyperglycemia and a common variant of GCKR are associated with the levels of eight amino acids in 9,369 finnish men. *Diabetes* **61,** 1895–1902 (2012).
149. Svingen, G. F. T. *et al.* Prospective associations of systemic and urinary choline metabolites with incident type 2 diabetes. *Clin. Chem.* **62,** 755–765 (2016).
150. Wang, T. J. *et al.* Metabolite profiles and the risk of developing diabetes. *Nat. Med.* **17,** 448–453 (2011).
151. Wang, T. J. *et al.* 2-Aminoadipic acid is a biomarker for diabetes risk. *J. Clin. Invest.* **123,** 4309–4317 (2013).
152. Yamakado, M. *et al.* Plasma Free Amino Acid Profiles Predict Four-Year Risk of Developing Diabetes, Metabolic Syndrome, Dyslipidemia, and Hypertension in Japanese Population. *Sci. Rep.* **5,** 11918 (2015).
153. Festa, A. *et al.* Nuclear magnetic resonance lipoprotein abnormalities in prediabetic subjects in the Insulin Resistance Atherosclerosis Study. *Circulation* **111,** 3465–3472 (2005).
154. Hodge, A. M., Jenkins, A. J., English, D. R., O'Dea, K. & Giles, G. G. NMR-determined lipoprotein subclass profile predicts type 2 diabetes. *Diabetes Res. Clin. Pract.* **83,** 132–9 (2009).
155. Mahendran, Y. *et al.* Glycerol and fatty acids in serum predict the development of hyperglycemia and type 2 diabetes in Finnish men. *Diabetes Care* **36,** 3732–3738 (2013).
156. Savolainen, O. *et al.* Biomarkers of food intake and nutrient status are associated with glucose tolerance status and development of type 2 diabetes in older Swedish women. *Am. J. Clin. Nutr.* **106,** 1302–1310 (2017).
157. Joyce, K. E. *et al.* Testosterone, Dihydrotestosterone, Sex Hormone Binding Globulin and Incident Diabetes among Older Men: the Cardiovascular Health Study. *J. Clin. Endocrinol. Metab.* **102,** 33–39 (2017).
158. Fenske, B. *et al.* Endogenous Androgens and Sex Hormone–Binding Globulin in Women and Risk of Metabolic Syndrome and Type 2 Diabetes. *J. Clin. Endocrinol. Metab.* **100,** 4595–4603 (2015).
159. Wigger, L. *et al.* Plasma Dihydroceramides Are Diabetes Susceptibility Biomarker Candidates in Mice and Humans. *Cell Rep.* **18,** 2269–2279 (2017).
160. Mwinyi, J. *et al.* Plasma 1-deoxysphingolipids are early predictors of incident type 2 diabetes mellitus. *PLoS One* **12,** 1–12 (2017).
161. Othman, A. *et al.* Plasma 1-deoxysphingolipids are predictive biomarkers for type 2 diabetes mellitus. *BMJ Open Diabetes Res. Care* **3,** e000073 (2015).
162. Wu, J. H. Y. *et al.* Omega-6 fatty acid biomarkers and incident type 2 diabetes: pooled analysis of individual-level data for 39 740 adults from 20 prospective cohort studies. *Lancet Diabetes Endocrinol.* **5,** 965–974 (2017).
163. de Mello, V. D. F. *et al.* Markers of cholesterol metabolism as biomarkers in predicting diabetes in the Finnish Diabetes Prevention Study. *Nutr. Metab. Cardiovasc. Dis.* **25,** 635–42 (2015).

164. Savolainen, O. *et al.* Biomarkers for predicting type 2 diabetes development - Can metabolomics improve on existing biomarkers? *PLoS One* **12,** e0177738 (2017).
165. Sun, L. *et al.* Early prediction of developing type 2 diabetes by plasma acylcarnitines: A population-based study. *Diabetes Care* **39,** 1563–1570 (2016).
166. Tillin, T. *et al.* Diabetes risk and amino acid profiles: cross-sectional and prospective analyses of ethnicity, amino acids and diabetes in a South Asian and European cohort from the SABRE (Southall And Brent REvisited) Study. *Diabetologia* **58,** 968–979 (2015).
167. Ruiz-Canela, M. *et al.* Plasma branched-chain amino acids and incident cardiovascular disease in the PREDIMED trial. *Clin. Chem.* **62,** 582–592 (2016).
168. Mayers, J. R. *et al.* Elevation of circulating branched-chain amino acids is an early event in human pancreatic adenocarcinoma development. *Nat. Med.* **20,** 1193–1198 (2014).
169. Shin, S.-Y. *et al.* An atlas of genetic influences on human blood metabolites. *Nat. Genet.* **46,** 543–50 (2014).
170. Kettunen, J. *et al.* Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat. Commun.* **7,** 11122 (2016).
171. Draisma, H. H. M. *et al.* Genome-wide association study identifies novel genetic variants contributing to variation in blood metabolite levels. *Nat. Commun.* **6,** 7208 (2015).
172. Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in ∼700 000 individuals of European ancestry. *Hum. Mol. Genet.* **27,** 3641–3649 (2018).
173. Lotta, L. A. *et al.* Genetic predisposition to an impaired metabolism of the branched-chain amino acids and risk of type 2 diabetes: A Mendelian randomisation analysis. *PLOS Med.* **13,** e1002179 (2016).
174. Wang, W. *et al.* Glycine metabolism in animals and humans: Implications for nutrition and health. *Amino Acids* **45,** 463–477 (2013).
175. Ding, Y. *et al.* Plasma Glycine and Risk of Acute Myocardial Infarction in Patients With Suspected Stable Angina Pectoris. *J. Am. Heart Assoc.* **5,** e002621 (2016).
176. El-Hafidi, M. *et al.* Glycine increases insulin sensitivity and glutathione biosynthesis and protects against oxidative stress in a model of sucrose-induced insulin resistance. *Oxid. Med. Cell. Longev.* **2018,** (2018).
177. Van den Eynden, J. *et al.* Glycine and glycine receptor signalling in non-neuronal cells. *Front. Mol. Neurosci.* **2,** 9 (2009).
178. Jackson, A. A., Dunn, R. L., Marchand, M. C. & Langley-Evans, S. C. Increased systolic blood pressure in rats induced by a maternal low-protein diet is reversed by dietary supplementation with glycine. *Clin. Sci.* **103,** 633–639 (2002).
179. Gall, W. E. *et al.* α-Hydroxybutyrate Is an Early Biomarker of Insulin Resistance and Glucose Intolerance in a Nondiabetic Population. *PLoS One* **5,** e10883 (2010).
180. Lustgarten, M. S., Lyn Price, L., Phillips, E. M. & Fielding, R. A. Serum glycine is associated with regional body fat and insulin resistance in functionally-limited older adults. *PLoS One* **8,** 8–14 (2013).
181. Michaliszyn, S. F. *et al.* Metabolomic profiling of amino acids and β-cell function relative to insulin sensitivity in youth. *J. Clin. Endocrinol. Metab.* **97,** 2119–2124 (2012).
182. Takashina, C. *et al.* Associations among the plasma amino acid profile, obesity, and glucose metabolism in Japanese adults with normal glucose tolerance. *Nutr. Metab. (Lond).* **13,** 5 (2016).
183. Xie, W. *et al.* Genetic variants associated with glycine metabolism and their role in insulin sensitivity and type 2 diabetes. *Diabetes* **62,** 2141–2150 (2013).

184. Long, T. *et al.* Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nat. Genet.* **49,** 568–578 (2017).

185. Brawley, L. *et al.* Glycine rectifies vascular dysfunction induced by dietary protein imbalance during pregnancy. *J. Physiol.* **554,** 497–504 (2004).

186. Hafidi, M. E. Glycine intake decreases plasma free fatty acids, adipose cell size, and blood pressure in sucrose-fed rats. *AJP Regul. Integr. Comp. Physiol.* **287,** R1387–R1393 (2004).

187. McCarty, M. F., Barroso-Aranda, J. & Contreras, F. The hyperpolarizing impact of glycine on endothelial cells may be anti-atherogenic. *Med. Hypotheses* **73,** 263–264 (2009).

188. Hartiala, J. A. *et al.* Genome-wide association study and targeted metabolomics identifies sex-specific association of CPS1 with coronary artery disease. *Nat. Commun.* **7,** 10558 (2016).

189. Mittelstrass, K. *et al.* Discovery of sexual dimorphisms in metabolic and genetic biomarkers. *PLoS Genet.* **7,** (2011).

190. Riboli, E. *et al.* European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr.* **5,** 1113–1124 (2002).

191. Day, N. *et al.* EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. *British journal of cancer* **80,** 95–103 (1999).

192. Mccarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48,** 1279–1283 (2016).

193. Huang, J. *et al.* Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat. Commun.* **6,** 8111 (2015).

194. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).

195. Howie, B. N., Donnelly, P. & Marchini, J. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLoS Genet.* **5,** e1000529 (2009).

196. Lassale, C. *et al.* Separate and combined associations of obesity and metabolic health with coronary heart disease: a pan-European case-cohort analysis. *Eur. Heart J.* **39,** 397–406 (2018).

197. Consortium, T. 1000 G. P. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–64 (2012).

198. Langenberg, C. *et al.* Design and cohort description of the InterAct Project: an examination of the interaction of genetic and lifestyle factors on the incidence of type 2 diabetes in the EPIC Study. *Diabetologia* **54,** 2272–82 (2011).

199. Moore, C. *et al.* The INTERVAL trial to determine whether intervals between blood donations can be safely and acceptably decreased to optimise blood supply: study protocol for a randomised controlled trial. *Trials* **15,** 363 (2014).

200. Soininen, P. *et al.* High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism. *Analyst* **134,** 1781–1785 (2009).

201. Inouye, M. *et al.* Metabonomic, transcriptomic, and genomic variation of a population cohort. *Mol. Syst. Biol.* **6,** 441 (2010).

202. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med.* **12,** e1001779 (2015).

203. Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47,** 284–290 (2015).

204. Shin, S. *et al.* Europe PMC Funders Group An atlas of genetic influences on human blood metabolites. **46,** 543–550 (2014).

205. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26,** 2190–2191 (2010).

206. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44,** 369–375 (2012).

207. Wolfgang Viechtbauer. Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* **36,** (2010).

208. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).

209. Nikpay, M. *et al.* A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47,** 1121–30 (2015).

210. Samani, N. J. *et al.* Genomewide association analysis of coronary artery disease. *N. Engl. J. Med.* **357,** 443–453 (2007).

211. Erdmann, J. *et al.* New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat. Genet.* **41,** 280–282 (2009).

212. Erdmann, J. *et al.* Genome-wide association study identifies a new locus for coronary artery disease on chromosome 10p11.23. *Eur. Heart J.* **32,** 158–168 (2011).

213. Bycroft, C. *et al.* Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv* 166298 (2017). doi:10.1101/166298

214. Willer, C. J. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45,** 1274–83 (2013).

215. Astle, W. J. *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* **167,** 1415–1429.e19 (2016).

216. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37,** 658–665 (2013).

217. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518,** 197–206 (2015).

218. Prokopenko, I. *et al.* A Central Role for GRB10 in Regulation of Islet Function in Man. *PLoS Genet.* **10,** e1004235 (2014).

219. Day, F. R. *et al.* Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat. Genet.* **3,** 834–841 (2017).

220. Dehghan, A. *et al.* Association of novel genetic loci with circulating fibrinogen levels a genome-wide association study in 6 population-based cohorts. *Circ. Cardiovasc. Genet.* **2,** 125–133 (2009).

221. Danik, J. S. *et al.* Novel loci, including those related to Crohn disease, psoriasis, and inflammation, identified in a genome-wide association study of Fibrinogen in 17686 women the Women's Genome Health Study. *Circ. Cardiovasc. Genet.* **2,** 134–141 (2009).

222. Sabater-Lleal, M. *et al.* Multiethnic meta-analysis of genome-wide association studies in >100 000 subjects identifies 23 fibrinogen-associated loci but no strong evidence of a causal association between circulating fibrinogen and cardiovascular disease. *Circulation* **128,** 1310–1324 (2013).

223. Tang, W. *et al.* Genetic associations for activated partial thromboplastin time and prothrombin time, their gene expression profiles, and risk of coronary artery disease. *Am. J. Hum. Genet.* **91,** 152–162 (2012).

224. Gaunt, T. R., Lowe, G. D. O., Lawlor, D. A., Casas, J. P. & Day, I. N. M. A gene-centric analysis of activated

partial thromboplastin time and activated protein C resistance using the HumanCVD focused genotyping array. *Eur. J. Hum. Genet.* **21,** 779–783 (2013).

225. Houlihan, L. M. *et al.* Common Variants of Large Effect in F12, KNG1, and HRG Are Associated with Activated Partial Thromboplastin Time. *Am. J. Hum. Genet.* **86,** 626–631 (2010).
226. Consortium, T. Cardi. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat. Genet.* **45,** 25–35 (2013).
227. Adeva-Andany, M. *et al.* Insulin resistance and glycine metabolism in humans. *Amino Acids* **50,** 11–27 (2018).
228. Applegarth, D. A. & Toone, J. R. Glycine encephalopathy (nonketotic hyperglycinaemia): Review and update. *J. Inherit. Metab. Dis.* **27,** 417–422 (2004).
229. Hoover-Fong, J. E. *et al.* Natural history of nonketotic hyperglycinemia in 65 patients. *Neurology* **63,** 1847–1853 (2004).
230. Hennermann, J. B., Berger, J.-M., Grieben, U., Scharer, G. & Van Hove, J. L. K. Prediction of long-term outcome in glycine encephalopathy: a clinical survey. *J. Inherit. Metab. Dis.* **35,** 253–261 (2012).
231. Narisawa, A. *et al.* Mutations in genes encoding the glycine cleavage system predispose to neural tube defects in mice and humans. *Hum. Mol. Genet.* **21,** 1496–1503 (2012).
232. Gannon, M. C., Nuttall, J. a. & Nuttall, F. Q. The metabolic response to ingested glycine. *Am. J. Clin. Nutr.* **76,** 1302–1307 (2002).
233. Gonzalez-Ortiz, M., Medina-Santillan, R., Martinez-Abundis, E. & van Drateln, C. R. Effect of glycine on insulin secretion and action in healthy first-degree relatives of type 2 diabetes mellitus patients. *Horm. Metab. Res.* **33,** 358–360 (2001).
234. Yan-Do, R. *et al.* A glycine-insulin autocrine feedback loop enhances insulin secretion from human β-cells and is impaired in type 2 diabetes. *Diabetes* **65,** 2311–2321 (2016).
235. Cruz, M., Maldonado-Bernal, C., Mondragón-Gonzalez, R. et al. Glycine treatment decreases proinflammatory cytokines and increases interferon-gamma in patients with type 2 diabetes. *Endocrinol Invest.* **31,** 694–699 (2008).
236. Tabák, A. G., Herder, C., Rathmann, W., Brunner, E. J. & Kivimäki, M. Prediabetes: A high-risk state for diabetes development. *Lancet* **379,** 2279–2290 (2012).
237. Yan-Do, R. & MacDonald, P. E. Impaired 'glycine'-mia in type 2 diabetes and potential mechanisms contributing to glucose homeostasis. *Endocrinology* **158,** 1064–1073 (2017).
238. Karpe, F., Dickmann, J. R. & Frayn, K. N. Fatty acids, obesity, and insulin resistance: Time for a reevaluation. *Diabetes* **60,** 2441–2449 (2011).
239. Webster, L. T., Siddiqui, U. A., Lucas, S. V., Strong, J. M. & Mieyal, J. J. Identification of Separate Acyl-CoA : Glycine and Activities in Fractions from Liver of Rhesus Monkey and. *J. Biol. Chem.* **251,** 3352–3358 (1976).
240. Gar, C. *et al.* Serum and plasma amino acids as markers of prediabetes, insulin resistance, and incident diabetes. *Crit. Rev. Clin. Lab. Sci.* **55,** 21–32 (2018).
241. Rapsomaniki, E. *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* **383,** 1899–1911 (2014).
242. Wang, L. *et al.* Reconstruction and analysis of correlation networks based on GC-MS metabolomics data for young hypertensive men. *Anal. Chim. Acta* **854,** 95–105 (2015).
243. Quan, H. *et al.* Hypertension and impaired glycine handling in mice lacking the orphan transporter XT2. *Mol. Cell. Biol.* **24,** 4166–73 (2004).

244. Margarita Díaz-Flores, Miguel Cruz, Genoveva Duran-Reyes, Catarina Munguia-Miranda, Hilda Loza-Rodríguez, Evelyn Pulido-Casas, Nayeli Torres-Ramírez, Olga Gaja-Rodriguez, Jesus Kumate, Luis Arturo Baiza-Gutman, D. H.-S. Oral supplementation with glycine reduces oxidative stress in patients with metabolic syndrome, improving their systolic blood pressure. *Can. J. Physiol. Pharmacol.* **91,** 855–860 (2013).

245. Schemmer, P. *et al.* Glycine reduces platelet aggregation. *Amino Acids* **44,** 925–931 (2013).

246. Van den Eynden, J. *et al.* Glycine and glycine receptor signalling in non-neuronal cells. *Front. Mol. Neurosci.* **2,** article 9 (2009).

247. Amelio, I., Cutruzzola, F., Antonov, A., Agostini, M. & Melino, G. Serine and glycine metabolism in cancer. *Trends Biochem. Sci.* **39,** 191–198 (2014).

248. Wien, T. N. *et al.* Cancer risk with folic acid supplements: A systematic review and meta-analysis. *BMJ Open* **2,** e000653 (2012).

249. Maddocks, O. D. K. *et al.* Modulating the therapeutic response of tumours to dietary serine and glycine starvation. *Nature* **544,** 372–376 (2017).

250. Pulit, S. L., de With, S. A. J. & de Bakker, P. I. W. Resetting the bar: Statistical significance in whole-genome sequencing-based association studies of global populations. *Genet. Epidemiol.* **41,** 145–151 (2017).

251. Nguyen D, Hsu JW, Jahoor F, S. R. Effect of Increasing Glutathione With Cysteine and Glycine Supplementation on Mitochondrial Fuel Oxidation, Insulin Sensitivity, and Body Composition in Older HIV-Infected Patients. *J. Clin. Endocrinol. Metab.* **99,** 169–177 (2014).

252. Ruiz-Ramírez, A., Ortiz-Balderas, E., Cardozo-Saldaña, G., Diaz-Diaz, E. & El-Hafidi, M. Glycine restores glutathione and protects against oxidative stress in vascular tissue from sucrose-fed rats. *Clin. Sci.* **126,** 19–29 (2014).

253. The GBD 2015 Obesity Collaborators. Health effects of overweight and obesity in 195 countries over 25 years. *N. Engl. J. Med.* **337,** 13–27 (2017).

254. Organisation, T. W. H. Global Health Observatory. (2017). Available at: www.who.int/gho/en.

255. Akiyama, M. *et al.* Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nat. Genet.* **49,** 1458–1467 (2017).

256. Wen, W. *et al.* Meta-analysis identifies common variants associated with body mass index in east Asians. *Nat. Genet.* **44,** 307–311 (2012).

257. Okorodudu, D. O. *et al.* Diagnostic performance of body mass index to identify obesity as defined by body adiposity: A systematic review and meta-analysis. *Int. J. Obes.* **34,** 791–799 (2010).

258. Romero-Corral, A. *et al.* Accuracy of body mass index in diagnosing obesity in the adult general population. *Int. J. Obes.* **32,** 959–966 (2008).

259. Jo, A. & Mainous III, A. G. Informational value of percent body fat with body mass index for the risk of abnormal blood glucose: a nationally representative cross-sectional study. *BMJ Open* **8,** e019200 (2018).

260. Ode, J. J., Pivarnik, J. M., Reeves, M. J. & Knous, J. L. Body mass index as a predictor of percent fat in college athletes and nonathletes. *Med. Sci. Sports Exerc.* **39,** 403–409 (2007).

261. Klop, B., Elte, J. W. F. & Cabezas, M. C. Dyslipidemia in Obesity: Mechanisms and Potential Targets. *Nutrients* **5,** 1218–1240 (2013).

262. Hardya, O. T., Czech, M. P. & Corvera, S. What causes the insulin resistance underlying obesity? *Curr Opin Endocrinol Diabetes Obes* **19,** 81–87 (2012).

263. Gómez-Ambrosi, J. *et al.* Body adiposity and type 2 diabetes: Increased risk with a high body fat percentage

even having a normal BMI. *Obesity* **19,** 1439–1444 (2011).

264. Patschan, S. & Scholze, J. Obesity-related hypertension. *Cardiol. Rev.* **24,** 32–35 (2007).
265. Lavie, C. J., Milani, R. V. & Ventura, H. O. Obesity and Cardiovascular Disease. Risk Factor, Paradox, and Impact of Weight Loss. *J. Am. Coll. Cardiol.* **53,** 1925–1932 (2009).
266. Kim, S. & Valdez, R. Metabolic risk factors in U.S. youth with low relative muscle mass. *Obes. Res. Clin. Pract.* **9,** 125–132 (2015).
267. Larsen, B. A. *et al.* Association of muscle mass, area, and strength with incident diabetes in older adults: The Health ABC Study. *J. Clin. Endocrinol. Metab.* **101,** 1847–1855 (2016).
268. Bower, J. K., Meadows, R. J., Foster, M. C., Foraker, R. E. & Shoben, A. B. The Association of Percent Body Fat and Lean Mass With HbA1c in US Adults. *J. Endocr. Soc.* **1,** 600–608 (2017).
269. Cheng, S. & Wiklund, P. The effects of muscle mass and muscle quality on cardio-metabolic risk in peripubertal girls: A longitudinal study from childhood to early adulthood. *Int. J. Obes.* **42,** 648–654 (2017).
270. Leong, D. P. *et al.* Prognostic value of grip strength: findings from the Prospective Urban Rural Epidemiology (PURE) study. *Lancet* **386,** 266–273 (2015).
271. Cohen, D. D. *et al.* Low muscle strength is associated with metabolic risk factors in Colombian children: The ACFIES study. *PLoS One* **9,** e93150 (2014).
272. Dorsey, K. B., Thornton, J. C., Heymsfield, S. B. & Gallagher, D. Greater lean tissue and skeletal muscle mass are associated with higher bone mineral content in children. *Nutr. Metab. (Lond).* **7,** (2010).
273. Davidson, L. E. *et al.* Skeletal muscle and organ masses differ in overweight adults with type 2 diabetes. *J. Appl. Physiol.* **117,** 377–382 (2014).
274. DeFronzo, R. A. & Tripathy, D. Skeletal muscle insulin resistance is the primary defect in type 2 diabetes. *Diabetes Care* **32 Suppl 2,** (2009).
275. Srikanthan, P. & Karlamangla, A. S. Relative muscle mass is inversely associated with insulin resistance and prediabetes. Findings from the Third National Health and Nutrition Examination Survey. *J. Clin. Endocrinol. Metab.* **96,** 2898–2903 (2011).
276. Holten, M. K. *et al.* Strength training increases insulin-mediated glucose uptake , GLUT4 content , and insulin signaling in skeletal muscle in patients with Type 2 Diabetes. *Diabetes* **53,** 294–305 (2004).
277. Kyle, U. G. *et al.* Bioelectrical impedance analysis - Part I: Review of principles and methods. *Clin. Nutr.* **23,** 1226–1243 (2004).
278. Ulbrich, E. J., Nanz, D., Leinhard, O. D., Marcon, M. & Fischer, M. A. Whole-body adipose tissue and lean muscle volumes and their distribution across gender and age: MR-derived normative values in a normal-weight Swiss population. *Magn. Reson. Med.* **79,** 449–458 (2018).
279. Völgyi, E. *et al.* Assessing body composition with DXA and bioimpedance: Effects of obesity, physical activity, and age. *Obesity* **16,** 700–705 (2008).
280. Turcot, V. *et al.* Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. *Nat. Genet.* **50,** 26–41 (2018).
281. Willems, S. M. *et al.* Large-scale GWAS identifies multiple loci for hand grip strength providing biological insights into muscular fitness. *Nat. Commun.* **8,** 16015 (2017).
282. Tikkanen, E. *et al.* Biological insights into muscular strength: Genetic findings in the UK Biobank. *Sci. Rep.* **8,** 6451 (2018).
283. Reed, R. L., Pearlmutter, L., Yochum, K., Meredith, K. E. & Mooradian, A. D. The Relationship between

Muscle Mass and Muscle Strength in the Elderly. *J. Am. Geriatr. Soc.* **39,** 555–561 (2018).

284. Gusev, A. *et al.* Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am. J. Hum. Genet.* **95,** 535–552 (2014).
285. Goldstein, J. I. *et al.* Zcall: A rare variant caller for array-based genotyping. *Bioinformatics* **28,** 2543–2545 (2012).
286. Grove, M. L. *et al.* Best practices and joint calling of the HumanExome BeadChip : the CHARGE consortium. *PLoS One* **8,** e68095 (2013).
287. Feng, S., Liu, D., Zhan, X., Wing, M. K. & Abecasis, G. R. RAREMETAL: Fast and powerful meta-analysis for rare variants. *Bioinformatics* **30,** 2828–2829 (2014).
288. Zhan, X., Hu, Y., Li, B., Abecasis, G. R. & Liu, D. J. RVTESTS: An efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics* **32,** 1423–1426 (2016).
289. Winkler, T. W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9,** 1192–1212 (2014).
290. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17,** (2016).
291. Marouli, E. *et al.* Rare and low-frequency coding variants alter human adult height. *Nature* (2017). doi:10.1038/nature21039
292. Justice, A. E. *et al.* Protein-coding variants implicate novel genes related to lipid homeostasis contributing to body fat distribution. *bioRxiv* 352674 (2018).
293. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed model association for biobank-scale data sets. *bioRxiv* 194944 (2018). doi:10.1101/194944
294. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47,** 291–295 (2015).
295. Gazal, S. *et al.* Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat. Genet.* **49,** 1421–1427 (2017).
296. Koch, E., Ristroph, M. & Kirkpatrick, M. Long range linkage disequilibrium across the human genome. *PLoS One* **8,** e80754 (2013).
297. Liu, D. J. *et al.* Exome-wide association study of plasma lipids in >300,000 individuals. *Nat. Genet.* **49,** 1758–1766 (2017).
298. Investigators, M. I. G. and Cardi. E. C. Coding Variation in ANGPTL4, LPL, and SVEP1 and the Risk of Coronary Disease. *N. Engl. J. Med.* **374,** 1134–1144 (2016).
299. Aschard, H., Vilhjálmsson, B. J., Joshi, A. D., Price, A. L. & Kraft, P. Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. *Am. J. Hum. Genet.* **96,** 329–339 (2015).
300. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47,** 1236–1241 (2015).
301. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46,** 310–315 (2014).
302. Zillikens, M. C. *et al.* Large meta-analysis of genome-wide association studies identifies five loci for lean body mass. *Nat. Commun.* **8,** 80 (2017).
303. Emdin, C. A. *et al.* Analysis of predicted loss-of-function variants in UK Biobank identifies variants protective for disease. *Nat. Commun.* **9,** 1613 (2018).
304. Zorzetto, M. *et al.* SERPINA1 gene variants in individuals from the general population with reduced α1-antitrypsin concentrations. *Clin. Chem.* **54,** 1331–1338 (2008).

305. Carrell, R. W. & Lomas, D. A. Alpha1-antitrypsin deficiency - a model for conformational diseases. *N. Engl. J. Med.* **346,** 45–53 (2002).
306. Tanash, H. A., Nilsson, P. M., Nilsson, J. Å. & Piitulainen, E. Survival in severe alpha-1-antitrypsin deficiency (PiZZ). *Respir. Res.* **11,** 44 (2010).
307. Li, C. *et al.* Sex differences in the relationships between BMI, WHR and incidence of cardiovascular disease: A population-based cohort study. *Int. J. Obes.* **30,** 1775–1781 (2006).
308. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518,** 187–196 (2015).
309. Pulit, S. L. *et al.* Meta-analysis of genome-wide association studies for body fat distribution in 694,649 individuals of European ancestry. *Hum. Mol. Genet.* ddy327 (2018). doi:https://doi.org/10.1093/hmg/ddy327
310. Sherman, R. M. *et al.* Assembly of a pan-genome from deep sequencing of 910 humans of African descent. *Nat. Genet.* **51,** 30–35 (2019).
311. Lee, D. H. *et al.* Predicted lean body mass, fat mass, and all cause and cause specific mortality in men: prospective US cohort study. *BMJ* **362,** k2575 (2018).
312. Tomlinson, D. J., Erskine, R. M., Morse, C. I., Winwood, K. & Onambélé-Pearson, G. The impact of obesity on skeletal muscle strength and structure through adolescence to old age. *Biogerontology* **17,** 467–483 (2016).
313. Kelwick, R., Desanlis, I., Wheeler, G. N. & Edwards, D. R. The ADAMTS (A Disintegrin and Metalloproteinase with Thrombospondin motifs) family. *Genome Biol.* **16,** 113 (2015).
314. Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46,** 1173–86 (2014).
315. Li, S. *et al.* Zbtb7b engages the long noncoding RNA Blnc1 to drive brown and beige fat development and thermogenesis. *Proc. Natl. Acad. Sci.* 201703494 (2017). doi:10.1073/pnas.1703494114
316. Matsushita, M. *et al.* Impact of brown adipose tissue on body fatness and glucose metabolism in healthy humans. *Int. J. Obes.* **38,** 812–817 (2014).
317. Seale, P. *et al.* PRDM16 controls a brown fat/skeletal muscle switch. *Nature* **454,** 961–967 (2008).
318. Widom, R. L., Culic, I., Lee, J. Y. & Korn, J. H. Cloning and characterization of hcKrox, a transcriptional regulator of extracellular matrix gene expression. *Gene* **198,** 407–20 (1997).
319. Kappes, D. J. Expanding roles for ThPOK in thymic development. *Immunol Rev.* **238,** 182–194 (2010).
320. Degerman, E. *et al.* From PDE3B to the regulation of energy homeostasis. *Curr. Opin. Pharmacol.* **11,** 676–682 (2011).
321. DiPilato, L. M. *et al.* The Role of PDE3B Phosphorylation in the Inhibition of Lipolysis by Insulin. *Mol. Cell. Biol.* **35,** 2752–2760 (2015).
322. Walz, H. A. *et al.* β-cell PDE3B regulates Ca2+-stimulated exocytosis of insulin. *Cell. Signal.* **19,** 1505–1513 (2007).
323. Chung, Y. W. *et al.* White to beige conversion in PDE3B KO adipose tissue through activation of AMPK signaling and mitochondrial function. *Sci. Rep.* **7,** 40445 (2017).
324. Duckers, J. M. *et al.* Cardiovascular and musculskeletal co-morbidities in patients with alpha 1 antitrypsin deficiency. *Respir. Res.* **11,** 173 (2010).
325. Heinzelmann, I. *et al.* Differences in training response between patients with alpha-1 antitrypsin deficiency and COPD patients. *Eur. Respir. J.* **46,** OA4974 (2015).
326. Müller, M. J. *et al.* Effect of constitution on mass of individual organs and their association with metabolic

rate in Humans-A detailed view on allometric scaling. *PLoS One* **6,** e22732 (2011).

327. Vega, G. L. *et al.* Influence of body fat content and distribution on variation in metabolic risk. *J. Clin. Endocrinol. Metab.* **91,** 4459–4466 (2006).
328. Hamer, M. & Stamatakis, E. Metabolically healthy obesity and risk of all-cause and cardiovascular disease mortality. *J. Clin. Endocrinol. Metab.* **97,** 2482–2488 (2012).
329. De Koning, L., Merchant, A. T., Pogue, J. & Anand, S. S. Waist circumference and waist-to-hip ratio as predictors of cardiovascular events: Meta-regression analysis of prospective studies. *Eur. Heart J.* **28,** 850–856 (2007).
330. Fox, C. S. *et al.* Abdominal visceral and subcutaneous adipose tissue compartments: Association with metabolic risk factors in the framingham heart study. *Circulation* **116,** 39–48 (2007).
331. Klein, S. *et al.* Waist circumference and cardiometabolic risk: A consensus statement from Shaping America's Health: Association for Weight Management and Obesity Prevention; NAASO, the Obesity Society; the American Society for Nutrition; and the American Diabetes Associat. *Diabetes Care* **30,** 1647–1652 (2007).
332. Després, J. P. *et al.* Abdominal Obesity and the Metabolic Syndrome: Contribution to global cardiometabolic risk. *Arterioscler. Thromb. Vasc. Biol.* **28,** 1039–1049 (2008).
333. Lissner, L., Björkelund, C., Heitmann, B. L., Seidell, J. C. & Bengtsson, C. Larger hip circumference independently predicts health and longevity in a Swedish female cohort. *Obes. Res.* **9,** 644–646 (2001).
334. Heitmann, B. L., Frederiksen, P. & Lissner, L. Hip circumference and cardiovascular morbidity and mortality in men and women. *Obes. Res.* **12,** 482–487 (2004).
335. Winkler, T. W. *et al.* A joint view on genetic variants for adiposity differentiates subtypes with distinct metabolic implications. *Nat. Commun.* **9,** 1946 (2018).
336. Lotta, L. A. *et al.* Association of Genetic Variants Related to Gluteofemoral vs Abdominal Fat Distribution With Type 2 Diabetes, Coronary Disease, and Cardiovascular Risk Factors. *Jama* **320,** 2553–2563 (2018).
337. Stoever, K., Heber, A., Eichberg, S. & Brixius, K. Sarcopenia and Predictors of Skeletal Muscle Mass in Elderly Men With and Without Obesity. *Gerontol. Geriatr. Med.* **3,** 233372141771363 (2017).
338. Hamasaki, H. *et al.* Associations between lower extremity muscle mass and metabolic parameters related to obesity in Japanese obese patients with type 2 diabetes. *PeerJ* **3,** e942 (2015).
339. Rahman, M. & Berenson, A. B. Racial difference in lean mass distribution among reproductive-aged women. *Ethn. Dis.* **20,** 346–352 (2010).
340. Geer, E. B. *et al.* MRI assessment of lean and adipose tissue distribution in female patients with Cushing's disease. *Clin. Endocrinol. (Oxf).* **73,** 469–475 (2011).
341. Rodríguez, M. A. *et al.* Differences in adipose tissue and lean mass distribution in patients with collagen VI related myopathies are associated with disease severity and physical ability. *Front. Aging Neurosci.* **9,** 268 (2017).
342. Janssen, I. *et al.* Skeletal muscle mass and distribution in 468 men and women aged 18 − 88 yr. *J. Appl. Physiol.* **89,** 81–88 (2000).
343. Borga, M. *et al.* Advanced body composition assessment: from body mass index to body composition profiling. *J. Investig. Med.* **66,** 887–895 (2018).
344. Fox, C. S. *et al.* Genome-wide association for abdominal subcutaneous and visceral adipose reveals a novel locus for visceral fat in women. *PLoS Genet.* **8,** e1002695 (2012).
345. Fox, C. S. *et al.* Genome-wide association of pericardial fat identifies a unique locus for Ectopic fat. *PLoS Genet.* **8,** e1002705 (2012).

346. Chu, A. Y. *et al.* Multiethnic genome-wide meta-analysis of ectopic fat depots identifies loci associated with adipocyte development and differentiation. *Nat. Genet.* **49,** 125–130 (2017).
347. Stults-Kolehmainen, M. A. *et al.* DXA estimates of fat in abdominal, trunk and hip regions varies by ethnicity in men. *Nutr. Diabetes* **3,** 4–9 (2013).
348. The DAPA measurement Toolkit. Available at: http://dapa-toolkit.mrc.ac.uk. (Accessed: 22nd June 2018)
349. Purcell, S. PLINK 1.9. Available at: http://pngu.mgh.harvard.edu/purcell/plink/.
350. McQuillan, R. *et al.* Runs of Homozygosity in European Populations. *Am. J. Hum. Genet.* **83,** 359–372 (2008).
351. Miller, K. L. *et al.* Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.* **19,** 1523–1536 (2016).
352. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39,** 906–913 (2007).
353. Aulchenko, Y. S., Ripke, S., Isaacs, A. & van Duijn, C. M. GenABEL: An R library for genome-wide association analysis. *Bioinformatics* **23,** 1294–1296 (2007).
354. Haller, T., Kals, M., Esko, T., Mägi, R. & Fischer, K. RegScan: A GWAS tool for quick estimation of allele effects on continuous traits and their combinations. *Brief. Bioinform.* **16,** 39–44 (2013).
355. Staley, J. R. *et al.* PhenoScanner: A database of human genotype-phenotype associations. *Bioinformatics* **32,** 3207–3209 (2016).
356. Kolde, R. R package ' pheatmap '. (2018). doi:10.1371/journal.pone.0030126.Author(s)Maintainer
357. Jay, M. A. & Ren, J. Peroxisome proliferator-activated receptor (PPAR) in metabolic syndrome and type 2 diabetes mellitus. *Curr. Diabetes Rev.* **3,** 33–9 (2007).
358. Hegele, R. A., Cao, H., Frankowski, C., Mathews, S. T. & Leff, T. PPARG F388L, a tranactivation-deficient mutant, in familial partial lipodystrophy. *Diabetes* **51,** 3586–3590 (2002).
359. Altshuler, D. *et al.* The common PPARγ Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat. Genet.* **26,** 76–80 (2000).
360. Fadista, J., Manning, A. K., Florez, J. C. & Groop, L. The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *Eur. J. Hum. Genet.* **24,** 1202–1205 (2016).
361. Biobank, T. U. Access matter: representativeness of the UK Biobank resource. Available at: http://www.ukbiobank.ac.uk/wp-content/uploads/2017/03/access-matters-representativeness-1.pdf.
362. Day, F. R., Loh, P. R., Scott, R. a., Ong, K. K. & Perry, J. R. B. A Robust Example of Collider Bias in a Genetic Association Study. *Am. J. Hum. Genet.* **98,** 392–393 (2016).
363. Burgess, S., Bowden, J., Fall, T., Ingelsson, E. & Thompson, S. G. Sensitivity analyses for robust causal inference from mendelian randomization analyses with multiple genetic variants. *Epidemiology* **28,** 30–42 (2017).
364. Holmes, M. V., Ala-Korpela, M. & Smith, G. D. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. *Nat. Rev. Cardiol.* **14,** 577–590 (2017).
365. Relton, C. L. & Davey Smith, G. Mendelian randomization: Applications and limitations in epigenetic studies. *Epigenomics* **7,** 1239–1243 (2015).
366. Vanderweele, T. J., Tchetgen Tchetgen, E. J., Cornelis, M. & Kraft, P. Methodological challenges in mendelian randomization. *Epidemiology* **25,** 427–35 (2014).
367. Murray, P., Chune, G. W. & Raghavan, V. A. Legacy effects from DCCT and UKPDS: What they mean and implications for future diabetes trials. *Curr. Atheroscler. Rep.* **12,** 432–439 (2010).

368. Bhaskaran, K. *et al.* Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5·24 million UK adults. *Lancet* **384,** 755–765 (2014).
369. Guo, Y. *et al.* Genetically predicted body mass index and breast cancer risk: Mendelian randomization analyses of data from 145,000 women of European descent. *PLoS Med.* **13,** e1002105 (2016).
370. Benn, M., Tybjærg-Hansen, A., Smith, G. D. & Nordestgaard, G. B. High body mass index and cancer risk—a Mendelian randomisation study. *Eur. J. Epidemiol.* **31,** 879–892 (2016).
371. Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. *Nature* **538,** 161–164 (2016).
372. Hindorff, L. A. *et al.* Prioritizing diversity in human genomics research. *Nat. Rev. Genet.* **19,** 175–185 (2018).
373. Langenberg, C. & Lotta, L. A. Genomic insights into the causes of type 2 diabetes. *Lancet* **391,** 2463–2474 (2018).
374. Ramachandran, A., Wan Ma, R. C. & Snehalatha, C. Diabetes in Asia. *Lancet* **375,** 408–418 (2010).
375. Krieger, N. *Epidemiology and the People's Health: Theory and Context.* (Oxford University Press, 2011).
376. Marmot, M. Social determinants of health inequalities. *Lancet* **365,** 1099–1104 (2005).
377. Berkman, L. F., Kawachi, I. & Glymour, M. M. *Social Epidemiology.* (Oxford University Press, 2014).
378. Marmot, M. G. *The Health Gap - The Challenge of an Unequal World.* (Bloomsbury Publishing, 2015).
379. The Lancet. UK life science research: time to burst the biomedical bubble. *Lancet* **392,** 187 (2018).
380. Pearson-Stuttard, J. *et al.* Worldwide burden of cancer attributable to diabetes and high body-mass index: A comparative risk assessment. *Lancet Diabetes Endocrinol.* **6,** 95–104 (2017).
381. Harpsøe, M. C. *et al.* Body mass index and risk of autoimmune diseases: A study within the Danish National Birth Cohort. *Int. J. Epidemiol.* **43,** 843–855 (2014).
382. Crane, P. K. *et al.* Glucose levels and risk of dementia. *N Engl J Med* **369,** 540–548 (2013).
383. Arnold, S. E. *et al.* Brain insulin resistance in type 2 diabetes and Alzheimer disease: Concepts and conundrums. *Nat. Rev. Neurol.* **14,** 168–181 (2018).
384. Brown, D. G. *et al.* Metabolomics and metabolic pathway networks from human colorectal cancers, adjacent mucosa, and stool. *Cancer Metab.* **4,** 11 (2016).
385. Varma, V. R. *et al.* Brain and blood metabolite signatures of pathology and progression in Alzheimer disease: A targeted metabolomics study. *PLoS Med.* **15,** e1002482 (2018).
386. Jansson, J. *et al.* Metabolomics reveals metabolic biomarkers of Crohn's disease. *PLoS One* **4,** e6386 (2009).
387. Steinberg, G. R., Kemp, B. E. & Watt, M. J. Adipocyte triglyceride lipase expression in human obesity. *Am J Physiol Endocrinol Metab* **293,** E958–E964 (2007).
388. Sun, K. *et al.* Dicotomous effects of VEGF-A on adipose tissue dysfunction. *PNAS* **109,** 5874–5879 (2012).
389. Kahn, B. B. & Mcgraw, T. E. Rosiglitazone, PPARG, and type 2 diabetes. *N Engl J Med* **363,** 2667–2669 (2010).
390. Home, P. D. *et al.* Rosiglitazone evaluated for cardiovascular outcomes in oral agent combination therapy for type 2 diabetes (RECORD): a multicentre, randomised, open-label trial. *Lancet* **373,** 2125–2135 (2009).
391. Harris, T. B. *et al.* Age, gene/environment susceptibility-reykjavik study: Multidisciplinary applied phenomics. *Am. J. Epidemiol.* **165,** 1076–1087 (2007).
392. Elliott, P. *et al.* The airwave health monitoring study of police officers and staff in great britain: Rationale, design and methods. *Environ. Res.* **134,** 280–285 (2014).
393. Robbins, J., Hirsch, C., Whitmer, R., Cauley, J. & Harris, T. The association of bone mineral density and depression in an older population. *J. Am. Geriatr. Soc.* **49,** 732–736 (2001).

394. Larsen, T. M. *et al.* The diet, obesity and genes (diogenes) dietary study in eight European countries - A comprehensive design for long-term intervention. *Obes. Rev.* **11,** 76–91 (2010).
395. Forouhi, N. G., Luan, J., Hennings, S. & Wareham, N. J. Incidence of Type 2 diabetes in England and its association with baseline impaired fasting glucose: The Ely study 1990-2000. *Diabet. Med.* **24,** 200–207 (2007).
396. Lind, L. *et al.* EpiHealth: A large population-based cohort study for investigation of gene-lifestyle interactions in the pathogenesis of common diseases. *Eur. J. Epidemiol.* **28,** 189–197 (2013).
397. Henneman, P. *et al.* Prevalence and heritability of the metabolic syndrome and its individual components in a Dutch isolate: the Erasmus Rucphen Family study. *J. Med. Genet.* **45,** 572–577 (2008).
398. Higgins, M. *et al.* NHLBI Family Heart Study: objectives and design. *Am. J. Epidemiol* **143,** 1219–1228 (1996).
399. Visser, M. *et al.* Body fat and skeletal muscle mass in relation to physical disability in very old men and women of the Framingham Heart Study. *Journals Gerontol. Med. Sci.* **53,** M214–M221 (1998).
400. Borodulin, K. *et al.* Cohort Profile: The National FINRISK Study. *Int. J. Epidemiol.* **47,** 696–696i (2017).
401. Harris, T. B. *et al.* Waist circumference and sagittal diameter reflect total body fat better than visceral fat in older men and women. The Health, Aging and Body Composition Study. *Ann. N. Y. Acad. Sci.* **904,** 462–73 (2000).
402. Thuesen, B. H. *et al.* Cohort profile: The Health2006 cohort, research centre for prevention and health. *Int. J. Epidemiol.* **43,** 568–575 (2014).
403. Wichmann, H. E., Gieger, C. & Illig, T. KORA-gen - Resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* **67,** S26–S30 (2005).
404. Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42,** 937–948 (2010).
405. Kooner, J. S. *et al.* Genome-wide scan identifies variation in MLXIPL associated with plasma triglycerides. *Nat. Genet.* **40,** 149–151 (2008).
406. Stancáková, A. *et al.* Changes in Insulin Sensitivity and Insulin Release in Relation to Glycemia and Glucose Tolerance in 6,414 Finnish Men. *Diabetes* **58,** 1212–1221 (2009).
407. De Mutsert, R. *et al.* The Netherlands epidemiology of obesity (NEO) study: Study design and data collection. *Eur. J. Epidemiol.* **28,** 513–523 (2013).
408. Lind, L., Fors, N., Hall, J., Marttala, K. & Stenborg, A. A comparison of three different methods to determine arterial compliance in the elderly: The Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) study. *J. Hypertens.* **25,** 2368–2375 (2005).
409. Straker, L. *et al.* Cohort Profile: The Western Australian Pregnancy Cohort (Raine) Study–Generation 2. *Int. J. Epidemiol.* **46,** 1384–1385j (2017).
410. Hills, S. A. *et al.* The EGIR-RISC Study (the European group for the study of insulin resistance: Relationship between insulin sensitivity and cardiovascular disease risk): I. Methodology and Objectives. *Diabetologia* **47,** 566–570 (2004).
411. Hofman, A. *et al.* The rotterdam study: 2014 objectives and design update. *Eur. J. Epidemiol.* **28,** 889–926 (2013).
412. Völzke, H. *et al.* Cohort profile: The study of health in Pomerania. *Int. J. Epidemiol.* **40,** 294–307 (2011).
413. Tönjes, A. *et al.* Association of FTO variants with BMI and fat mass in the self-contained population of Sorbs in Germany. *Eur. J. Hum. Genet.* **18,** 104–110 (2010).
414. Moayyeri, A., Hammond, C. J., Valdes, A. M. & Spector, T. D. Cohort profile: Twinsuk and healthy ageing

twin study. *Int. J. Epidemiol.* **42,** 76–85 (2013).

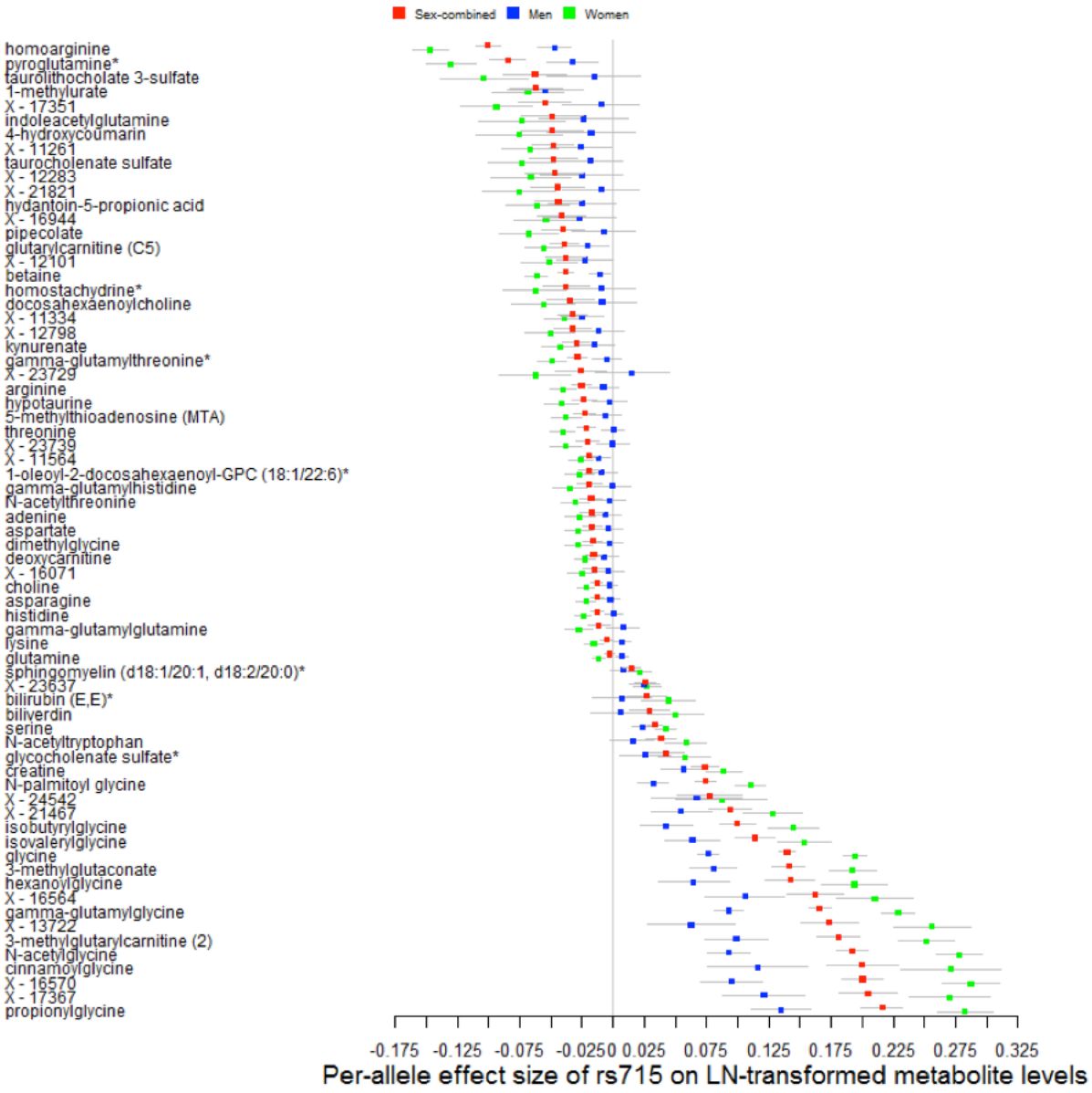
415. Petersen, E. R. B. *et al.* Vejle diabetes Biobank – A resource for studies of the etiologies of diabetes and its comorbidities. *Clin. Epidemiol.* **8,** 393–413 (2016).
416. Carty, C. L. *et al.* Genome-wide association study of body height in African Americans: The Women's Health Initiative SNP Health Association Resource (SHARe). *Hum. Mol. Genet.* **21,** 711–720 (2012).
417. Kan, M. *et al.* Rare variant associations with waist-to-hip ratio in European-American and African-American women from the NHLBI-Exome Sequencing Project. *Eur. J. Hum. Genet.* **24,** 1181–1187 (2016).
418. Raitakari, O. T. *et al.* Cohort profile: The cardiovascular risk in young Finns study. *Int. J. Epidemiol.* **37,** 1220–1226 (2008).

Supplementary Materials

Supplementary Figures

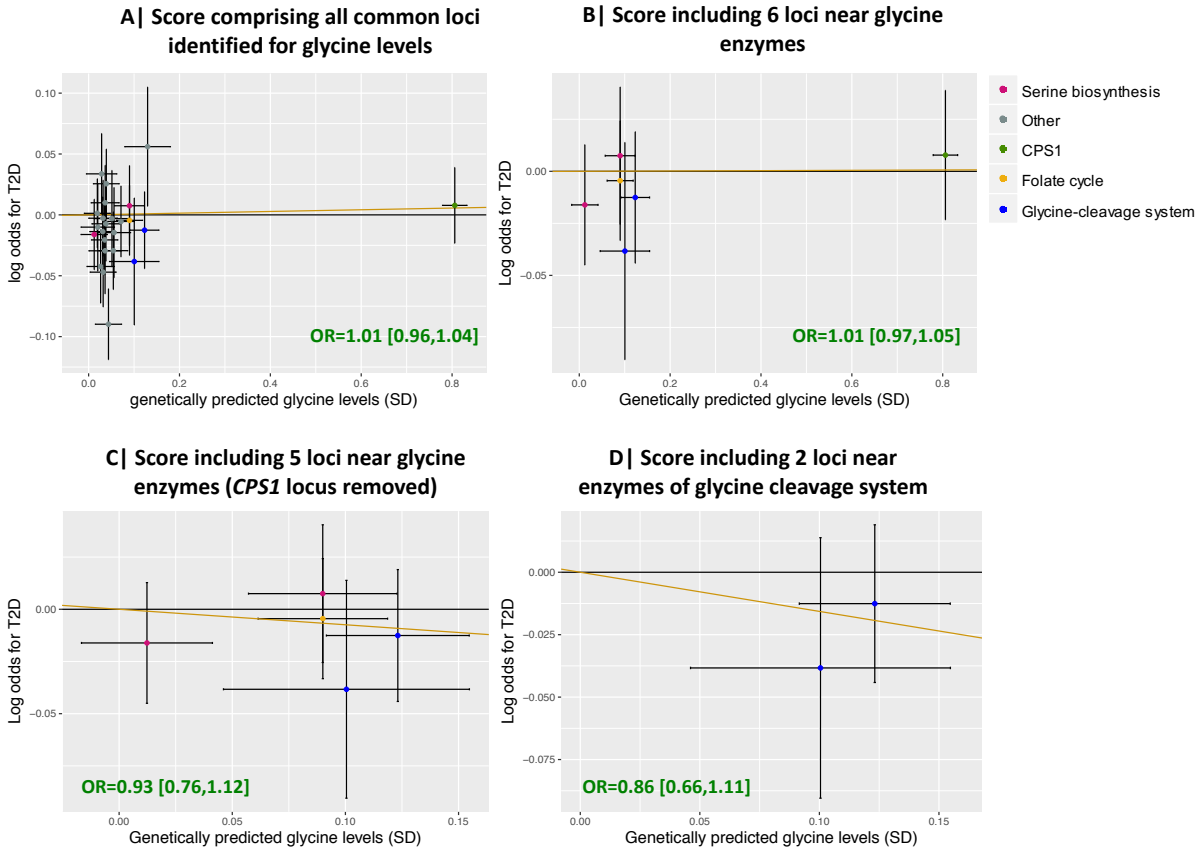
Chapter 2

Supplementary Figure 2.1: Sex-combined and sex-specific per-allele effect sizes of rs715 (CPS1) on metabolite levels.

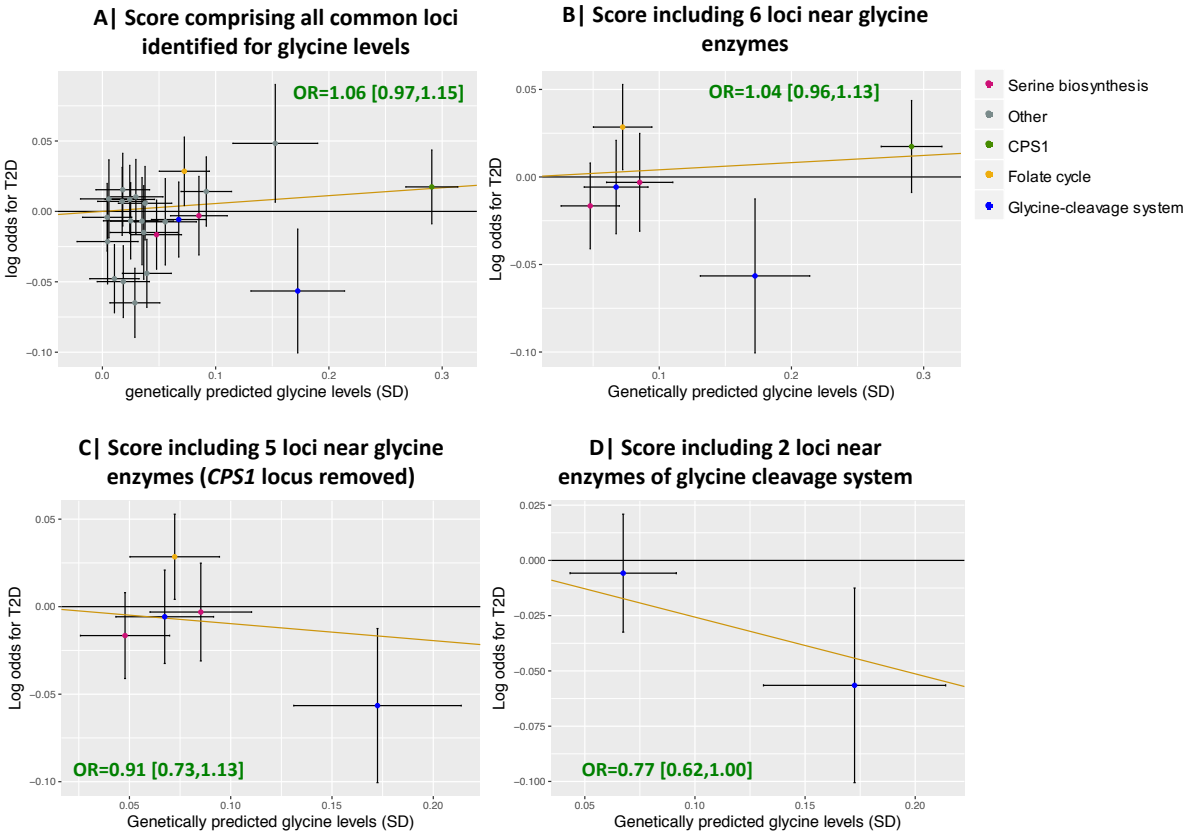


Analyses based on 5,706 women and 5,086 men of the EPIC-Norfolk study. Metabolites were included in the plot if they were associated with rs715 at $p<5.6\times10^{-4}$ in men or women, i.e., 69 out of 894 metabolites measured in both batches and at least 50% of the total sample size.

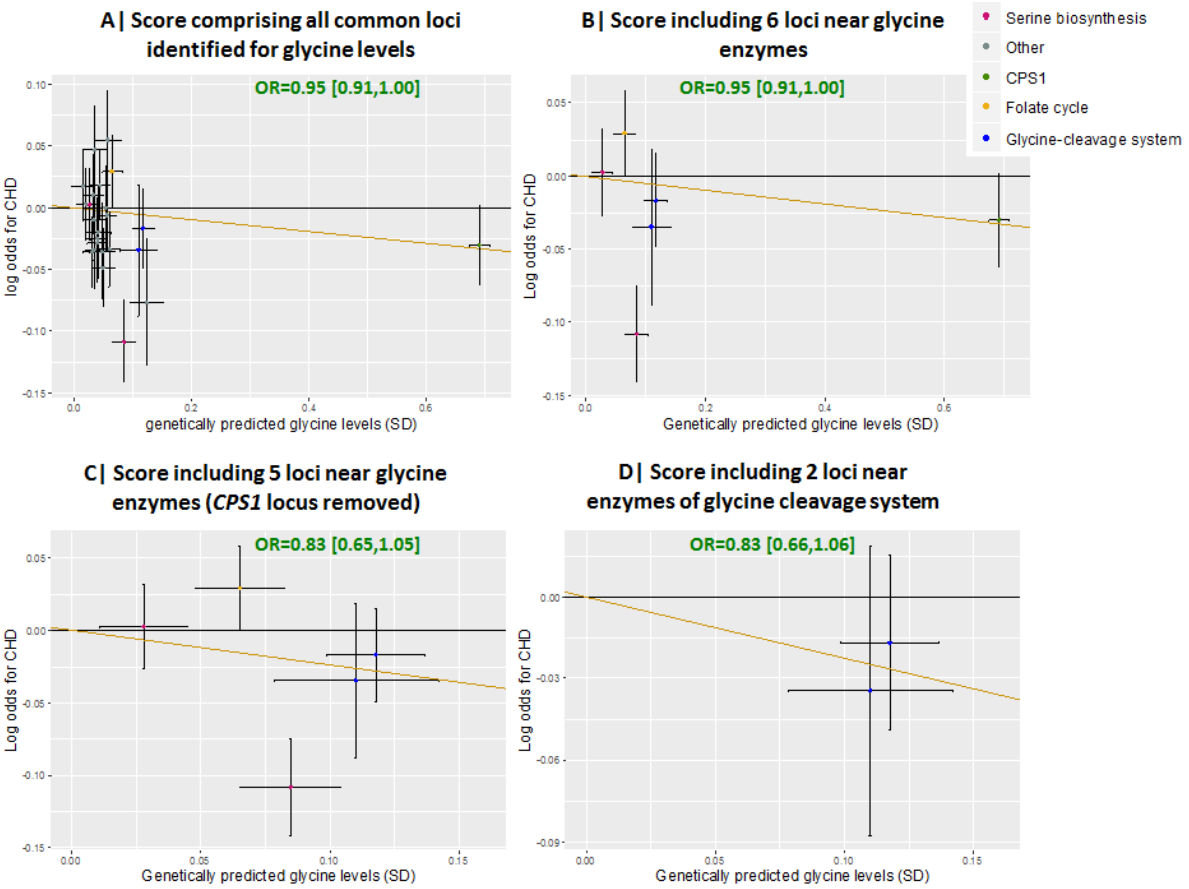
Supplementary Figure 2.2: Dosage plots of the effect sizes of genetic variants for glycine on standard deviations of glycine levels versus the log odds for type 2 diabetes for women-only analyses



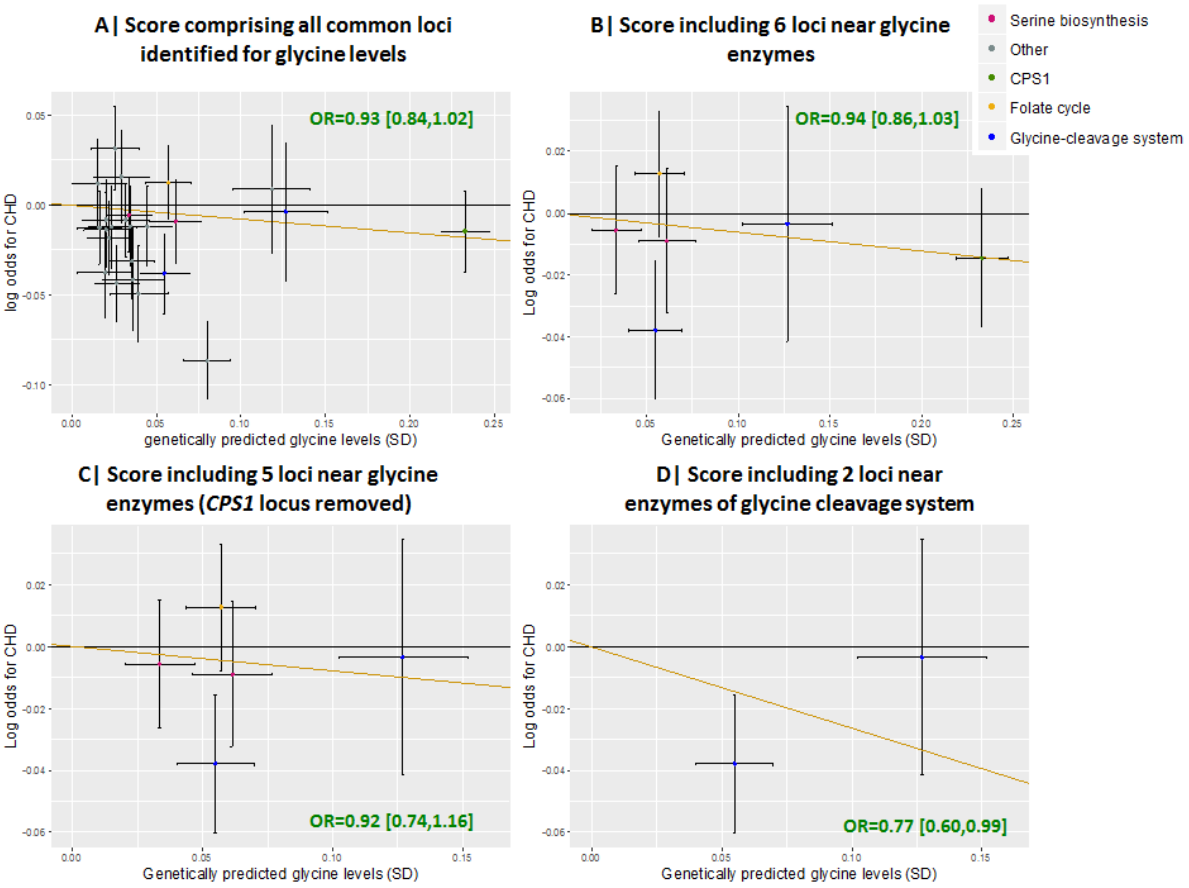
Supplementary Figure 2.3: Dosage plots of the effect sizes of genetic variants for glycine on standard deviations of glycine levels versus the log odds for type 2 diabetes for men-only analyses



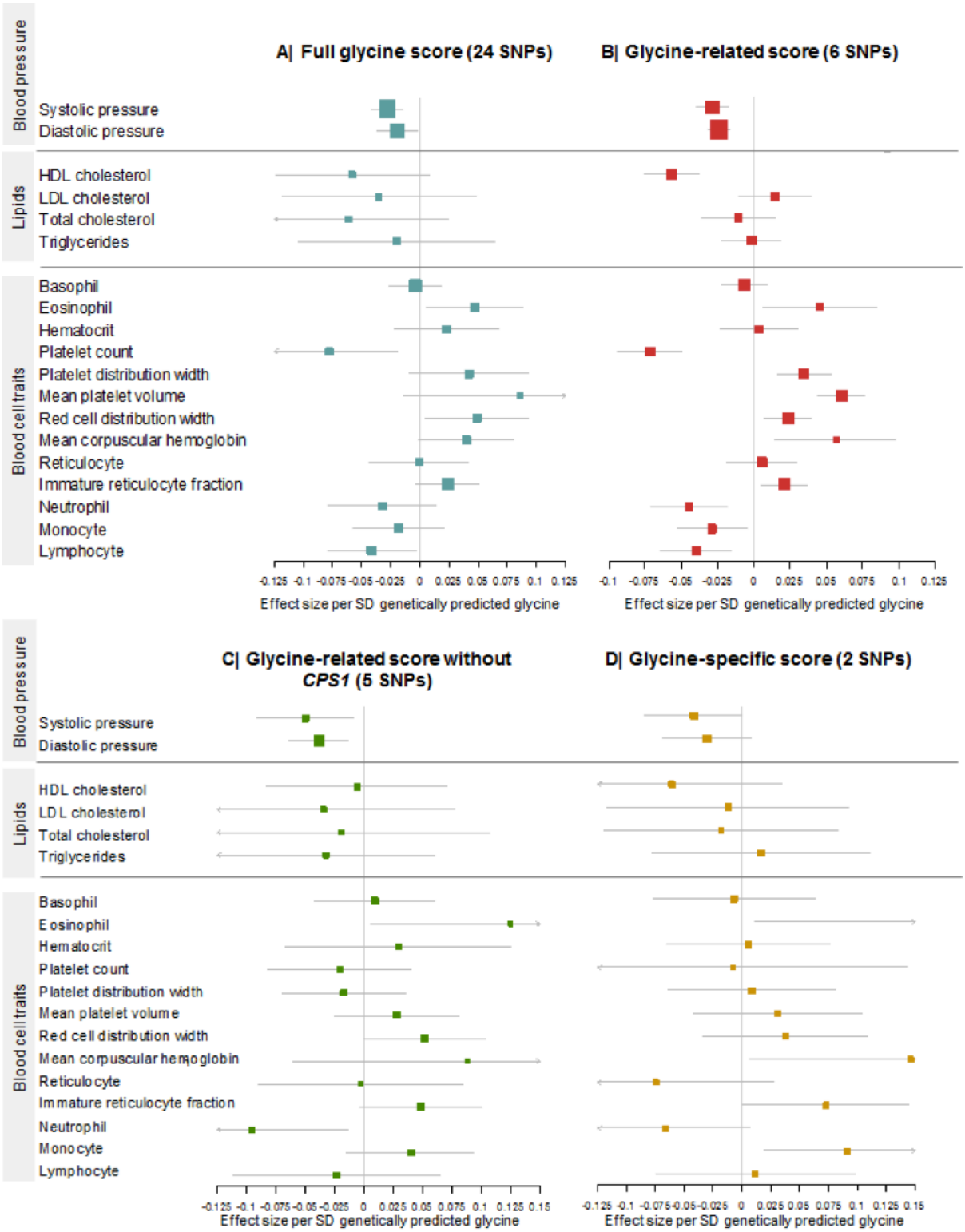
Supplementary Figure 2.4: Dosage plots of the effect sizes of genetic variants for glycine on standard deviations of glycine levels versus the log odds for coronary heart disease for women-only analyses



Supplementary Figure 2.5: Dosage plots of the effect sizes of genetic variants for glycine on standard deviations of glycine levels versus the log odds for coronary heart disease for men-only analyses

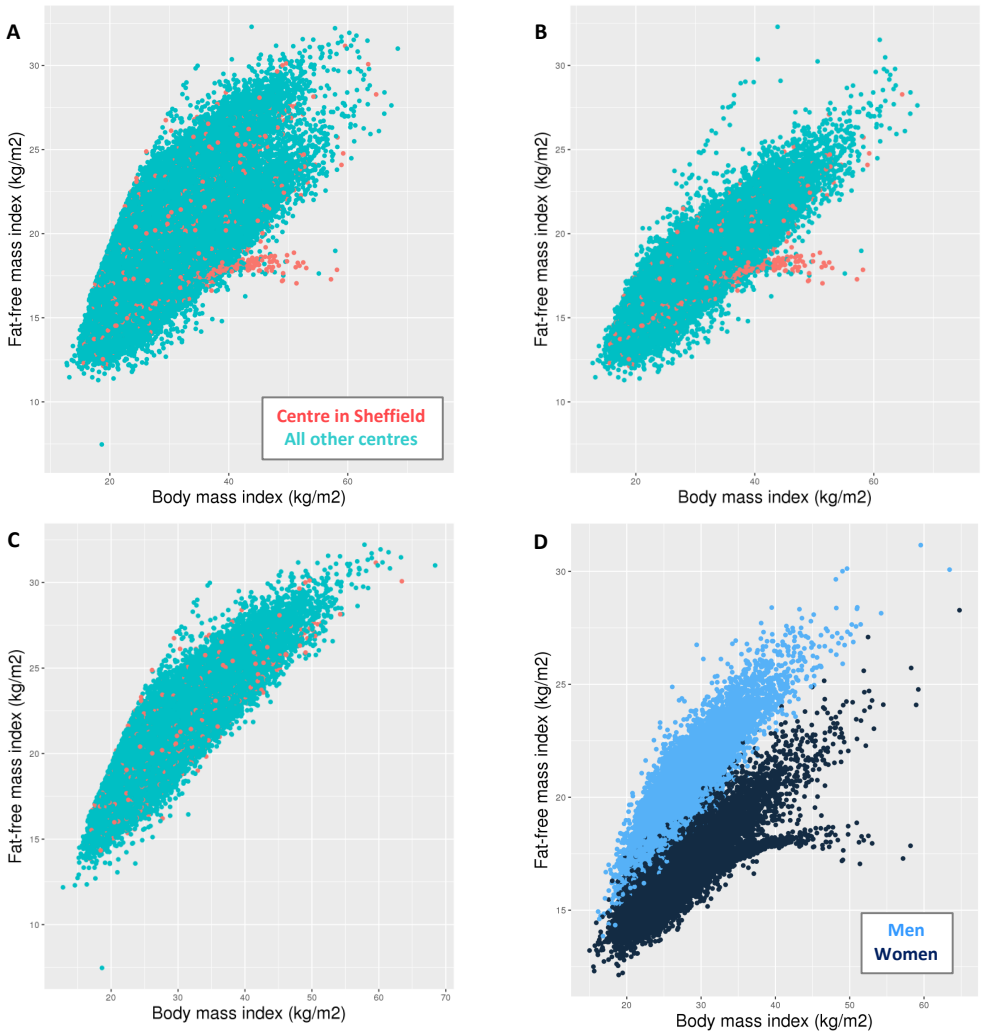


Supplementary Figure 2.6: Forest plots showing the effect sizes of genetically predicted glycine levels on 19 risk factors of coronary heart disease, using 4 genetic instruments for glycine

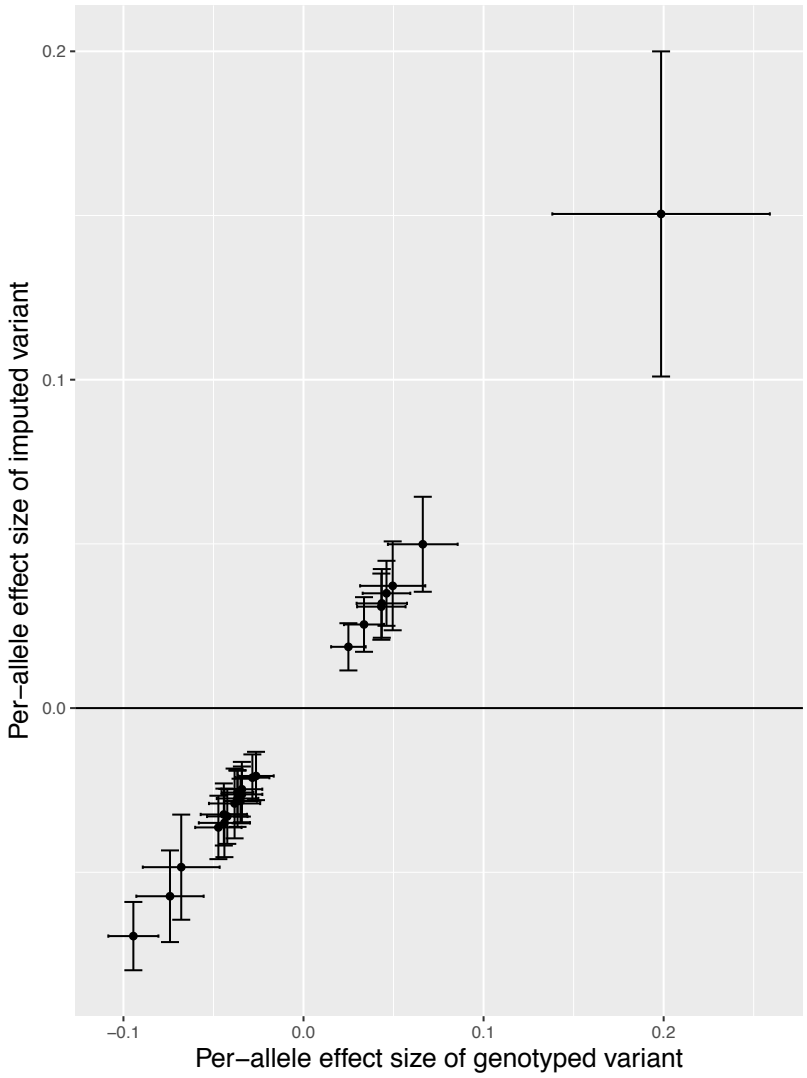


Chapter 3

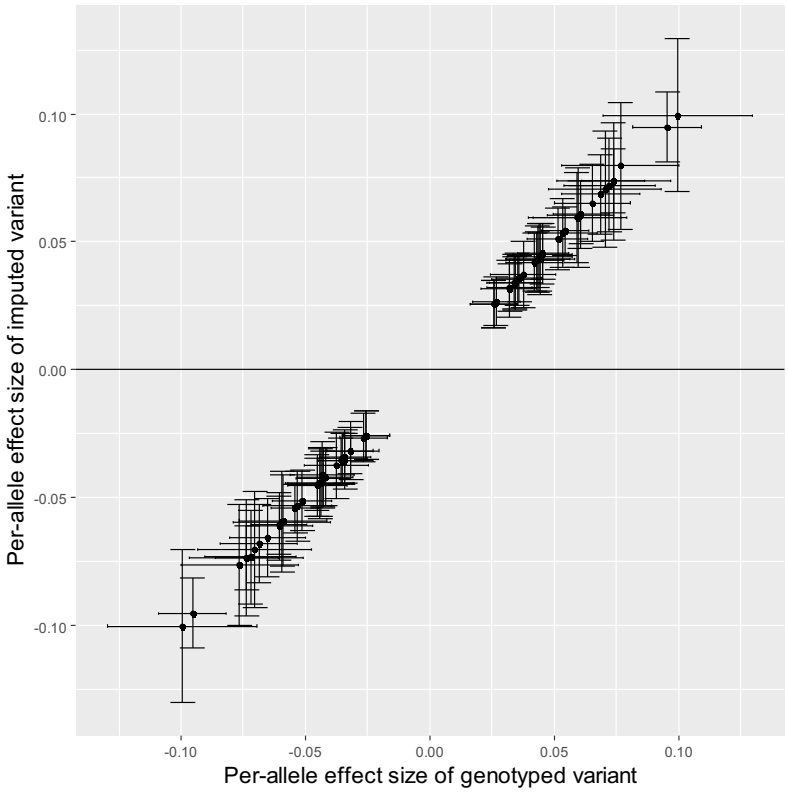
Supplementary Figure 3.1: Scatterplots of BMI versus BF% based on all assessment centres of the UK Biobank (A, B and C) and for Sheffield alone (D)



Supplementary Figure 3.2: Effect sizes of the low-frequency coding variants reaching significance for BF%: Imputed versus genotyped variants



Supplementary Figure 3.3: Effect sizes of the low-frequency coding variants reaching significance for FFMI: Imputed versus genotyped variants



Supplementary Tables

Chapter 1

Supplementary Table 1.1: Search terms for systematic literature search on metabolite patterns associated with T2D

	Search Terms	Number of Titles
Observational epidemiological studies #1	"Case-Control Studies"[MeSH] OR Case-Control Study OR Case-Control Studies OR Case-control Study OR Case-control Studies OR Studies, Case-Control OR Study, Case-Control OR Case-Comparison Studies OR Case Comparison Studies OR Case-Comparison Study OR Studies, Case-Comparison OR Study, Case- Comparison OR Case-Referent Studies OR Case Referent Studies OR Case- Referent Study OR Studies, Case-Referent OR Study, Case-Referent OR Case- Referent Studies OR Case Referent Studies OR Case-Referent Study OR Studies, Case-Referent OR Study, Case-Referent OR Case Control Studies OR Case Control Study OR Studies, Case Control OR Study, Case Control OR Matched Case-Control Studies OR Case-Control Studies, Matched OR Case-Control Study, Matched OR Matched Case Control Studies OR Matched Case-Control Study OR Studies, Matched Case-Control OR Study, Matched Case-Control OR Nested Case-Control Studies OR Case-Control Studies, Nested OR Case-Control Study, Nested OR Nested Case Control Studies OR Nested Case-Control Study OR Studies, Nested Case-Control OR Study, Nested Case-Control OR "Cross-Sectional Studies"[MeSH] OR Cross-sectional Studies OR Cross Sectional Studies OR Cross Sectional Study OR Cross-Sectional Study OR Studies, Cross-Sectional OR Study, Cross-Sectional OR Cross Sectional Analysis OR Analyses, Cross Sectional OR Cross Sectional Analyses OR Disease Frequency Surveys OR Surveys, Disease Frequency OR Disease Frequency Survey OR Survey, Disease Frequency OR Analysis, Cross- Sectional OR Analyses, Cross-Sectional OR Analysis, Cross Sectional OR Cross- Sectional Analyses OR Cross-Sectional Analysis OR Cross-Sectional Survey OR Cross Sectional Survey OR Cross-Sectional Surveys OR Survey, Cross-Sectional OR Surveys, Cross-Sectional OR Prevalence Studies OR Prevalence Study OR Studies, Prevalence OR Study, Prevalence OR "Cohort Studies"[Mesh] OR Cohort Study OR Studies, Cohort OR Study, Cohort OR Studies, Historical Cohort OR Cohort Studies, Historical Cohort Study, Historical OR Historical Cohort Study OR Study, Historical OR Analysis, Cohort OR Analyses,	3,837,948

	Search Terms	**Number of Titles**
	Cohort OR Cohort Analyses OR Cohort Analysis OR Incidence Studies OR Incidence Study OR Studies, Incidence OR Study, Incidence OR "Incidence"[Mesh] OR "Cohort Studies"[Mesh] OR Case- cohort study OR Nested case-control study OR Survey OR cross-sectional OR case- control	
Type 2 Diabetes #2	"Diabetes mellitus, Type 2"[Mesh] OR NIDDM OR Maturity-Onset Diabetes OR Diabetes Mellitus, Noninsulin-Dependent OR Diabetes Mellitus, Adult-Onset OR Adult-Onset Diabetes Mellitus OR Diabetes Mellitus, Adult Onset OR Diabetes Mellitus, Maturity-Onset OR Diabetes Mellitus, Maturity Onset OR Diabetes Mellitus, Non-Insulin Dependent OR Diabetes Mellitus, Non-Insulin-Dependent OR Non-Insulin-Dependent Diabetes Mellitus OR Diabetes Mellitus, Noninsulin Dependent OR Diabetes Mellitus, Slow-Onset OR Diabetes Mellitus, Slow Onset OR Slow-Onset Diabetes Mellitus OR Diabetes Mellitus, Stable OR Stable Diabetes Mellitus OR Diabetes Mellitus, Type II OR Maturity-Onset Diabetes Mellitus OR Maturity Onset Diabetes Mellitus OR MODY OR Type 2 Diabetes Mellitus OR Noninsulin-Dependent Diabetes Mellitus OR "Type 2 Diabetes"[tiab] OR "Diabetes"[ti]	289,004
Metabolomics #3	metabolomics OR metabolome OR metabonomics OR metabolic profiling OR metabolite profiling OR metabolites OR metabolome OR lipid profiling OR amino acid profiling OR lipidomics OR metabolite profiles OR metabolite profile OR "mass spectrometry" [MeSH] OR mass spectrometry OR NMR OR nuclear magnetic resonance OR "NMR spectroscopy" [MeSH] OR LC-MS OR GC-MS	1,190,288
	#1 AND #2 AND #3	**1,805**

Chapter 2

Supplementary Table 2.1: Results of approximate conditional analyses: stepwise selection model to identify additional independent signals for glycine levels

Locus	rsid	chr:pos	EA	OA	EAF	Nearest gene	Type	$Z_{marginal}$	$p_{marginal}$	Z_{joint}	P_{joint}
1	rs4646961	1:76217169	A	G	0.297	*ACADM*	intronic	8.854	8.4E-19	8.854	9.1E-19
2	rs561931	1:120254506	G	A	0.593	*PHGDH*	UTR5	7.478	7.6E-14	7.478	7.9E-14
3	rs10184004	2:165508389	T	C	0.400	*COBLL1*	intergenic	6.041	1.5E-09	6.041	1.6E-09
4	rs1861072	2:210851634	T	C	0.591	*UNC80*	intronic	3.932	8.4E-05	7.250	2.8E-12
4	rs62202111	2:211285782	C	A	0.024	*LANCL1-AS1*	ncRNA intronic	13.236	5.4E-40	6.496	6.4E-10
4	rs1509820	2:211386484	A	G	0.530	*CPS1*	intronic	11.194	4.4E-29	29.075	2.1E-52
4	rs715	2:211543055	C	T	0.313	*CPS1*	UTR3	86.626	######	36.940	1.9E-105
4	rs62177991	2:211884231	G	A	0.517	*CPS1*	intergenic	8.318	9.0E-17	7.506	6.9E-14
4	rs115683961	2:212385920	A	G	0.938	*ERBB4*	intronic	3.89	1.0E-04	5.571	2.8E-08
5	rs141015398	3:125618725	A	G	0.989	*LOC101927056*	intergenic	5.558	2.7E-08	5.530	3.2E-08
5	rs2276724	3:125854409	T	C	0.841	*ALDH1L1*	exonic	6.554	5.6E-11	7.162	1.6E-11
5	rs4646705	3:125875831	A	G	0.497	*ALDH1L1*	intronic	6.187	6.1E-10	5.920	2.3E-08
5	rs9862438	3:125910381	T	C	0.416	*ALDH1L1-AS2*	ncRNA intronic	11.514	1.1E-30	13.317	6.3E-38
5	rs79064296	3:125985069	T	G	0.045	*ALDH1L1-AS2*	intergenic	8.236	1.8E-16	6.448	1.6E-10
6	rs148685782	4:155533035	G	C	0.996	*FGG*	exonic	Rare variant: not included in conditional analyses			
7	rs1912826	4:187149540	A	G	0.522	*KLKB1*	intronic	6.175	6.6E-10	6.175	6.8E-10
8	rs156380	5:53378450	C	T	0.807	*ARL15*	intronic	5.47	4.5E-08	5.470	4.5E-08
9	rs3105793	5:90226061	A	G	0.273	*ADGRV1*	intronic	5.489	4.0E-08	5.489	4.1E-08
10	rs10900807	5:131757480	G	C	0.805	*C5orf56*	ncRNA intronic	6.073	1.3E-09	6.073	1.3E-09
11	rs2545801	5:176841339	C	T	0.747	*F12*	intergenic	7.484	7.2E-14	7.484	7.5E-14
12	rs543159	6:160776017	A	C	0.482	*SLC22A3*	intronic	6.246	4.2E-10	6.246	4.3E-10
13	rs4689	7:56067016	T	G	0.771	*NIPSNAP2*	UTR3	11.743	7.7E-32	8.236	3.2E-12
13	rs62457261	7:56080426	G	C	0.570	*PSPH*	intronic	10.98	4.8E-28	6.604	2.3E-08
14	rs7012637	8:9173209	G	A	0.525	*LOC157273*	intergenic	10.387	2.9E-25	8.088	1.8E-15
14	rs9987289	8:9183358	A	G	0.100	*LOC157273*	ncRNA intronic	14.788	1.7E-49	13.402	1.0E-39
15	rs55748921	8:126492927	C	A	0.693	*TRIB1*	intergenic	6.542	6.1E-11	6.646	3.1E-11
15	rs28601761	8:126500031	G	C	0.416	*TRIB1*	intergenic	11.338	8.5E-30	11.399	5.4E-30
16	rs13299380	9:6078779	T	C	0.053	*RANBP6*	intergenic	8.276	1.3E-16	6.662	3.6E-11
16	rs113626350	9:6490491	A	G	0.046	*UHRF2*	intronic	9.359	8.1E-21	6.976	5.4E-12
16	rs17591030	9:6550024	C	T	0.715	*GLDC*	intronic	13.316	1.9E-40	11.612	2.7E-30
16	rs143119940	9:6554781	A	C	0.011	*GLDC*	exonic	10.537	5.8E-26	9.140	1.4E-19
16	rs67506027	9:6657690	A	C	0.607	*GLDC*	intergenic	8.571	1.0E-17	6.792	1.9E-11
17	rs676996	9:136146077	T	G	0.668	*ABO*	intronic	7.843	4.4E-15	7.843	4.6E-15
18	rs190595610	10:32274880	A	G	0.997	*KIF5B*	intergenic	Rare variant: not included in conditional analyses			
19	rs6479877	10:64878118	C	G	0.532	*NRBF2*	intergenic	3.519	4.3E-04	-14.647	1.0E-10

Locus	rsid	chr:pos	EA	OA	EAF	Nearest gene	Type	$Z_{marginal}$	$p_{marginal}$	Z_{joint}	P_{joint}
19	rs10740134	10:65315433	T	C	0.515	*REEP3*	intronic	7.108	1.2E-12	20.257	4.0E-19
20	rs12297321	12:47109387	T	C	0.152	*SLC38A4*	intergenic	7.172	7.4E-13	7.238	4.6E-13
21	rs2638314	12:56866334	A	T	0.182	*GLS2*	intronic	5.659	1.5E-08	5.743	9.4E-09
22	rs9514191	13:104520138	C	G	0.312	*LINC01309*	intergenic	5.536	3.1E-08	5.536	3.1E-08
23	rs201393666	15:43677979	A	C	0.029	*TUBGCP4*	intronic	Did not reach significance in conditional analyses			
24	rs2280195	15:58467095	A	G	0.441	*AQP9*	intronic	5.923	3.2E-09	5.923	3.2E-09
25	rs9923732	16:81110903	A	G	0.914	*C16orf46*	upstream	13.518	1.2E-41	11.615	2.3E-28
25	rs4520846	16:81142358	A	C	0.353	*PKD1L2*	intronic	2.586	9.7E-03	8.070	2.6E-08
25	rs11647428	16:81150559	G	A	0.490	*PKD1L2*	intronic	8.398	4.5E-17	22.827	1.5E-28
25	rs62051533	16:81152279	C	A	0.323	*PKD1L2*	intronic	3.416	6.4E-04	-12.816	3.0E-11
25	rs16954688	16:81156808	G	A	0.953	*PKD1L2*	intronic	8.602	7.8E-18	7.724	4.7E-14
26	rs8078686	17:45735706	C	T	0.509	*KPNB1*	intronic	6.617	3.7E-11	6.617	3.8E-11
27	rs273510	19:18223350	A	G	0.708	*MAST3*	intronic	5.903	3.6E-09	5.903	3.6E-09

EA: effect allele, OA: other allele, EAF: effect allele frequency, chr:pos: chromosome and base pair position according to genome build GRCh37, $Z_{marginal}$: Z score from p-value-based meta-analysis of 5 studies, $p_{marginal}$: p-value from p-value-based meta-analyses of 5 studies, Z_{joint}: Z from the joint model fitted by GCTA-COJO, p_{joint}: p-value from the joint model fitted by GCTA-COJO.

Supplementary Table 2.2: Reported associations of glycine loci with other traits and metabolites

Locus	rsid	Traits associated with variant	Traits associated with variants in high LD (R2>0.6)	Reported metabolite associations for lead variant	Reported metabolite associations for variants in high LD (R2>0.8)	Reported metabolite associations for variants within 1 MB of lead variant
1	rs4646961			hexanoylcarnitine [24816252]; octanoylcarnitine[24816252]; cis-4-decenoyl carnitine [24816252]; decanoylcarnitine [24816252]; X-23293[28263315]; X-12824 [28263315]; X-12855 [24816252]	hexanoylcarnitine [24816252]; octanoylcarnitine [24816252]; cis-4-decenoyl carnitine [24816252]; decanoylcarnitine [24816252]; Decenoylcarnitine [26068415]; Acylcarnitine levels [26068415]; X-23293 [28263315]; C6(C4:1-DC) [20037589]; C8 [20037589]; C10 [20037589]; X-12824 [28263315]; hexanoylglycine [28263315]; caprylate (8:0) [28263315]; X-12855 [24816252]; X-17327 [28263315]; X-12101 [28263315] ; X-12472 [28263315];	hexanoylcarnitine [24816252]; octanoylcarnitine [24816252]; cis-4-decenoyl carnitine [24816252]; Decenoylcarnitine [26068415]; Decanoylcarnitine [26068415]; Acylcarnitine levels [26068415]; X-23293 [28263315]; C6(C4:1-DC) [20037589]; Serum concentration of decanoylcarnitine [21886157]; C8 [20037589]; C10 [20037589]; X-12824 [28263315]; hexanoylglycine [28263315]; C10:1 [20037589]; LDL cholesterol [23063622]; caprylate (8:0) [28263315]; X-12855 [24816252]; X-17327 [28263315]; X-12101 [28263315]; malonate [28263315];
2	rs561931			Serine[26068415]	serine [24816252]	serine[24816252]; alpha-hydroxyisovalerate [24816252]; 2-hydroxy-3-methylvalerate [28263315]
3	rs10184004	Type II diabetes (-) [26551672]; hip (+) [25673412]; waist/hip (-) [25673412]; triglycerides (-) [24097068];	HDL (+) [24097068]; LDL (-) [24097068]; fasting insulin (-) [22885924]			1-palmitoylglycerophosphocholine [21886157]; HDL cholesterol lipoprotein fraction concentration [19936222]
4	rs715	eGFR creatinine (-) [26831199], fibrinogen (-) [23969696], serum creatinine [20383146],	HDL (-) [24097068]	glycine [24816252]; N-acetylglycine[24816252]; X-16570 [28263315]; propionylglycine [28263315]; N-palmitoylglycine [28263315]; X-08988 [24816252];	glycine [24816252]; N-acetylglycine [28263315]; propionylglycine [28263315]; X-16570 [28263315]; N-palmitoylglycine [28263315]; Serum glycine (females) [21852955]; 3-methylglutaconate	glycine [24816252]; nonanoylcarnitine [24816252]; Nonaylcarnitine [26068415]; N-acetylglycine [28263315]; propionylglycine [28263315]; X-16570 [28263315]; N-palmitoylglycine [28263315]; 3-methylglutaconate [28263315]; hexanoylglycine

Locus	rsid	Traits associated with variant	Traits associated with variants in high LD (R2>0.6)	Reported metabolite associations for lead variant	Reported metabolite associations for variants in high LD (R2>0.8)	Reported metabolite associations for variants within 1 MB of lead variant
		chronic kidney disease (-) [26831199]		hexanoylglycine [28263315]; 3-methylglutaconate [28263315]; homoarginine [28263315]; cinnamoylglycine [28263315]; 3-methylglutarylcarnitine[28263315]; creatine [24816252]; serine [24816252]; betaine [24816252]; gamma-glutamylglycine [28263315]; creatine (urine) [26352407]; pyroglutamine* [24816252]; glutaroyl carnitine [24816252]; guanidinosuccinate [28263315]; creatinine [20383146];glutarylcarnitine (C5) [28263315]; gamma-glutamylthreonine [28263315]	[28263315]; hexanoylglycine [28263315]; X-08988 [24816252]; 3-methylglutarylcarnitine (2) [28263315]; homoarginine [28263315]; cinnamoylglycine [28263315]; creatine [24816252]; serine [24816252]; glycine (urine) [26352407]; betaine [24816252]; gamma-glutamylglycine [28263315]; creatine (urine) [26352407]; pyroglutamine [24816252]; glutaroyl carnitine [24816252]; guanidinosuccinate [28263315]; Serum creatinine [20383146]; HDL cholesterol [24097068]; Serum glycine (males) [21852955]; gamma-glutamylthreonine [28263315]; glutarylcarnitine (C5) [28263315)	[28263315]; X-08988 [24816252]; 3-methylglutarylcarnitine (2) [28263315]; homoarginine [28263315]; cinnamoylglycine [28263315]; Serum 2,6-dimethylheptanoylcarnitine [23281178]; serine [24816252]; glycine (urine) [26352407]; betaine [24816252]; gamma-glutamylglycine [28263315]; creatine (urine) 26352407]; pyroglutamine [24816252]; glutaroyl carnitine [24816252]; Serine [26068415]; Serum creatinine [20383146]; HDL cholesterol [24097068]; pyroglutamine [28263315]; gamma-glutamylthreonine [28263315]; Serum concentration of creatine [21886157]; citrulline [24816252]; glutarylcarnitine (C5) [28263315]
5	rs9862438					Glycine [26068415)
6	rs148685782	Fibrinogen (-) [28107422]		X-21628 [28263315]	X-21628 [28263315]	Bilirubin levels [20639394]; HDL cholesterol [23063622]
7	rs71640034		Activated thromboplastin time (-) [22703881]		prolylproline [28263315]; glycylphenylalanine [28263315]; X-11792 [24816252]; isoleucylvaline [28263315]; Serum His [22286219]; HWESASXX [28263315]; X-24071 [28263315)	prolylproline [28263315]; glycylphenylalanine [28263315]; X-11792 [24816252]; isoleucylvaline [28263315]; ADSGEGDFXAEGGGVR [24816252];Serum His [22286219]; HWESASXX [28263315]; X-24071 [28263315]
8	rs156380					citrulline [24816252]
9	rs3105793					Formate:N,N-Dimethylglycine [21572414]

Locus	rsid	Traits associated with variant	Traits associated with variants in high LD (R2>0.6)	Reported metabolite associations for lead variant	Reported metabolite associations for variants in high LD (R2>0.8)	Reported metabolite associations for variants within 1 MB of lead variant
10	rs10900807	height (+) [25282103], fibrinogen (+) [20031576]		isovalerylcarnitine [24816252]; 3-dehydrocarnitine [24816252]; Octadecadienylcarnitine [26068415]; Octadecenoylcarnitine [26068415]	isovalerylcarnitine [24816252]; 3-dehydrocarnitine [24816252]; Octadecadienylcarnitine [26068415]; Octadecenoylcarnitine [26068415]; glutaroyl carnitine [24816252]; palmitoylcarnitine [24816252]	isovalerylcarnitine [24816252]; tryptophan betaine [24816252]; 3-dehydrocarnitine [24816252]; Octadecadienylcarnitine [26068415]; carnitine [24816252]; Acylcarnitine levels [26068415]; propionylcarnitine [24816252]; Serum concentration of Isovalerylcarnitine [21886157]; Octadecenoylcarnitine [26068415]; Acetylcarnitine [26068415]; Serum concentration of 3-dehydrocarnitine [21886157]; tryptophan betaine [28263315]; hexanoylcarnitine [24816252]; oleoylcarnitine [24816252]; homostachydrine [24816252]; myristoylcarnitine [28263315]; stearoylcarnitine [24816252]; X-21365 [28263315]; palmitoylcarnitine [24816252]; serum concentration of propionylcarnitine [21886157]; 2-tetradecenoyl carnitine [24816252]; hydroxyisovaleroyl carnitine [24816252]; glutaroyl carnitine [24816252]; butyrylcarnitine [24816252];linoleoylcarnitine [28263315]
11	rs2545801	activated thromboplastin time (-) [22703881], C-terminal pro-endothelin-1 (-) [23381795], midregional proadrenomedulin (-) [23381795]		prolylproline [28263315]; glycylphenylalanine [28263315]; X-24071 [28263315]; isoleucylvaline [28263315]; Serum Phe [22286219]; ADSGEGDFXAEGGGVR [24816252]	leucylphenylalanine/isoleucylphenylalanine [28263315]; prolylproline [28263315]; glycylphenylalanine [28263315]; X-24071 [28263315]; isoleucylvaline [28263315]; ADSGEGDFXAEGGGVR [24816252]; Serum Phe [22286219]; X-11792 [24816252]	leucylphenylalanine/isoleucylphenylalanine [28263315]; prolylproline [28263315]; glycylphenylalanine [28263315]; X-24071 [28263315]; isoleucylvaline [28263315]; Serum creatinine [20383146]; ADSGEGDFXAEGGGVR [24816252]; 3-methylhistidine [28263315]; Serum Phe (p=4.56e-11 [22286219]; X-14838 [28263315]; X-11792 [24816252]

Locus	rsid	Traits associated with variant	Traits associated with variants in high LD (R2>0.6)	Reported metabolite associations for lead variant	Reported metabolite associations for variants in high LD (R2>0.8)	Reported metabolite associations for variants within 1 MB of lead variant
12	rs543159			total cholesterol (-) [24097068], total triglycerides (-) [24097068]	X-12798 [24816252]	isobutyrylcarnitine [24816252]; X-11261 [24816252]; Octenoylcarnitine [26068415]; propionylcarnitine [24816252]; Serum creatinine [20383146]; 3-methylglutarylcarnitine (2) [28263315]; beta-hydroxyisovalerate [24816252]; N1-methyladenosine [24816252]; 3-dehydrocarnitine* [24816252]; 2-methylbutyroylcarnitine [24816252]; 3-phenylpropionate (hydrocinnamate) [21886157]
13	rs4947534			Serine [26068415]	Serine [26068415]	Serine [26068415]; Metabolites in glutathione and glycine biosynthesis pathways [23378610]; erythronate [24816252]; X-11317 [24816252]
14	rs9987289	HDL (-) [24097068], LDL (-) [24097068], total cholesterol (-) [24097068], fasting glucose (+) [22885924], fasting insulin (+) [22885924], CRP (+) [21300955]	fatty liver (+)[21423719], liver enzymes (+) [22001757]			VLDL cholesterol mean particle size lipoprotein fraction concentration in fasting sample [19936222]; HDL Cholesterol - Triglycerides (HDLC-TG) [21386085]
15	rs28601761		HDL (-) [24097068], LDL (+) [24097068], total cholesterol (+) [24097068], total triglycerides (+) [24097068], CAD (+) [28530674]			LDL cholesterol small lipoprotein fraction concentration [19936222]; LDL cholesterol total lipoprotein fraction concentration [19936222]; Triglycerides by NMR lipoprotein fraction [19936222]; LDL cholesterol mean size lipoprotein fraction concentration [19936222]; X-16946 [28263315]
16	rs17591030					3-methylglutaconate [28263315]; 3-methylglutarylcarnitine (2) [28263315]

Locus	rsid	Traits associated with variant	Traits associated with variants in high LD (R2>0.6)	Reported metabolite associations for lead variant	Reported metabolite associations for variants in high LD (R2>0.8)	Reported metabolite associations for variants within 1 MB of lead variant
17	rs676996		MI (-) [26343387], venous thrombosis (-) [22675575], LDL (-) [24097068], total cholesterol (-) [24097068], Activated partial thromboplastin time (+) [22703881], Factor VIII antigen (-) [23381943], von Willebrand factor (-) [23381943]		Circulating galectin-3 levels [23056639]; Serum concentration of ADpSGEGDFXAEGGGVR* [21886157]; Serum phytosterol (campesterol) [20529992]; leucylalanine [28263315]	Circulating galectin-3 levels [23056639]; ADpSGEGDFXAEGGGVR* [24816252];Serum phytosterol (campesterol) [20529992]; X-17178 [28263315]; leucylalanine [28263315]; O-sulfo-L-tyrosine [24816252]; Serum phytosterol [20529992]; glycylglycine [28263315];alpha-glutamylglycine [28263315]
18		rs190595610				
19	rs10740134	triglycerides (+) [24097068], HDL (-) [24097068]			VLDL cholesterol total lipoprotein fraction concentration [19936222]; VLDL cholesterol medium lipoprotein fraction concentration [19936222]	VLDL cholesterol medium lipoprotein fraction concentration [19936222]
20	rs12297321					alanine [24816252]; ibuprofen [24816252]; carnitine [24816252]
21	rs2638314		fasting blood glucose (+) [22885924]	glutamine [28263315]; gamma-glutamylglutamine [28263315)	gamma-glutamylglutamine [24816252]; glutamine [24816252]	Serum Gln [22286219]; Urate levels [23263486]; gamma-glutamylglutamine [24816252]; 1,7-dimethylurate [28263315]; caffeine [28263315]
22	rs9514191					X-12850 [24816252]; taurolithocholate 3-sulfate [28263315]
23		rs201393666				

Locus	rsid	Traits associated with variant	Traits associated with variants in high LD (R2>0.6)	Reported metabolite associations for lead variant	Reported metabolite associations for variants in high LD (R2>0.8)	Reported metabolite associations for variants within 1 MB of lead variant
24	rs2280195					HDL cholesterol large lipoprotein fraction concentration [19936222]; Serum XL-HDL-TG [22286219]; HDL cholesterol mean size lipoprotein fraction concentration [19936222]; LDL cholesterol large lipoprotein fraction concentration [19936222]; HDL cholesterol large lipoprotein fraction concentration in fasting sample [19936222]; PE 36:4 [22359512]; Serum HDL-D [22286219]; HDL cholesterol mean size lipoprotein fraction concentration in fasting sample [19936222]; Serum LDL-D [22286219]; HDL cholesterol lipoprotein fraction concentration by NMR [19936222]; Serum XL-HDL-PL [22286219]; LDL cholesterol large lipoprotein fraction concentration in fasting sample [19936222]; Serum XL-HDL-P [22286219]; LDL cholesterol mean size lipoprotein fraction concentration [19936222]; Serum L-HDL-P [22286219]; HDL cholesterol (female) [23063622]; Serum L-HDL-L [22286219]; Serum XL-HDL-L [22286219]; etc.
25	rs9923732			Glycine [26068415]	Glycine [26068415]	Beta-carotene plasma levels [19185284]; Carotenoid and tocopherol levels [19185284]; Lutein plasma levels [19185284]; caprylate (8:0) [24816252]; caprate (10:0) [24816252]
26	rs8078686	height (+) [25282103], LDL (-) [24097068], total cholesterol (-) [24097068]	optic cup area (-) [25631615]		carnitine [24816252]	carnitine [24816252]
27	rs273510					myo-inositol [24816252]

Supplementary Table 2.3: Sex-specific effect sizes of the 27 glycine loci on log-transformed-only levels of glycine in the EPIC-Norfolk study

		Men			Women			P-value for sex
Locus	**SNP**	**Beta**	**SE**	***P***	**Beta**	**SE**	***P***	**difference**
1	rs4646961	0.009	0.003	2.30E-03	0.015	0.004	1.01E-04	0.214
2	rs561931	0.016	0.003	4.28E-09	0.003	0.004	3.72E-01	0.004
3	rs10184004	0.004	0.003	1.03E-01	0.015	0.004	3.49E-05	0.021
4	rs715	0.076	0.003	4.59E-159	0.194	0.003	0.00E+00	3.184E-158
5	rs9862438	0.017	0.003	1.97E-10	0.031	0.004	9.21E-18	0.003
6	rs148685782	0.026	0.024	2.89E-01	0.057	0.031	6.68E-02	0.436
7	rs71640034	0.000	0.003	8.84E-01	0.008	0.004	3.45E-02	0.110
8	rs156380	0.002	0.003	5.85E-01	0.010	0.004	2.72E-02	0.151
9	rs3105793	0.007	0.003	2.66E-02	0.006	0.004	1.44E-01	0.877
10	rs10900807	0.009	0.003	7.66E-03	0.006	0.005	2.17E-01	0.554
11	rs2545801	0.003	0.003	3.24E-01	0.001	0.004	8.96E-01	0.628
12	rs543159	0.005	0.003	8.37E-02	0.002	0.004	6.00E-01	0.528
13	rs4947534	0.020	0.003	2.34E-10	0.014	0.004	5.37E-04	0.305
14	rs9987289	0.033	0.005	1.16E-12	0.021	0.006	7.43E-04	0.123
15	rs28601761	0.024	0.003	3.67E-19	0.015	0.004	7.30E-05	0.030
16	rs17591030	0.014	0.003	2.64E-06	0.032	0.004	6.39E-16	0.000
17	rs676996	0.003	0.003	3.44E-01	0.008	0.004	4.51E-02	0.307
18	rs190595610	0.030	0.025	2.30E-01	0.027	0.036	4.54E-01	0.939
19	rs10740134	0.002	0.003	3.58E-01	0.010	0.004	6.43E-03	0.104
20	rs12297321	0.008	0.004	3.09E-02	0.021	0.005	9.77E-06	0.026
21	rs2638314	0.004	0.003	2.47E-01	0.015	0.005	9.90E-04	0.057
22	rs9514191	0.003	0.003	3.33E-01	0.005	0.004	1.72E-01	0.616
23	rs201393666	0.036	0.008	3.08E-06	-0.004	0.011	7.25E-01	0.003
24	rs2280195	0.010	0.003	1.35E-04	0.001	0.004	7.34E-01	0.044
25	rs9923732	0.042	0.005	1.22E-16	0.016	0.007	1.72E-02	0.002
26	rs8078686	0.010	0.003	1.33E-04	0.006	0.004	8.36E-02	0.354
27	rs273510	0.007	0.003	1.77E-02	0.006	0.004	1.16E-01	0.865

Supplementary Table 2.4: Associations of the 4 glycine scores and rs715 with the 84 metabolites in EPIC-Norfolk significantly associated with one or more scores

Metabolite	R obs glycine[a]	Full score (24 SNPs)		Glycine-related score (6 SNPs)		Glycine-related score without *CPS1* (5 SNPS)		Glycine-specific score (2 SNPs)		rs715	
		Beta	*p*	Beta	*p*	Beta	*p*	Beta	*p*	Beta	*p*
glycine	1.000	1.203	<1E-320	1.260	<1E-320	1.156	1.7E-32	1.112	3.4E-16	0.564	1E-320
gamma-glutamylglycine	0.903	1.056	2.22E-317	1.117	3.7E-308	0.991	2.1E-23	0.968	3.4E-12	0.501	4.6E-284
N-acetylglycine	0.477	0.831	1.4E-197	0.913	3.4E-208	0.879	3.5E-19	0.869	2.4E-10	0.407	3.0E-189
propionylglycine	0.451	0.796	5.4E-145	0.871	4.7E-152	0.663	1.8E-09	0.736	1.8E-06	0.395	4.9E-144
X - 16570	0.404	0.679	1.5E-114	0.753	1.6E-122	0.618	4.1E-09	0.663	6.7E-06	0.340	9.1E-115
3-methylglutaconate	0.180	0.611	2.4E-96	0.681	2.2E-104	0.319	2.0E-03	0.652	5.9E-06	0.319	3.8E-105
3-methylglutarylcarnitine	0.247	0.564	7.5E-85	0.642	2.3E-95	0.354	4.9E-04	0.751	1.2E-07	0.298	1.2E-94
homoarginine	-0.306	-0.544	5.2E-82	-0.605	4.5E-88	-0.255	1.0E-02	-0.286	3.9E-02	-0.285	1.2E-89
N-palmitoyl glycine	0.235	0.512	2.7E-68	0.554	1.2E-69	0.530	2.5E-07	0.531	2.2E-04	0.247	1.0E-63
X - 17367	0.286	0.540	1.8E-63	0.585	6.0E-65	0.252	2.6E-02	0.129	4.2E-01	0.275	6.1E-66
X - 13722	0.300	0.526	6.4E-58	0.539	5.9E-53	0.564	8.2E-07	0.590	2.4E-04	0.237	2.4E-47
isovalerylglycine	0.373	0.521	3.0E-54	0.537	1.8E-50	0.507	1.8E-05	0.709	1.8E-05	0.240	2.0E-46
serine	0.560	0.427	6.2E-50	0.418	6.6E-42	0.978	1.6E-22	0.488	5.0E-04	0.159	1.8E-28
cinnamoylglycine	0.268	0.420	3.5E-47	0.448	9.6E-47	0.448	1.0E-05	0.488	6.0E-04	0.199	2.3E-42
X - 16564	0.227	0.411	3.9E-44	0.452	2.8E-46	0.252	1.5E-02	0.327	2.3E-02	0.210	6.1E-46
isobutyrylglycine	0.275	0.469	1.3E-43	0.504	5.9E-44	0.435	2.3E-04	0.490	3.0E-03	0.227	5.1E-41
creatine	0.253	0.353	1.0E-42	0.376	4.3E-42	0.370	4.0E-05	0.348	5.7E-03	0.167	2.5E-38
hexanoylglycine	0.221	0.523	1.1E-41	0.610	2.6E-49	0.548	5.8E-05	0.661	6.4E-04	0.273	1.2E-45
pyroglutamine*	-0.165	-0.279	8.9E-34	-0.308	8.9E-36	-0.384	1.7E-06	-0.372	9.2E-04	-0.133	6.6E-31
betaine	0.030	-0.312	8.2E-29	-0.378	2.1E-36	-0.301	2.1E-03	-0.200	1.4E-01	-0.171	1.9E-34
X - 21467	0.165	0.324	1.1E-25	0.367	1.6E-28	0.274	1.1E-02	0.304	4.3E-02	0.167	2.6E-27
X - 23637	0.468	0.250	9.8E-18	0.250	1.2E-15	0.691	8.8E-12	0.438	2.0E-03	0.090	5.8E-10
X - 24542	0.159	0.205	4.7E-11	0.211	2.3E-10	0.259	1.7E-02	0.297	5.0E-02	0.092	3.9E-09
5-methylthioadenosine (MTA)	-0.142	-0.189	3.4E-10	-0.201	3.9E-10	-0.391	1.7E-04	-0.251	8.4E-02	-0.080	8.9E-08
X - 11261	-0.047	-0.182	5.7E-10	-0.199	2.5E-10	-0.312	2.3E-03	-0.433	2.5E-03	-0.083	1.5E-08
X - 11564	-0.133	-0.172	1.5E-09	-0.205	1.5E-11	-0.402	4.9E-05	-0.306	2.7E-02	-0.082	7.9E-09
glutarylcarnitine (C5)	-0.088	-0.172	2.9E-09	-0.205	4.0E-11	-0.038	7.0E-01	0.043	7.6E-01	-0.099	9.2E-12
X - 17351	-0.045	-0.181	4.9E-09	-0.182	3.9E-08	-0.243	2.4E-02	0.088	5.6E-01	-0.078	4.4E-07
taurocholenate sulfate	-0.154	-0.171	5.9E-09	-0.169	8.8E-08	-0.176	8.6E-02	-0.123	3.9E-01	-0.075	4.1E-07
1-stearoyl-2-docosapentaenoyl-GPC (18:0/22:5n6)*	-0.134	-0.171	6.6E-09	-0.153	1.3E-06	-0.217	3.5E-02	-0.270	6.0E-02	-0.065	1.1E-05
gamma-glutamylthreonine*	0.196	-0.172	6.7E-09	-0.225	1.7E-12	-0.147	1.6E-01	-0.084	5.6E-01	-0.103	3.4E-12
X - 11334	-0.070	-0.163	1.5E-08	-0.187	1.3E-09	-0.204	4.1E-02	-0.015	9.1E-01	-0.082	1.1E-08
hydantoin-5-propionic acid	-0.123	-0.170	1.8E-08	-0.175	7.1E-08	-0.325	2.0E-03	-0.298	4.2E-02	-0.071	3.2E-06

Metabolite	R obs glycine[a]	Full score (24 SNPs)		Glycine-related score (6 SNPs)		Glycine-related score without *CPS1* (5 SNPS)		Glycine-specific score (2 SNPs)		rs715	
		Beta	*p*	Beta	*p*	Beta	*p*	Beta	*p*	Beta	*p*
1-stearoyl-2-docosahexaenoyl-GPC (18:0/22:6)	-0.074	-0.161	4.0E-08	-0.130	3.6E-05	-0.180	7.8E-02	-0.272	5.7E-02	-0.055	1.6E-04
1-methylurate	-0.014	-0.164	4.9E-08	-0.184	1.2E-08	-0.048	6.5E-01	0.057	6.9E-01	-0.088	5.2E-09
1-docosapentaenoyl-GPC (22:5n6)*	-0.088	-0.170	5.0E-08	-0.159	2.0E-06	-0.163	1.4E-01	-0.186	2.2E-01	-0.070	6.7E-06
arginine	0.092	-0.157	1.0E-07	-0.194	9.2E-10	-0.037	7.2E-01	-0.144	3.2E-01	-0.093	2.6E-10
aspartate	-0.091	-0.150	3.3E-07	-0.153	1.2E-06	-0.284	5.4E-03	-0.256	7.3E-02	-0.062	2.6E-05
glycocholenate sulfate*	0.035	0.146	3.8E-07	0.180	5.4E-09	0.143	1.5E-01	0.048	7.3E-01	0.082	1.4E-08
deoxycarnitine	0.032	-0.134	4.1E-07	-0.152	8.2E-08	-0.173	6.0E-02	-0.196	1.3E-01	-0.066	5.0E-07
taurolithocholate 3-sulfate	-0.028	-0.146	7.7E-07	-0.157	7.1E-07	-0.122	2.3E-01	-0.108	4.5E-01	-0.071	1.4E-06
X - 21821	-0.050	-0.154	9.3E-07	-0.157	3.5E-06	-0.250	2.3E-02	0.076	6.2E-01	-0.065	3.6E-05
X - 13431	-0.042	-0.143	9.3E-07	-0.143	4.7E-06	0.011	9.2E-01	-0.108	4.5E-01	-0.071	1.3E-06
1-stearoyl-2-docosapentaenoyl-GPC (18:0/22:5n3)*	-0.032	-0.144	1.1E-06	-0.122	1.2E-04	-0.210	4.0E-02	-0.356	1.3E-02	-0.050	7.1E-04
1-oleoyl-2-docosahexaenoyl-GPC (18:1/22:6)*	0.098	-0.139	1.3E-06	-0.130	2.5E-05	-0.106	2.9E-01	-0.126	3.7E-01	-0.059	4.3E-05
choline	0.022	-0.140	1.5E-06	-0.179	1.0E-08	-0.126	2.1E-01	-0.106	4.6E-01	-0.082	2.0E-08
sphingomyelin (d18:1/20:1, d18:2/20:0)*	0.241	0.132	2.0E-06	0.135	5.8E-06	0.168	8.3E-02	0.162	2.3E-01	0.059	2.6E-05
X - 12283	-0.047	-0.160	2.3E-06	-0.162	8.2E-06	-0.223	5.8E-02	-0.082	6.2E-01	-0.069	4.7E-05
kynurenate	-0.045	-0.136	2.9E-06	-0.172	3.4E-08	-0.128	2.1E-01	-0.050	7.2E-01	-0.078	7.1E-08
adenine	-0.104	-0.137	3.3E-06	-0.144	4.7E-06	-0.270	8.3E-03	-0.076	5.9E-01	-0.058	7.7E-05
octanoylcarnitine	-0.103	-0.138	3.4E-06	-0.068	3.2E-02	-0.103	3.2E-01	-0.117	4.2E-01	-0.029	5.3E-02
dimethylglycine	-0.084	-0.133	3.7E-06	-0.145	2.6E-06	-0.232	2.0E-02	-0.319	2.2E-02	-0.060	2.8E-05
N-acetylthreonine	-0.021	-0.135	3.9E-06	-0.139	9.3E-06	-0.142	1.6E-01	-0.107	4.5E-01	-0.061	2.6E-05
X - 16944	0.015	-0.137	4.3E-06	-0.143	7.7E-06	-0.126	2.2E-01	-0.129	3.7E-01	-0.064	1.7E-05
docosahexaenoylcholine	0.021	-0.133	5.5E-06	-0.113	3.3E-04	-0.030	7.7E-01	-0.064	6.5E-01	-0.054	2.4E-04
biliverdin	-0.021	0.131	5.9E-06	0.122	8.2E-05	0.180	7.4E-02	0.147	3.0E-01	0.052	3.7E-04
4-hydroxycoumarin	-0.057	-0.146	7.1E-06	-0.155	8.8E-06	-0.274	1.5E-02	-0.234	1.4E-01	-0.063	1.0E-04
homostachydrine*	0.060	-0.132	8.8E-06	-0.136	1.8E-05	-0.165	1.1E-01	0.049	7.3E-01	-0.059	6.5E-05
X - 16071	-0.087	-0.128	9.0E-06	-0.113	2.6E-04	-0.123	2.2E-01	-0.012	9.3E-01	-0.050	5.7E-04
decanoylcarnitine	-0.105	-0.132	9.0E-06	-0.079	1.3E-02	-0.074	4.7E-01	-0.061	6.7E-01	-0.035	1.7E-02
1-adrenoyl-GPC (22:4)*	-0.059	-0.131	9.5E-06	-0.112	3.9E-04	-0.106	3.0E-01	-0.023	8.7E-01	-0.050	7.0E-04
bilirubin (Z,Z)	0.075	0.127	9.9E-06	0.122	6.7E-05	0.138	1.7E-01	0.063	6.5E-01	0.054	1.8E-04
1-arachidonoyl-GPI (20:4)*	-0.048	-0.128	1.2E-05	-0.108	5.8E-04	-0.138	1.7E-01	-0.230	1.1E-01	-0.047	1.5E-03
histidine	0.186	-0.126	1.6E-05	-0.166	1.1E-07	-0.006	9.5E-01	-0.091	5.2E-01	-0.081	2.6E-08
imidazole propionate	-0.080	-0.130	1.7E-05	-0.137	2.2E-05	-0.091	3.9E-01	0.049	7.3E-01	-0.063	2.9E-05
1-stearoyl-2-dihomo-linolenoyl-GPC (18:0/20:3n3 or 6)*	-0.135	-0.125	2.1E-05	-0.093	3.1E-03	-0.289	4.5E-03	-0.343	1.6E-02	-0.032	2.9E-02
1-docosahexaenoyl-GPC (22:6)*	0.002	-0.124	2.3E-05	-0.094	2.7E-03	0.011	9.1E-01	0.089	5.3E-01	-0.047	1.4E-03
hexanoylcarnitine	-0.121	-0.125	2.3E-05	-0.031	3.3E-01	-0.108	3.0E-01	-0.073	6.1E-01	-0.010	5.0E-01
guanidinoacetate	0.140	0.121	2.5E-05	0.133	1.3E-05	0.231	2.0E-02	0.112	4.2E-01	0.055	1.3E-04

Metabolite	R obs glycine[a]	Full score (24 SNPs)		Glycine-related score (6 SNPs)		Glycine-related score without *CPS1* (5 SNPS)		Glycine-specific score (2 SNPs)		rs715	
		Beta	*p*	Beta	*p*	Beta	*p*	Beta	*p*	Beta	*p*
gamma-glutamylhistidine	0.120	-0.122	3.0E-05	-0.138	1.0E-05	-0.150	1.4E-01	-0.184	1.9E-01	-0.061	3.2E-05
X - 23739	0.187	-0.120	4.2E-05	-0.142	6.9E-06	-0.104	3.1E-01	-0.160	2.6E-01	-0.065	1.1E-05
X - 24452	0.090	0.117	4.5E-05	0.111	2.8E-04	-0.044	6.6E-01	-0.164	2.4E-01	0.057	7.4E-05
X - 14939	-0.019	-0.120	4.6E-05	-0.121	1.3E-04	-0.056	5.9E-01	-0.094	5.1E-01	-0.057	1.2E-04
N-acetyltryptophan	0.071	0.122	5.5E-05	0.184	1.2E-08	-0.066	5.3E-01	0.024	8.7E-01	0.093	6.0E-10
pipecolate	-0.048	-0.119	5.8E-05	-0.145	5.4E-06	-0.057	5.8E-01	-0.108	4.6E-01	-0.068	4.1E-06
threonine	0.276	-0.117	6.8E-05	-0.185	4.3E-09	0.026	8.0E-01	-0.095	5.1E-01	-0.092	4.0E-10
indoleacetylglutamine	-0.029	-0.134	1.0E-04	-0.150	5.5E-05	-0.154	2.0E-01	-0.192	2.5E-01	-0.066	1.3E-04
O-methylcatechol sulfate	-0.006	-0.114	1.2E-04	-0.130	4.0E-05	-0.333	1.2E-03	-0.256	7.4E-02	-0.048	1.1E-03
hypotaurine	0.074	-0.103	3.7E-04	-0.146	2.5E-06	-0.072	4.7E-01	-0.135	3.3E-01	-0.068	2.4E-06
X - 12739	-0.001	-0.114	4.0E-04	-0.161	2.9E-06	-0.115	3.0E-01	-0.146	3.5E-01	-0.074	4.5E-06
X - 24527	0.006	-0.103	8.1E-04	-0.139	2.3E-05	-0.096	3.6E-01	-0.084	5.7E-01	-0.064	3.4E-05
asparagine	0.292	-0.098	8.8E-04	-0.131	3.4E-05	0.120	2.4E-01	0.065	6.5E-01	-0.070	2.2E-06
X - 12101	0.052	-0.092	2.0E-03	-0.155	1.2E-06	-0.167	1.1E-01	-0.211	1.4E-01	-0.068	4.8E-06
X - 12798	0.005	-0.083	4.8E-03	-0.129	4.8E-05	0.060	5.6E-01	0.154	2.8E-01	-0.066	8.1E-06

Supplementary Table 2.5: Results of Mendelian randomisation analyses for glycine to T2D

Results of MR analyses based on 4 different genetic instruments for glycine and 4 different methods are shown: inverse-variance-weighted (IVW), EGGER, weighted median (WM) and penalized weighted median (PWM). OR: odds ratio, LCI: lower confidence interval, UCI: upper confidence interval, CochQ *p*: *p* for Cochran's Q statistic, interEGGER: intercept of MR EGGER, EGGER *p*: *p* for EGGER intercept.

Full Score	Sex-combined				Women				Men			
	IVW	**EGGER**	**WM**	**PWM**	**IVW**	**EGGER**	**WM**	**PWM**	**IVW**	**EGGER**	**WM**	**PWM**
OR	0.972	1.054	1.004	1.005	0.994	1.032	1.007	1.007	1.009	1.135	1.058	1.063
LCI	0.877	0.934	0.969	0.967	0.928	0.960	0.969	0.969	0.875	0.954	0.970	0.973
UCI	1.078	1.189	1.040	1.044	1.065	1.109	1.046	1.046	1.163	1.349	1.154	1.162
p	0.594	0.397	0.840	0.808	0.871	0.391	0.731	0.729	0.902	0.152	0.203	0.178
CochQ *p*	0.020				0.139				2.0E-04			
interEGGER		-0.013				-0.014				-0.015		
EGGER *p*		0.032				0.026				0.038		

6 SNP score	Sex-combined				Women				Men			
	IVW	**EGGER**	**WM**	**PWM**	**IVW**	**EGGER**	**WM**	**PWM**	**IVW**	**EGGER**	**WM**	**PWM**
OR	0.990	1.021	1.001	1.003	1.005	1.025	1.008	1.008	1.022	1.051	1.042	1.049
LCI	0.935	0.948	0.964	0.966	0.968	0.978	0.971	0.972	0.894	0.823	0.957	0.963
UCI	1.049	1.099	1.039	1.041	1.043	1.074	1.046	1.045	1.169	1.342	1.134	1.142
p	0.743	0.588	0.966	0.887	0.801	0.310	0.683	0.675	0.752	0.689	0.342	0.271
CochQ *p*	0.929				0.846				0.807			
interEGGER		-0.009				-0.011				-0.005		
EGGER *p*		0.228				0.084				0.776		

5 SNP score	Sex-combined				Women				Men			
	IVW	**EGGER**	**WM**	**PWM**	**IVW**	**EGGER**	**WM**	**PWM**	**IVW**	**EGGER**	**WM**	**PWM**
OR	0.869	0.758	0.868	0.868	0.919	1.054	0.928	0.928	0.918	0.708	0.908	0.831
LCI	0.755	0.489	0.757	0.756	0.778	0.735	0.767	0.763	0.704	0.338	0.730	0.675
UCI	1.001	1.173	0.996	0.997	1.085	1.511	1.123	1.130	1.197	1.483	1.129	1.024
p	0.052	0.214	0.044	0.046	0.318	0.775	0.445	0.460	0.527	0.360	0.385	0.082
CochQ *p*	0.93				0.792				0.724			
interEGGER		0.010				-0.014				0.023		
EGGER *p*		0.510				0.354				0.458		

2 SNP score	Sex-combined				Women				Men			
	IVW	**EGGER**	**WM**	**PWM**	**IVW**	**EGGER**	**WM**	**PWM**	**IVW**	**EGGER**	**WM**	**PWM**
OR	0.792	NA	0.792	0.792	0.919	NA	0.855	0.855	0.774	NA	0.774	0.774
LCI	0.671	NA	0.682	0.680	0.778	NA	0.661	0.664	0.623	NA	0.620	0.613
UCI	0.933	NA	0.919	0.921	1.085	NA	1.106	1.100	0.960	NA	0.966	0.976
p	0.005	NA	0.002	0.003	0.318	NA	0.233	0.223	0.020	NA	0.024	0.031
CochQ *p*	0.713				0.657	NA			0.674			
interEGGER		NA				NA				NA		
EGGER *p*		NA				NA				NA		

Supplementary Table 2.6: reverse Mendelian randomisation analyses to assess the causality of T2D risk factors on glycine levels

These are the results of the MR analyses with T2D risk factors as the exposure and glycine levels as the outcome. Analyses were conducted for body mass index (BMI), fasting insulin adjusted for BMI as a marker for insulin resistance (IR) and insulin levels at 30 minutes during an oral glucose tolerance test, as a marker for early-phase insulin secretion (IS). Results of 4 different methods are shown: inverse-variance-weighted (IVW), EGGER, weighted median (WM) and penalized weighted median (PWM). OR: odds ratio, LCI: lower confidence interval, UCI: upper confidence interval, CochQ *p*: *p* for Cochran's Q statistic, interEGGER: intercept of MR EGGER, EGGER *p*: *p* for EGGER intercept.

BMI	**IVW**	**MR-Egger**	**WM**	**PWM**
Beta	-0.041	-0.041	-0.009	-0.007
SE	0.030	0.074	0.041	0.040
pvalue	0.169	0.581	0.825	0.870
CochQp	1.000			
interEGGER		-1.34E-05		
interEGGER p		0.995		

IR	**IVW**	**EGGER**	**WM**	**PWM**
beta	-0.960	-1.179	-0.863	-0.822
SE	0.212	1.253	0.211	0.211
p	5.98E-06	3.47E-01	4.29E-05	9.56E-05
CochQ *p*	0.393			
interEGGER		0.003		
interEGGER p		0.859		

IS	**IVW**	**EGGER**	**MW**	**PMW**
Beta	0.022	-0.030	-0.011	-0.010
SE	0.033	0.087	0.038	0.037
pvalue	0.516	0.733	0.762	0.786
CochQ *p*	0.999			
interEGGER		-0.003		
EGGER *p*		0.524		

Supplementary Table 2.7: Results of Mendelian randomisation analyses for glycine to CHD

Results of MR analyses based on 4 different genetic instruments for glycine and 4 different methods are shown: inverse-variance-weighted (IVW), EGGER, weighted median (WM) and penalized weighted median (PWM). OR: odds ratio, LCI: lower confidence interval, UCI: upper confidence interval, CochQ p: p for Cochran's Q statistic, interEGGER: intercept of MR EGGER, EGGER p: p for EGGER intercept.

Full Score	Sex-combined				Women				Men			
	IVW	**EGGER**	**WM**	**PWM**	**IVW**	**EGGER**	**WM**	**PWM**	**IVW**	**EGGER**	**WM**	**PWM**
OR	0.904	0.977	0.948	0.881	0.928	0.957	0.953	0.954	0.817	0.954	0.926	0.937
LCI	0.836	0.897	0.919	0.793	0.848	0.864	0.910	0.912	0.692	0.763	0.842	0.816
UCI	0.978	1.064	0.978	0.979	1.014	1.060	0.997	0.998	0.964	1.193	1.019	1.076
p	0.012	0.587	0.001	0.019	0.100	0.401	0.039	0.041	0.017	0.677	0.118	0.360
CochQ *p*	0.507				0.174				0.092			
interEGGER		-0.013				-0.010				0.007		
EGGER *p*		0.003				0.226				0.057		

6 SNP score	Sex-combined				Women				Men			
	IVW	**EGGER**	**WM**	**PWM**	**IVW**	**EGGER**	**WM**	**PWM**	**IVW**	**EGGER**	**WM**	**PWM**
OR	0.945	0.967	0.953	0.958	0.942	0.975	0.954	0.954	0.926	0.972	0.941	0.942
LCI	0.887	0.885	0.924	0.929	0.826	0.816	0.911	0.913	0.814	0.777	0.860	0.857
UCI	1.006	1.056	0.984	0.988	1.075	1.165	0.998	0.998	1.054	1.216	1.029	1.036
p	0.076	0.452	0.003	0.006	0.375	0.779	0.039	0.040	0.245	0.804	0.181	0.218
CochQ *p*	0.845				0.127				0.800			
interEGGER		-0.007				-0.017				-0.007		
EGGER *p*		0.441				0.540				0.585		

5 SNP score	Sex-combined				Women				Men			
	IVW	**EGGER**	**WM**	**PWM**	**IVW**	**EGGER**	**WM**	**PWM**	**IVW**	**EGGER**	**WM**	**PWM**
OR	0.836	0.657	0.803	0.767	0.735	0.590	0.827	0.840	0.885	1.043	0.924	0.957
LCI	0.699	0.389	0.710	0.677	0.424	0.128	0.649	0.658	0.663	0.432	0.736	0.756
UCI	0.999	1.110	0.908	0.870	1.273	2.727	1.054	1.072	1.182	2.518	1.161	1.212
p	0.049	0.116	4.8E-4	3.7E-5	0.272	0.500	0.124	0.162	0.408	0.926	0.498	0.716
CochQ *p*	0.789				0.104				0.686			
interEGGER		0.018				0.019				-0.011		
EGGER *p*		0.338				0.759				0.695		

2 SNP score	Sex-combined				Women				Men			
	IVW	**EGGER**	**WM**	**PWM**	**IVW**	**EGGER**	**WM**	**PWM**	**IVW**	**EGGER**	**WM**	**PWM**
OR	0.798	NA	0.798	0.798	0.832	NA	0.832	0.832	0.769	NA	0.769	0.798
LCI	0.670	NA	0.693	0.701	0.656	NA	0.655	0.652	0.413	NA	0.595	0.615
UCI	0.950	NA	0.919	0.908	1.055	NA	1.056	1.062	1.433	NA	0.993	1.036
p	0.011	NA	0.002	0.001	0.130		0.131	0.139	0.408	NA	0.044	0.091
CochQ *p*	0.677				0.784				0.269			
interEGGER		NA				NA				NA		
EGGER *p*		NA				NA				NA		

Supplementary Table 2.8: Results of Cox proportional hazards models for the association of glycine levels with CHD, myocardial infarction and stroke, including stroke sub-types

1. Coronary heart disease	**sex-combined**		**women**		**men**	
	HR [95% CI]	p	HR [95% CI]	p	HR [95% CI]	p
Standardized glycine levels	0.92 [0.87,0.96]	4.7E-4	0.94 [0.88,1.01]	0.083	0.88 [0.82,0.95]	7.2E-4
Log-transformed glycine levels	0.71 [0.58,0.86]	4.8E-4	0.79 [0.61,1.03]	0.086	0.60 [0.45,0.81]	7.1E-4
2. Myocardial infarction	**sex-combined**		**women**		**men**	
	HR [95% CI]	p	HR [95% CI]	p	HR [95% CI]	p
Standardized glycine levels	0.89 [0.82,0.97]	5.50E-3	0.95 [0.85,1.07]	0.39	0.82 [0.73,0.93]	1.6E-3
Log-transformed glycine levels	0.62 [0.45,0.87]	5.80E-3	0.82 [0.51,1.31]	0.4	0.46 [0.28,0.75]	1.7E-3
3. Stroke (all types)	**sex-combined**		**women**		**men**	
	HR [95% CI]	p	HR [95% CI]	p	HR [95% CI]	p
Standardized glycine levels	0.99 [0.93,1.05]	0.70	0.98 [0.91,1.06]	0.68	1.00 [0.90,1.10]	0.92
Log-transformed glycine levels	0.95 [0.75,1.21]	0.70	0.94 [0.70,1.26]	0.68	0.98 [0.65,1.49]	0.93
4. Haemorrhagic stroke	**sex-combined**		**women**		**men**	
	HR [95% CI]	p	HR [95% CI]	p	HR [95% CI]	p
Standardized glycine levels	1.11 [0.96,1.29]	0.15	1.07 [0.89,1.28]	0.49	1.20 [0.94,1.53]	0.13
Log-transformed glycine levels	1.54 [0.86,2.75]	0.15	1.30 [0.63,2.69]	0.48	2.11 [0.80,5.59]	0.13
5. Ischemic stroke	**sex-combined**		**women**		**men**	
	HR [95% CI]	p	HR [95% CI]	p	HR [95% CI]	p
Standardized glycine levels	1.00 [0.92,1.10]	0.95	1.02 [0.91,1.14]	0.74	0.97 [0.84, 1.13]	0.73
Log-transformed glycine levels	1.01 [0.71,1.46]	0.94	1.08 [0.69,1.69]	0.74	0.90 [0.49,1.67]	0.74

Supplementary table 2.9: Associations of 4 genetic scores for glycine with blood pressure, lipid and blood cell phenotypes

CHD risk factors	Full glycine score (24 SNPs)			Glycine-related score (6 SNPs)			Glycine-related score without *CPS1* (5 SNP score)			Glycine-specific score (2 SNPs)		
	beta	SE	p-value	beta	SE	p-value	beta	SE	p-value	beta	SE	p-value
Systolic blood pressure	-0.028	0.007	1.49E-05	-0.029	0.006	2.56E-06	-0.050	0.021	1.93E-02	-0.042	0.022	5.71E-02
Diastolic blood pressure	-0.019	0.009	3.87E-02	-0.024	0.004	4.32E-11	-0.038	0.013	3.09E-03	-0.030	0.020	1.28E-01
HDL cholesterol	-0.058	0.034	8.89E-02	-0.056	0.010	7.37E-09	-0.006	0.039	8.76E-01	-0.061	0.049	2.16E-01
LDL cholesterol	-0.035	0.043	4.11E-01	0.015	0.013	2.53E-01	-0.034	0.057	5.53E-01	-0.012	0.053	8.23E-01
Total cholesterol	-0.061	0.044	1.65E-01	-0.011	0.013	4.18E-01	-0.019	0.065	7.68E-01	-0.018	0.052	7.30E-01
Triglycerides	-0.020	0.043	6.48E-01	-0.002	0.010	8.69E-01	-0.032	0.048	4.98E-01	0.017	0.048	7.31E-01
Basophil	-0.004	0.012	7.52E-01	-0.006	0.008	4.25E-01	0.010	0.026	7.11E-01	-0.007	0.036	8.56E-01
Eosinophil	0.047	0.021	2.75E-02	0.046	0.020	2.52E-02	0.125	0.061	4.12E-02	0.213	0.103	3.92E-02
Hematocrit	0.023	0.023	3.14E-01	0.004	0.014	7.96E-01	0.030	0.049	5.49E-01	0.006	0.036	8.76E-01
Platelet count	-0.078	0.030	9.94E-03	-0.072	0.011	3.16E-10	-0.020	0.031	5.16E-01	-0.008	0.077	9.17E-01
Platelet distribution width	0.042	0.026	1.06E-01	0.035	0.010	2.88E-04	-0.017	0.027	5.25E-01	0.009	0.037	8.15E-01
Mean platelet volume	0.086	0.051	9.25E-02	0.061	0.008	5.35E-13	0.028	0.027	3.04E-01	0.031	0.037	4.06E-01
Red cell distribution width	0.049	0.023	3.15E-02	0.023	0.008	4.37E-03	0.052	0.027	5.29E-02	0.038	0.037	3.04E-01
Mean corpuscular hemoglobin	0.040	0.021	5.63E-02	0.057	0.021	7.78E-03	0.088	0.076	2.45E-01	0.146	0.071	4.00E-02
Reticulocyte	-0.001	0.022	9.80E-01	0.006	0.012	6.48E-01	-0.003	0.045	9.52E-01	-0.074	0.052	1.56E-01
Immature reticulocyte fraction	0.024	0.014	7.83E-02	0.021	0.008	9.42E-03	0.049	0.027	6.93E-02	0.072	0.037	4.79E-02
Neutrophil	-0.032	0.024	1.80E-01	-0.045	0.014	1.03E-03	-0.095	0.042	2.26E-02	-0.066	0.037	7.52E-02
Monocyte	-0.018	0.020	3.73E-01	-0.029	0.013	2.24E-02	0.040	0.028	1.52E-01	0.091	0.037	1.35E-02
Lymphocyte	-0.041	0.019	3.32E-02	-0.040	0.013	1.87E-03	-0.023	0.045	6.11E-01	0.012	0.044	7.93E-01

Chapter 3

Supplementary Table 3.1: Overview of cohorts included in the exome chip meta-analyses for BF% and FFMI

Study	Study design	Ethnicity	N (women) BF%	N (women) FFMI	Genotyping chip	Software	N variants analysed	Technique	Instrument	Reference
AGES	Population-based	European	3988 (2234)	3988 (2234)	Illumina HumanExome v1.0 and v1.2	RareMetalWorker	234,708	BIA	A Xitron HYDRA ECF/ICF, Model 4200	[391]Harris 2007
Airwave	Population-based	European	14,3477 (5390)	--	Illumina Infinium HumanExome12 v1.1	RareMetalWorker	233,034	BIA	Tanita BC-418MA body composition analyser	[392]Elliott 2014
CHS	Population-based	European	4129 (2316)	4135 (2320)	Illumina ExomeChip v1.0	RareMetalWorker	227,009	DEXA	Hologic QDR-2000 Bone densitometer	[393]Robbins 2001
CHS	Population-based	Afro-American	764 (476)	766 (477)	Illumina ExomeChip v1.0	RareMetalWorker	227,009	DEXA	Hologic QDR-2000 Bone densitometer	[393]Robbins 2001
Diogenes	Population-based	European	2159 (1068)	2159 (1068)	Illumina ExomeChip	RareMetalWorker	224,702	DEXA/BIA	GE Lunar Prodigy/Tanita body composition analyser	[394]Larsen 2010
Ely	Population-based	European	1432 (772)	1432 (772)	Illumina HumanCoreExome	RareMetalWorker	231,367	BIA	BODY STAT 1500	[395]Forouhi 2007
EPIC-Norfolk	Population-based	European	13554 (7426)	13542 (7417)	UK Biobank Axiom Array	RareMetalWorker	57,858	BIA	Tanita TBF300	[191]Day 1999
EPIHEALTH	Population-based	European	2350 (1179)	2350 (1179)	Illumina HumanCoreExome	RVTest	233,191	BIA	BC-418, Tanita	[396]Lind 2013
Erasmus Rucphen Family study	Family-based	European	1144 (608)	1144 (608)	Illumina Exome BeadChip array v1.0	RVTest	240,035	DEXA	GE Lunar Prodigy	[397]Henneman 2008
FamHS	Population-based	European	2337 (1294)	2374 (1229)	Illumina ExomeChip v1.0	RareMetalWorker	237,395	BIA	Weight Manager Version 2.05a	[398]Higgins 1996
Fenland-CE	Population-based	European	1006 (549)	1006 (549)	Illumina HumanCoreExome	RareMetalWorker	234,201	DEXA	GE Lunar Prodigy	[79]Lotta 2017
Fenland-EC	Population-based	European	1315 (706)	1314 (706)	Illumina ExomeChip v1.0	RareMetalWorker	240,881	DEXA	GE Lunar Prodigy	[79]Lotta 2017
Fenland-OMICS	Population-based	European	7363 (3867)	7363 (3867)	UK Biobank Axiom Array	RareMetalWorker	58,262	DEXA	GE Lunar Prodigy	[79]Lotta 2017
FHS	Population-based	European	5674 (3281)	5671 (3278)	Illumina ExomeChip v1.0	RareMetalWorker	246,691	DEXA	GE Lunar DPX-L	[399]Visser 1998

Study	Study design	Ethnicity	N (women) BF%	N (women) FFMI	Genotyping chip	Software	N variants analysed	Technique	Instrument	Reference
FINRISK07	Population-based	European	2919 (1635)	2919 (1635)	HumanExome12 v1.1A	RVTest	501,249	BIA	TANITA TBF-300MA bioelectrical impedance device	[400]Borodulin 2017
HealthABC	Longitudinal cohort study	European	1572 (747)	1572 (747)	Illumina ExomeChip v1.0	RVTest	228,582	DEXA	Hologic QDR-4500	[401]Harris 2000
HealthABC	Longitudinal cohort study	Afro-American	1060 (603)	1060 (603)	Illumina ExomeChip v1.0	RVTest	228,571	DEXA	Hologic QDR-4500	[401]Harris 2000
Health2006	Population-based	European	2919 (1614)	2919 (1614)	Illumina HumanExome 12v1	RareMetalWorker	227,684	BIA	foot-to-foot Tanita Body Composition Analyzer	[402]Thuesen 2014
Health2008	Population-based	European	747 (422)	747 (422)	Illumina HumanExome 12v1	RareMetalWorker	227,684	BIA	foot-to-foot Tanita Body Composition Analyzer	[402]Thuesen 2014
KORA-F4	Population-based	European	2881 (1477)	2881 (1477)	Illumina ExomeChip v1.0	RVTest	245,163	BIA	GmbH BIA 2000-S	[403]Wichmann 2005
Leipzig adults	Population-based	European	608 (395)	608 (395)	Illumina HumanExome 12v1A	RareMetalWorker	230,972	BIA	BIA-Nutriguard-MS with the software Nutri3	[404]Speliotes 2010
LOLIPOP-Exome	Population-based with some enrichment	Indian Asian	1466 (421)	1466 (421)	Illumina Human Exome BeadChip	RareMetalWorker	240,292	BIA	Tanita TBF-401	[405]Kooner 2008
LOLIPOP-OmniEE	Population-based with some enrichment	Indian Asian	864 (417)	864 (417)	Illumina OmniExpressExome BeadChip	RareMetalWorker	240,137	BIA	Tanita TBF-401	[405]Kooner 2008
METSIM	Population-based	European	10019 (0)	10019 (0)	HumanExome 12v1A	RVTest	241,780	BIA	Bioimpedance Analyzer Model BIA101 (Akern Srl, Florence Italy)	[406]Stancakova 2009
NEO	Population-based	European	6096 (3176)	6096 (3176)	Illumina HumanCoreExomeChip 24 v1.0	RVTest	209,892	DEXA	Hologic Discovery A, Tromp Medical BV, Castricum, The Netherlands	[407]de Mutsert 2013
PIVUS	Population-based	European	824 (419)	824 (419)	Illumina HumanExome-12v1_A	RareMetalWorker	233,166	DEXA	GE Lunar Prodigy	[408]Lind 2005

Study	Study design	Ethnicity	N (women) BF%	N (women) FFMI	Genotyping chip	Software	N variants analysed	Technique	Instrument	Reference
RAINE	Population-based	European	995 (475)	995 (475)	Illuminia HumanExome 12 v1A	RVTest	240,080	DEXA	Norland XR-36 densitometer	[409]Straker 2017
RISC	Population-based	European	313 (157)	313 (157)	Illumina Human Exome Beadchip v1	RareMetalWorker	236,893	BIA	Tanita TBF-300 Body Composition Analyser	[410]Hills 2004
Rotterdam Study 1	Population-based	European	1213 (560)	1198 (555)	Illumina ExomeChip v1.1	RVTest	237,783	DEXA	GE Lunar Prodigy	[411]Hofman 2013
SHIP	Population-based	European	6309 (3295)	6309 (3295)	Illumina ExomeChip v1.0	RVTest	237,442	BIA	BIA-Nutriguard-MS with the software Nutri3	[412]Völzke 2011
SORBS	Population-based	European	997 (595)	997 (595)	Illumina HumanExome 12 v1A	RareMetalWorker	230,972	BIA	BIA-2000-S (Data Input GmbH, Darmstadt, Germany)	[413]Tönjes 2010
TwinsUK	Twins study	European	1864 (1864)	1864 (1864)	Illumina HumanExome 12 v1A	RareMetalWorker	221,836	DEXA	Hologic Discovery W' - QDR software version 12.6	[414]Moayyeri 2013
Vejle cases	Case-control	European	1971 (757)	1971 (757)	Illumina HumanExome 12 v1	RareMetalWorker	227,684	BIA	Tanita Body Composition Analyzer 418, IL, USA	[415]Pedersen 2016
Vejle controls	Case-control	European	430 (286)	430 (286)	Illumina HumanExome 12 v1	RareMetalWorker	227,684	BIA	Tanita Body Composition Analyzer 418, IL, USA	[415]Pedersen 2016
Women's Health initiative	Population-based	European	4109 (4109)	4093 (4093)	Illumina ExomeChip v1.0	RVTest	246,321	DEXA	DXA; QDR2000, 2000+, or 4500W; Hologic Inc.	[416]Carty 2012 [417]Kan 2016
Women's Health initiative	Population-based	Afro-American	387 (387)	385 (385)	Illumina ExomeChip v1.0	RVTest	246,321	DEXA	DXA; QDR2000, 2000+, or 4500W; Hologic Inc.	[416]Carty 2012 [417]Kan 2016
YFS	Population-based	European	288 (145)	288 (145)	Illumina CoreExome v1.0b	RVTest	237,099	BIA	TBF 300P balance from Tanita (Neuilly-sur-Seine, France)	[418]Raitakari 2008

Supplementary Table 3.2: Regions of long-range linkage disequilibrium

Regions of long-range LD reported by Koch *et al.*[296]

Chr	Start position (bp)	End position (bp)
1	48,000,000	52,000,000
2	86,000,000	100,500,000
2	134,500,000	138,000,000
2	183,000,000	190,000,000
3	47,500,000	50,000,000
3	83,500,000	87,000,000
3	89,000,000	97,500,000
5	44,000,000	51,500,000
5	98,000,000	100,500,000
5	129,000,000	132,000,000
5	135,500,000	138,500,000
6	25,000,000	33,500,000
6	57,000,000	64,000,000
6	140,000,000	142,500,000
7	55,000,000	66,000,000
8	8,000,000	12,000,000
8	43,000,000	50,000,000
10	37,000,000	43,000,000
11	45,000,000	57,000,000
11	87,500,000	90,500,000
12	33,000,000	40,000,000
12	109,500,000	112,000,000
20	32,000,000	34,500,000

Supplementary Table 3.3: Variants reaching exome-wide significance for BF% in European and all ancestry exome chip meta-analyses

rsid	Chr	Pos	Gene	Consequence	OA	EA	EAF	Most significant analysis	beta	p	N	Other significant analyses
rs1384588816	1	72,765,116	*NEGR1*	intronic	A	G	0.612	Eur sex-comb add	0.025	3.7E-08	107,902	All anc sex-comb add
rs1308970187	1	72,812,440	*NEGR1*	intergenic	A	G	0.611	Eur sex-comb add	0.025	7.4E-08	102,633	All anc sex-comb add
rs1423754764	1	72,954,611	*NEGR1*	intergenic	C	A	0.358	Eur sex-comb add	-0.027	1.5E-07	86,985	All anc sex-comb add
rs1003699594	1	177,889,480	*SEC16B*	intergenic	G	A	0.194	All anc sex-comb add	0.033	2.9E-09	109,524	Eur sex-comb add
rs2947411	2	614,168	*TMEM18*	intergenic	G	A	0.830	Eur sex-comb add	0.035	5.6E-09	104,983	All anc sex-comb add, all anc women add
rs2867125	2	622,827	*TMEM18*	intergenic	C	T	0.832	Eur sex-comb add	0.035	1.3E-08	104,983	All anc sex-comb add
rs6548238	2	634,905	*TMEM18*	intergnic	C	T	0.833	Eur sex-comb add	0.034	1.5E-08	107,902	All anc sex-comb add
rs11676272	2	25,141,538	*ADCY3*	missense	G	A	0.455	Eur sex-comb add	0.025	4.4E-08	103,105	All anc sex-comb add
rs713586	2	25,158,008	*DNAJC27*	intergenic	C	T	0.472	Eur sex-comb add	0.025	1.3E-08	107,902	All anc sex-comb add, All anc sex-comb rec, Eur sex-comb rec, All anc women rec, Eur women add, Eur women rec
rs4665736	2	25,187,599	*DNAJC27*	intronic	T	C	0.533	Eur sex-comb add	-0.026	7.6E-09	107,902	All anc sex-comb add, Eur women add
rs10179126	2	165,511,794	*COBLL1*	intergenic	G	C	0.371	All anc sex-comb add	0.034	8.3E-11	83,438	Eur sex-comb add
rs10195252	2	165,513,091	*COBLL1*	intergenic	C	T	0.399	All anc sex-comb add	0.036	1.1E-11	81,987	Eur sex-comb add, All anc men add, Eur men add
rs13389219	2	165,528,876	*COBLL1*	intergenic	T	C	0.399	All anc sex-comb add	0.033	2.7E-11	91,526	Eur sex-comb add, All anc men add, Eur men add
rs7578326	2	227,020,653	*LOC646736*	intronic	G	A	0.351	All anc sex-comb add	0.031	8.6E-12	112,443	All anc sex-comb rec, Eur sex-comb add, Eur sex-comb rec, All anc men add, Eur men add
rs2943634	2	227,068,080	*LOC646736*	intergenic	C	A	0.659	All anc sex-comb add	-0.031	9.7E-12	112,443	All anc sex-comb rec, Eur sex-comb add, Eur sex-comb rec, All anc men add, All anc men rec, Eur men add, Eur men rec
rs2943641	2	227,093,745	*LOC646736*	intergenic	C	T	0.640	All anc sex-comb add	-0.032	3.6E-12	112,443	All anc sex-comb rec, Eur sex-comb add, Eur sex-comb rec, All anc men add, All anc men rec, Eur men add, Eur men rec
rs2972146	2	227,100,698	*LOC646736*	intergenc	T	G	0.639	All anc men add	-0.045	2.5E-12	56,037	All anc sex-comb add, All anc sex-comb rec, Eur sex-comb add, Eur sex-comb rec, All anc men rec, Eur men add, Eur men rec
rs17036101	3	12,277,845	*PPARG*	intergenic	A	G	0.068	All anc sex-comb add	0.048	2.0E-07	98,889	
rs1801282	3	12,393,125	*PPARG*	missense	G	C	0.128	All anc sex-comb add	0.050	5.3E-14	106,347	Eur sex-comb add, All anc men add, Eur men add
rs10938397	4	45,182,527	*GNPDA2*	intergenic	G	A	0.438	All anc sex-comb add	0.028	1.4E-08	90,937	Eur sex-comb add, All anc men add
rs459193	5	55,806,751	*C5orf67*	downstream	G	A	0.735	All anc sex-comb add	-0.026	1.2E-07	109,137	
rs10968576	9	28,414,339	*LINGO2*	intronic	G	A	0.316	All anc sex-comb add	0.030	1.7E-10	112,443	Eur sex-comb add, Eur men add
rs11057401	12	124,427,306	*CCDC92*	missense	A	T	0.314	all anc women add	0.037	1.2E-07	50,082	
rs2904880	16	28,944,396	*CD19*	missense	G	C	0.720	All anc sex-comb rec	-0.053	6.5E-08	51,224	Eur sex-comb rec
rs9930333	16	53,799,977	*FTO*	intronic	G	T	0.432	Eur sex-comb add	0.052	1.6E-23	86,985	All anc sex-comb add, All anc sex-comb rec, Eur sex-comb rec, All anc men add, All anc men rec, Eur men add, Eur men rec, all anc women add, All anc women rec, Eur women add, Eur women rec

rsid	Chr	Pos	Gene	Consequence	OA	EA	EAF	Most significant analysis	beta	p	N	Other significant analyses
rs1421085	16	53,800,954	*FTO*	intronic	C	T	0.406	All anc sex-comb add	0.055	2.9E-32	112,443	All anc sex-comb rec, Eur sex-comb add, Eur sex-comb rec, All anc men add, All anc men rec, Eur men add, Eur men rec, all anc women add, All anc women rec, Eur women add, Eur women rec
rs1558902	16	53,803,574	*FTO*	intronic	A	T	0.405	All anc sex-comb add	0.053	8.7E-24	85,430	All anc sex-comb rec, Eur sex-comb add, Eur sex-comb rec, All anc men add, All anc men rec, Eur men add, Eur men rec, all anc women add, All anc women rec, Eur women add
rs1121980	16	53,809,247	*FTO*	intronic	A	G	0.431	Eur sex-comb add	0.052	3.3E-29	107,902	All anc sex-comb add, All anc sex-comb rec, Eur sex-comb rec, All anc men add, All anc men rec, Eur men add, Eur men rec, all anc women add, All anc women rec, Eur women add, Eur women rec
rs17817449	16	53,813,367	*FTO*	intronic	G	T	0.400	All anc sex-comb add	0.052	2.9E-29	112,443	All anc sex-comb rec, Eur sex-comb add, Eur sex-comb rec, All anc men add, All anc men rec, Eur men add, Eur men rec, all anc women add, All anc women rec, Eur women add, Eur women rec
rs11075987	16	53,815,161	*FTO*	intronic	G	T	0.499	Eur sex-comb add	0.042	7.3E-17	86,985	All anc sex-comb add, All anc sex-comb rec, Eur sex-comb rec, All anc men add, All anc men rec, Eur men add, Eur men rec, Eur women add
rs8050136	16	53,816,275	*FTO*	intronic	A	C	0.400	All anc sex-comb add	0.052	3.7E-29	112,443	All anc sex-comb rec, Eur sex-comb add, Eur sex-comb rec, All anc men add, All anc men rec, Eur men add, Eur men rec, all anc women add, All anc women rec, Eur women add, Eur women rec
rs9939609	16	53,820,527	*FTO*	intronic	A	T	0.401	All anc sex-comb add	0.051	1.6E-27	106,347	All anc sex-comb rec, Eur sex-comb add, Eur sex-comb rec, All anc men add, All anc men rec, Eur men add, Eur men rec, all anc women add, Eur women add
rs9941349	16	53,825,488	*FTO*	intronic	T	C	0.417	Eur sex-comb add	0.051	1.8E-27	107,902	All anc sex-comb add, All anc sex-comb rec, Eur sex-comb rec, All anc men add, All anc men rec, Eur men add, Eur men rec, all anc women add, All anc women rec, Eur women add, Eur women rec
rs9930506	16	53,830,465	*FTO*	intronic	G	A	0.435	Eur sex-comb add	0.048	1.4E-22	97,726	All anc sex-comb add, All anc sex-comb rec, Eur sex-comb rec, All anc men add, All anc men rec, Eur men add, Eur men rec, all anc women add, Eur women add
rs571312	18	57,839,769	*MC4R*	intergenic	A	C	0.235	All anc sex-comb add	0.028	5.2E-08	112,443	
rs17782313	18	57,851,097	*MC4R*	intergenic	C	T	0.235	All anc sex-comb add	0.028	1.5E-07	109,524	
rs10871777	18	57,851,763	*MC4R*	intergenic	G	A	0.236	All anc sex-comb add	0.027	1.8E-07	112,443	
rs1800437	19	46,181,392	*GIPR*	downstream	C	G	0.207	All anc sex-comb add	-0.032	4.1E-09	104,742	Eur sex-comb add
rs2284063	22	38,544,298	*PLA2G6*	intronic	G	A	0.354	All anc sex-comb add	0.029	2.1E-10	112,443	All anc sex-comb rec, Eur sex-comb add
rs738322	22	38,569,006	*PLA2G6*	intronic	G	A	0.469	All anc sex-comb add	0.035	1.6E-15	112,443	All anc sex-comb rec, Eur sex-comb add, Eur sex-comb rec, All anc men add, all anc women add, Eur women add

Supplementary Table 3.4: Variants reaching exome-wide significance for FFMI in European or all ancestry exome chip meta-analyses

rsid	Chr	Pos	Gene	Type	OA	EA	EAF	Most significant analysis	beta	p	N	Other significant analyses
rs141845046	1	154,987,704	*ZBTB7B*	missense	C	T	0.031	all anc sex-comb add	0.074	3.0E-08	98,035	Eur sex-comb add
rs2282301	1	155,868,625	*RIT1*	3' UTR	G	A	0.247	Eur sex-comb add	-0.037	3.5E-09	72,589	all anc sex-comb add
rs633715	1	177,852,580	*SEC16B*	intergenic	T	C	0.200	all anc sex-comb add	0.045	5.5E-12	76,740	Eur sex-comb add, all anc women add, Eur women add
rs543874	1	177,889,480	*SEC16B*	intergenic	A	G	0.193	all anc sex-comb add	0.050	4.9E-17	95,116	all anc sex-comb rec, Eur sex-comb add, Eur sex-comb rec, all anc women add, all anc women rec, Eur women add, Eur women rec
rs591120	1	177,902,753	*SEC16B*	missense	G	C	0.435	all anc sex-comb add	0.025	2.0E-07	91,939	
rs10913469	1	177,913,519	*SEC16B*	intronic	T	C	0.194	Eur women add	0.056	1.1E-11	47,404	all anc sex-comb add, Eur sex-comb add, all anc women add
rs2947411	2	614,168	*TMEM18*	intergenic	A	G	0.830	Eur sex-comb add	0.050	1.6E-14	90,575	all anc sex-comb add, all anc sex-comb rec, Eur sex-comb rec, all anc men add, all anc women add, all anc women rec, Eur women add, Eur women rec
rs2867125	2	622,827	*TMEM18*	intergenic	T	C	0.832	Eur sex-comb add	0.051	5.6E-15	90,575	all anc sex-comb add, all anc sex-comb rec, Eur sex-comb rec, all anc men add, Eur men add, all anc women add, all anc women rec, Eur women add, Eur women rec
rs6548238	2	634,905	*TMEM18*	intergenic	T	C	0.833	Eur sex-comb add	0.053	2.7E-16	93,494	all anc sex-comb add, all anc sex-comb rec, Eur sex-comb rec, all anc men add, all anc men rec, Eur men add, all anc women add, all anc women rec, Eur women add, Eur women rec
rs7561317	2	644,953	*TMEM18*	intergenic	A	G	0.839	Eur sex-comb add	0.050	5.9E-12	78,996	all anc sex-comb add, all anc sex-comb rec, Eur sex-comb rec
rs1062633	3	49,924,940	*MST1R*	missense	T	C	0.484	Eur sex-comb add	0.026	1.9E-07	93,494	
rs206936	6	34,302,869	*NUDT3*	intronic	A	G	0.212	all anc sex-comb add	0.030	1.2E-07	98,035	
rs2206277	6	50,798,526	*TFAP2B*	intronic	C	T	0.184	all anc sex-comb add	0.039	1.2E-10	96,171	Eur sex-comb add, all anc men add, Eur men add
rs987237	6	50,803,050	*TFAP2B*	intronic	A	G	0.183	all anc sex-comb add	0.036	1.3E-09	98,035	Eur sex-comb add, all anc men add, Eur men add
rs1167796	7	75,173,180	*HIP1*	intronic	A	G	0.567	Eur sex-comb add	0.029	5.8E-09	90,575	all anc sex-comb add
rs1167800	7	75,176,196	*HIP1*	intronic	G	A	0.569	all anc sex-comb add	0.026	3.7E-08	98,035	Eur sex-comb add
rs925946	11	27,667,202	*BDNF*	intronic	T	G	0.691	all anc sex-comb add	-0.029	2.0E-08	93,662	all anc sex-comb rec, Eur sex-comb add, Eur sex-comb rec
rs6265	11	27,679,916	*BDNF*	missense	C	T	0.183	all anc sex-comb add	-0.042	1.4E-11	93,252	Eur sex-comb add, all anc men add, Eur men add
rs4483821	15	84,488,636	*ADAMTSL3*	missense	A	G	0.447	Eur sex-comb add	-0.028	7.2E-09	93,494	all anc sex-comb add
rs7183263	15	84,573,041	*ADAMTSL3*	intronic	T	G	0.515	Eur sex-comb add	-0.031	5.0E-10	93,494	all anc sex-comb add, all anc sex-comb rec, Eur sex-comb rec
rs11259936	15	84,580,582	*ADAMTSL3*	intronic	A	C	0.515	Eur sex-comb add	-0.031	4.3E-10	93,494	all anc sex-comb add, all anc sex-comb rec, Eur sex-comb rec
rs4842838	15	84,582,124	*ADAMTSL3*	missense	G	T	0.516	Eur sex-comb add	-0.029	1.4E-08	88,225	all anc sex-comb add, all anc sex-comb rec, Eur sex-comb rec
rs34047645	15	84,611,367	*ADAMTSL3*	missense	G	C	0.183	Eur sex-comb add	0.041	7.0E-09	72,284	all anc sex-comb add

rsid	Chr	Pos	Gene	Type	OA	EA	EAF	Most significant analysis	beta	p	N	Other significant analyses
rs7191155	16	19,800,213	*IQCK*	missense	T	C	0.162	Eur sex-comb add	-0.039	5.1E-09	88,225	all anc sex-comb add
rs12444979	16	19,933,600	*GPRC5B*	upstream	C	T	0.134	Eur sex-comb add	-0.048	3.7E-12	93,494	all anc sex-comb add, Eur women add
rs9930333	16	53,799,977	*FTO*	intronic	T	G	0.435	Eur sex-comb add	0.061	4.5E-27	72,589	all anc sex-comb add, all anc sex-comb rec, Eur sex-comb rec, all anc men add, all anc men rec, Eur men add, Eur men rec, all anc women add, all anc women rec, Eur women add, Eur women rec
rs1421085	16	53,800,954	*FTO*	intronic	T	C	0.406	all anc sex-comb add	0.060	3.2E-34	98,035	all anc sex-comb rec, Eur sex-comb add, Eur sex-comb rec, all anc men add, all anc men rec, Eur men add, Eur men rec, all anc women add, all anc women rec, Eur women add, Eur women rec
rs1558902	16	53,803,574	*FTO*	intronic	T	A	0.407	all anc sex-comb add	0.064	6.1E-29	71,034	all anc sex-comb rec, Eur sex-comb add, Eur sex-comb rec, all anc men add, all anc men rec, Eur men add, Eur men rec, all anc women add, all anc women rec, Eur women add, Eur women rec
rs1121980	16	53,809,247	*FTO*	intronic	G	A	0.433	Eur sex-comb add	0.057	2.0E-30	93,494	all anc sex-comb add, all anc sex-comb rec, Eur sex-comb rec, all anc men add, all anc men rec, Eur men add, Eur men rec, all anc women add, all anc women rec, Eur women add, Eur women rec
rs17817449	16	53,813,367	*FTO*	intronic	T	G	0.401	all anc sex-comb add	0.058	5.5E-32	98,035	all anc sex-comb rec, Eur sex-comb add, Eur sex-comb rec, all anc men add, all anc men rec, Eur men add, Eur men rec, all anc women add, all anc women rec, Eur women add, Eur women rec
rs11075987	16	53,815,161	*FTO*	intronic	T	G	0.501	Eur sex-comb add	0.051	2.2E-20	72,589	all anc sex-comb add, all anc sex-comb rec, Eur sex-comb rec, all anc men add, all anc men rec, Eur men add, all anc women add, Eur women add, Eur women rec, ,
rs8050136	16	53,816,275	*FTO*	intronic	C	A	0.401	all anc sex-comb add	0.059	2.3E-32	98,035	all anc sex-comb rec, Eur sex-comb add, Eur sex-comb rec, all anc men add, all anc men rec, Eur men add, Eur men rec, all anc women add, all anc women rec, Eur women add, Eur women rec
rs9939609	16	53,820,527	*FTO*	intronic	T	A	0.403	all anc sex-comb add	0.059	9.4E-31	91,939	all anc sex-comb rec, Eur sex-comb add, Eur sex-comb rec, all anc men add, all anc men rec, Eur men add, Eur men rec, all anc women add, all anc women rec, Eur women add, Eur women rec
rs9941349	16	53,825,488	*FTO*	intronic	C	T	0.413	all anc sex-comb add	0.055	2.1E-29	98,035	all anc sex-comb rec, Eur sex-comb add, Eur sex-comb rec, all anc men add, all anc men rec, Eur men add, Eur men rec, all anc women add, all anc women rec, Eur women add, Eur women rec
rs9930506	16	53,830,465	*FTO*	intronic	A	G	0.430	all anc sex-comb add	0.051	8.4E-23	87,875	all anc sex-comb rec, Eur sex-comb add, Eur sex-comb rec, all anc men add, all anc men rec, Eur men add, Eur men rec, all anc women add, all anc women rec, Eur women add, Eur women rec
rs571312	18	57,839,769	*MC4R*	intergenic	C	A	0.235	all anc sex-comb add	0.042	3.8E-14	98,035	Eur sex-comb add, all anc men add, Eur men add
rs17782313	18	57,851,097	*MC4R*	intergenic	T	C	0.235	all anc sex-comb add	0.042	1.8E-13	95,116	all anc sex-comb rec, Eur sex-comb add, all anc men add, Eur men add
rs10871777	18	57,851,763	*MC4R*	intergenic	A	G	0.237	all anc sex-comb add	0.042	5.8E-14	98,035	all anc sex-comb rec, Eur sex-comb add, all anc men add, Eur men add
rs489693	18	57,882,787	*MC4R*	intergenic	C	A	0.312	all anc sex-comb add	0.032	3.3E-10	98,035	all anc sex-comb rec, Eur sex-comb add, all anc men add, Eur men add, Eur men rec

rsid	Chr	Pos	Gene	Type	OA	EA	EAF	Most significant analysis	beta	p	N	Other significant analyses
rs12970134	18	57,884,750	*MC4R*	intergenic	G	A	0.260	all anc sex-comb add	0.035	3.8E-10	92,766	Eur sex-comb add, all anc men add, Eur men add

Supplementary Table 3.5: Independent variants associated with BF% at genome-wide significance in a European meta-GWAS

rsid	Chr	Pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
Sex-combined										
rs72634814	1	1,538,046	*C1orf233*	intergenic	A	G	0.35	-0.013±0.0016	1.1E-15	442,278
rs3762444	1	2,427,712	*PLCH2*	intronic	T	C	0.46	-0.0084±0.0014	7.1E-09	506,306
rs12124126	1	6,660,349	*KLHL21*	intronic	A	G	0.67	0.0113±0.0016	6.9E-13	442,278
rs4908676	1	7,737,099	*CAMTA1*	intronic	A	G	0.54	-0.0101±0.0014	2.3E-12	507,570
rs301798	1	8,488,565	*LOC102724552*	ncRNA exonic	A	G	0.66	-0.0106±0.0015	2.1E-12	506,306
rs1318408	1	11,925,781	*NPPB*	intergenic	A	G	0.88	-0.0139±0.0022	2.1E-10	530,194
rs7540681	1	16,832,446	*CROCCP3*	intergenic	T	C	0.52	-0.0104±0.0015	1.5E-12	492,344
rs779765342	1	23,353,906	*KDM1A*	intronic	I	D	0.83	-0.0138±0.002	4.2E-12	442,278
rs747686078	1	29,453,545	*TMEM200B*	intergenic	I	D	0.11	-0.0154±0.0024	2.4E-10	442,278
rs909001	1	32,196,647	*ADGRB2*	exonic	C	G	0.83	-0.0124±0.002	2.8E-10	442,278
rs12022461	1	33,232,525	*KIAA1522*	intronic	A	G	0.17	-0.0111±0.0019	4.8E-09	508,103
rs9426003	1	34,602,870	*CSMD2*	intronic	A	G	0.30	-0.0106±0.0016	1.4E-11	506,310
rs4660590	1	42,441,288	*HIVEP3*	intergenic	A	G	0.46	0.0114±0.0014	1.5E-15	508,070
rs56158851	1	42,784,893	*FOXJ3*	intronic	I	D	0.69	0.008±0.0016	7.2E-07	442,278
rs115092994	1	46,185,726	*IPP*	intronic	A	G	0.96	0.0284±0.0038	4.4E-14	442,278
rs11211481	1	47,694,167	*TAL1*	intronic	A	G	0.41	0.0116±0.0014	1.4E-15	508,101
rs7531656	1	49,828,663	*AGBL4*	intronic	A	G	0.32	0.0134±0.0015	4.9E-19	529,754
rs630602	1	54,728,864	*SSBP3*	intronic	C	G	0.61	0.0091±0.0015	8.0E-10	505,839
rs10493217	1	57,820,637	*DAB1*	intronic	T	C	0.09	0.015±0.0025	2.0E-09	531,556
rs41313250	1	62,483,623	*INADL*	intronic	A	G	0.88	0.0215±0.0023	3.7E-20	442,278
rs1013293	1	62,570,321	*INADL*	intronic	A	G	0.43	-0.0143±0.0014	5.2E-23	508,072
rs7519259	1	66,434,743	*PDE4B*	intronic	A	G	0.53	0.0102±0.0014	1.3E-12	500,064
rs1993709	1	72,838,529	*NEGR1*	intergenic	A	G	0.20	-0.02±0.0018	2.1E-29	531,385
rs1514173	1	74,995,110	*FPGT-TNNI3K,TNNI3K*	intronic	T	C	0.40	0.0084±0.0014	4.8E-09	530,219
rs71658797	1	77,967,507	*AK5*	intronic	A	T	0.12	0.0237±0.0023	1.2E-25	442,278
rs17391694	1	78,623,626	*GIPC2*	intergenic	T	C	0.14	0.0204±0.0021	4.3E-23	529,785
rs6688826	1	80,812,329	*LOC101927412*	intergenic	T	C	0.70	-0.01±0.0016	1.4E-10	508,105
rs2642183	1	84,603,487	*PRKACB*	intronic	A	G	0.24	-0.01±0.0017	1.9E-09	508,099
rs11165643	1	96,924,097	*LOC101928241*	intergenic	T	C	0.59	0.015±0.0014	1.3E-25	531,517
rs12072739	1	98,315,893	*DPYD*	intronic	A	G	0.77	-0.0141±0.0018	1.9E-15	442,278
rs2077569	1	103,350,876	*COL11A1*	intronic	A	G	0.61	-0.0093±0.0015	8.5E-10	442,278
rs11434203	1	107,623,489	*PRMT6*	intergenic	D	I	0.34	0.0104±0.0016	3.8E-11	442,278
rs7550711	1	110,082,886	*GPR61*	intronic	T	C	0.03	0.0458±0.0044	1.7E-25	525,851
rs3768486	1	110,123,971	*GNAI3*	intronic	A	G	0.82	-0.0114±0.0018	5.2E-10	529,292
rs12120956	1	113,202,571	*CAPZA1*	intronic	A	G	0.20	-0.0147±0.0018	7.4E-17	508,065
rs11205303	1	149,906,413	*MTMR11*	missense	T	C	0.59	-0.0182±0.0014	3.2E-36	526,146
rs114529840	1	150,457,517	*TARS2*	intergenic	T	G	0.94	0.0293±0.0032	1.7E-20	442,278
rs2305813	1	150,981,267	*PRUNE*	intronic	C	G	0.12	-0.016±0.0022	1.2E-12	442,278
rs79100766	1	151,690,804	*CELF3*	intergenic	I	D	0.48	-0.0094±0.0015	3.0E-10	442,278
rs35154152	1	155,172,725	*THBS3*	missense	T	C	0.89	0.0195±0.0024	4.2E-16	442,278
rs76102184	1	156,171,486	*SLC25A44*	intronic	T	C	0.02	0.0368±0.005	1.3E-13	442,278
rs12046534	1	170,709,593	*PRRX1*	intergenic	T	G	0.45	-0.0091±0.0015	1.8E-09	442,278
rs77779886	1	173,740,028	*KLHL20*	intronic	A	G	0.02	-0.0311±0.005	3.4E-10	442,278
rs189194497	1	174,347,661	*RABGAP1L*	intronic	T	C	0.99	0.0568±0.0082	3.7E-12	442,278
rs77560793	1	175,001,179	*MRPS14*	intergenic	A	G	0.03	-0.0298±0.0043	5.8E-12	442,278

rsid	Chr	Pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs543874	1	177,889,480	*SEC16B*	intergenic	A	G	0.80	-0.0283±0.0017	2.8E-59	531,565
rs746336168	1	184,706,212	*EDEM3*	intronic	D	I	0.48	0.0099±0.0015	4.5E-11	442,278
rs491055	1	190,308,834	*BRINP3*	intronic	A	G	0.44	0.0088±0.0015	3.2E-09	442,278
1:195029618_C_CT	1	195,029,618	*KCNT2*	intergenic	I	D	0.31	-0.0116±0.0017	3.7E-12	442,278
rs2820295	1	201,800,868	*IPO9*	intronic	A	G	0.34	0.0134±0.0015	6.5E-19	508,032
rs16849710	1	202,106,797	*ARL8A*	intronic	A	G	0.47	0.0081±0.0014	1.9E-08	505,010
rs2802774	1	203,527,812	*OPTC*	intergenic	A	C	0.55	0.0108±0.0015	7.2E-13	442,278
rs2356722	1	209,524,734	*MIR205HG*	intergenic	A	G	0.71	-0.0093±0.0016	3.9E-09	508,063
rs17015701	1	210,337,691	*SYT14*	downstream	A	G	0.19	0.0114±0.0018	2.0E-10	531,566
rs528296052	1	213,287,698	*RPS6KC1*	intronic	I	D	0.66	-0.0098±0.0016	6.7E-10	442,278
rs1377184	1	215,246,432	*KCNK2*	intronic	A	T	0.25	-0.0125±0.0017	2.8E-13	442,278
rs11118310	1	219,637,671	*LYPLAL1*	intergenic	A	T	0.41	0.0206±0.0015	1.5E-45	508,106
rs11118893	1	222,085,672	*LOC101929771*	intergenic	A	G	0.70	0.0104±0.0016	2.7E-11	505,804
rs10915840	1	225,668,524	*ENAH*	intergenic	A	G	0.27	-0.0096±0.0016	2.2E-09	507,936
rs11122450	1	230,301,811	*GALNT2*	intronic	T	G	0.39	0.0095±0.0015	2.2E-10	465,732
rs12042959	1	243,533,273	*SDCCAG8*	intronic	A	G	0.85	0.0133±0.002	5.9E-11	508,044
rs7549780	1	243,712,455	*AKT3*	intronic	A	C	0.32	0.0071±0.0015	4.4E-06	507,190
rs62106258	2	417,167	*FAM150B*	intergenic	T	C	0.95	0.0504±0.0034	1.7E-48	442,278
rs13415094	2	653,093	*TMEM18*	intergenic	T	C	0.83	0.0291±0.0019	1.7E-54	530,246
rs4669869	2	12,898,460	*TRIB2*	intergenic	T	C	0.55	-0.0092±0.0014	1.4E-10	508,027
rs78265103	2	24,468,191	*ITSN2*	intronic	A	T	0.95	0.0205±0.0034	9.5E-10	442,278
rs6752378	2	25,150,116	*ADCY3*	intergenic	A	C	0.48	0.0228±0.0014	8.6E-59	530,088
rs934778	2	25,389,224	*POMC*	intronic	A	G	0.68	-0.0188±0.0015	2.0E-35	531,497
rs1627854	2	26,953,540	*KCNK3*	UTR3	A	G	0.53	-0.0115±0.0014	6.9E-16	505,091
rs3770799	2	36,788,616	*FEZ2*	intronic	A	G	0.63	-0.0122±0.0015	1.6E-16	506,313
rs7580551	2	41,718,026	*LOC388942*	intergenic	A	G	0.64	-0.0123±0.0015	1.2E-16	508,081
rs786420	2	44,719,893	*CAMKMT*	intronic	T	C	0.76	0.0099±0.0017	3.3E-09	507,160
rs58569862	2	46,896,921	*SOCS5*	intergenic	T	C	0.79	0.0144±0.0018	3.4E-15	442,278
rs58120873	2	47,313,562	*C2orf61*	downstream	A	G	0.09	-0.0163±0.0026	6.3E-10	442,278
rs555219016	2	50,215,757	*NRXN1*	intronic	D	I	0.60	-0.0105±0.0015	1.4E-11	442,278
2:50759450_AT_A	2	50,759,450	*NRXN1*	intronic	D	I	0.45	0.0098±0.0016	3.7E-10	442,278
rs10176888	2	50,861,049	*NRXN1*	intronic	A	T	0.75	-0.0098±0.0016	2.7E-09	508,000
rs7601895	2	55,281,901	*RTN4*	intergenic	C	G	0.69	0.0101±0.0016	9.2E-11	503,609
rs13432055	2	56,603,985	*CCDC85A*	intronic	T	C	0.71	-0.0096±0.0016	9.8E-10	531,000
rs6754296	2	58,002,882	*VRK2*	intergenic	T	G	0.92	-0.0161±0.0027	1.8E-09	505,116
rs10197655	2	58,791,420	*LINC01122*	ncRNA intronic	A	G	0.43	-0.0114±0.0015	2.2E-14	442,278
rs6545714	2	59,307,725	*LINC01122*	intergenic	A	G	0.60	-0.0143±0.0014	2.6E-23	531,512
rs6545966	2	62,861,353	*EHBP1*	intergenic	A	G	0.52	0.0089±0.0015	1.8E-09	442,278
rs12477088	2	67,841,326	*LOC101927701*	intergenic	T	C	0.59	0.0111±0.0014	2.0E-14	508,083
rs4453725	2	69,659,126	*NFU1*	missense	A	T	0.59	0.0127±0.0014	2.3E-18	508,101
rs396354	2	86,850,022	*RNF103*	UTR5	T	C	0.28	0.01±0.0016	2.5E-10	507,230
rs2871344	2	100,839,430	*LINC01104*	ncRNA intronic	A	G	0.60	0.0129±0.0014	2.7E-19	531,486
rs6707445	2	104,420,858	*LOC100287010*	intergenic	A	G	0.45	0.0092±0.0014	1.9E-10	506,310
rs445077	2	105,404,701	*LINC01158*	intergenic	A	T	0.48	0.0113±0.0015	2.9E-14	442,278
rs4676084	2	110,010,962	*SH3RF3*	intronic	A	G	0.42	0.009±0.0014	4.3E-10	507,511
rs4848281	2	112,945,188	*FBLN7*	UTR3	A	T	0.74	-0.0106±0.0016	1.1E-10	507,196
rs10204422	2	133,528,953	*NCKAP5*	intronic	C	G	0.73	-0.0107±0.0017	1.9E-10	442,278
rs10496731	2	135,597,628	*ACMSD*	intronic	T	G	0.63	0.0108±0.0015	3.6E-13	508,084
rs4988235	2	136,608,646	*MCM6*	intronic	A	G	0.74	0.0097±0.0016	2.1E-09	529,763
rs4662318	2	144,016,529	*ARHGAP15*	intronic	T	C	0.16	-0.0117±0.002	2.6E-09	508,036

rsid	Chr	Pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs453520	2	147,907,202	*PABPC1P2*	intergenic	T	C	0.58	-0.0115±0.0014	1.4E-15	507,146
rs62169721	2	151,471,941	*LOC101929282*	intergenic	T	G	0.03	0.0298±0.0047	2.3E-10	442,278
rs3771653	2	159,483,811	*PKP4*	intronic	T	C	0.63	0.0088±0.0015	7.2E-09	442,278
rs6711375	2	161,090,873	*LOC100505984*	intergenic	A	G	0.68	0.0103±0.0015	1.8E-11	508,038
rs939643	2	165,387,977	*GRB14*	intronic	T	G	0.43	0.0097±0.0014	2.1E-11	508,068
rs1128249	2	165,528,624	*COBLL1*	intergenic	T	G	0.39	0.0176±0.0014	3.3E-34	531,493
rs3754963	2	166,185,707	*SCN2A*	intronic	A	T	0.74	0.0105±0.0016	1.2E-10	508,073
rs17694506	2	171,632,225	*ERICH2*	intronic	T	C	0.61	-0.0098±0.0015	3.6E-11	496,879
rs12997093	2	172,564,438	*DYNC1I2*	intronic	T	C	0.41	-0.0094±0.0014	7.2E-11	507,624
rs34234296	2	175,166,636	*LINC01305*	intergenic	A	G	0.39	-0.0117±0.0015	2.5E-14	442,278
rs374500261	2	175,521,358	*WIPF1*	intronic	A	G	0.99	0.050±0.009	4.48E-09	442,278
rs4972476	2	176,426,297	*KIAA1715*	intergenic	A	G	0.11	-0.017±0.0023	2.8E-13	508,078
rs11679338	2	181,606,895	*SCHLAP1*	ncRNA intronic	T	C	0.66	0.0117±0.0015	9.2E-15	495,637
rs12478299	2	193,811,641	*PCGEM1*	intergenic	T	C	0.75	0.0109±0.0016	4.1E-11	508,020
rs2043016	2	198,146,381	*ANKRD44-IT1*	ncRNA intronic	T	C	0.38	0.0104±0.0015	1.7E-12	506,311
rs12997625	2	202,970,250	*KIAA2012*	intronic	T	C	0.53	0.0098±0.0014	6.4E-12	508,096
rs11693128	2	204,115,901	*CYP20A1*	intronic	A	G	0.45	-0.0092±0.0014	9.1E-11	530,216
rs4482463	2	205,375,909	*PARD3B*	intergenic	A	C	0.92	-0.0224±0.0027	1.7E-16	494,260
rs7569010	2	206,084,518	*PARD3B*	intronic	C	G	0.43	0.01±0.0014	4.1E-12	507,146
rs16846136	2	212,291,371	*ERBB4*	intronic	A	C	0.27	0.0107±0.0016	3.6E-11	508,094
rs7599312	2	213,413,231	*ERBB4*	intergenic	A	G	0.27	-0.0097±0.0016	1.2E-09	531,505
rs2712169	2	217,671,349	*TNP1*	intergenic	A	G	0.58	-0.0108±0.0015	1.2E-13	508,046
rs952227	2	227,062,080	*LOC646736*	intergenic	A	G	0.33	0.0171±0.0015	2.2E-30	530,620
rs200882902	2	228,996,202	*SPHKAP*	intronic	I	D	0.33	0.0133±0.0016	1.2E-16	442,278
rs3791837	2	230,681,905	*TRIP12*	intronic	A	G	0.67	-0.0141±0.0015	2.0E-20	508,090
rs71036299	2	236,819,547	*AGAP1*	intronic	D	I	0.28	0.0106±0.0018	3.4E-09	442,278
rs3085964	2	242,021,008	*SNED1*	intronic	D	I	0.86	-0.0132±0.0022	2.9E-09	442,278
rs11542009	3	9,517,369	*SETD5*	missense	T	C	0.10	0.0152±0.0024	4.5E-10	442,278
rs7649970	3	12,392,272	*PPARG*	intronic	T	C	0.12	0.0289±0.0022	3.6E-41	530,258
rs6809832	3	12,414,420	*PPARG*	intronic	T	C	0.85	0.019±0.002	1.3E-21	531,557
rs17819328	3	12,489,342	*PPARG*	intergenic	T	G	0.57	0.0068±0.0014	1.8E-06	529,267
rs2600224	3	12,933,866	*IQSEC1*	intergenic	T	C	0.68	-0.0083±0.0015	7.3E-08	506,310
rs4619804	3	18,674,644	*SATB1-AS1*	intergenic	A	C	0.26	-0.0131±0.0016	1.1E-15	506,312
rs17016133	3	25,313,167	*RARB*	intronic	T	C	0.87	0.0125±0.0021	3.9E-09	508,101
rs602543	3	27,547,083	*SLC4A7*	intergenic	T	C	0.40	0.0095±0.0015	6.4E-11	507,190
rs10490871	3	35,667,761	*ARPP21*	intergenic	A	G	0.63	-0.0102±0.0015	5.0E-12	508,098
rs9816029	3	41,311,362	*ULK4*	intronic	C	G	0.34	0.0102±0.0015	1.3E-11	508,041
rs33485	3	42,417,982	*LYZL4*	intergenic	T	C	0.74	-0.0112±0.0016	5.8E-12	508,062
rs7637852	3	44,041,777	*MIR138-1*	intergenic	A	G	0.30	0.0123±0.0016	3.0E-15	507,567
rs33807	3	45,334,692	*TMEM158*	intergenic	A	G	0.12	0.0134±0.0022	1.9E-09	502,248
rs62259939	3	49,386,047	*USP4*	intergenic	A	G	0.44	0.0131±0.0015	1.3E-18	442,278
rs2681780	3	49,897,830	*CAMKV*	intronic	T	C	0.51	0.0155±0.0014	3.5E-27	508,075
rs2612012	3	53,745,625	*CACNA1D*	intronic	A	C	0.75	0.0105±0.0017	2.3E-10	505,829
rs1916801	3	61,187,046	*FHIT*	intronic	A	T	0.59	0.0108±0.0014	5.4E-14	531,322
rs9968060	3	62,471,282	*CADPS*	intronic	T	C	0.64	0.0112±0.0015	1.5E-13	505,028
rs9864077	3	64,704,891	*ADAMTS9-AS2*	ncRNA intronic	T	C	0.70	-0.0118±0.0015	2.5E-14	530,071
rs3856595	3	66,492,604	*LRIG1*	intronic	T	G	0.45	-0.0091±0.0015	1.0E-09	442,278
rs1491592	3	70,654,244	*FOXP1*	intergenic	T	C	0.55	0.0088±0.0015	3.6E-09	442,278
rs17007949	3	70,920,041	*FOXP1*	intergenic	C	G	0.31	0.0091±0.0016	5.2E-09	506,312
rs9866090	3	71,585,144	*FOXP1*	intronic	A	G	0.23	0.0123±0.0017	6.2E-13	507,576
rs13080520	3	81,850,742	*GBE1*	intergenic	T	C	0.64	-0.0105±0.0015	1.6E-12	507,604
rs76345589	3	84,185,140	*LINC00971*	intergenic	C	G	0.93	0.0201±0.003	1.2E-11	442,278

rsid	Chr	Pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs12495178	3	85,886,077	*CADM2*	intronic	T	C	0.64	0.0116±0.0015	3.9E-15	529,741
rs9826925	3	88,206,392	*C3orf38*	UTR3	T	G	0.88	0.016±0.0022	9.1E-13	506,312
rs1495824	3	88,733,658	*EPHA3*	intergenic	A	G	0.50	-0.0105±0.0014	2.0E-13	508,063
rs4857329	3	94,036,952	*NSUN3*	intergenic	A	G	0.48	0.013±0.0014	9.6E-20	508,060
rs4577503	3	99,657,922	*MIR548G*	ncRNA intronic	A	G	0.59	-0.0088±0.0015	1.6E-09	508,095
rs2141180	3	101,179,057	*SENP7*	intronic	T	C	0.38	-0.0095±0.0015	6.0E-11	531,559
rs1436348	3	104,612,668	*ALCAM*	intergenic	A	G	0.42	-0.0117±0.0014	6.8E-16	508,076
rs1078455	3	108,129,348	*MYH15*	intronic	T	C	0.69	-0.0105±0.0016	1.4E-11	506,313
rs2124499	3	123,093,541	*ADCY5*	intronic	C	G	0.37	-0.0131±0.0015	4.8E-19	517,754
rs9820766	3	123,264,017	*HACD2*	intronic	T	C	0.69	-0.0103±0.0015	3.0E-11	508,085
rs9847672	3	131,618,541	*CPNE4*	intronic	T	C	0.28	0.0141±0.0016	1.0E-18	508,079
rs10935143	3	134,665,159	*EPHB1*	intronic	A	G	0.45	-0.0083±0.0014	8.5E-09	507,618
rs2293251	3	138,124,114	*MRAS*	UTR3	T	G	0.84	0.0132±0.002	2.1E-11	508,104
rs11917587	3	141,207,575	*RASA2*	intronic	A	G	0.43	0.0121±0.0014	3.2E-17	530,992
rs62271373	3	150,066,540	*LINC01214*	intergenic	A	T	0.06	-0.024±0.0032	4.5E-14	442,278
rs7652639	3	153,183,198	*C3orf79*	intergenic	T	C	0.55	0.0091±0.0014	2.6E-10	508,102
rs1568488	3	153,657,951	*ARHGEF26-AS1*	intergenic	C	G	0.60	0.0115±0.0015	6.3E-15	506,312
rs9289970	3	156,301,324	*SSR3*	intergenic	A	C	0.85	0.0124±0.002	8.1E-10	497,414
rs6802443	3	157,888,438	*RSRC1*	intronic	A	G	0.44	0.0098±0.0014	1.4E-11	496,946
rs8192675	3	170,724,883	*SLC2A2*	intronic	T	C	0.71	-0.0095±0.0016	1.1E-09	531,548
rs4894808	3	171,833,266	*FNDC3B*	intronic	C	G	0.40	-0.0093±0.0015	5.1E-10	506,739
rs488029	3	173,119,378	*NLGN1*	intronic	A	G	0.53	0.0103±0.0014	8.5E-13	503,321
rs11926346	3	173,659,090	*NLGN1*	intronic	A	G	0.79	0.0115±0.0018	5.9E-11	497,425
3:183513243_ATTTTTTTTTTTTT_A	3	183,513,243	*YEATS2*	intronic	D	I	0.53	-0.0099±0.0015	6.0E-11	442,278
rs1516725	3	185,824,004	*ETV5*	intronic	T	C	0.14	-0.0142±0.0021	5.1E-12	530,218
rs200915692	3	196,115,344	*UBXN7*	intronic	D	I	0.64	-0.0099±0.0016	2.2E-10	442,278
rs6782581	3	196,979,106	*DLG1*	intronic	C	G	0.56	0.0102±0.0014	1.6E-12	496,887
rs2279178	4	906,647	*GAK*	intronic	A	G	0.18	0.0132±0.0019	2.8E-12	502,923
rs2798297	4	3,064,004	*HTT-AS*	downstream	A	G	0.37	0.0102±0.0016	5.1E-11	442,278
rs2192527	4	18,329,824	*LCORL*	intergenic	A	G	0.54	-0.0139±0.0015	7.9E-21	442,278
rs1485554	4	20,120,863	*SLIT2*	intergenic	A	G	0.12	0.0167±0.0023	1.8E-13	442,278
rs7664617	4	21,862,428	*KCNIP4*	intronic	C	G	0.49	0.0083±0.0014	6.4E-09	508,106
4:25355782_AAAG_A	4	25,355,782	*ZCCHC4*	intronic	D	I	0.22	-0.0126±0.0018	4.3E-12	442,278
rs17644283	4	26,308,792	*RBPJ*	intergenic	A	G	0.37	0.0089±0.0015	1.5E-09	529,134
4:28526950_AAAAC_A	4	28,526,950	*MIR4275*	intergenic	D	I	0.17	-0.0153±0.002	2.2E-14	442,278
rs13137372	4	34,928,521	*LOC101928622*	intergenic	T	C	0.70	-0.0111±0.0016	6.8E-12	442,278
rs10938397	4	45,182,527	*GNPDA2*	intergenic	A	G	0.57	-0.0204±0.0014	2.3E-46	531,498
rs411261	4	48,354,057	*SLAIN2*	intronic	T	C	0.62	0.0088±0.0015	1.6E-09	530,595
rs62645069	4	49,125,019	*CWH43*	intergenic	A	T	0.69	-0.0098±0.0017	4.3E-09	442,278
rs3113509	4	52,932,825	*SPATA18*	intronic	T	C	0.73	-0.01±0.0017	2.3E-09	442,278
rs11943456	4	56,276,334	*TMEM165*	intronic	T	C	0.54	-0.0114±0.0014	2.3E-15	508,064
rs4694127	4	73,552,093	*ADAMTS3*	intergenic	T	C	0.06	0.022±0.003	1.0E-13	508,102
rs7692075	4	78,808,669	*MRPL1*	intronic	T	G	0.85	-0.0133±0.0021	8.0E-11	486,266
rs72649373	4	80,609,966	*LINC00989*	intergenic	T	C	0.86	-0.0153±0.0022	1.3E-12	442,278
rs564988630	4	83,192,564	*HNRNPD*	intergenic	I	D	0.50	-0.0096±0.0015	1.5E-10	442,278
rs2276936	4	89,726,283	*FAM13A*	intronic	A	C	0.53	0.0127±0.0014	2.2E-19	530,402
rs2865384	4	96,145,120	*UNC5C*	intronic	A	G	0.63	-0.0102±0.0015	5.9E-12	506,303
rs1229984	4	100,239,319	*ADH1B*	missense	T	C	0.02	-0.0324±0.0048	1.9E-11	442,278
rs546865674	4	102,197,018	*PPP3CA*	intronic	D	I	0.85	-0.0153±0.0021	5.8E-13	442,278
rs13107325	4	103,188,709	*SLC39A8*	missense	T	C	0.07	0.0304±0.0027	5.9E-29	528,344

rsid	Chr	Pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs1511022	4	120,126,233	*LOC101929762*	ncRNA intronic	T	C	0.12	0.0129±0.0022	2.7E-09	507,580
rs4864201	4	130,731,284	*LOC101927282*	intergenic	T	C	0.35	0.0094±0.0015	2.3E-10	529,748
rs1296328	4	137,083,193	*LINC00613*	intergenic	A	C	0.44	0.0115±0.0015	2.3E-15	506,308
rs57800857	4	140,863,365	*MAML3*	intronic	A	C	0.63	0.0135±0.0015	3.8E-18	442,278
rs113079574	4	147,354,089	*SLC10A7*	intronic	T	C	0.19	-0.0114±0.0019	1.8E-09	442,278
rs112323962	5	284,768	*PDCD6*	intronic	D	I	0.91	-0.017±0.0027	4.7E-10	442,278
rs13177679	5	42,915,470	*FLJ32255*	intergenic	T	C	0.34	-0.0105±0.0015	5.0E-12	528,730
rs17820010	5	50,384,662	*LOC100287592*	intergenic	T	G	0.70	-0.0109±0.0016	4.3E-12	505,722
rs12189178	5	50,914,726	*ISL1*	intergenic	T	C	0.03	0.0279±0.004	2.3E-12	531,323
rs256108	5	52,874,253	*NDUFS4*	intronic	T	C	0.77	-0.0126±0.0017	5.6E-14	531,496
rs11744289	5	53,283,329	*ARL15*	intronic	A	G	0.94	0.0195±0.0032	1.0E-09	504,520
rs40271	5	55,796,319	*C5orf67*	intergenic	T	C	0.74	-0.0125±0.0016	1.1E-14	531,434
rs13173241	5	55,861,359	*C5orf67*	intronic	A	G	0.21	-0.0103±0.0017	3.8E-09	531,510
rs10805383	5	63,034,606	*HTR1A*	intergenic	A	G	0.48	0.0119±0.0014	4.9E-17	530,112
5:63940500_CA_C	5	63,940,500	*RGS7BP*	intergenic	I	D	0.47	-0.0102±0.0015	6.5E-12	442,278
rs6893495	5	64,409,378	*ADAMTS6*	intergenic	T	C	0.21	0.0117±0.0018	1.2E-10	442,278
rs27218	5	66,207,261	*MAST4*	intronic	T	C	0.28	-0.0103±0.0016	1.2E-10	508,077
rs4976033	5	67,714,246	*PIK3R1*	intergenic	A	G	0.60	0.012±0.0015	3.7E-16	529,123
rs11954242	5	74,411,629	*ANKRD31*	intronic	C	G	0.66	0.0142±0.0016	1.9E-19	442,278
rs2112347	5	75,015,242	*POC5*	intergenic	T	G	0.64	0.0175±0.0015	1.2E-32	531,525
5:80811501_AT_A	5	80,811,501	*SSBP2*	intronic	D	I	0.23	-0.0108±0.0018	7.7E-10	442,278
rs115912456	5	82,815,158	*VCAN*	intronic	A	G	0.96	0.0278±0.0037	9.6E-14	442,278
rs11951885	5	86,727,566	*CCNH*	intergenic	T	C	0.98	-0.0321±0.0049	5.0E-11	501,024
rs6870983	5	87,697,533	*TMEM161B-AS1*	ncRNA intronic	T	C	0.21	-0.0148±0.0017	1.2E-17	531,532
rs7733438	5	87,986,284	*LINC00461*	intergenic	T	G	0.86	-0.0244±0.0021	3.7E-31	508,058
rs618741	5	88,064,012	*MEF2C*	intronic	T	C	0.58	-0.0104±0.0014	6.9E-13	508,079
rs4869417	5	92,544,460	*NR2F1-AS1*	intergenic	T	C	0.27	-0.0103±0.0016	8.7E-11	531,526
rs159032	5	94,206,202	*MCTP1*	intronic	T	C	0.25	0.0102±0.0017	1.0E-09	506,312
rs10476546	5	95,552,596	*LOC101929710*	ncRNA intronic	T	C	0.67	0.011±0.0015	5.7E-13	507,146
rs28119	5	96,131,733	*ERAP1*	intronic	A	G	0.79	0.0109±0.0018	5.7E-10	508,042
rs10059133	5	103,901,424	*RAB9BP1*	intergenic	T	C	0.54	-0.0101±0.0014	2.2E-12	508,092
rs11742930	5	105,774,098	*LOC102467213*	intergenic	T	C	0.57	0.0086±0.0014	3.2E-09	508,049
rs10623997	5	107,478,679	*FBXL17*	intronic	D	I	0.76	0.0148±0.0018	1.8E-16	442,278
rs11953203	5	112,446,930	*MCC*	intronic	T	G	0.57	-0.0089±0.0014	7.0E-10	507,583
rs3203922	5	118,728,953	*TNFAIP8*	synonymous	C	G	0.27	0.0102±0.0016	2.6E-10	508,091
rs1948325	5	119,374,787	*FAM170A*	intergenic	C	G	0.53	-0.0091±0.0014	1.9E-10	518,858
rs6860245	5	127,367,998	*LINC01184*	ncRNA intronic	C	G	0.25	-0.0116±0.0017	1.4E-11	442,278
rs34912177	5	127,932,957	*FBN2*	intergenic	A	C	0.86	0.0104±0.0021	9.6E-07	442,278
rs17768640	5	128,769,387	*ADAMTS19-AS1*	intergenic	A	G	0.91	-0.0156±0.0026	2.0E-09	502,772
rs329120	5	133,861,756	*JADE2*	intronic	T	C	0.42	-0.0086±0.0015	4.4E-09	496,529
rs13174863	5	139,080,745	*CXXC5*	intergenic	A	G	0.85	-0.0135±0.002	1.9E-11	528,061
rs2190788	5	144,484,261	*KCTD16*	intergenic	T	G	0.32	0.0101±0.0015	4.5E-11	508,064
rs114285050	5	145,895,394	*GPR151*	stopgain	A	G	0.008	-0.050±0.008	5.52E-10	550,180
rs1438945	5	152,510,937	*LINC01470*	intergenic	A	T	0.71	-0.0095±0.0016	9.1E-09	442,278
rs2126165	5	153,167,595	*GRIA1*	intronic	A	G	0.49	0.0101±0.0015	1.1E-11	442,278
rs6580054	5	153,548,134	*MFAP3*	intergenic	T	C	0.35	0.0113±0.0015	3.6E-14	508,082

rsid	Chr	Pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs10044492	5	157,931,500	*LOC101927697*	intergenic	T	C	0.27	0.0137±0.0016	9.2E-18	529,986
rs1122080	5	158,015,903	*EBF1*	intergenic	A	G	0.19	-0.0149±0.0018	4.6E-16	531,494
rs7730898	5	170,459,675	*RANBP17*	intronic	A	G	0.73	0.0136±0.0016	7.6E-18	531,558
rs758654273	5	170,694,204	*RANBP17*	intronic	D	I	0.09	-0.0152±0.0028	8.2E-08	442,278
rs7733087	5	176,164,941	*LINC01574*	intergenic	A	G	0.38	-0.009±0.0015	2.0E-09	489,496
rs6899218	5	178,988,608	*RUFY1*	intronic	A	T	0.51	0.0098±0.0015	4.0E-11	442,278
rs369386134	5	179,134,664	*CANX*	intronic	D	I	0.20	0.0112±0.0019	3.6E-09	442,278
rs4959613	6	1,835,403	*GMDS*	intronic	A	C	0.59	0.0097±0.0016	5.5E-10	442,278
rs2228213	6	12,124,855	*HIVEP1*	missense	A	G	0.35	-0.0118±0.0015	2.2E-15	531,538
rs10807323	6	12,795,031	*PHACTR1*	intronic	A	G	0.43	-0.0079±0.0014	4.1E-08	531,536
rs796970069	6	13,177,596	*PHACTR1*	intronic	D	I	0.68	0.0106±0.0016	5.6E-11	442,278
rs1322842	6	20,488,897	*E2F3*	intronic	A	G	0.39	0.0117±0.0015	9.9E-15	463,355
rs9366627	6	25,686,405	*SCGN*	intronic	T	C	0.29	-0.0094±0.0016	2.1E-09	531,478
rs10946808	6	26,233,387	*HIST1H1D*	intergenic	A	G	0.72	0.0209±0.0016	9.6E-40	529,743
rs2130655	6	26,789,628	*GUSBP2*	intergenic	T	C	0.75	0.0157±0.0017	1.8E-19	442,278
rs147930667	6	31,676,702	*LY6G6F*	intronic	D	I	0.87	0.0187±0.0022	2.4E-17	442,278
rs368213214	6	32,378,647	*BTNL2*	intergenic	D	I	0.43	-0.0118±0.0015	2.0E-14	442,278
6:33249617_CCTTTT_C	6	33,249,617	*WDR46*	intronic	I	D	0.65	0.0092±0.0016	3.3E-09	442,278
rs2281819	6	33,771,673	*MLN*	intronic	A	T	0.23	-0.0137±0.0017	9.8E-16	507,155
rs2815005	6	34,638,847	*C6orf106*	intronic	A	G	0.14	0.0245±0.002	1.2E-33	531,534
rs11755266	6	35,162,141	*SCUBE3*	intergenic	T	C	0.10	0.0256±0.0024	3.5E-26	527,708
rs2766535	6	35,691,782	*FKBP5*	intronic	A	G	0.47	0.0112±0.0014	5.3E-15	505,821
rs9471333	6	40,362,023	*LRFN2*	intronic	T	C	0.55	-0.0143±0.0015	9.3E-22	442,278
rs998584	6	43,757,896	*VEGFA*	intergenic	A	C	0.48	-0.0095±0.0014	2.8E-11	527,142
rs1867015	6	46,321,089	*LOC101926915*	ncRNA intronic	T	G	0.48	-0.0091±0.0014	1.9E-10	508,077
rs4391262	6	50,378,717	*TFAP2D*	intergenic	T	C	0.05	0.0234±0.0032	3.1E-13	507,243
rs4715213	6	50,911,091	*PKHD1*	intergenic	T	C	0.17	0.0229±0.0019	5.2E-34	530,185
rs2474896	6	51,760,527	*PKHD1*	intronic	T	C	0.55	0.0121±0.0015	7.8E-16	442,278
rs796663884	6	52,426,285	*TRAM2*	intronic	I	D	0.42	0.0094±0.0015	7.6E-10	442,278
rs7452651	6	55,135,513	*HCRTR2*	intronic	A	G	0.25	-0.0101±0.0017	9.9E-10	508,094
rs6910590	6	70,359,738	*LMBRD1*	intergenic	A	G	0.21	-0.0114±0.0018	1.1E-10	506,309
rs9294260	6	83,433,228	*UBE3D*	intergenic	A	G	0.48	0.0089±0.0014	4.2E-10	531,042
rs4839906	6	97,761,611	*LOC101927314,MIR548H3*	ncRNA intronic	T	C	0.29	-0.0108±0.0016	8.7E-12	508,087
rs1487441	6	98,553,894	*MIR2113*	intergenic	A	G	0.48	-0.0152±0.0014	2.8E-26	508,076
rs57989773	6	100,629,078	*MCHR2-AS1*	intergenic	T	C	0.76	-0.0123±0.0018	4.8E-12	442,278
rs314288	6	105,434,078	*LIN28B*	intronic	T	C	0.11	0.0157±0.0023	1.5E-11	442,278
rs768023	6	108,876,002	*FOXO3*	intergenic	A	G	0.63	0.0112±0.0015	2.1E-14	531,523
rs77043842	6	111,647,203	*REV3L*	intronic	D	I	0.91	-0.0178±0.0027	6.4E-11	442,278
rs13218383	6	120,173,501	*MIR3144*	intergenic	C	G	0.66	0.0093±0.0015	6.8E-10	508,094
rs1361108	6	126,767,600	*MIR588*	intergenic	T	C	0.46	-0.0113±0.0014	4.1E-15	507,187
rs72959041	6	127,454,893	*RSPO3*	intronic	A	G	0.05	-0.0223±0.0034	6.4E-11	461,206
rs9321191	6	130,165,691	*TMEM244*	intronic	T	C	0.80	0.011±0.0018	8.2E-10	507,576
rs1970341	6	141,361,290	*MIR4465*	intergenic	A	G	0.05	0.0195±0.0033	4.4E-09	442,278
6:143187255_CA_C	6	143,187,255	*HIVEP2*	intronic	I	D	0.51	0.0096±0.0015	8.5E-11	442,278
rs9390370	6	146,357,569	*GRM1*	intronic	T	C	0.31	0.0088±0.0015	1.6E-08	508,052
rs2063347	6	160,882,029	*LPAL2*	intergenic	A	G	0.39	-0.0084±0.0014	7.5E-09	531,529
rs6461115	7	2,103,668	*MAD1L1*	intronic	A	G	0.77	0.0121±0.0017	8.2E-13	531,563
rs1182197	7	2,863,289	*GNA12*	intronic	A	C	0.62	0.0091±0.0015	2.0E-09	465,729
rs4722398	7	3,125,220	*CARD11*	intergenic	T	C	0.14	0.0137±0.0021	5.4E-11	508,076
rs11408091	7	6,387,863	*FAM220A*	intronic	D	I	0.87	-0.0132±0.0022	2.5E-09	442,278
rs4307239	7	24,354,300	*NPY*	intergenic	A	G	0.54	-0.0084±0.0014	4.5E-09	508,062
rs6967749	7	26,713,194	*SKAP2*	intronic	T	C	0.91	-0.0163±0.0026	2.3E-10	505,064

rsid	Chr	Pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs4722672	7	27,231,762	*HOXA11-AS*	intergenic	T	C	0.82	-0.0156±0.0019	4.3E-17	529,277
rs215634	7	32,369,148	*PDE1C*	intergenic	A	G	0.39	0.0121±0.0015	3.8E-16	497,301
rs4549685	7	39,326,478	*POU6F2*	intronic	T	C	0.33	-0.0114±0.0016	6.0E-13	442,278
rs2289379	7	44,804,225	*ZMIZ2*	intronic	T	C	0.40	-0.0089±0.0015	1.3E-09	507,136
rs17172722	7	46,620,312	*IGFBP3*	intergenic	T	C	0.42	-0.0098±0.0015	1.7E-11	508,044
rs10259490	7	49,616,420	*VWC2*	intergenic	A	G	0.62	-0.0088±0.0015	2.4E-09	506,312
rs17449259	7	50,733,714	*GRB10*	intronic	T	G	0.07	0.0188±0.003	2.0E-10	506,722
rs10263467	7	69,179,509	*AUTS2*	intronic	A	G	0.05	0.0191±0.0032	2.4E-09	496,510
rs10237317	7	70,045,941	*AUTS2*	intronic	A	G	0.58	-0.0114±0.0014	3.4E-15	529,753
rs10950290	7	71,439,590	*CALN1*	intronic	A	T	0.74	0.0111±0.0016	1.2E-11	495,636
rs116903188	7	74,347,183	*PMS2P5*	intergenic	T	G	0.87	0.0194±0.0027	1.3E-12	442,278
rs236660	7	75,050,086	*POM121C*	intronic	T	C	0.43	-0.0141±0.0016	1.8E-19	442,278
rs6970595	7	77,426,482	*TMEM60*	intronic	T	C	0.60	-0.0106±0.0015	4.0E-13	508,104
rs3840590	7	77,827,064	*MAGI2*	intronic	D	I	0.65	0.0113±0.0016	5.6E-13	442,278
rs17704028	7	95,140,031	*ASB4*	intronic	T	C	0.15	-0.0131±0.002	9.7E-11	503,677
rs73232637	7	96,613,246	*DLX6-AS1*	ncRNA intronic	A	G	0.96	-0.0248±0.004	7.5E-10	442,278
rs1011024	7	99,197,401	*GS1-259H13.2*	intronic	A	G	0.85	0.0149±0.002	6.0E-14	508,039
rs12375196	7	103,416,541	*RELN*	intronic	A	C	0.42	0.0104±0.0015	6.0E-12	442,278
7:104885481_AACACAC_A	7	104,885,481	*SRPK2*	intronic	D	I	0.38	0.0093±0.0015	1.4E-09	442,278
rs7788008	7	112,972,483	*LINC00998*	intergenic	A	G	0.44	-0.0109±0.0014	3.2E-14	508,074
rs6944092	7	113,483,569	*PPP1R3A*	intergenic	A	G	0.63	-0.0092±0.0015	6.0E-10	508,022
rs1530680	7	114,407,396	*FOXP2*	intergenic	T	G	0.20	-0.0109±0.0018	1.3E-09	508,070
rs7777351	7	121,962,953	*CADPS2*	intronic	A	C	0.37	0.0094±0.0015	9.4E-10	442,278
rs2283093	7	126,721,231	*GRM8*	intronic	T	C	0.20	0.0107±0.0018	1.5E-09	508,077
rs972283	7	130,466,854	*KLF14*	intergenic	A	G	0.49	0.0148±0.0014	1.0E-25	531,549
rs6977081	7	150,542,515	*AOC1*	intergenic	T	G	0.33	-0.0113±0.0015	9.6E-14	530,577
rs11782341	8	4,813,459	*CSMD1*	intronic	A	G	0.81	-0.0107±0.0018	6.4E-09	508,039
rs4841218	8	9,653,902	*TNKS*	intergenic	T	C	0.38	0.0098±0.0015	2.7E-11	506,310
rs775959252	8	14,190,443	*SGCZ*	intronic	I	D	0.26	0.0107±0.0018	1.5E-09	442,278
rs11786089	8	21,975,521	*HR*	intronic	A	G	0.54	-0.0104±0.0015	3.7E-12	442,278
rs11781222	8	23,389,571	*SLC25A37*	intronic	T	C	0.86	0.0154±0.0021	9.3E-14	531,549
rs117176448	8	27,261,138	*PTK2B*	intronic	C	G	0.90	-0.017±0.0025	1.5E-11	442,278
rs11775287	8	30,864,339	*PURG*	intronic	T	C	0.48	-0.0102±0.0014	1.3E-12	508,085
rs12681990	8	36,859,186	*KCNU1*	intergenic	T	C	0.84	0.0124±0.0019	1.2E-10	531,492
rs1808629	8	73,435,964	*KCNB2*	intergenic	A	G	0.69	-0.0167±0.0016	2.9E-25	442,278
rs1992974	8	76,653,245	*HNF4G*	intergenic	C	G	0.41	-0.0145±0.0015	3.3E-23	508,101
rs1594447	8	77,227,907	*LINC01111*	intergenic	A	G	0.43	-0.0148±0.0014	3.4E-25	531,553
rs17481178	8	78,882,007	*LOC102724874*	intergenic	A	G	0.37	0.0086±0.0015	4.6E-09	529,759
rs111884404	8	87,342,875	*WWP1*	intergenic	A	G	0.64	-0.0104±0.0017	6.8E-10	442,278
rs2957446	8	106,355,864	*ZFPM2*	intronic	A	G	0.47	-0.0087±0.0014	7.2E-10	531,408
rs4433183	8	112,501,602	*LINC01609*	intergenic	T	C	0.27	-0.0103±0.0016	2.8E-10	508,043
rs3808477	8	116,670,347	*TRPS1*	intronic	T	C	0.28	-0.0176±0.0016	2.0E-28	507,945
rs4466418	8	126,323,787	*NSMCE2*	intronic	A	G	0.56	0.01±0.0014	3.8E-12	507,592
8:126504383_CCACCAT_C	8	126,504,383	*TRIB1*	intergenic	I	D	0.40	-0.0126±0.0016	1.2E-14	442,278
rs1106761	8	142,619,234	*MROH5*	intergenic	A	G	0.39	0.0103±0.0015	2.9E-11	442,278
rs4072917	8	143,300,279	*TSNARE1*	intronic	A	G	0.48	0.0086±0.0014	2.8E-09	505,828
rs2717609	8	143,769,252	*PSCA*	intergenic	A	T	0.53	0.0098±0.0015	5.8E-11	442,278
8:144505206_ACCT_A	8	144,505,206	*MAFA-AS1*	intergenic	D	I	0.21	0.0115±0.0019	6.2E-10	442,278
rs10124645	9	11,321,119	*PTPRD-AS2*	intergenic	A	G	0.59	0.0092±0.0015	3.3E-10	507,608

rsid	Chr	Pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs1948080	9	11,852,043	*TYRP1*	intergenic	T	G	0.63	0.0102±0.0015	4.7E-12	508,073
rs424539	9	14,442,595	*NFIB*	intergenic	C	G	0.62	-0.0104±0.0015	2.3E-12	506,668
rs7046483	9	14,777,395	*FREM1*	intronic	A	G	0.61	-0.0098±0.0015	1.4E-10	442,278
rs16933350	9	15,515,940	*PSIP1*	intergenic	T	G	0.96	0.0327±0.0035	1.0E-20	504,263
rs10756713	9	15,880,555	*CCDC171*	intronic	A	G	0.56	0.0193±0.0015	9.7E-38	442,278
rs10116857	9	16,595,649	*BNC2*	intronic	A	C	0.06	-0.0237±0.003	3.2E-15	508,077
rs10756798	9	16,739,763	*BNC2*	intronic	T	C	0.65	-0.0138±0.0016	6.6E-19	442,278
rs10968577	9	28,415,512	*LINGO2*	intronic	T	C	0.32	0.0158±0.0015	2.4E-25	521,832
rs13292491	9	29,688,257	*LINGO2*	intergenic	T	C	0.76	0.0108±0.0018	7.4E-10	442,278
rs16916303	9	30,823,761	*LINC01242*	intergenic	A	G	0.88	0.015±0.0023	3.1E-11	493,875
rs10973160	9	36,994,969	*PAX5*	intronic	T	C	0.32	-0.0101±0.0016	2.7E-10	442,278
rs725959	9	81,349,608	*LOC101927450*	intergenic	T	G	0.41	-0.0091±0.0015	4.3E-10	508,096
rs12349699	9	86,640,377	*RMI1*	intergenic	T	C	0.52	0.0087±0.0014	1.8E-09	506,156
rs7030895	9	92,189,513	*GADD45G*	intergenic	A	C	0.46	-0.0107±0.0014	1.0E-13	508,057
rs7024028	9	96,395,022	*PHF2*	intronic	T	C	0.32	0.0091±0.0015	3.5E-09	508,097
9:103113656_AC_A	9	103,113,656	*TEX10*	intronic	D	I	0.38	0.0107±0.0016	5.8E-12	442,278
rs41307479	9	116,082,647	*WDR31*	missense	C	G	0.78	-0.0121±0.0018	1.3E-11	442,278
rs10982576	9	117,920,248	*01-Dec*	intronic	T	C	0.76	0.0112±0.0017	2.7E-11	506,313
rs10982884	9	118,462,904	*LOC101928775*	intergenic	T	C	0.09	0.0152±0.0025	9.0E-10	505,068
rs1928295	9	120,378,483	*LOC101928797*	intergenic	T	C	0.57	0.0085±0.0014	2.1E-09	531,540
rs7863807	9	124,625,886	*TTLL11*	intronic	A	G	0.68	-0.0091±0.0015	3.2E-09	508,067
rs3829849	9	129,390,800	*LMX1B*	intronic	T	C	0.37	0.0113±0.0015	1.0E-14	531,538
rs11790436	9	129,929,189	*RALGPS1*	intronic	T	G	0.40	-0.01±0.0015	1.1E-11	506,313
rs1147345	9	132,212,511	*LINC00963*	intergenic	A	T	0.47	0.0088±0.0015	6.0E-09	442,278
rs4740383	9	133,783,566	*FIBCD1*	intronic	A	G	0.41	0.0089±0.0015	1.4E-09	496,638
rs674302	9	136,146,664	*ABO*	intronic	A	T	0.32	0.0088±0.0015	9.0E-09	508,086
rs10993907	9	136,928,644	*BRD3*	intronic	T	C	0.67	0.0098±0.0015	1.0E-10	508,058
rs35475612	10	16,750,948	*RSU1*	intronic	I	D	0.69	0.0108±0.0016	2.0E-11	442,278
rs7084454	10	21,821,274	*MLLT10*	intergenic	A	G	0.32	0.0159±0.0015	4.2E-25	503,323
rs12762056	10	33,969,962	*LINC00838*	intergenic	T	C	0.84	-0.0129±0.002	5.4E-11	508,043
rs10998304	10	70,342,775	*TET1*	intronic	T	C	0.55	-0.0098±0.0014	4.9E-12	531,499
rs12359330	10	72,414,845	*ADAMTS14*	intergenic	T	C	0.27	0.0159±0.0017	3.6E-21	442,278
rs3088142	10	76,854,564	*DUSP13*	missense	T	C	0.41	0.0091±0.0015	3.7E-10	508,041
rs200800095	10	77,630,139	*C10orf11*	intronic	I	D	0.73	0.0109±0.0017	8.5E-11	442,278
rs2114824	10	88,119,015	*GRID1*	intronic	A	G	0.50	-0.0094±0.0014	6.8E-11	500,061
rs2274224	10	96,039,597	*PLCE1*	missense	C	G	0.43	-0.0163±0.0014	3.6E-30	531,015
rs4110517	10	96,650,328	*CYP2C19*	intergenic	A	G	0.21	-0.0123±0.0018	4.2E-12	508,088
rs522110	10	99,772,885	*CRTAC1*	intronic	A	G	0.44	-0.0129±0.0014	3.0E-19	508,096
rs41310284	10	102,447,647	*PAX2*	intergenic	A	C	0.10	-0.0209±0.0025	2.5E-17	442,278
rs11191017	10	103,213,564	*BTRC*	intronic	A	G	0.19	-0.0106±0.0018	3.2E-09	531,513
rs7096641	10	107,579,493	*LOC101927549*	ncRNA intronic	A	C	0.49	0.0082±0.0014	1.1E-08	507,145
rs7903146	10	114,758,349	*TCF7L2*	intronic	T	C	0.29	-0.0111±0.0016	1.2E-12	531,494
rs10886017	10	118,672,531	*SHTN1*	intronic	A	C	0.25	0.0119±0.0016	6.0E-13	531,405
rs11199266	10	122,061,612	*PLPP4*	intergenic	A	G	0.64	-0.009±0.0015	1.6E-09	506,309
rs2172071	10	122,968,030	*MIR5694*	intergenic	T	C	0.68	0.009±0.0015	4.7E-09	508,093
rs72828935	10	126,587,488	*ZRANB1*	intergenic	C	G	0.28	0.0116±0.0017	2.8E-12	442,278
rs374873119	10	128,836,112	*DOCK1*	intronic	I	D	0.54	-0.0097±0.0016	8.8E-10	442,278
rs11017772	10	132,954,247	*TCERG1L*	intronic	T	C	0.21	-0.0117±0.0018	3.5E-11	508,098
rs4880341	10	133,992,689	*JAKMIP3*	intronic	T	C	0.58	-0.0106±0.0014	2.9E-13	507,959
rs6578412	11	1,482,582	*BRSK2*	UTR3	T	C	0.06	0.019±0.0032	4.2E-09	442,278
rs10128597	11	8,694,830	*RPL27A*	intergenic	A	G	0.27	-0.0137±0.0016	1.3E-17	529,114

rsid	Chr	Pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs2923093	11	10,360,934	*CAND1.11*	ncRNA intronic	A	T	0.59	-0.0104±0.0014	6.0E-13	528,445
rs6416051	11	10,409,041	*CAND1.11*	ncRNA intronic	T	C	0.50	0.0108±0.0014	6.3E-14	494,258
rs4757144	11	13,331,226	*ARNTL*	intronic	A	G	0.59	0.0116±0.0015	2.1E-15	508,048
rs7938266	11	14,709,324	*PDE3B*	intronic	A	G	0.59	0.0078±0.0014	6.3E-08	531,485
rs150090666	11	14,865,399	*PDE3B*	stop gain	T	C	0.001	0.150±0.025	2.56E-09	442,278
rs6265	11	27,679,916	*BDNF*	missense	T	C	0.19	-0.0212±0.0018	6.1E-32	531,552
rs962369	11	27,734,420	*BDNF*	intronic	T	C	0.69	-0.0173±0.0015	1.9E-29	529,352
rs6484478	11	30,306,440	*ARL14EP*	intergenic	A	G	0.23	-0.0105±0.0017	7.2E-10	506,313
rs10838123	11	43,551,572	*MIR670*	intergenic	C	G	0.49	-0.0094±0.0014	8.4E-11	507,930
rs2862996	11	43,653,833	*HSD17B12*	intergenic	T	G	0.69	-0.0177±0.0015	5.0E-31	531,483
rs71474196	11	46,977,160	*C11orf49*	intronic	T	C	0.11	-0.0158±0.0024	6.0E-11	442,278
rs7124681	11	47,529,947	*CELF1*	intronic	A	C	0.41	0.0222±0.0014	4.1E-54	531,501
rs2081361	11	57,411,742	*YPEL4*	downstream	T	C	0.28	0.0095±0.0016	3.2E-09	505,841
rs71458418	11	62,318,081	*AHNAK*	intergenic	D	I	0.64	0.0088±0.0016	2.3E-08	442,278
rs2516633	11	62,371,773	*EML3*	intronic	A	C	0.31	-0.0114±0.0016	3.4E-13	506,308
rs7947143	11	64,090,422	*PRDX5*	intergenic	A	G	0.16	-0.0161±0.0019	1.6E-16	506,312
rs3200401	11	65,271,832	*MALAT1*	ncRNA exonic	T	C	0.22	0.0111±0.0017	1.9E-10	506,155
rs801738	11	65,924,217	*PACS1*	intronic	C	G	0.64	0.0133±0.0015	9.5E-18	442,278
rs35099456	11	66,649,527	*PC*	intronic	C	G	0.06	-0.0213±0.0031	5.1E-12	442,278
rs4930427	11	67,200,819	*RPS6KB2*	synonymous	T	C	0.44	-0.0095±0.0014	4.5E-11	519,452
rs7102705	11	69,143,284	*MYEOV*	intergenic	A	G	0.20	-0.0113±0.0018	3.8E-10	496,218
rs112326082	11	76,479,391	*TSKU*	intergenic	D	I	0.79	0.0112±0.0019	3.2E-09	442,278
rs393308	11	85,135,100	*DLG2*	intronic	T	C	0.16	-0.0127±0.002	3.3E-10	442,278
rs61903695	11	89,922,417	*NAALAD2*	intronic	A	G	0.74	-0.0109±0.0017	1.9E-10	442,278
rs2155645	11	112,912,947	*NCAM1*	intronic	T	C	0.25	-0.0102±0.0016	3.9E-10	529,755
rs719802	11	113,234,679	*TTC12*	intronic	T	C	0.39	0.0102±0.0015	4.5E-12	506,308
rs2046149	11	117,054,709	*SIDT2*	intronic	C	G	0.94	-0.0185±0.0031	2.0E-09	508,103
rs583893	11	118,904,233	*SLC37A4*	intergenic	T	G	0.43	-0.0103±0.0015	1.1E-11	442,278
rs607824	11	119,754,895	*LOC102724301*	intergenic	A	G	0.47	-0.0087±0.0014	1.3E-09	531,037
rs12419776	11	122,528,863	*UBASH3B*	intronic	A	T	0.37	0.0093±0.0015	1.7E-09	442,278
rs11603783	11	122,752,954	*C11orf63*	upstream	T	C	0.75	-0.0115±0.0017	3.7E-11	442,278
rs3829271	11	130,753,912	*SNX19*	intronic	A	T	0.41	0.0117±0.0015	1.0E-15	508,102
rs12788343	11	131,452,912	*NTM*	intronic	T	C	0.59	-0.011±0.0015	3.9E-13	442,278
rs11533200	11	131,955,594	*NTM*	intronic	A	G	0.31	-0.0092±0.0016	3.1E-09	508,048
rs4936175	11	132,641,959	*OPCML*	intronic	T	C	0.56	-0.0087±0.0014	1.7E-09	508,088
rs329651	11	133,767,622	*MIR4697HG*	ncRNA exonic	T	G	0.81	0.0105±0.0018	6.1E-09	528,019
rs61910767	11	134,515,899	*LOC283177*	intergenic	T	C	0.17	-0.0147±0.002	2.1E-13	442,278
rs55726687	12	991,306	*WNK1*	intronic	A	G	0.21	0.0132±0.0018	3.5E-13	442,278
rs2108635	12	2,159,556	*CACNA1C-IT2*	downstream	A	G	0.66	-0.0093±0.0016	3.1E-09	442,278
rs3782800	12	3,350,276	*TSPAN9*	intronic	A	G	0.09	-0.0191±0.0025	3.0E-14	508,035
rs7980313	12	3,353,862	*TSPAN9*	intronic	T	C	0.94	0.016±0.003	1.4E-07	500,676
rs1526573	12	18,202,053	*RERGL*	intergenic	A	C	0.31	0.01±0.0016	5.8E-10	442,278
rs883528	12	19,241,157	*PLEKHA5*	intergenic	T	C	0.30	-0.0091±0.0016	2.6E-08	442,278
rs11045172	12	20,470,221	*PDE3A*	intergenic	A	C	0.80	-0.0109±0.0018	2.0E-09	505,685
rs12811752	12	20,577,805	*PDE3A*	intronic	T	C	0.59	0.0087±0.0015	4.2E-09	505,682
rs1350429	12	41,819,131	*PDZRN4*	intronic	A	G	0.52	-0.0109±0.0014	2.8E-14	508,087
rs3730071	12	49,168,798	*ADCY6*	missense	A	C	0.03	-0.0266±0.0042	3.7E-10	493,910
rs7138803	12	50,247,468	*BCDIN3D*	intergenic	A	G	0.37	0.0184±0.0015	2.3E-36	531,530
rs10876528	12	54,421,476	*HOXC4,HOXC6*	intronic	A	C	0.37	0.0102±0.0015	4.1E-12	518,947
rs4759276	12	57,526,646	*LRP1*	intronic	A	G	0.39	0.0098±0.0015	1.0E-10	442,278

rsid	Chr	Pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs11173522	12	60,953,472	*SLC16A7*	intergenic	A	C	0.21	0.011±0.0018	4.0E-10	508,068
rs1031580	12	62,263,995	*FAM19A2*	intronic	T	C	0.76	0.01±0.0017	1.2E-08	442,278
rs12819035	12	64,236,002	*SRGAP1*	intergenic	A	T	0.08	-0.017±0.0029	5.3E-09	442,278
rs650198	12	69,674,595	*CPSF6*	intergenic	T	C	0.72	-0.0094±0.0016	5.4E-09	495,637
rs61754230	12	72,179,446	*RAB21*	missense	T	C	0.02	0.0319±0.0053	2.3E-09	442,278
rs10506971	12	89,757,937	*DUSP6*	intergenic	A	G	0.55	-0.015±0.0014	1.7E-25	508,090
rs1054807	12	89,913,419	*GALNT4*	UTR3	T	C	0.24	0.0134±0.0017	1.0E-15	531,388
rs12426970	12	90,436,372	*LINC00936*	intergenic	A	C	0.95	0.0282±0.0033	2.2E-17	504,634
rs11105846	12	91,257,268	*LINC00615*	intergenic	T	G	0.37	-0.0104±0.0015	4.3E-12	508,056
rs10745785	12	97,586,257	*RMST*	intergenic	T	C	0.66	-0.009±0.0015	3.2E-09	508,090
rs7137989	12	103,689,076	*C12orf42*	intergenic	T	C	0.34	-0.0102±0.0015	1.4E-11	507,612
rs2287214	12	108,090,518	*PWP1*	intronic	A	G	0.60	-0.0131±0.0015	4.0E-19	508,084
rs11609659	12	108,296,260	*LOC728739*	downstream	T	C	0.75	0.0134±0.0017	1.5E-15	505,745
rs3764002	12	108,618,630	*WSCD2*	missense	T	C	0.26	-0.0179±0.0016	1.4E-28	518,172
rs12300276	12	110,697,965	*ATP2A2*	intergenic	A	G	0.24	0.011±0.0018	3.3E-10	442,278
rs10492021	12	113,197,315	*RPH3A*	intergenic	T	C	0.87	-0.0131±0.0021	7.4E-10	508,069
rs7963783	12	118,409,363	*KSR2*	intergenic	T	G	0.71	0.0094±0.0016	2.8E-09	508,077
rs61945796	12	120,384,155	*CCDC64*	intergenic	T	G	0.03	-0.0255±0.0041	4.1E-10	442,278
rs145350287	12	120,907,309	*SRSF9*	missense	A	T	0.04	-0.0307±0.0038	5.1E-16	442,278
rs75412871	12	121,709,430	*CAMKK2*	intronic	T	C	0.05	-0.0208±0.0033	4.5E-10	442,278
rs371575888	12	122,293,634	*HPD*	intronic	A	C	0.04	-0.0341±0.0041	3.5E-17	442,278
rs12369179	12	122,963,550	*ZCCHC8*	intronic	T	C	0.09	-0.0286±0.0025	4.7E-30	504,173
rs147530811	12	123,817,477	*SBNO1*	intronic	A	C	0.04	-0.0288±0.0039	2.1E-13	442,278
rs7133378	12	124,409,502	*DNAH10*	intronic	A	G	0.32	0.0196±0.0015	7.3E-38	529,731
rs863750	12	124,505,444	*ZNF664-FAM101A*	intronic	T	C	0.60	-0.0152±0.0014	1.1E-25	531,553
rs1023229	13	20,270,925	*PSPC1*	intergenic	A	G	0.86	-0.0145±0.002	1.0E-12	508,076
rs9579775	13	20,616,557	*ZMYM2*	intronic	A	C	0.86	-0.0141±0.0023	3.9E-10	442,278
rs9512696	13	28,012,527	*MTIF3*	intronic	A	G	0.34	-0.0121±0.0016	1.7E-14	442,278
rs2026752	13	31,012,459	*LINC01058*	intergenic	T	G	0.75	0.0117±0.0017	1.2E-11	442,278
rs7328213	13	33,377,830	*LINC00423*	intergenic	T	C	0.55	0.0116±0.0015	7.6E-15	442,278
rs75983170	13	40,788,838	*LINC00548*	ncRNA intronic	T	G	0.66	-0.0097±0.0016	1.6E-09	442,278
rs12429545	13	54,102,206	*LINC00558*	intergenic	A	G	0.13	0.0198±0.0021	1.9E-20	527,719
rs2446169	13	54,520,877	*LINC00558*	intergenic	A	G	0.15	0.013±0.002	2.1E-10	505,832
rs7995015	13	54,828,961	*MIR1297*	intergenic	T	G	0.32	0.0104±0.0015	1.6E-11	508,087
rs6561937	13	58,257,667	*PCDH17*	intronic	A	T	0.75	-0.0112±0.0017	2.5E-11	502,695
rs4055791	13	59,266,053	*LINC00374*	intergenic	T	C	0.42	-0.009±0.0015	2.1E-09	442,278
rs9527958	13	59,841,918	*DLAPH3*	intergenic	A	G	0.66	-0.0094±0.0015	4.4E-10	508,054
rs629443	13	76,386,075	*LMO7*	intronic	T	G	0.24	0.0109±0.0017	7.1E-11	530,424
rs1441264	13	79,580,919	*LINC00331*	intergenic	A	G	0.59	0.0117±0.0015	1.9E-15	526,774
rs693839	13	80,958,288	*SPRY2*	intergenic	T	C	0.70	-0.0112±0.0015	4.5E-13	531,051
rs112108364	13	86,490,590	*SLITRK6*	intergenic	T	G	0.72	-0.0122±0.0017	2.3E-13	442,278
rs9634490	13	97,049,100	*HS6ST3*	intronic	A	G	0.46	0.0082±0.0014	1.1E-08	506,313
rs4294654	13	99,119,607	*STK24*	intronic	A	G	0.29	-0.0134±0.0016	2.6E-17	508,092
rs7992207	13	104,090,183	*LINC01309*	intergenic	C	G	0.15	-0.0119±0.002	3.1E-09	505,111
rs9522183	13	111,977,280	*TEX29*	intronic	T	G	0.56	-0.0106±0.0015	1.7E-12	442,278
rs1183668	13	112,191,837	*TEX29*	intergenic	C	G	0.63	0.011±0.0015	1.2E-12	442,278
rs17256211	14	23,754,580	*HOMEZ*	intronic	A	G	0.35	-0.0092±0.0015	1.6E-09	505,106
rs9788550	14	29,681,138	*MIR548AI*	intergenic	C	G	0.25	-0.0154±0.0017	4.0E-19	442,278
rs1959430	14	30,184,162	*PRKD1*	intronic	T	C	0.59	-0.0092±0.0015	1.0E-09	442,278
rs61979560	14	30,727,033	*G2E3*	intergenic	A	C	0.29	0.0107±0.0017	1.6E-10	442,278
rs17522122	14	33,302,882	*AKAP6*	downstream	T	G	0.47	0.0123±0.0014	7.6E-18	525,539
rs796950072	14	33,403,525	*NPAS3*	intergenic	I	D	0.46	-0.0083±0.0015	6.4E-08	442,278
rs1451963	14	41,350,367	*LOC644919*	intergenic	T	G	0.08	0.0161±0.0026	8.9E-10	505,075
rs34688745	14	47,298,299	*MDGA2*	intergenic	D	I	0.50	0.0098±0.0015	9.9E-11	442,278
rs72681869	14	50,655,357	*SOS2*	missense	C	G	0.01	-0.0573±0.0071	8.4E-16	442,278

rsid	Chr	Pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs72681698	14	51,207,741	*NIN*	intronic	T	C	0.99	0.0541±0.0072	4.8E-14	442,278
rs2481899	14	56,460,686	*PELI2*	intergenic	A	G	0.45	-0.0089±0.0015	3.0E-09	442,278
rs7154973	14	59,408,754	*LINC01500*	ncRNA intronic	A	T	0.83	0.0134±0.0019	3.1E-12	508,092
rs217672	14	62,361,021	*SNAPC1*	intergenic	A	C	0.73	-0.0132±0.0017	2.8E-15	442,278
rs4430672	14	63,094,407	*KCNH5*	intergenic	T	C	0.20	0.0112±0.0018	5.6E-10	506,309
rs8003790	14	64,947,181	*ZBTB25*	intronic	T	C	0.52	-0.0088±0.0015	3.9E-09	442,278
rs12890931	14	69,753,369	*GALNT16*	intronic	T	G	0.64	-0.01±0.0015	2.1E-11	507,151
rs10138360	14	75,293,386	*YLPM1*	intronic	A	G	0.54	0.0094±0.0014	2.9E-11	529,754
rs745614299	14	79,563,356	*NRXN3*	intronic	D	I	0.63	0.0085±0.0015	3.4E-08	442,278
rs2370901	14	79,572,031	*NRXN3*	intronic	T	C	0.63	0.0087±0.0015	5.3E-09	508,055
rs7144011	14	79,940,383	*NRXN3*	intronic	T	G	0.22	0.0192±0.0017	2.1E-29	531,469
rs11624512	14	93,111,120	*RIN3*	intronic	T	C	0.19	-0.0147±0.0018	1.6E-15	505,934
rs12434217	14	94,014,337	*UNC79*	intronic	A	G	0.37	-0.0131±0.0015	2.2E-17	442,278
rs12885251	14	99,670,791	*BCL11B*	intronic	A	G	0.46	-0.0095±0.0015	3.0E-10	442,278
rs7161194	14	101,529,005	*MIR377*	downstream	A	G	0.34	0.0107±0.0016	1.7E-11	524,667
rs10431745	14	102,302,372	*PPP2R5C*	intronic	A	G	0.93	0.0182±0.0028	3.5E-11	508,098
rs3803286	14	103,246,470	*TRAF3*	intronic	A	G	0.33	0.0124±0.0015	2.6E-16	507,941
rs2010281	14	103,862,322	*MARK3*	intronic	A	G	0.34	-0.0096±0.0015	1.3E-10	529,377
rs12441543	15	31,689,543	*KLF13*	intronic	A	G	0.29	0.0119±0.0016	6.2E-14	506,309
rs316611	15	41,751,678	*RTF1*	intronic	T	C	0.25	0.0111±0.0017	2.4E-11	505,093
rs1559677	15	47,738,063	*SEMA6D*	intronic	A	G	0.60	-0.0098±0.0014	1.5E-11	531,502
rs199764228	15	48,745,490	*FBN1*	intronic	D	I	0.81	0.0117±0.002	2.7E-09	442,278
rs8031903	15	51,838,161	*DMXL2*	intronic	T	C	0.52	-0.0105±0.0014	4.1E-13	497,425
rs35697691	15	52,353,498	*MAPK6*	missense	C	G	0.91	-0.0212±0.0026	1.1E-15	442,278
rs2440315	15	53,084,353	*ONECUT1*	intergenic	T	G	0.74	-0.0114±0.0016	3.0E-12	508,088
rs7169205	15	59,080,203	*FAM63B*	intronic	A	G	0.85	-0.0124±0.002	4.7E-10	508,077
rs340025	15	60,908,307	*RORA-AS1*	ncRNA intronic	T	C	0.42	-0.0089±0.0014	7.5E-10	529,735
rs12595158	15	62,316,035	*VPS13C*	missense	T	C	0.02	-0.0295±0.005	3.8E-09	525,103
rs2245586	15	66,878,900	*LINC01169*	ncRNA intronic	A	G	0.35	-0.0102±0.0015	6.0E-12	531,467
rs2119261	15	67,011,980	*SMAD6*	intronic	T	C	0.35	0.0104±0.0015	4.7E-12	508,078
rs8038652	15	67,552,064	*IQCH*	intronic	A	G	0.23	-0.0158±0.0017	2.1E-20	508,083
rs16951304	15	68,089,618	*MAP2K5*	intronic	T	C	0.79	0.0206±0.0018	2.6E-31	508,074
rs7168991	15	68,605,612	*ITGA11*	intronic	T	C	0.08	-0.0162±0.0027	2.0E-09	507,169
rs7164727	15	73,093,991	*ADPGK-AS1*	intergenic	T	C	0.67	0.0146±0.0015	5.2E-22	531,550
rs3826043	15	73,618,238	*HCN4*	intronic	T	C	0.43	-0.0135±0.0015	5.8E-19	442,278
rs5742915	15	74,336,633	*PML*	missense	T	C	0.54	0.0087±0.0014	9.7E-10	529,126
rs936227	15	75,131,959	*ULK3*	synonymous	A	G	0.37	-0.0091±0.0015	5.0E-10	530,166
rs139469059	15	75,922,603	*SNUPN*	intergenic	A	G	0.009	0.050±0.008	3.36E-09	442,278
rs11856579	15	78,012,688	*LINGO1*	intronic	A	G	0.27	-0.0101±0.0016	5.7E-10	495,634
rs12914623	15	80,993,570	*ABHD17C*	intronic	C	G	0.27	-0.0104±0.0016	1.6E-10	506,153
rs11853949	15	84,516,904	*ADAMTSL3*	intronic	A	G	0.29	-0.0224±0.0016	3.9E-43	463,363
rs11631096	15	85,100,929	*UBE2Q2P1*	ncRNA intronic	C	G	0.28	0.0155±0.0016	6.3E-22	503,325
rs3817428	15	89,415,247	*ACAN*	missense	C	G	0.73	0.0148±0.0016	7.7E-20	508,098
rs16946314	15	92,572,682	*SLCO3A1*	intronic	A	G	0.22	-0.0107±0.0018	8.5E-10	508,054
rs11073382	15	95,274,277	*LOC440311*	intergenic	A	G	0.49	-0.0082±0.0014	1.1E-08	508,048
rs12908437	15	99,287,375	*IGF1R*	intronic	T	C	0.37	0.011±0.0015	8.0E-14	531,509
rs2581348	15	100,514,063	*ADAMTS17*	UTR3	T	C	0.65	0.0088±0.0016	1.6E-08	442,278
rs72755233	15	100,692,953	*ADAMTS17*	missense	A	G	0.11	-0.0156±0.0023	3.2E-11	442,278
rs412243	16	339,672	*AXIN1*	intronic	T	C	0.62	0.0104±0.0015	1.3E-11	442,278
rs7203729	16	2,140,010	*PKD1*	synonymous	A	G	0.81	0.0129±0.0019	1.8E-11	442,278
rs188242066	16	2,803,298	*SRRM2*	intronic	T	C	0.01	0.0515±0.0075	5.2E-12	442,278
rs879620	16	4,015,729	*ADCY9*	UTR3	T	C	0.61	0.0138±0.0015	9.8E-21	508,053

rsid	Chr	Pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs79431809	16	4,299,089	*LINC01569*	ncRNA intronic	A	G	0.22	0.0125±0.0018	2.7E-12	442,278
rs2456779	16	4,925,817	*UBN1*	intronic	C	G	0.35	0.0108±0.0015	5.7E-13	508,060
rs2534760	16	6,509,009	*RBFOX1*	intronic	A	T	0.70	-0.0093±0.0016	3.8E-09	508,015
rs4780837	16	20,047,093	*GPR139*	intronic	A	G	0.78	-0.0095±0.0017	6.3E-08	508,086
rs4238585	16	20,255,097	*GP2*	intergenic	T	C	0.87	0.015±0.0021	1.7E-12	528,579
rs4483850	16	20,375,776	*PDILT*	intronic	A	T	0.50	0.0103±0.0014	5.3E-13	507,581
rs2343606	16	24,793,741	*TNRC6A*	intronic	T	C	0.27	-0.0134±0.0016	1.2E-16	508,055
rs2466826	16	28,348,344	*NPIPB6*	intergenic	A	G	0.40	0.0213±0.0015	3.0E-44	442,278
rs7187776	16	28,857,645	*TUFM*	UTR5	A	G	0.60	-0.0228±0.0014	3.3E-56	531,482
rs72798148	16	29,926,552	*KCTD13*	intronic	T	C	0.78	0.0156±0.0018	1.0E-17	442,278
rs111739641	16	30,510,984	*ITGAL*	intronic	D	I	0.50	0.0088±0.0015	3.6E-09	442,278
rs4889606	16	31,011,183	*STX1B*	intronic	A	G	0.62	0.0145±0.0015	2.9E-23	531,410
rs9923167	16	51,817,229	*LINC01571*	intergenic	C	G	0.91	0.0167±0.0027	3.3E-10	442,278
rs16951015	16	52,317,810	*TOX3*	intergenic	A	G	0.17	-0.0114±0.0019	1.6E-09	508,079
rs1558902	16	53,803,574	*FTO*	intronic	A	T	0.40	0.0422±0.0014	9.9E-189	530,202
rs12927987	16	66,720,206	*CMTM4*	intronic	A	T	0.48	-0.0088±0.0015	4.5E-09	442,278
rs192042440	16	69,285,510	*SNTB2*	intronic	T	C	0.03	-0.0312±0.0048	1.0E-10	442,278
rs3790085	16	69,887,720	*WWP2*	intronic	A	G	0.55	-0.0151±0.0014	3.1E-26	531,503
rs936994	16	70,442,408	*ST3GAL2*	intronic	T	C	0.50	0.0094±0.0014	5.3E-11	507,566
rs7191938	16	71,407,530	*CALB2*	intronic	A	G	0.30	-0.0098±0.0016	5.3E-10	507,143
rs811054	16	72,251,132	*PMFBP1*	intergenic	T	C	0.54	0.0096±0.0014	1.8E-11	529,756
rs756717	16	72,996,162	*ZFHX3*	intronic	A	G	0.40	-0.0096±0.0015	4.6E-11	531,517
rs4500770	16	74,658,430	*RFWD3*	intronic	A	T	0.64	0.0093±0.0015	2.0E-09	442,278
rs3751859	16	81,735,012	*CMIP*	intronic	A	G	0.16	-0.0126±0.002	1.8E-10	504,604
rs7206608	16	82,872,628	*CDH13*	intronic	C	G	0.68	-0.0102±0.0015	3.7E-11	508,091
rs2966859	16	85,324,544	*LINC00311*	intergenic	A	G	0.21	0.011±0.0018	1.9E-09	442,278
rs4843612	16	87,479,339	*ZCCHC14*	intronic	A	G	0.14	-0.0117±0.002	1.1E-08	506,494
rs3743860	16	89,818,491	*FANCA*	intronic	T	C	0.58	0.0109±0.0015	8.4E-14	494,742
rs8082551	17	1,310,661	*YWHAE*	intergenic	T	C	0.12	-0.0144±0.0023	2.6E-10	442,278
rs546133832	17	1,837,168	*RTN4RL1*	downstream	D	I	0.85	0.021±0.0021	5.0E-24	442,278
17:3986931_CT_C	17	3,986,931	*ZZEF1*	intronic	I	D	0.36	0.0104±0.0016	3.8E-11	442,278
rs3026101	17	5,280,440	*RABEP1*	synonymous	T	C	0.70	-0.0108±0.0015	2.5E-12	530,658
rs12940684	17	7,453,919	*TNFSF12,TNFSF12-TNFSF13*	intronic	T	C	0.70	-0.0118±0.0016	7.3E-14	508,030
rs7213608	17	21,279,289	*KCNJ12*	upstream	T	C	0.68	-0.0128±0.0016	1.6E-16	506,308
rs3115086	17	28,025,949	*SSH2*	intronic	T	C	0.50	-0.0104±0.0014	2.3E-13	528,345
rs3794809	17	28,531,258	*SLC6A4*	intronic	T	C	0.44	-0.012±0.0015	1.0E-15	442,278
17:31471175_CAAA_C	17	31,471,175	*ASIC2*	intronic	I	D	0.78	0.0116±0.0018	1.7E-10	442,278
rs12150665	17	34,914,787	*GGNBP2*	intronic	T	C	0.59	0.011±0.0014	2.1E-14	531,069
rs12945575	17	40,713,071	*COASY*	intergenic	T	C	0.25	0.0103±0.0017	2.2E-09	442,278
rs8176166	17	41,240,277	*BRCA1*	intronic	T	C	0.85	0.0142±0.0021	1.5E-11	442,278
rs2071167	17	42,287,519	*UBTF*	synonymous	T	C	0.24	0.0161±0.0017	8.2E-22	531,529
rs149053776	17	43,130,624	*DCAKD*	intronic	D	I	0.38	0.0112±0.0015	2.6E-13	442,278
rs200442988	17	44,566,824	*ARL17A*	intergenic	D	I	0.31	0.0128±0.0017	1.5E-13	442,278
rs7221720	17	46,168,892	*CBX1*	intronic	A	G	0.89	0.0072±0.0024	2.7E-03	442,278
rs208015	17	46,252,346	*SKAP1*	intronic	T	C	0.07	0.0253±0.0028	5.8E-19	507,186
rs138774094	17	46,275,597	*SKAP1*	intronic	T	C	0.02	0.011±0.0051	3.2E-02	442,278
rs11079849	17	47,090,785	*IGF2BP1*	intronic	T	C	0.33	-0.0146±0.0015	1.8E-21	506,310
rs4794538	17	52,925,502	*TOM1L1*	intergenic	T	C	0.83	-0.012±0.0019	5.2E-10	507,192
rs16947005	17	61,702,989	*MAP3K3*	intronic	A	G	0.72	0.0105±0.0016	6.4E-11	508,102
17:63960821_AAGGGTGC_A	17	63,960,821	*CEP112*	intronic	D	I	0.44	0.0089±0.0015	4.0E-09	442,278

rsid	Chr	Pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs12602912	17	65,870,073	*BPTF*	intronic	T	C	0.20	0.0201±0.0018	1.5E-29	528,553
rs142186653	17	73,879,851	*TRIM65*	intergenic	A	C	0.77	-0.0125±0.0018	4.4E-12	442,278
rs11150745	17	78,757,626	*RPTOR*	intronic	A	G	0.68	0.0126±0.0016	4.0E-15	442,278
rs9319615	17	79,072,594	*BALAP2*	intronic	A	C	0.45	0.0099±0.0015	3.3E-11	442,278
rs8086819	18	1,663,417	*LINC00470*	intergenic	A	G	0.66	-0.0097±0.0015	1.8E-10	508,096
rs11664106	18	2,846,812	*EMILIN2*	upstream	A	T	0.63	-0.0093±0.0016	4.3E-09	442,278
rs35939555	18	13,065,174	*CEP192*	intronic	D	I	0.67	-0.0103±0.0016	1.1E-10	442,278
rs891386	18	21,103,971	*C18orf8*	intronic	T	G	0.54	0.0171±0.0014	1.5E-32	508,086
rs16940859	18	22,184,421	*LOC729950*	intergenic	A	G	0.81	0.0135±0.0018	2.2E-13	508,054
rs16975921	18	39,914,870	*LINC00907*	ncRNA intronic	A	T	0.68	-0.0108±0.0015	2.6E-12	508,029
rs2052607	18	40,788,387	*SYT4*	intergenic	A	G	0.34	-0.0136±0.0015	2.4E-19	507,568
rs7233512	18	42,595,076	*SETBP1*	intronic	A	G	0.30	-0.0107±0.0016	9.1E-12	508,051
rs7239114	18	45,921,214	*CTIF*	intergenic	A	G	0.54	0.0088±0.0014	1.0E-09	526,306
rs3764512	18	52,495,324	*RAB27B*	upstream	T	G	0.23	0.01±0.0017	4.2E-09	508,101
rs663129	18	57,838,401	*MC4R*	intergenic	A	G	0.23	0.0261±0.0017	1.0E-54	530,028
rs10503034	18	57,977,741	*MC4R*	intergenic	T	C	0.69	-0.018±0.0015	2.0E-31	527,245
rs112099489	18	58,417,964	*MC4R*	intergenic	T	C	0.02	-0.0513±0.0056	5.7E-20	442,278
rs2012927	18	63,297,672	*CDH7*	intergenic	A	G	0.32	0.0108±0.0015	1.5E-12	531,481
rs8103489	19	3,508,499	*FZR1*	intronic	A	G	0.27	0.0103±0.0017	8.3E-10	488,184
rs188955288	19	4,067,314	*ZBTB7A*	upstream	T	C	0.19	-0.015±0.0019	5.1E-15	442,278
rs2620829	19	5,150,934	*KDM4B*	intronic	T	G	0.74	0.01±0.0017	4.6E-09	442,278
rs62621197	19	8,670,147	*ADAMTS10*	missense	T	C	0.04	-0.0247±0.0041	1.7E-09	442,278
rs11666430	19	8,673,593	*ADAMTS10*	intronic	T	C	0.36	-0.0077±0.0015	5.9E-07	442,278
rs12979274	19	10,000,925	*OLFM2*	intronic	T	C	0.48	0.0093±0.0014	9.9E-11	508,103
rs4808762	19	18,326,222	*PDE4C*	intronic	T	C	0.71	-0.0173±0.0016	1.0E-27	505,826
rs9636202	19	18,449,238	*PGPEP1*	intergenic	A	G	0.27	-0.014±0.0016	5.5E-18	526,695
rs10221489	19	18,826,975	*CRTC1*	intronic	T	C	0.34	-0.0138±0.0015	1.5E-19	506,311
rs3218036	19	30,305,684	*CCNE1*	intronic	A	G	0.33	0.0133±0.0015	2.8E-18	519,965
rs7258937	19	33,938,800	*PEPD*	intronic	T	C	0.51	0.0154±0.0014	6.5E-27	508,057
rs11670463	19	34,023,146	*PEPD*	intergenic	T	C	0.11	-0.0049±0.0024	4.1E-02	442,278
rs6857	19	45,392,254	*PVRL2*	UTR3	T	C	0.17	-0.0206±0.0019	1.2E-26	500,149
rs7412	19	45,412,079	*APOE*	missense	T	C	0.08	0.0201±0.0027	1.0E-13	465,689
rs1800437	19	46,181,392	*GIPR*	missense	C	G	0.20	-0.0232±0.0018	1.1E-37	511,388
rs3810291	19	47,569,003	*ZC3H4*	UTR3	A	G	0.68	0.0119±0.0015	5.9E-15	529,039
rs16996657	20	15,816,236	*MACROD2*	intronic	T	C	0.87	-0.0139±0.0022	1.2E-10	505,104
rs6083828	20	25,369,918	*ABHD12*	intronic	T	C	0.44	0.0119±0.0014	8.4E-17	531,545
rs2252934	20	25,957,535	*LOC101926955*	intergenic	C	G	0.56	0.0089±0.0015	1.0E-09	500,322
rs2881138	20	36,834,862	*KIAA1755*	intergenic	A	G	0.54	0.0086±0.0014	2.9E-09	507,191
rs6029180	20	39,178,923	*MAFB*	intergenic	A	G	0.67	-0.0098±0.0015	2.6E-10	506,305
rs17265513	20	39,832,628	*ZHX3*	missense	T	C	0.80	-0.0123±0.0018	5.7E-12	529,771
rs6103254	20	41,990,761	*SRSF6*	intergenic	T	C	0.87	0.0157±0.0023	3.7E-12	442,278
rs6031847	20	43,514,203	*YWHAB*	upstream	T	C	0.27	0.0098±0.0017	4.7E-09	442,278
rs56218501	20	46,365,636	*SULF2*	missense	T	C	0.21	-0.0144±0.0018	2.8E-15	442,278
rs112852122	20	47,498,117	*ARFGEF2*	intergenic	A	G	0.16	-0.0161±0.0021	5.9E-15	442,278
rs13043475	20	51,209,603	*LINC01524*	intergenic	A	G	0.81	0.016±0.0018	3.4E-18	508,105
rs6023649	20	53,470,583	*DOK5*	intergenic	A	G	0.26	0.0108±0.0017	8.8E-11	507,511
rs62217799	20	62,347,191	*ZGPAT*	intronic	T	G	0.66	0.0126±0.0016	1.6E-15	442,278
rs9976423	21	18,093,210	*MIR99AHG*	intergenic	A	G	0.45	-0.0097±0.0015	1.2E-10	442,278
rs73197346	21	36,770,189	*LOC100506403*	ncRNA intronic	T	C	0.87	0.0141±0.0022	1.7E-10	442,278
rs394608	21	46,581,798	*ADARB1*	intronic	T	C	0.46	-0.0124±0.0015	1.3E-16	442,278
rs424708	22	18,214,192	*BCL2L13*	downstream	C	G	0.83	-0.0119±0.0019	3.7E-10	507,902
rs2267373	22	38,600,542	*MAFF*	intronic	T	C	0.58	-0.0174±0.0015	8.5E-33	506,309
rs202657	22	41,844,786	*TOB2*	intergenic	T	G	0.80	-0.0145±0.0018	5.4E-16	507,603
rs8137409	22	42,359,793	*LINC00634*	intergenic	A	G	0.11	0.0137±0.0023	2.1E-09	505,547

rsid	Chr	Pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs4253755	22	46,615,376	*PPARA*	intronic	A	G	0.13	0.0162±0.0022	6.8E-14	506,308
rs9615905	22	48,875,699	*FAM19A5*	intergenic	T	C	0.46	0.0103±0.0014	9.7E-13	506,307
rs5934505	23	8,913,826	*FAM9B*	intergenic	T	C	0.73	0.0105±0.0014	4.1E-14	442,278
rs1379871	23	31,854,782	*DMD*	intronic	C	G	0.34	0.0133±0.0013	3.5E-24	442,278
rs201906093	23	53,108,771	*GPR173*	UTR3	I	D	0.15	-0.0114±0.0019	2.2E-09	442,278
rs78801842	23	53,652,838	*HUWE1*	intronic	A	G	0.49	0.0082±0.0013	9.8E-11	442,278
X:66222744_CAA_C	23	66,222,744	*EDA2R*	intergenic	I	D	0.84	0.0123±0.0017	9.1E-13	442,278
rs41303733	23	67,652,748	*OPHN1*	missense	T	C	0.08	0.0133±0.0022	1.9E-09	442,278
rs780579062	23	85,580,098	*DACH2*	intronic	D	I	0.33	-0.0091±0.0013	3.8E-12	442,278
rs2248999	23	117,902,604	*IL13RA1*	intronic	C	G	0.17	0.0136±0.0017	3.1E-16	442,278
rs210034	23	129,251,559	*ELF4*	intergenic	A	G	0.59	0.0088±0.0013	2.3E-12	442,278
rs5975828	23	135,986,549	*RBMX*	intergenic	T	C	0.28	0.0089±0.0014	1.3E-10	442,278
rs6633949	23	137,059,126	*ZIC3*	intergenic	C	G	0.36	0.0078±0.0013	1.1E-09	442,278
Women										
rs530311131	3	10,096,319	*FANCD2*	intronic	I	D	0.06	0.035±0.0058	1.6E-09	240,589
rs28412751	8	61,639,875	*CHD7*	intronic	T	C	0.45	0.016±0.0027	2.4E-09	240,589
rs1915792	8	64,703,342	*LINC01289*	intergenic	T	C	0.48	-0.016±0.0027	2.3E-09	240,589
rs10823885	10	53,674,590	*PRKG1*	intronic	A	T	0.42	0.016±0.0027	4.0E-09	240,589
rs78378788	11	89,075,754	*NOX4*	intronic	A	G	0.93	-0.0312±0.0053	5.3E-09	240,589
rs11116176	12	84,434,919	*SLC6A15*	intergenic	T	G	0.24	-0.0179±0.003	3.0E-09	268,571
rs12872889	13	28,674,628	*FLT3*	missens	T	C	0.77	-0.0202±0.0032	3.2E-10	240,589
rs1423798	16	65,289,733	*LINC00922*	intergenic	A	T	0.64	-0.0155±0.0027	4.9E-09	274,156
rs1075901	17	15,943,910	*NCOR1*	intronic	T	C	0.44	-0.0154±0.0025	7.1E-10	286,089
X:72227619_CATATATATATATAT_C	23	72,227,619	*PABPC1L2B*	intergenic	I	D	0.25	-0.0189±0.0032	3.3E-09	240,589
Men										
rs6545039	2	48,646,493	*PPP1R21*	intergenic	T	C	0.46	-0.0168±0.0027	7.3E-10	231,788
rs45449699	2	113,955,233	*PSD4*	intronic	T	C	0.15	0.0251±0.0041	7.7E-10	201,689
2:158119170_CAACAAT_C	2	158,119,170	*GALNT5*	intronic	I	D	0.89	0.0275±0.0047	4.9E-09	201,689
rs17058066	5	160,046,184	*ATP10B*	intronic	T	G	0.93	-0.0316±0.0053	2.8E-09	228,084
rs71231793	7	76,682,610	*PMS2P9*	downstream	T	C	0.84	-0.0287±0.0043	1.6E-11	201,689
rs2506141	10	33,468,014	*NRP1*	UTR3	T	C	0.54	0.0166±0.0028	2.6E-09	227,007
rs10766451	11	2,917,510	*SLC22A18AS*	intronic	T	C	0.32	0.0183±0.0031	2.8E-09	210,385
rs10875746	12	48,518,264	*PFKM*	intronic	A	C	0.76	-0.0187±0.0032	5.3E-09	233,131
rs770387	13	51,143,755	*DLEU1*	intergenic	T	C	0.79	0.0194±0.0033	6.3E-09	232,636
rs7170961	15	39,493,179	*C15orf54*	intergenic	T	C	0.69	-0.0178±0.0029	7.7E-10	246,094
rs12608143	18	36,492,188	*MIR924HG*	intergenic	T	C	0.53	-0.0168±0.0028	1.1E-09	232,634
rs143384	20	34,025,756	*GDF5*	UTR5	A	G	0.59	0.019±0.0027	3.3E-12	245,216
rs372038686	23	109,852,772	*CHRDL1*	intergenic	T	C	0.61	-0.0124±0.0021	6.5E-09	201,689
rs1150185	23	133,112,360	*GPC3*	intronic	T	C	0.12	0.0192±0.0032	2.1E-09	201,689

Supplementary Table 3.6: Independent variants associated with FFMI at genome-wide significance in meta-analyses of GWAS in UK Biobank and exome chip meta-analyses

rsid	chr	pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
Sex-combined										
rs4648729	1	1,808,769	*GNB1*	intron	T	C	0.506	-0.014±0.0014	3.0E-23	442,391
rs7535528	1	2,444,414	*PANK4*	missense	A	G	0.372	-0.0106±0.0014	2.0E-13	442,391
rs9435184	1	9,343,551	*SPSB1*	downstream	T	C	0.134	-0.0159±0.002	7.7E-15	442,391
rs9442571	1	9,349,611	*SPSB1*	upstream	A	T	0.130	-0.0164±0.0021	2.0E-15	442,391
rs12137140	1	9,454,067	*SPSB1*	intergenic	T	G	0.410	-0.0087±0.0014	1.0E-09	442,391
rs17396340	1	10,286,176	*KIF1B*	intron	A	G	0.127	-0.0134±0.0021	1.2E-10	442,391
rs12046278	1	10,799,577	*CASZ1*	intron	T	C	0.653	-0.01±0.0014	8.8E-13	535,885
rs36083532	1	10,839,085	*CASZ1*	intron	T	C	0.690	-0.0128±0.0015	2.2E-17	442,391
1:11254006_AC_A	1	11,254,006	*ANGPTL7, MTOR*	intron	D	I	0.782	-0.0118±0.0017	4.0E-12	442,391
rs3170740	1	17,312,743	*ATP13A2*	missense	T	C	0.529	-0.0142±0.0013	6.8E-26	507,518
rs12407400	1	19,964,620	*MINOS1-NBL1*	intron	A	G	0.238	-0.0114±0.0016	2.5E-12	442,391
rs945211	1	32,191,798	*ADGRB2*	downstream	C	G	0.616	0.0098±0.0014	7.3E-12	442,391
rs10753280	1	33,774,752	*A3GALT2*	intron	A	C	0.317	-0.0109±0.0015	3.2E-13	442,391
rs61744853	1	35,453,651	*ZMYM6*	missense	A	C	0.991	-0.0437±0.007	4.1E-10	535,885
rs569405229	1	36,581,053	*COL8A2*	intron	T	C	0.007	-0.0687±0.0089	1.0E-14	442,391
1:39738803_AT_A	1	39,738,803	*MACF1*	intron	D	I	0.207	0.0143±0.0017	2.3E-16	442,391
rs151224948	1	39,940,430	*MACF1*	intron	A	G	0.031	-0.0335±0.004	4.5E-17	442,391
1:46154575_AT_A	1	46,154,575	*TMEM69*	intron	D	I	0.466	-0.0086±0.0015	3.7E-09	442,391
rs6669341	1	47,678,458	*TAL1*	downstream	A	G	0.418	0.0101±0.0014	5.2E-13	442,391
rs11205538	1	49,237,594	*AGBL4, BEND5*	intron	T	C	0.714	-0.012±0.0015	5.9E-15	442,391
rs11205836	1	51,804,731	*TTC39A*	intron	T	C	0.192	0.0106±0.0017	5.9E-10	514,980
rs9970807	1	56,965,664	*PLPP3*	intron	T	C	0.092	0.0139±0.0023	1.8E-09	514,980
rs12140153	1	62,579,891	*PATJ*	missense	T	G	0.095	-0.0213±0.0024	2.6E-19	487,987
rs61779814	1	65,939,443	*LEPR*	intron	C	G	0.393	0.0113±0.0014	1.1E-15	442,391
rs2568958	1	72,765,116	*NEGR1*	upstream	A	G	0.603	0.015±0.0014	2.0E-28	535,885
rs1514175	1	74,991,644	*FPGT-TNNI3K,TNNI3K*	intron	A	G	0.426	0.0139±0.0013	4.1E-25	530,616
rs11162351	1	77,944,732	*AK5*	intron	C	G	0.604	-0.0102±0.0014	6.4E-13	442,391
rs140681455	1	78,444,764	*FUBP1*	5' UTR,NMD transcript	I	D	0.127	0.0236±0.0021	1.3E-28	442,391
rs34353539	1	79,331,762	*ADGRL4*	intergenic	D	I	0.272	0.0099±0.0016	2.3E-10	442,391
rs412378	1	82,374,494	*ADGRL2*	intron	A	G	0.755	-0.0094±0.0016	6.2E-09	442,391
rs12089815	1	91,189,933	*BARHL2*	intergenic	A	G	0.548	-0.0087±0.0014	5.2E-10	442,391
rs143792309	1	92,505,309	*EPHX4*	intron	T	C	0.064	0.0174±0.0029	1.2E-09	442,391
rs2391199	1	93,160,902	*EVI5*	missense	T	C	0.100	0.0194±0.0022	1.9E-18	535,885
rs1342396	1	96,281,460	*LOC102723661*	intergenic	T	C	0.670	0.0123±0.0015	7.4E-17	442,391
rs201973270	1	97,074,111	*PTBP2*	intergenic	A	G	0.310	0.0144±0.0015	9.7E-22	442,391

rsid	chr	pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs2811218	1	98,027,316	*DPYD*	intron	A	G	0.186	0.0109±0.0018	9.4E-10	442,391
rs2622848	1	103,421,003	*COL11A1*	intron	T	C	0.729	0.0104±0.0016	1.9E-11	442,391
rs76210342	1	107,967,807	*NTNG1*	intron	D	I	0.616	0.0111±0.0016	9.2E-13	442,391
rs1933182	1	109,999,838	*SYPL2*	intergenic	A	C	0.297	0.0137±0.0015	8.7E-21	512,061
rs7550711	1	110,082,886	*GPR61*	intron NMD transcript	T	C	0.027	0.0383±0.0042	3.4E-20	514,980
rs197402	1	112,302,591	*DDX20*	intron	A	G	0.656	-0.0103±0.0015	1.1E-12	442,391
rs34611728	1	113,255,456	*PPM1J*	missense	A	C	0.129	0.0121±0.002	9.3E-10	535,885
rs6674967	1	118,854,698	*SPAG17*	intergenic	A	G	0.255	0.012±0.0016	3.9E-14	442,391
rs984222	1	119,503,843	*TBX15*	intron	C	G	0.386	0.0141±0.0014	2.3E-24	508,884
rs10802069	1	119,517,357	*TBX15*	intron	T	C	0.387	0.0143±0.0014	8.3E-24	442,391
rs11205303	1	149,906,413	*MTMR11*	missense	T	C	0.594	0.0199±0.0014	3.8E-49	535,885
rs2902811	1	151,010,521	*BNIPL*	intron	C	G	0.768	-0.0126±0.0016	1.4E-14	442,391
rs1194610	1	154,296,076	*AQP10*	synonymous	T	C	0.765	-0.0097±0.0016	2.3E-09	442,391
rs3753639	1	154,986,091	*ZBTB7B*	intron	T	C	0.756	-0.0202±0.0016	8.3E-36	442,391
rs146554560	1	155,661,719	*DAP3*	intron	T	C	0.023	0.0508±0.0047	2.7E-27	442,391
rs2297792	1	156,011,444	*UBQLN4*	missense	T	C	0.373	-0.0085±0.0014	6.5E-10	535,885
rs35208023	1	156,415,673	*C1orf61*	downstream	T	C	0.669	0.0115±0.0015	5.9E-15	442,391
1:170705685_CTATATATATA_C	1	170,705,685	*PRRX1*	3' UTR	I	D	0.379	0.0125±0.0014	3.7E-18	442,391
rs10919515	1	170,869,758	*MROH9*	intergenic	A	T	0.095	-0.0197±0.0026	4.1E-14	442,391
rs4916229	1	171,443,368	*PRRC2C*	intergenic	C	G	0.904	-0.0185±0.0024	3.8E-15	442,391
rs139303240	1	173,901,373	*RC3H1*	3' UTR	D	I	0.888	-0.0165±0.0022	6.5E-14	442,391
1:174455489_CT_C	1	174,455,489	*RABGAP1L*	intron	I	D	0.846	-0.0171±0.0021	2.2E-16	442,391
rs12087196	1	174,962,071	*RABGAP1L*	3' UTR	A	G	0.888	-0.0152±0.0022	4.1E-12	442,391
rs543874	1	177,889,480	*SEC16B*	downstream	A	G	0.795	-0.0331±0.0016	1.5E-89	532,966
rs1046934	1	184,023,529	*TSEN15*	missense	A	C	0.651	0.0097±0.0014	2.6E-12	535,885
rs655598	1	190,287,713	*BRINP3*	intron	A	G	0.562	-0.012±0.0014	5.9E-18	442,391
rs778404403	1	191,434,186	*RGS18*	intergenic	D	I	0.292	-0.0077±0.0015	6.1E-07	442,391
rs34781011	1	194,890,316	*LINC01724*	intergenic	I	D	0.292	-0.0087±0.0015	1.0E-08	442,391
1:197133836_AATATATTAT_A	1	197,133,836	*ZBTB41*	intron	D	I	0.280	-0.0101±0.0016	2.3E-10	442,391
rs3850625	1	201,016,296	*CACNA1S*	missense	A	G	0.119	-0.0165±0.002	8.4E-16	535,885
rs2820312	1	201,869,257	*LMOD1*	missense	A	G	0.339	0.0163±0.0014	2.4E-31	535,885
rs4971212	1	203,491,392	*ATP2B4*	intergenic	T	G	0.799	-0.0106±0.0017	7.2E-10	442,391
rs1354853	1	209,557,991	*LINC01698*	downstream	A	G	0.867	-0.0126±0.002	6.4E-10	442,391
rs2605100	1	219,644,224	*LOC102723886*	intergenic	A	G	0.315	-0.0086±0.0014	1.8E-09	535,885
rs10863578	1	221,213,997	*HLX*	intergenic	C	G	0.670	0.0134±0.0015	9.4E-20	442,391
rs12079987	1	221,715,102	*DUSP10*	intergenic	A	G	0.024	-0.0317±0.0045	2.7E-12	442,391
rs7518221	1	225,561,346	*DNAH14*	intron NMD transcript	T	C	0.353	0.0089±0.0014	6.7E-10	442,391
rs2154367	1	227,153,930	*COQ8A*	intron	A	G	0.213	0.0136±0.0017	1.2E-15	442,391
rs12141801	1	227,961,403	*SNAP47*	intron	T	C	0.235	0.0113±0.0016	5.5E-12	442,391

rsid	chr	pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs287615	1	228,489,090	*OBSCN*	non-coding transcript exon variant	A	G	0.091	-0.0184±0.0024	3.3E-14	442,391
rs147929768	1	235,473,509	*ARID4B*	intron	T	G	0.028	0.0251±0.0042	2.6E-09	442,391
1:243815349_CT_C	1	243,815,349	*AKT3*	intron	I	D	0.230	0.0113±0.0018	1.7E-10	442,391
rs300789	2	89,910	*FAM110C*	intergenic	A	G	0.798	-0.0103±0.0017	3.5E-09	442,391
rs28715684	2	410,501	*LINC01865*	intron	T	C	0.267	0.0161±0.0017	4.8E-21	442,391
rs62106258	2	417,167	*LINC01874*	upstream	T	C	0.952	0.0678±0.0032	3.0E-98	442,391
rs546003868	2	558,371	*LOC105373352*	intron	A	C	0.064	-0.0297±0.0031	1.2E-21	442,391
rs2867125	2	622,827	*TMEM18*	intergenic	T	C	0.172	-0.0421±0.0018	2.2E-126	532,966
rs60996394	2	644,512	*TMEM18*	intergenic	A	C	0.462	0.0236±0.0018	2.0E-39	442,391
rs2867115	2	651,382	*TMEM18*	intergenic	T	C	0.882	0.0494±0.0024	5.9E-96	442,391
rs7578605	2	1,643,208	*PXDN*	intron	A	G	0.937	0.0225±0.0029	3.4E-15	442,391
rs2609181	2	6,161,899	*LOC400940*	intergenic	C	G	0.285	-0.0111±0.0015	6.9E-13	442,391
rs10203320	2	9,771,620	*YWHAQ*	upstream	T	C	0.671	-0.0092±0.0015	4.9E-10	442,391
rs713586	2	25,158,008	*DNAJC27*	intergenic	T	C	0.512	-0.0208±0.0013	3.6E-55	535,885
2:25365642_TAATA_T	2	25,365,642	*EFR3B*	intron	D	I	0.654	0.0193±0.0015	5.1E-40	442,391
rs1631026	2	26,953,850	*KCNK3*	3' UTR	T	C	0.472	0.0099±0.0014	1.1E-12	442,391
rs202049562	2	27,779,565	*C2orf16*	intron	A	T	0.179	0.0178±0.0019	8.5E-22	442,391
rs6547820	2	28,281,545	*BABAM2*	intron	A	G	0.788	-0.012±0.0017	9.4E-13	442,391
rs404108	2	33,301,389	*LTBP1*	intron	A	G	0.765	-0.0098±0.0016	2.1E-09	442,391
rs9678810	2	33,398,399	*LTBP1*	intron	A	G	0.448	-0.0102±0.0014	1.5E-12	442,391
rs2305575	2	33,745,336	*RASGRP3*	intron	T	C	0.219	-0.0104±0.0017	5.0E-10	442,391
rs10204994	2	35,443,726	*MIR548AD*	downstream	A	G	0.229	-0.011±0.0016	2.2E-11	442,391
rs62132389	2	36,478,876	*LOC100288911*	intergenic	C	G	0.074	0.0156±0.0027	4.7E-09	442,391
rs7589518	2	37,995,422	*LINC00211*	regulatory	A	T	0.486	0.0124±0.0015	2.4E-17	442,391
rs6713781	2	40,291,940	*SLC8A1-AS1*	intron	C	G	0.400	-0.0097±0.0014	1.1E-11	442,391
rs6742827	2	40,640,326	*SLC8A1*	intron	T	C	0.169	0.0117±0.0018	2.4E-10	442,391
rs11690012	2	43,542,480	*THADA*	intron	C	G	0.230	-0.0096±0.0016	5.9E-09	442,391
rs35809007	2	47,019,521	*LINC01118*	intergenic	A	G	0.363	-0.0111±0.0014	1.5E-14	442,391
rs72618637	2	48,953,979	*LHCGR,STON1-GTF2A1L*	intron	A	T	0.189	-0.0137±0.0018	2.7E-14	442,391
rs9636391	2	50,201,110	*NRXN1*	missense	A	G	0.163	0.0151±0.0019	8.4E-16	442,391
rs6724631	2	50,713,331	*NRXN1*	intron	C	G	0.536	-0.0103±0.0014	9.9E-14	442,391
rs6760379	2	55,279,536	*RTN4*	intron	T	C	0.693	0.0098±0.0015	8.3E-11	442,391
rs3791679	2	56,096,892	*EFEMP1*	intron	A	G	0.773	-0.016±0.0016	3.8E-24	535,885
rs11688767	2	57,988,194	*VRK2*	intron	A	T	0.514	0.0096±0.0014	3.5E-12	442,391
rs2708149	2	58,951,187	*LINC01122*	intron	A	G	0.431	-0.0155±0.0014	2.6E-28	442,391
rs7424120	2	59,313,974	*LINC01122*	intergenic	T	C	0.601	-0.0107±0.0014	5.1E-14	442,391
rs12476772	2	60,203,917	*MIR4432HG*	intron	A	C	0.331	0.0099±0.0015	1.9E-11	442,391
rs76511385	2	61,734,552	*XPO1*	intron	T	C	0.987	-0.0377±0.0062	9.2E-10	442,391
rs10192894	2	62,838,936	*EHBP1*	intron	A	G	0.560	-0.0087±0.0014	4.5E-10	442,391

rsid	chr	pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs2861691	2	67,838,694	*LINC01812*	intron	C	G	0.389	-0.0104±0.0014	2.4E-13	442,391
rs34355307	2	86,800,506	*RNF103-CHMP3*	intron	A	G	0.610	-0.0111±0.0014	8.5E-15	442,391
rs11692215	2	100,746,573	*AFF3*	intron	T	C	0.242	-0.0119±0.0016	1.6E-13	442,391
rs10489971	2	102,527,827	*MAP4K4*	intergenic	T	C	0.307	-0.009±0.0015	2.1E-09	442,391
2:104130139_AT_A	2	104,130,139	*LINC01935*	intergenic	D	I	0.497	0.0093±0.0014	3.4E-11	442,391
rs6739701	2	105,132,332	*LINC01102*	intron	A	G	0.568	0.0102±0.0014	3.4E-13	442,391
rs10197031	2	105,454,590	*LINC01158*	intron	T	C	0.714	-0.0132±0.0015	5.7E-18	442,391
2:144001846_GA_G	2	144,001,846	*ARHGAP15*	intron	D	I	0.140	0.0127±0.002	3.4E-10	442,391
rs497418	2	147,842,671	*PABPC1P2*	intergenic	A	C	0.385	0.0113±0.0014	2.5E-15	442,391
2:157597637_GA_G	2	157,597,637	*GPD2*	intergenic	D	I	0.398	0.0086±0.0014	2.1E-09	442,391
rs13023477	2	161,111,887	*LINC02478*	intron	T	C	0.186	0.0127±0.0018	9.7E-13	442,391
rs767585131	2	164,541,449	*FIGN*	intron	D	I	0.147	0.0133±0.002	3.3E-11	442,391
rs765335847	2	172,425,405	*CYBRD1*	intergenic	D	I	0.250	-0.0142±0.0016	8.5E-19	442,391
rs34234296	2	175,166,636	*LINC01305*	upstream	A	G	0.393	-0.0086±0.0014	1.9E-09	442,391
rs72923454	2	176,984,544	*HOXD10*	3' UTR	T	C	0.025	0.0305±0.0044	5.2E-12	442,391
rs9630985	2	181,607,676	*SCHLAP1*	regulatory	A	C	0.334	-0.0097±0.0015	3.3E-11	442,391
rs7422875	2	190,163,961	*COL5A2*	intergenic	T	C	0.956	-0.02±0.0034	2.7E-09	442,391
rs61327788	2	198,338,222	*COQ10B*	intron	A	G	0.499	-0.015±0.0014	6.2E-27	442,391
rs1064213	2	198,950,240	*PLCL1*	missense	A	G	0.478	0.0172±0.0013	2.2E-38	532,966
rs10497820	2	199,612,796	*SATB2*	intron	A	C	0.341	0.0107±0.0015	4.0E-13	442,391
rs6705250	2	200,307,014	*SATB2*	intron	A	G	0.862	-0.0132±0.002	6.1E-11	442,391
rs4673627	2	200,848,896	*MAIP1*	intron	A	G	0.560	0.0081±0.0014	5.9E-09	442,391
rs4482463	2	205,375,909	*PARD3B*	downstream	A	C	0.924	-0.0208±0.0026	1.7E-15	442,391
rs13000757	2	207,145,906	*ZDBF2*	intron	T	G	0.507	0.01±0.0014	6.0E-13	442,391
rs1263615	2	207,959,238	*KLF7*	intron	T	C	0.758	-0.0107±0.0016	3.3E-11	442,391
rs1047891	2	211,540,507	*CPS1*	missense	A	C	0.316	0.0158±0.0015	2.4E-26	442,391
rs6435622	2	212,296,094	*ERBB4*	intron	C	G	0.311	0.0089±0.0015	2.5E-09	442,391
rs13427822	2	213,414,265	*ERBB4*	intergenic	A	G	0.728	0.0131±0.0016	7.7E-17	442,391
rs2712184	2	217,682,779	*TNP1*	intron	A	C	0.574	0.0119±0.0014	2.7E-18	508,913
rs2230115	2	219,509,618	*ZNF142*	missense	A	C	0.571	-0.0127±0.0013	4.2E-21	535,885
rs11683846	2	220,205,929	*RESP18*	intergenic	A	C	0.151	-0.0158±0.0019	3.9E-16	442,391
rs55760516	2	220,354,108	*SPEG*	missense	A	G	0.676	0.0114±0.0014	7.3E-16	535,885
rs2197780	2	223,960,207	*KCNE4*	intron	T	C	0.280	-0.0093±0.0015	1.4E-09	442,391
rs1515098	2	227,073,854	*LOC646736*	intergenic	T	C	0.681	0.0101±0.0015	1.1E-11	442,391
rs10195674	2	229,016,094	*SPHKAP*	intron	A	G	0.339	0.009±0.0015	7.8E-10	442,391
rs111828944	2	230,250,876	*DNER*	intron	A	G	0.092	0.014±0.0024	4.7E-09	442,391
2:230754796_CT_C	2	230,754,796	*TRIP12*	intron	I	D	0.248	-0.011±0.0016	2.3E-11	442,391
rs2101435	2	233,115,158	*DIS3L2*	intron	C	G	0.722	-0.0103±0.0015	2.8E-11	442,391
rs7568228	2	236,848,488	*AGAP1*	intron	C	G	0.526	-0.0089±0.0014	1.2E-10	442,391
rs7340253	2	239,374,836	*ASB1*	intergenic	A	G	0.560	-0.0086±0.0014	6.6E-10	442,391

rsid	chr	pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs7559460	2	241,341,322	*GPC1*	intergenic	T	C	0.613	0.0088±0.0014	8.2E-10	442,391
rs2108485	2	242,021,742	*SNED1*	missense	T	G	0.155	-0.0144±0.0018	3.3E-15	535,885
rs4428159	3	11,196,516	*HRH1*	intron	C	G	0.095	0.0158±0.0025	1.1E-10	442,391
rs17776482	3	11,633,951	*VGLL4*	intron	A	G	0.140	0.0157±0.002	7.1E-15	442,391
rs3732671	3	13,407,556	*NUP210*	missense	T	C	0.433	0.0085±0.0013	2.2E-10	535,885
rs2167196	3	13,593,475	*FBLN2*	intron	T	G	0.813	0.0135±0.0018	3.4E-14	442,391
rs2607744	3	14,235,285	*LSM3*	intron	C	G	0.510	-0.0084±0.0014	1.4E-09	442,391
rs2470537	3	15,716,898	*ANKRD28*	intron	T	C	0.598	-0.0085±0.0014	1.5E-09	442,391
rs34373881	3	20,432,033	*LOC101927829*	non-coding exon	A	G	0.277	-0.0109±0.0016	2.2E-12	442,391
rs7619139	3	25,110,415	*RARB*	intron	A	T	0.589	0.0099±0.0014	3.0E-12	442,391
rs11921432	3	35,117,776	*LOC101928135*	intergenic	T	C	0.886	-0.0129±0.0022	5.0E-09	442,391
rs6599205	3	38,529,440	*ACVR2B*	3' UTR	C	G	0.386	-0.0146±0.0014	9.1E-25	442,391
rs111768603	3	42,329,113	*CCK*	intergenic	T	G	0.110	-0.0148±0.0022	2.2E-11	442,391
rs123509	3	42,733,468	*KLHL40*	missense	T	C	0.248	0.0117±0.0015	4.4E-14	531,394
rs11711770	3	43,066,768	*FAM198A*	intron NMD transcript	A	G	0.829	-0.0108±0.0018	5.0E-09	442,391
rs9852062	3	45,373,442	*LARS2*	intergenic	A	T	0.556	-0.0095±0.0014	1.4E-11	442,391
rs62246403	3	47,095,177	*SETD2*	intron NMD transcript	T	C	0.036	0.0237±0.0037	2.3E-10	442,391
rs72906474	3	47,817,007	*SMARCC1*	intron	T	G	0.584	-0.0124±0.0014	2.9E-18	442,391
rs55786114	3	48,982,335	*ARIH2*	intron	T	C	0.074	0.0211±0.0027	1.7E-15	442,391
rs55754265	3	49,687,486	*BSN*	intron	T	C	0.139	0.0212±0.002	4.0E-26	442,391
rs2526389	3	50,192,826	*SEMA3F-AS1*	5' UTR	T	C	0.426	0.0209±0.0014	7.7E-50	442,391
3:50785304_AG_A	3	50,785,304	*DOCK3*	intron	D	I	0.864	0.0208±0.002	1.8E-24	442,391
rs10433609	3	51,311,574	*DOCK3*	intron	A	T	0.821	0.0199±0.0018	5.7E-28	442,391
rs141434554	3	52,208,046	*POC1A*	intergenic	D	I	0.780	-0.0128±0.0017	1.5E-13	442,391
rs772585020	3	52,725,718	*GNL3*	intron	D	I	0.490	0.0177±0.0014	7.3E-37	442,391
rs150638279	3	53,426,450	*DCP1A*	upstream	D	I	0.931	0.0192±0.0027	2.7E-12	442,391
rs1564907	3	56,193,500	*ERC2*	intron	T	C	0.107	0.014±0.0022	4.5E-10	442,391
rs6445198	3	61,219,865	*FHIT*	intron	T	G	0.416	-0.0112±0.0014	1.6E-15	442,391
rs56358876	3	61,544,965	*PTPRG*	upstream	T	G	0.888	-0.0145±0.0022	4.2E-11	442,391
rs925018	3	62,713,143	*CADPS*	intron	C	G	0.685	-0.0108±0.0015	4.1E-13	442,391
rs4142723	3	66,683,941	*LOC105377143*	upstream	A	T	0.236	-0.0101±0.0016	7.6E-10	442,391
rs17732997	3	70,470,834	*SAMMSON*	intergenic	C	G	0.569	-0.0097±0.0014	4.9E-12	442,391
rs68093214	3	70,657,785	*FOXP1*	intergenic	T	C	0.751	0.0111±0.0016	5.2E-12	442,391
rs9809116	3	72,397,279	*RYBP*	regulatory	A	G	0.593	-0.0098±0.0014	5.5E-12	442,391
rs6805432	3	78,870,517	*ROBO1*	intron	T	C	0.703	0.0108±0.0015	1.1E-12	442,391
rs116631513	3	84,603,308	*LINC00971*	intergenic	A	T	0.946	0.0246±0.0031	1.2E-15	442,391
3:85660399_GT_G	3	85,660,399	*CADM2*	intron	D	I	0.640	-0.0161±0.0015	3.9E-28	442,391
rs9851777	3	88,267,467	*C3orf38*	intergenic	T	C	0.116	-0.0144±0.0022	2.6E-11	442,391
rs1454687	3	94,038,085	*NSUN3*	intergenic	C	G	0.484	0.0146±0.0014	6.5E-26	442,391
3:104629856_CT_C	3	104,629,856	*ALCAM*	intergenic	I	D	0.236	-0.0112±0.0017	1.7E-11	442,391

rsid	chr	pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs2399213	3	107,395,302	*BBX*	intron	T	C	0.819	-0.014±0.0018	1.2E-14	442,391
rs1471093	3	108,031,094	*HHLA2*	intron	A	G	0.617	0.0089±0.0014	6.0E-10	442,391
rs17619973	3	114,417,675	*ZBTB20*	intron	A	G	0.925	0.0158±0.0026	2.2E-09	442,391
rs9881641	3	117,602,500	*LINC02024*	intron	T	C	0.687	-0.009±0.0015	1.9E-09	442,391
rs1870931	3	119,757,013	*GSK3B*	intron	C	G	0.828	0.0105±0.0018	1.2E-08	442,391
rs17282078	3	124,481,760	*ITGB5*	3' UTR	A	T	0.131	-0.0146±0.0021	1.5E-12	442,391
rs76594121	3	128,189,391	*DNAJB8-AS1*	intron	T	G	0.953	0.0239±0.0034	1.1E-12	442,391
rs1225004	3	131,626,991	*CPNE4*	intron	T	C	0.722	-0.016±0.0015	5.7E-25	442,391
rs138628404	3	131,771,462	*CPNE4*	intron	T	C	0.833	-0.0164±0.0019	3.5E-17	442,391
rs9827728	3	133,971,558	*RYK*	upstream	A	T	0.880	-0.0149±0.0021	3.1E-12	442,391
rs9844666	3	135,974,216	*PCCB*	5' UTR	A	G	0.240	0.017±0.0016	1.1E-27	532,966
rs13080786	3	136,543,708	*SLC35G2*	intron	A	G	0.762	-0.0163±0.0016	1.3E-23	442,391
rs59714050	3	141,267,294	*RASA2*	intron	A	T	0.066	0.0213±0.0028	2.9E-14	442,391
rs9438	3	154,018,887	*DHX36*	missense	C	G	0.394	0.0112±0.0014	2.7E-16	529,789
rs4383468	3	156,880,688	*CCNL1*	non-coding transcript exon variant	A	G	0.663	-0.0113±0.0015	1.1E-14	442,391
rs62644679	3	157,946,897	*RSRC1*	intron	A	G	0.323	-0.0105±0.0016	9.8E-11	442,391
rs67481408	3	170,734,443	*SLC2A2*	intron	D	I	0.726	-0.0126±0.0016	9.0E-16	442,391
rs13085472	3	171,129,859	*TNIK*	intron	T	C	0.332	0.0091±0.0015	8.7E-10	442,391
rs556464	3	173,113,680	*NLGN1*	upstream	C	G	0.472	-0.0104±0.0014	1.0E-13	442,391
rs9841616	3	181,167,585	*SOX2-OT*	intron	A	T	0.170	0.0116±0.0019	3.7E-10	442,391
rs11546878	3	183,976,103	*ECE2*	missense	T	C	0.176	-0.0162±0.0018	2.9E-20	535,885
rs1001817	3	183,995,341	*ECE2*	intron	T	C	0.493	-0.012±0.0013	1.7E-19	535,885
rs9816226	3	185,834,499	*ETV5*	intron	A	T	0.185	-0.0256±0.0017	1.2E-48	498,816
rs74935402	3	195,747,247	*TFRC*	intergenic	T	C	0.970	0.0246±0.0042	4.3E-09	442,391
rs2051559	4	3,298,800	*RGS12*	intron	T	C	0.867	-0.0146±0.0021	1.1E-12	442,391
rs11377284	4	4,800,758	*LOC101928279*	intergenic	D	I	0.306	-0.0093±0.0015	1.2E-09	442,391
rs11722554	4	5,016,883	*CYTL1*	missense	A	G	0.038	0.0259±0.0035	1.4E-13	522,343
rs11721374	4	8,599,520	*CPZ*	intron	A	G	0.435	0.0093±0.0014	3.4E-11	442,391
rs6833878	4	10,005,555	*SLC2A9*	intron	A	T	0.732	-0.0095±0.0016	1.3E-09	442,391
rs5856175	4	12,918,813	*RAB28*	intergenic	I	D	0.858	0.0125±0.002	3.8E-10	442,391
rs6449393	4	18,507,692	*LCORL*	intergenic	A	C	0.667	-0.009±0.0015	1.0E-09	442,391
rs544200874	4	20,124,826	*SLIT2*	intergenic	T	C	0.082	0.0193±0.0027	2.0E-12	442,391
rs34811474	4	25,408,838	*ANAPC4*	missense	A	G	0.231	-0.023±0.0016	5.6E-48	535,885
rs73245728	4	26,206,514	*RBPJ*	intron	A	C	0.936	0.0206±0.0029	5.1E-13	442,391
rs9999056	4	26,880,639	*STIM2*	intron	T	C	0.774	-0.0103±0.0017	5.2E-10	442,391
4:28526950_AAAAC_A	4	28,526,950	*MIR4275*	intron	D	I	0.166	-0.0127±0.0019	1.2E-11	442,391
rs7434610	4	30,839,793	*PCDH7*	intron	A	G	0.460	-0.0126±0.0014	1.3E-19	442,391
rs139430988	4	31,031,755	*PCDH7*	intron	D	I	0.645	-0.0125±0.0015	6.0E-16	442,391
rs3209570	4	38,699,657	*KLF3*	3' UTR	A	G	0.399	-0.0095±0.0014	2.8E-11	442,391
rs10938397	4	45,182,527	*GNPDA2*	intergenic	A	G	0.565	-0.018±0.0014	4.2E-40	515,854

rsid	chr	pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
4:53738084_AT_A	4	53,738,084	*SCFD2*	downstream	D	I	0.356	0.0103±0.0016	3.3E-10	442,391
rs147565558	4	54,330,689	*LNX1*	intron	D	I	0.848	0.0125±0.0019	1.3E-10	442,391
rs34772064	4	55,495,948	*LINC02260*	intergenic	T	G	0.443	0.0106±0.0014	4.5E-14	442,391
rs781659	4	57,779,851	*REST*	intron	A	G	0.573	-0.0092±0.0014	6.4E-11	442,391
rs78862524	4	73,171,190	*ADAMTS3*	intron	A	C	0.055	0.0135±0.0031	1.6E-05	442,391
rs141374503	4	73,179,445	*ADAMTS3*	missense	T	C	0.004	0.0747±0.0107	2.6E-12	535,885
rs10518106	4	73,519,842	*ADAMTS3*	intergenic	T	C	0.064	-0.0389±0.0028	1.1E-42	442,391
rs56091759	4	73,678,173	*COX18*	downstream	T	C	0.611	-0.0051±0.0014	3.6E-04	442,391
rs139457833	4	77,126,502	*SCARB2*	intron	D	I	0.849	-0.0111±0.0019	1.1E-08	442,391
rs994014	4	82,165,790	*PRKG2*	intergenic	T	C	0.694	0.0106±0.0015	2.4E-12	442,391
rs1054627	4	88,732,692	*IBSP*	missense	A	G	0.313	0.0098±0.0014	1.0E-11	532,966
rs2199936	4	89,045,331	*ABCG2*	intron	A	G	0.114	-0.0148±0.0022	1.2E-11	442,391
rs9993839	4	95,307,863	*HPGDS*	intergenic	T	C	0.804	0.0115±0.0017	5.9E-11	442,391
rs17620626	4	101,086,037	*DDIT4L*	intron	A	G	0.056	0.0174±0.003	9.3E-09	442,391
rs201081507	4	102,681,041	*BANK1*	intron	A	G	0.943	-0.022±0.0033	1.4E-11	442,391
rs13107325	4	103,188,709	*SLC39A8*	missense	T	C	0.074	0.0289±0.0026	1.4E-29	535,885
4:103981698_GT_G	4	103,981,698	*SLC9B2*	intron	D	I	0.404	-0.0103±0.0014	4.7E-13	442,391
rs75544266	4	104,584,997	*TACR3*	intron	T	C	0.056	-0.0215±0.003	1.5E-12	442,391
rs77096252	4	105,537,183	*CXXC4-AS1*	intron	A	G	0.956	0.0201±0.0034	4.2E-09	442,391
rs17509519	4	109,192,790	*LEF1-AS1*	intergenic	T	C	0.423	-0.0091±0.0014	9.2E-11	442,391
rs2595105	4	111,552,761	*PITX2*	intron	T	C	0.302	0.0102±0.0015	1.8E-11	442,391
rs9991259	4	112,686,354	*C4orf32*	intergenic	A	G	0.633	0.0086±0.0014	2.4E-09	442,391
rs57590313	4	113,323,430	*ALPK1*	intron	A	C	0.193	0.0128±0.0018	1.5E-12	442,391
rs28454448	4	115,125,392	*ARSJ*	intergenic	A	G	0.280	0.0098±0.0015	2.8E-10	442,391
rs7671759	4	120,248,094	*FABP2*	upstream	A	C	0.661	-0.0142±0.0015	4.0E-22	442,391
rs7688177	4	120,760,872	*LINC01365*	intergenic	T	C	0.770	-0.0112±0.0017	1.5E-11	442,391
rs2952862	4	130,758,713	*LINC02466*	intron	T	C	0.344	0.0094±0.0015	1.6E-10	442,391
rs1283032	4	137,098,470	*LINC00613*	intron	T	C	0.412	0.0119±0.0014	6.9E-17	442,391
rs35107973	4	143,125,948	*INPP4B*	intron	T	C	0.119	0.0126±0.0021	4.0E-09	442,391
rs7666785	4	144,060,464	*LOC105377623*	intron	A	G	0.400	0.0094±0.0014	9.6E-12	514,980
rs11933087	4	145,722,862	*HHIP*	intergenic	A	T	0.629	0.012±0.0014	1.1E-16	442,391
rs28605564	4	147,388,846	*SLC10A7*	intron	C	G	0.752	0.0095±0.0016	3.2E-09	442,391
rs35016840	4	151,239,774	*LRBA*	intron	T	C	0.641	-0.0096±0.0014	3.0E-11	442,391
rs6843852	4	162,132,758	*FSTL5*	intergenic	T	C	0.508	0.0104±0.0014	6.7E-14	442,391
rs148636479	4	164,359,678	*KTL2*	intergenic	A	T	0.971	0.0266±0.0044	1.9E-09	442,391
rs111598585	4	171,635,471	*LINC02382*	intergenic	T	C	0.208	-0.0122±0.0017	1.1E-12	442,391
rs7683836	4	180,167,906	*LINC01098*	intergenic	A	G	0.558	-0.0088±0.0014	3.0E-10	442,391
rs10866554	5	3,489,595	*LINC01019*	intron	T	C	0.481	0.0082±0.0014	3.4E-09	442,391
rs3792752	5	32,768,634	*NPR3*	intron	A	G	0.742	0.0141±0.0016	6.9E-19	442,391
rs11952835	5	42,331,880	*GHR*	intergenic	T	G	0.101	0.0126±0.0023	7.3E-08	442,391
rs13176429	5	43,152,216	*ZNF131*	intron	T	C	0.313	-0.0115±0.0015	1.2E-14	442,391
rs114263339	5	50,932,343	*ISL1*	intergenic	T	C	0.026	0.0242±0.0044	3.0E-08	442,391

rsid	chr	pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs6871876	5	53,120,750	*LINC02105*	intergenic	A	G	0.463	0.0114±0.0014	3.6E-16	442,391
rs4865796	5	53,272,664	*ARL15*	intron	A	G	0.692	-0.0101±0.0015	2.2E-11	442,391
rs700910	5	53,439,777	*ARL15*	intron	T	C	0.838	-0.0105±0.0019	2.9E-08	442,391
rs459193	5	55,806,751	*C5orf67*	downstream	A	G	0.254	-0.0132±0.0015	7.3E-18	532,966
rs13179413	5	55,868,097	*C5orf67*	intron	T	C	0.285	0.0104±0.0016	2.5E-11	442,391
rs60593437	5	56,253,216	*MIER3*	intron	I	D	0.355	0.0132±0.0015	1.2E-19	442,391
5:60708793_AT_A	5	60,708,793	*ZSWIM6*	intron	D	I	0.362	-0.0115±0.0014	2.8E-15	442,391
rs12523594	5	63,967,230	*FAM159B*	upstream	A	G	0.389	0.0106±0.0014	1.9E-13	442,391
rs764487108	5	64,488,059	*ADAMTS6*	intron	D	I	0.402	0.0097±0.0014	1.1E-11	442,391
rs55963623	5	64,555,615	*ADAMTS6*	intron NMD transcript	T	C	0.456	0.013±0.0014	1.3E-20	442,391
rs62360508	5	65,177,737	*ERBIN*	intergenic	T	C	0.913	0.0154±0.0025	4.9E-10	442,391
rs4703855	5	71,693,899	*PTCD2*	intergenic	T	C	0.295	0.0108±0.0015	1.9E-12	442,391
rs4704187	5	74,480,288	*ANKRD31*	intron	T	C	0.530	0.0137±0.0014	9.4E-23	442,391
rs2307111	5	75,003,678	*POC5*	missense	T	C	0.607	0.0174±0.0014	3.2E-36	516,336
rs59893724	5	80,830,788	*SSBP2*	intron	A	G	0.756	0.0111±0.0016	7.3E-12	442,391
rs61749613	5	82,815,170	*VCAN*	missense splice region	A	G	0.959	-0.0379±0.0033	8.4E-30	535,885
rs10942491	5	86,382,726	*MIR4280*	intron	C	G	0.553	-0.0095±0.0014	1.3E-11	442,391
rs7442885	5	87,682,877	*TMEM161B-AS1*	intron	C	G	0.789	0.0168±0.0017	7.7E-23	442,391
rs17561711	5	88,248,730	*MEF2C-AS1*	intron	A	C	0.917	0.0162±0.0025	1.3E-10	442,391
rs35843836	5	88,798,726	*MEF2C-AS1*	intergenic	A	T	0.633	-0.0145±0.0014	8.2E-24	442,391
rs6235	5	95,728,898	*PCSK1*	missense	C	G	0.732	-0.0113±0.0015	7.6E-14	522,505
rs2611742	5	95,856,501	*LOC101929710*	intron	T	C	0.607	-0.0132±0.0014	1.9E-20	442,391
rs200755293	5	102,072,999	*PAM*	intergenic	D	I	0.425	0.0108±0.0014	1.8E-14	442,391
rs252818	5	106,725,691	*EFNA5*	intron	T	C	0.816	0.0117±0.0018	1.1E-10	442,391
rs149457	5	107,438,057	*FBXL17*	intron	T	C	0.170	-0.0189±0.0019	2.3E-24	442,391
rs5019041	5	108,179,863	*FER*	intron	A	G	0.762	0.015±0.0016	9.5E-20	442,391
rs1402025	5	113,987,898	*LOC101927078*	intron	T	C	0.774	-0.0108±0.0017	6.5E-11	442,391
rs72800433	5	115,033,709	*LOC102467217*	intergenic	T	C	0.722	0.0091±0.0015	4.7E-09	442,391
rs11438541	5	122,080,320	*LINC02201*	intergenic	I	D	0.420	0.0077±0.0014	5.4E-08	442,391
rs34984032	5	122,692,653	*CEP120*	intron	D	I	0.536	0.0129±0.0014	2.2E-20	442,391
rs152120	5	126,656,500	*MEGF10*	intron	T	C	0.570	-0.008±0.0014	1.4E-08	442,391
rs2250127	5	127,388,844	*LINC01184*	intron	A	G	0.247	0.0184±0.0016	4.4E-30	442,391
rs78727187	5	127,668,685	*FBN2*	missense	T	G	0.006	-0.0569±0.0087	5.8E-11	535,885
rs329120	5	133,861,756	*JADE2*	intron	T	C	0.419	-0.0116±0.0014	1.6E-16	442,391
rs757647	5	137,707,315	*KDM3B*	intron	A	G	0.204	0.0109±0.0017	4.1E-11	535,885
rs13174863	5	139,080,745	*CXXC5*	intron	A	G	0.852	-0.0147±0.002	8.3E-14	442,391
rs7268	5	139,712,550	*HBEGF*	3' UTR	A	C	0.439	-0.0108±0.0014	1.1E-14	442,391
rs1479585	5	140,976,528	*DIAPH1*	intron	C	G	0.607	0.0111±0.0014	6.7E-15	442,391
rs72801007	5	152,196,711	*LINC01470*	intron	C	G	0.715	0.0077±0.0015	6.3E-07	442,391
rs1438945	5	152,510,937	*LINC01470*	intron	A	T	0.714	-0.0098±0.0015	2.2E-10	442,391
rs4958702	5	153,544,512	*GALNT10*	intron	T	C	0.427	0.01±0.0014	8.8E-13	442,391

rsid	chr	pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs62385430	5	158,465,102	*EBF1*	intron	C	G	0.573	0.0105±0.0014	9.7E-14	442,391
rs6868089	5	159,241,494	*LINC01847*	intron	A	G	0.951	0.0194±0.0032	1.7E-09	442,391
rs34259768	5	166,179,473	*LINC01947(dist =152753)*	intergenic	D	I	0.589	-0.0086±0.0014	1.3E-09	442,391
rs1014194	5	168,192,944	*SLIT3*	intron	A	C	0.641	0.0094±0.0015	9.6E-11	442,391
rs35239894	5	168,312,829	*SLIT3*	intron	A	T	0.456	0.0091±0.0014	1.0E-10	442,391
rs245769	5	170,529,335	*RANBP17*	intron NMD transcript	T	C	0.729	0.0111±0.0016	1.2E-12	442,391
rs351855	5	176,520,243	*FGFR4*	missense	A	G	0.297	0.0163±0.0015	4.1E-28	504,952
rs3822593	5	178,551,299	*ADAMTS2*	intron	T	G	0.417	-0.0108±0.0014	2.4E-14	442,391
rs11382360	6	2,094,804	*GMDS*	intron	D	I	0.456	-0.0094±0.0014	2.2E-11	442,391
rs220025	6	12,154,333	*HIVEP1*	intron	A	G	0.707	-0.0091±0.0015	3.0E-09	442,391
rs6916983	6	18,575,300	*MIR548A1*	downstream	T	C	0.719	-0.0112±0.0015	5.0E-13	442,391
rs9350100	6	19,076,417	*LOC101928519*	intergenic	T	C	0.801	-0.0137±0.0017	3.7E-15	442,391
rs9466957	6	23,874,157	*NRSN1*	intergenic	T	G	0.348	0.0104±0.0015	6.5E-12	442,391
6:26316118_TTTTCTTTC_T	6	26,316,118	*BTN3A2*	intergenic	D	I	0.609	0.0098±0.0014	6.4E-12	442,391
rs210140	6	33,544,293	*BAK1*	intron	T	C	0.365	-0.0082±0.0014	1.3E-08	442,391
rs79984746	6	34,201,127	*HMGA1*	upstream	A	C	0.047	-0.0262±0.0033	9.4E-16	442,391
rs206936	6	34,302,869	*NUDT3,RPS10-NUDT3*	intron	A	G	0.813	-0.0193±0.0017	1.8E-29	535,885
rs113344808	6	34,701,981	*SNRPC*	intergenic	I	D	0.314	0.0186±0.0015	1.9E-34	442,391
rs35104496	6	39,273,821	*KCNK17*	intron	I	D	0.176	0.0114±0.0019	3.4E-09	442,391
rs34045288	6	40,369,081	*LRFN2*	intron	T	C	0.336	0.0146±0.0015	3.4E-23	442,391
rs776152132	6	41,902,508	*CCND3*	downstream	I	D	0.749	0.011±0.0016	7.0E-12	442,391
rs9394951	6	43,350,753	*ZNF318*	intergenic	T	C	0.567	-0.0093±0.0014	3.7E-11	442,391
rs35679149	6	43,604,167	*MAD2L1BP*	synonymous	A	G	0.973	0.031±0.0043	5.7E-13	442,391
6:44961164_ATTAACTGAAAAATGCCTTCTCCAGAAGGC_A	6	44,961,164	*SUPT3H*	intron	D	I	0.289	-0.012±0.0016	1.2E-14	442,391
rs71568416	6	47,474,426	*CD2AP*	intron	C	G	0.690	-0.0112±0.0015	1.1E-13	442,391
rs2206277	6	50,798,526	*TFAP2B*	intron	T	C	0.180	0.0285±0.0017	3.4E-60	534,021
rs2635727	6	50,820,940	*TFAP2B*	downstream	T	C	0.245	-0.0187±0.0016	7.5E-31	442,391
rs7771607	6	51,803,390	*PKHD1*	intron	T	C	0.301	0.0098±0.0015	1.6E-10	442,391
rs9475170	6	55,011,619	*HCRTR2*	intergenic	A	T	0.658	0.0106±0.0015	6.1E-13	442,391
rs78856780	6	56,939,238	*ZNF451*	intergenic	A	G	0.856	-0.0121±0.002	1.2E-09	442,391
rs147380831	6	81,150,794	*BCKDHB*	upstream	D	I	0.900	0.0142±0.0024	1.4E-09	442,391
rs2917460	6	83,343,317	*UBE3D*	intergenic	A	C	0.523	-0.01±0.0014	1.0E-12	442,391
rs41271629	6	86,257,229	*SNX14*	missense	T	G	0.971	0.0252±0.004	3.0E-10	512,183
6:97806419_GGA_G	6	97,806,419	*LOC101927314, MIR548H3*	intergenic	D	I	0.347	-0.0097±0.0015	4.9E-11	442,391
rs11961595	6	100,052,886	*PRDM13*	upstream	A	T	0.268	0.0092±0.0016	6.3E-09	442,391
rs240164	6	101,044,487	*ASCC3*	intron	A	T	0.529	-0.0119±0.0014	1.6E-17	442,391
rs270689	6	104,790,532	*HACE1*	intergenic	A	T	0.198	-0.0126±0.0017	3.9E-13	442,391
rs314274	6	105,412,932	*LIN28B*	intron	A	C	0.336	-0.0145±0.0014	3.0E-24	514,980

rsid	chr	pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs2153960	6	108,988,184	*FOXO3*	intron	A	G	0.711	0.0143±0.0015	2.9E-22	532,966
rs1064583	6	116,446,576	*COL10A1*	missense	A	G	0.604	0.0086±0.0014	2.5E-10	534,313
6:120064242_TA_T	6	120,064,242	*LOC105377975*	intergenic	D	I	0.137	0.0154±0.0023	7.3E-12	442,391
rs12332861	6	123,795,554	*TRDN*	intron	T	C	0.826	0.0122±0.0018	3.2E-11	442,391
rs13209968	6	126,089,285	*HEY2*	intergenic	C	G	0.526	0.0089±0.0014	1.3E-10	442,391
rs6899976	6	130,358,428	*L3MBTL3*	intron	A	G	0.698	-0.0106±0.0015	4.1E-13	514,980
rs4896582	6	142,703,877	*ADGRG6*	intron	A	G	0.301	0.0146±0.0015	2.0E-23	514,980
rs11380190	6	147,357,397	*STXBP5-AS1*	intron	I	D	0.367	0.009±0.0014	5.6E-10	442,391
rs9371672	6	153,422,140	*RGS17*	intron	C	G	0.711	-0.0098±0.0015	1.9E-10	442,391
rs369809361	6	154,335,717	*OPRM1*	intron	D	I	0.180	0.0114±0.0018	3.1E-10	442,391
rs13191362	6	163,033,350	*PRKN*	intron	A	G	0.875	0.016±0.0021	3.8E-14	442,391
rs4719730	7	929,123	*GET4*	intron	T	C	0.175	-0.011±0.0018	1.9E-09	442,391
rs756700851	7	1,271,233	*UNCX*	upstream	I	D	0.203	-0.0129±0.0017	8.1E-14	442,391
rs4721089	7	1,872,921	*MAD1L1*	intron	T	C	0.784	0.0107±0.0017	2.3E-10	442,391
rs62441156	7	5,347,446	*TNRC18*	3' UTR	A	G	0.773	-0.0102±0.0017	3.0E-09	442,391
7:6411992_CA_C	7	6,411,992	*RAC1*	upstream	I	D	0.528	-0.0085±0.0014	3.0E-09	442,391
rs12669977	7	13,132,400	*ARL4A*	intergenic	T	G	0.502	0.0082±0.0014	3.1E-09	442,391
rs9638713	7	14,645,949	*DGKB*	intron	A	G	0.025	0.0317±0.0045	1.6E-12	442,391
rs17141862	7	19,770,756	*TMEM196*	intron	T	C	0.137	-0.0139±0.002	6.4E-12	442,391
rs73086541	7	20,376,103	*ITGB8*	intron	A	C	0.395	0.0104±0.0014	2.8E-13	442,391
rs28423374	7	20,579,647	*ABCB5*	intergenic	T	C	0.736	-0.0114±0.0016	6.8E-13	442,391
rs2711111	7	24,529,055	*MPP6*	intergenic	A	G	0.434	0.0082±0.0014	1.0E-08	442,391
rs2067087	7	27,241,660	*HOTTIP*	non-coding transcript exon variant	C	G	0.715	0.0111±0.0016	1.2E-12	442,391
rs7384844	7	28,025,308	*JAZF1*	intron	A	T	0.556	0.0086±0.0014	8.1E-10	442,391
rs864745	7	28,180,556	*JAZF1*	intron	T	C	0.495	-0.0155±0.0013	4.9E-31	535,885
rs554794335	7	32,931,486	*KBTBD2*	5' UTR	D	I	0.689	0.0095±0.0015	4.0E-10	442,391
rs329270	7	35,075,619	*DPY19L1*	intron	A	G	0.515	0.0101±0.0014	5.0E-13	442,391
rs3778934	7	39,445,385	*POU6F2-AS1*	intron	A	C	0.659	0.0095±0.0015	9.0E-11	442,391
rs799449	7	44,784,697	*ZMIZ2*	upstream	T	C	0.558	0.0108±0.0014	1.9E-14	442,391
rs80077929	7	46,094,089	*IGFBP3*	intergenic	T	C	0.119	0.0218±0.0022	1.3E-23	442,391
rs2881198	7	46,634,506	*LOC730338*	intergenic	C	G	0.533	-0.0126±0.0014	2.3E-19	442,391
rs12718611	7	50,593,906	*DDC*	intron	A	G	0.860	-0.0137±0.002	8.5E-12	442,391
rs2866719	7	70,106,061	*AUTS2*	intron	T	C	0.369	0.0088±0.0014	1.2E-09	442,391
rs861491	7	71,586,049	*CALN1*	intron	T	C	0.660	0.0091±0.0015	6.2E-10	442,391
rs139709472	7	72,140,581	*TYW1B*	intron	T	C	0.963	-0.027±0.0042	1.9E-10	442,391
rs3812316	7	73,020,337	*MLXIPL*	missense	C	G	0.872	-0.018±0.002	2.9E-19	523,409
rs6955671	7	74,489,486	*RCC1L*	missense	T	C	0.540	-0.0134±0.0014	8.9E-22	442,391
rs236660	7	75,050,086	*POM121C*	non-coding transcript exon variant	T	C	0.429	-0.016±0.0015	5.0E-28	442,391
rs1167800	7	75,176,196	*HIP1*	intron	A	G	0.560	0.0147±0.0013	6.5E-28	535,885

rsid	chr	pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs200778979	7	76,686,511	*SPDYE18*	intron	A	G	0.129	0.0242±0.0022	1.7E-27	442,391
rs740158	7	77,055,836	*LOC101927243*	upstream	T	C	0.512	0.0089±0.0014	3.6E-11	514,980
rs1852006	7	77,829,768	*MAGI2*	intron	A	G	0.357	-0.0088±0.0015	1.2E-09	442,391
7:78141093_TA_T	7	78,141,093	*MAGI2*	intron	D	I	0.263	0.0096±0.0016	1.4E-09	442,391
7:93148206_TA_T	7	93,148,206	*CALCR*	intron	D	I	0.477	0.0109±0.0014	8.8E-15	442,391
rs3763469	7	94,021,475	*COL1A2-AS1*	upstream	T	C	0.778	-0.0123±0.0017	5.3E-13	442,391
rs1207728	7	96,638,021	*DLX6-AS1*	intron	C	G	0.225	0.01±0.0017	2.1E-09	442,391
rs28566086	7	99,025,490	*ATP5J2-PTCD1,PTCD1*	intron	C	G	0.114	-0.0176±0.0022	1.1E-15	442,391
rs1133790	7	99,632,982	*ZKSCAN1*	3' UTR	T	C	0.676	-0.0096±0.0015	1.7E-10	442,391
rs12669944	7	100,446,365	*SLC12A9-AS1*	intron	A	G	0.350	-0.0076±0.0015	2.9E-07	442,391
rs2239538	7	100,798,728	*AP1S1*	intron	A	G	0.147	0.0127±0.002	1.1E-10	442,391
rs39328	7	103,444,978	*RELN*	intron	T	C	0.424	0.0138±0.0014	1.1E-22	442,391
rs4730494	7	111,340,760	*DOCK4*	intergenic	T	C	0.216	0.0104±0.0017	1.0E-09	442,391
7:112946725_CTGTT_C	7	112,946,725	*LINC00998*	intergenic	I	D	0.637	0.0094±0.0014	1.1E-10	442,391
rs6962980	7	113,452,183	*PPP1R3A*	intergenic	A	C	0.442	0.0122±0.0014	3.3E-18	442,391
rs10500039	7	114,351,267	*FOXP2*	intergenic	T	C	0.578	-0.0109±0.0014	1.1E-14	442,391
rs62621812	7	127,015,083	*ZNF800*	missense	A	G	0.022	0.0457±0.0047	5.5E-22	535,885
rs112320081	7	127,519,243	*SND1*	intron	T	C	0.988	-0.0402±0.0065	5.5E-10	442,391
7:129965811_CAT_C	7	129,965,811	*CPA4*	downstream	I	D	0.768	-0.0117±0.0017	4.2E-12	442,391
rs273957	7	137,600,690	*CREB3L2*	missense	T	C	0.614	-0.0103±0.0014	4.8E-14	535,885
rs11525873	7	138,817,193	*TTC26*	upstream	T	C	0.902	0.0202±0.0023	5.6E-18	442,391
rs822531	7	148,629,759	*EZH2*	intergenic	T	C	0.795	-0.0123±0.0017	1.2E-12	442,391
rs10236214	7	150,668,070	*KCNH2*	intron	T	C	0.642	0.0144±0.0015	4.2E-23	442,391
rs759544745	7	150,759,219	*SLC4A2*	intron	I	D	0.423	0.0085±0.0014	1.9E-09	442,391
rs2407940	8	4,134,874	*CSMD1*	intron	T	C	0.503	-0.0083±0.0014	3.7E-09	442,391
rs1658820	8	4,288,577	*CSMD1*	intron	T	G	0.245	0.0097±0.0016	3.2E-09	442,391
rs6985109	8	10,761,585	*XKR6*	intron	A	G	0.542	-0.0185±0.0014	7.2E-40	442,391
8:12184258_A_G	8	12,184,258	*DEFB130B*	intergenic	A	G	0.699	0.0191±0.0023	7.7E-17	442,391
rs7010322	8	13,196,821	*DLC1*	intron	T	C	0.360	-0.009±0.0015	7.4E-10	442,391
rs13263601	8	14,095,900	*SGCZ*	intron	A	C	0.652	-0.011±0.0015	4.8E-14	442,391
rs146032132	8	15,136,469	*SGCZ*	intergenic	I	D	0.113	0.0127±0.0022	9.5E-09	442,391
rs71211068	8	15,538,727	*TUSC3*	intron	I	D	0.777	0.0118±0.0018	4.8E-11	442,391
rs2616143	8	20,632,022	*SNORD3F*	intergenic	A	G	0.321	-0.0098±0.0015	4.9E-11	442,391
rs3174040	8	23,431,407	*SLC25A37*	missense	A	G	0.795	0.0103±0.0017	2.4E-09	442,391
rs7823498	8	32,403,573	*NRG1*	intron	T	C	0.780	0.0112±0.0017	2.9E-11	442,391
rs6996013	8	33,276,170	*FUT10*	intron	A	G	0.765	0.0106±0.0017	2.1E-10	442,391
rs10091344	8	34,132,075	*LINC01288*	intron	A	G	0.328	-0.0101±0.0015	1.1E-11	442,391
rs4082204	8	38,328,902	*FGFR1*	upstream	A	G	0.402	0.0097±0.0014	1.0E-11	442,391
rs35392772	8	57,026,229	*MOS*	missense	A	C	0.157	-0.0107±0.0018	6.2E-09	534,021
rs13264909	8	64,702,385	*LINC01289*	intron	A	T	0.571	0.0099±0.0014	2.1E-12	442,391
rs7006364	8	67,199,420	*LINC00967*	intergenic	A	G	0.273	-0.011±0.0015	4.1E-13	514,980

rsid	chr	pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs35957544	8	73,440,371	*KCNB2*	intergenic	T	G	0.575	-0.0102±0.0014	4.1E-13	442,391
rs28611482	8	74,490,314	*STAU2*	intron	A	C	0.798	-0.0125±0.0017	7.0E-13	442,391
rs2977324	8	76,716,737	*HNF4G*	intergenic	T	G	0.300	-0.0125±0.0015	3.3E-16	442,391
rs960923	8	77,228,925	*LINC01111*	intergenic	A	G	0.433	-0.0125±0.0014	7.9E-19	442,391
rs406629	8	81,458,387	*ZBTB10*	downstream	T	G	0.310	-0.0099±0.0015	4.6E-11	442,391
rs7821259	8	85,081,619	*RALYL*	upstream	T	G	0.265	-0.0115±0.0016	3.6E-13	442,391
rs28676726	8	85,619,884	*RALYL*	intron	T	G	0.228	-0.0117±0.0017	1.9E-12	442,391
rs2664371	8	89,056,962	*MMP16*	intron	A	T	0.690	0.0085±0.0015	2.0E-08	442,391
rs2134963	8	89,407,821	*MMP16*	intron	A	C	0.306	-0.0163±0.0015	3.9E-27	442,391
rs2114210	8	95,595,162	*VIRMA*	intergenic	A	G	0.335	0.0114±0.0015	1.1E-14	442,391
rs71506350	8	101,962,901	*YWHAZ*	intron	A	G	0.387	-0.0088±0.0014	8.7E-10	442,391
rs3808424	8	116,518,399	*TRPS1*	intron	T	C	0.781	0.0119±0.0017	2.0E-12	442,391
rs147009081	8	118,882,739	*EXT1*	intron	I	D	0.106	0.0157±0.0023	5.4E-12	442,391
rs4144738	8	130,760,850	*GSDMC*	missense	A	G	0.539	-0.0107±0.0013	1.4E-15	535,885
rs4507813	8	141,796,344	*PTK2*	intron	T	C	0.422	0.0082±0.0014	7.0E-09	442,391
rs7838717	8	145,504,343	*BOP1*	intron	T	C	0.365	-0.0097±0.0015	3.6E-11	442,391
rs10815306	9	5,914,727	*KIAA2026*	intron NMD transcript	A	C	0.200	0.0104±0.0017	2.5E-09	442,391
rs13292799	9	6,448,764	*UHRF2*	intron	A	G	0.807	-0.0107±0.0018	1.1E-09	442,391
rs10960276	9	11,819,686	*TYRP1*	intergenic	A	C	0.356	-0.0093±0.0015	1.8E-10	442,391
rs12335987	9	13,235,045	*MPDZ*	intron	A	C	0.870	0.0129±0.0021	5.8E-10	442,391
rs35249758	9	13,933,542	*LINC00583*	intron	T	C	0.184	-0.0116±0.0018	1.2E-10	442,391
rs927520	9	14,347,235	*NFIB*	intron	T	G	0.336	0.0092±0.0015	5.9E-10	442,391
rs4741546	9	15,846,112	*CCDC171*	intron	T	C	0.396	-0.0082±0.0014	1.0E-08	442,391
rs10962552	9	16,723,742	*BNC2*	intron	T	C	0.166	0.0163±0.0019	3.4E-18	442,391
rs539864802	9	17,023,805	*CNTLN*	intergenic	D	I	0.847	-0.0124±0.002	3.6E-10	442,391
rs66801939	9	19,024,042	*SAXO1*	intron	T	C	0.618	0.0098±0.0014	1.1E-11	442,391
rs1541104	9	23,348,762	*LOC101929563*	intergenic	A	G	0.414	-0.0087±0.0014	8.4E-10	442,391
9:27621063_AC_A	9	27,621,063	*C9orf72*	intergenic	D	I	0.251	-0.009±0.0016	2.7E-08	442,391
rs3922980	9	27,803,737	*LINGO2*	intergenic	T	G	0.502	-0.0098±0.0014	2.6E-12	442,391
rs10968576	9	28,414,339	*LINGO2*	intron	A	G	0.677	-0.0168±0.0014	4.2E-32	535,885
rs10969334	9	29,717,279	*LINGO2*	intergenic	A	C	0.394	-0.0083±0.0014	4.9E-09	442,391
rs544957562	9	33,978,015	*UBAP2*	intron	A	T	0.845	0.0164±0.002	1.1E-15	442,391
rs13290451	9	34,581,819	*CNTFR-AS1*	intron	A	T	0.524	-0.0089±0.0014	1.6E-10	442,391
rs10687055	9	37,083,295	*EBLN3P*	intron	D	I	0.464	0.0101±0.0014	5.6E-13	442,391
rs148434637	9	73,811,782	*TRPM3*	intron	D	I	0.557	-0.0119±0.0014	1.3E-16	442,391
rs35307904	9	78,511,889	*PCSK5*	intron	A	G	0.122	0.0211±0.0021	6.1E-23	442,391
rs958225	9	78,759,705	*PCSK5*	intron	A	T	0.057	-0.0167±0.003	3.3E-08	442,391
rs10870005	9	80,626,343	*GNAQ*	intron	T	C	0.532	-0.0093±0.0014	2.9E-11	442,391
rs9314675	9	81,371,636	*LOC101927450*	intergenic	T	G	0.447	-0.0085±0.0014	1.5E-09	442,391
rs7874181	9	85,858,511	*FRMD3*	intron	T	G	0.847	-0.012±0.0019	6.4E-10	442,391
rs1982151	9	86,617,265	*RMI1*	missense	A	G	0.257	-0.0114±0.0015	8.5E-14	532,966
rs353834	9	89,073,689	*ZCCHC6*	intergenic	A	G	0.507	0.0083±0.0014	4.0E-09	442,391

rsid	chr	pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs35870355	9	92,039,730	*SEMA4D*	intron	A	T	0.235	0.0101±0.0016	7.3E-10	442,391
rs10820852	9	94,186,973	*NFIL3*	upstream	A	C	0.275	-0.0122±0.0016	5.5E-15	442,391
rs10761129	9	94,486,321	*ROR2*	missense	T	C	0.671	0.0138±0.0014	4.5E-22	524,369
rs10761155	9	95,153,729	*CENPP,OGN*	intron	C	G	0.731	-0.0146±0.0016	2.1E-20	442,391
rs9650755	9	96,484,342	*PHF2*	intergenic	A	G	0.737	-0.0138±0.0016	3.2E-18	442,391
rs72753485	9	96,673,230	*BARX1*	regulatory	C	G	0.097	0.019±0.0025	7.1E-14	442,391
rs370727606	9	98,268,396	*PTCH1*	intron	A	G	0.011	-0.0412±0.0068	1.5E-09	442,391
rs2672813	9	98,781,179	*ERCC6L2,LINC00092*	downstream	A	G	0.440	-0.0101±0.0014	4.8E-13	442,391
rs62565259	9	102,162,570	*NAMA*	regulatory	T	C	0.173	-0.011±0.0019	3.1E-09	442,391
9:103113656_AC_A	9	103,113,656	*TEX10*	intron	D	I	0.379	0.0105±0.0015	7.2E-13	442,391
rs13294145	9	109,138,289	*MIR8081*	intron	T	C	0.172	0.013±0.0018	2.0E-12	442,391
rs2417998	9	111,958,746	*EPB41L4B*	intron	C	G	0.293	0.0093±0.0015	1.2E-09	442,391
rs10441737	9	114,301,585	*ZNF483*	intron	T	C	0.652	0.0098±0.0015	2.3E-11	442,391
rs2274159	9	117,166,246	*WHRN*	missense	A	G	0.506	0.0079±0.0013	3.2E-09	535,885
rs3838333	9	117,787,666	*TNC*	intron	I	D	0.429	0.0089±0.0014	2.7E-10	442,391
rs80031633	9	120,687,550	*TLR4*	intergenic	A	T	0.457	-0.009±0.0014	2.8E-10	442,391
rs10760277	9	126,093,999	*CRB2*	intergenic	T	C	0.386	0.0094±0.0014	5.6E-11	442,391
rs7024161	9	127,011,695	*NEK6*	intergenic	T	C	0.385	-0.0131±0.0014	6.0E-20	442,391
rs77752857	9	127,085,612	*NEK6*	intron	T	G	0.099	0.0184±0.0023	2.7E-15	442,391
rs396015	9	128,028,603	*GAPVD1*	intron	A	T	0.466	-0.0088±0.0014	4.6E-10	442,391
rs7030609	9	129,415,775	*LMX1B*	intron	A	G	0.910	-0.0191±0.0024	5.2E-15	442,391
rs10987417	9	129,462,501	*LMX1B*	3' UTR	T	G	0.385	0.0094±0.0014	8.5E-11	442,391
rs2267958	9	131,015,279	*DNM1*	intron	A	G	0.512	-0.0088±0.0014	7.2E-10	442,391
rs8176746	9	136,131,322	*ABO*	non-coding transcript exon variant	T	G	0.064	-0.0188±0.0027	6.4E-12	535,885
rs10858334	9	137,989,785	*OLFM1*	3' UTR	C	G	0.856	-0.012±0.002	2.1E-09	442,391
rs36094334	9	139,140,538	*QSOX2*	upstream	C	G	0.661	0.011±0.0015	1.7E-13	442,391
9:140265782_C_T	9	140,265,782	*EXD3*	intron	T	C	0.124	-0.015±0.0021	1.3E-12	442,391
rs35741360	10	12,945,180	*CCDC3*	intron	A	G	0.283	0.0105±0.0015	1.1E-11	442,391
rs202121703	10	16,753,335	*RSU1*	intron	I	D	0.371	-0.0082±0.0014	1.5E-08	442,391
rs73601548	10	18,549,889	*CACNB2*	5' UTR	T	C	0.115	0.015±0.0022	8.0E-12	442,391
rs35043348	10	21,993,540	*MLLT10*	intron	I	D	0.581	-0.0111±0.0015	8.3E-14	442,391
rs112276510	10	25,056,538	*ARHGAP21*	intergenic	D	I	0.269	0.0108±0.0016	8.6E-12	442,391
rs117799729	10	27,434,230	*YME1L1*	intron	A	C	0.071	0.0188±0.0028	1.5E-11	442,391
rs112157102	10	29,083,096	*LINC00837,LINC01517*	intron	A	T	0.579	0.0091±0.0014	1.6E-10	442,391
rs2247538	10	34,444,740	*PARD3*	intron	T	C	0.316	-0.011±0.0015	2.3E-13	442,391
rs796709989	10	35,051,218	*PARD3*	intron	I	D	0.709	0.0092±0.0016	4.9E-09	442,391
rs2435381	10	43,678,796	*CSGALNACT2*	missense	T	C	0.277	-0.0098±0.0016	3.5E-10	442,391
rs11006229	10	52,350,006	*SGMS1*	splice region,5' UTR	T	C	0.200	0.01±0.0017	2.3E-09	535,885
rs4595495	10	53,673,286	*PRKG1*	intron	A	G	0.579	-0.0085±0.0014	1.8E-09	442,391

rsid	chr	pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs181130148	10	63,204,424	*TMEM26*	intron	A	G	0.966	0.0235±0.0039	2.3E-09	442,391
rs12761779	10	63,782,043	*ARID5B*	intron	C	G	0.652	0.0104±0.0015	2.4E-12	442,391
rs7924036	10	65,191,645	*JMJD1C*	intron	T	G	0.503	-0.0122±0.0014	1.9E-18	442,391
rs68049170	10	72,432,047	*ADAMTS14*	upstream	A	G	0.277	-0.014±0.0016	2.9E-19	442,391
10:76097518_AT_A	10	76,097,518	*ADK*	intron	D	I	0.388	0.0102±0.0015	8.9E-12	442,391
rs4255484	10	77,215,583	*LRMDA*	intergenic	C	G	0.441	0.0111±0.0014	2.3E-15	442,391
10:87354315_AGCT_A	10	87,354,315	*GRID1-AS1*	downstream	D	I	0.050	0.0247±0.0032	1.2E-14	442,391
rs2631681	10	93,032,943	*PCGF5*	intron	T	C	0.323	-0.0114±0.0014	2.9E-15	514,980
rs7916961	10	94,997,875	*MYOF*	intergenic	T	C	0.859	0.0117±0.002	5.0E-09	442,391
rs2274224	10	96,039,597	*PLCE1*	missense	C	G	0.432	0.0194±0.0014	9.7E-46	511,156
rs35808802	10	97,143,826	*SORBS1*	missense	T	C	0.939	0.0185±0.0028	3.1E-11	535,885
rs150407151	10	97,784,002	*ENTPD1-AS1*	intron	D	I	0.649	-0.0092±0.0015	6.4E-10	442,391
rs11595634	10	99,001,779	*ARHGAP19-SLIT1*	intron	A	T	0.586	-0.0088±0.0015	2.2E-09	442,391
rs1983864	10	100,017,453	*LOXL4*	missense	T	G	0.661	0.0153±0.0014	1.5E-27	535,885
rs603424	10	102,075,479	*PKD2L1*	intron	A	G	0.170	0.0098±0.0019	1.2E-07	442,391
rs41310284	10	102,447,647	*PAX2*	intergenic	A	C	0.102	-0.0154±0.0023	2.8E-11	442,391
rs10883560	10	102,673,707	*SLF2*	intron	C	G	0.571	-0.0106±0.0014	5.9E-14	442,391
rs3915773	10	103,366,837	*DPCD*	intron	A	C	0.713	-0.0092±0.0016	3.2E-09	442,391
rs372582961	10	103,958,317	*ELOVL3*	intergenic	D	I	0.847	0.0146±0.0019	5.8E-14	442,391
rs12413409	10	104,719,096	*CNNM2*	intron	A	G	0.078	0.0232±0.0025	1.4E-20	535,885
rs3824780	10	105,384,826	*SH3PXD2A*	intron	T	C	0.356	0.0094±0.0015	1.2E-10	442,391
rs6585201	10	114,768,783	*TCF7L2*	intron	A	G	0.454	-0.0101±0.0014	5.9E-13	442,391
rs11197866	10	118,777,633	*SHTN1*	intron	T	C	0.747	-0.0101±0.0016	6.1E-10	442,391
rs4336954	10	120,976,553	*GRK5*	intron	T	C	0.386	0.0085±0.0014	3.6E-09	442,391
rs845085	10	125,217,896	*GPR26*	intron	A	G	0.472	-0.0109±0.0014	7.0E-16	514,980
rs72828952	10	126,629,608	*ZRANB1*	intron	A	G	0.576	0.0114±0.0014	2.0E-15	442,391
rs11245187	10	128,538,476	*DOCK1*	intergenic	T	C	0.796	0.0101±0.0017	4.7E-09	442,391
rs117118217	10	131,783,328	*EBF3*	intergenic	C	G	0.018	0.0327±0.0055	2.9E-09	442,391
rs11146237	10	134,004,746	*DPYSL4*	intron	A	G	0.507	0.009±0.0014	6.0E-10	442,391
rs76560824	11	370,252	*B4GALNT4*	intron	T	C	0.898	0.0144±0.0023	5.0E-10	442,391
rs7103389	11	881,639	*CHID1*	intron	T	C	0.381	-0.0122±0.0014	2.6E-17	442,391
rs10840606	11	2,234,690	*MIR4686*	regulatory	A	G	0.822	-0.013±0.0018	1.8E-12	442,391
rs10741669	11	2,643,480	*KCNQ1OT1*	non-coding transcript exon variant	A	G	0.880	0.0158±0.0022	3.1E-13	442,391
rs4929923	11	8,639,200	*TRIM66*	3' UTR	T	C	0.354	-0.011±0.0014	4.2E-15	535,885
rs2957694	11	10,369,014	*CAND1.11*	intron	A	G	0.416	-0.0103±0.0014	4.2E-13	442,391
rs11824377	11	11,787,253	*MIR8070*	intergenic	A	G	0.441	0.0087±0.0014	4.8E-10	442,391
rs900144	11	13,294,268	*ARNTL*	upstream	T	C	0.568	0.0098±0.0014	2.3E-12	442,391
rs72632979	11	16,615,883	*SOX6*	intron	A	G	0.829	-0.0125±0.0019	1.3E-11	442,391
rs5215	11	17,408,630	*KCNJ11*	missense	T	C	0.640	0.0118±0.0014	2.3E-17	535,885

rsid	chr	pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs6265	11	27,679,916	*BDNF*	missense	T	C	0.189	-0.0299±0.0017	1.6E-68	531,102
rs10835367	11	28,642,593	*MIR8068*	intergenic	A	C	0.386	-0.0094±0.0014	8.0E-11	442,391
rs1607227	11	28,808,617	*MIR8068*	intron	T	G	0.295	-0.0125±0.0015	3.5E-16	442,391
rs35381476	11	29,309,284	*LINC01616*	intergenic	I	D	0.261	-0.0102±0.0016	2.2E-10	442,391
rs7941828	11	30,430,331	*MPPED2*	intron	T	C	0.361	-0.0124±0.0014	1.3E-17	442,391
rs1806153	11	31,850,105	*PAUPAR*	intron	T	G	0.231	-0.0103±0.0017	5.1E-10	442,391
rs9651614	11	43,849,913	*HSD17B12*	intron	T	C	0.646	-0.0136±0.0015	8.1E-20	442,391
rs7115013	11	43,934,592	*ALKBH3-AS1*	intron	T	C	0.443	-0.0091±0.0014	1.2E-10	442,391
rs34042421	11	45,420,233	*LOC399886*	intron	A	G	0.695	0.0111±0.0015	2.1E-13	442,391
rs3817334	11	47,650,993	*MTCH2*	intron	T	C	0.408	0.0103±0.0014	4.9E-14	535,885
rs78287937	11	64,773,101	*ARL2-SNX15*	intergenic	T	G	0.909	-0.0159±0.0024	5.1E-11	442,391
rs2234458	11	65,639,374	*EFEMP2*	non-coding transcript exon variant	T	C	0.639	-0.0165±0.0014	4.0E-30	442,391
rs4930394	11	66,680,543	*PC*	intron	T	C	0.450	0.0109±0.0014	7.6E-15	442,391
rs61734601	11	67,184,725	*CARNS1*	non-coding transcript exon variant	A	G	0.084	0.0167±0.0025	2.8E-11	442,391
11:68271686_GT_G	11	68,271,686	*PPP6R3*	intron	D	I	0.276	-0.0138±0.0016	1.0E-18	442,391
rs597539	11	68,601,974	*CPT1A*	intron	C	G	0.686	-0.0109±0.0015	3.8E-13	442,391
rs4980661	11	69,306,579	*LINC01488*	intergenic	A	G	0.528	-0.0145±0.0014	2.4E-25	442,391
rs1192925	11	69,476,279	*ORAOV1*	intron	T	C	0.631	0.0132±0.0014	7.9E-20	442,391
rs584961	11	75,277,628	*SERPINH1*	synonymous	A	G	0.114	-0.0139±0.0022	2.8E-10	442,391
11:77923601_CCTAT_C	11	77,923,601	*USP35*	intron	I	D	0.836	0.0136±0.0019	4.0E-13	442,391
11:84790528_ATT_A	11	84,790,528	*DLG2*	intron	D	I	0.526	-0.011±0.0014	1.4E-14	442,391
rs610476	11	86,184,102	*ME3*	intron	A	G	0.410	-0.0085±0.0014	1.9E-09	442,391
rs61903695	11	89,922,417	*NAALAD2*	intron	A	G	0.744	-0.0099±0.0016	4.9E-10	442,391
rs680071	11	103,088,414	*DYNC2H1*	intron	T	C	0.119	-0.0134±0.0022	4.1E-10	442,391
rs1801516	11	108,175,462	*ATM*	missense	A	G	0.146	-0.0115±0.0019	9.7E-10	535,885
rs1048932	11	115,044,850	*CADM1*	3' UTR	A	C	0.413	-0.0134±0.0014	2.7E-21	442,391
rs1064939	11	118,396,331	*LOC101929089*	3' UTR	A	T	0.977	0.0335±0.0047	1.6E-12	442,391
rs1784302	11	118,940,957	*VPS11*	synonymous	C	G	0.597	-0.0081±0.0014	3.4E-09	529,789
rs12574203	11	119,901,007	*TRIM29*	intergenic	A	G	0.837	0.0122±0.0019	9.9E-11	442,391
rs11218510	11	121,922,587	*MIR100HG*	intron	A	G	0.401	-0.009±0.0014	2.3E-10	442,391
rs10893502	11	126,254,018	*ST3GAL4*	intron	T	C	0.408	0.0082±0.0014	1.0E-08	442,391
rs7936928	11	130,279,168	*ADAMTS8*	intron	T	C	0.397	0.0087±0.0014	1.1E-09	442,391
rs7933085	11	130,796,248	*SNX19*	intergenic	A	G	0.492	-0.0088±0.0014	3.6E-10	442,391
rs12788343	11	131,452,912	*NTM*	intron	T	C	0.589	-0.0086±0.0014	1.0E-09	442,391
rs7480395	11	133,667,674	*LOC646522*	intron	A	G	0.523	-0.0091±0.0014	8.6E-11	442,391
rs12364470	11	134,601,012	*LOC283177*	upstream	T	G	0.835	-0.0124±0.0019	4.0E-11	442,391
rs55726687	12	991,306	*WNK1*	intron	A	G	0.211	0.0181±0.0017	2.8E-26	442,391
rs10774018	12	2,157,925	*CACNA1C-IT2*	intron	C	G	0.219	0.0103±0.0017	1.3E-09	442,391

rsid	chr	pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
12:3338920_GCACAGCTGGGGCTGA_G	12	3,338,920	*TSPAN9*	intron	D	I	0.214	0.012±0.0017	3.3E-12	442,391
rs76895963	12	4,384,844	*CCND2-AS1*	intron	T	G	0.979	-0.0316±0.0054	4.1E-09	442,391
rs778329432	12	9,100,432	*M6PR*	intron	D	I	0.647	0.0091±0.0015	3.4E-09	442,391
rs10744137	12	16,993,026	*SKP1P2*	intergenic	T	C	0.231	0.0103±0.0017	4.9E-10	442,391
rs10846458	12	17,169,641	*SKP1P2*	intergenic	A	T	0.302	-0.0072±0.0015	2.1E-06	442,391
rs10771041	12	24,058,135	*SOX5*	intron	T	C	0.114	0.0147±0.0022	1.9E-11	442,391
rs74567249	12	28,453,440	*CCDC91*	intron	D	I	0.694	-0.0113±0.0015	9.7E-14	442,391
rs2730829	12	41,831,174	*PDZRN4*	intron	T	C	0.477	0.0099±0.0014	2.4E-13	514,980
rs7134130	12	42,861,370	*PRICKLE1*	intron	T	C	0.440	0.0095±0.0014	1.3E-11	442,391
12:46752872_CA_C	12	46,752,872	*SLC38A2*	3' UTR	I	D	0.421	-0.0118±0.0014	2.1E-16	442,391
rs145878042	12	48,143,315	*RAPGEF3*	missense	A	G	0.989	-0.0548±0.0064	9.8E-18	534,021
rs12228854	12	48,396,920	*COL2A1*	intron	T	G	0.223	-0.015±0.0017	3.0E-19	442,391
rs1126930	12	49,399,132	*PRKAG1*	missense	C	G	0.036	0.0313±0.0036	4.7E-18	529,789
rs7132908	12	50,263,148	*FAIM2*	3' UTR	A	G	0.383	0.0174±0.0014	3.2E-35	489,220
rs10783377	12	50,983,046	*DIP2B*	intron	A	C	0.346	-0.0121±0.0015	1.3E-16	442,391
12:51589293_TA_T	12	51,589,293	*POU6F1*	intron	D	I	0.784	-0.0081±0.0017	3.2E-06	442,391
rs36120387	12	53,831,064	*PRR13*	upstream	T	C	0.110	-0.0123±0.0022	4.5E-08	442,391
rs17110109	12	54,668,908	*CBX5*	intron	T	C	0.610	0.0114±0.0014	1.6E-15	442,391
rs4759228	12	56,508,409	*PA2G4*	upstream	C	G	0.296	-0.0144±0.0015	3.6E-21	442,391
rs730560	12	57,938,565	*DCTN2*	intron	A	G	0.582	0.01±0.0014	1.9E-12	442,391
rs113221802	12	58,624,856	*LINC02403*	intergenic	A	G	0.126	0.0131±0.0021	4.8E-10	442,391
rs1154752	12	68,089,328	*LINC02421*	intergenic	T	C	0.606	-0.0085±0.0014	2.2E-09	442,391
rs490872	12	69,645,864	*CPSF6*	splice acceptor,NMD transcript	A	G	0.274	0.0109±0.0016	2.5E-12	442,391
12:90599240_TA_T	12	90,599,240	*LINC02392*	intergenic	D	I	0.735	-0.0104±0.0016	7.2E-11	442,391
rs10860392	12	99,498,187	*LOC101928937*	intron	T	C	0.633	-0.0104±0.0014	1.2E-13	514,980
rs35159593	12	103,651,399	*C12orf42*	intron NMD transcript	D	I	0.676	0.0145±0.0015	2.8E-22	442,391
12:110289655_CT_C	12	110,289,655	*GLTP*	3' UTR	I	D	0.416	-0.0085±0.0014	3.6E-09	442,391
rs4766500	12	110,993,791	*PPTC7*	intron	A	G	0.634	-0.0087±0.0014	2.1E-09	442,391
rs653178	12	112,007,756	*ATXN2*	intron	T	C	0.518	0.0101±0.0013	3.8E-14	535,885
rs3825200	12	113,503,897	*DTX1*	intron	T	C	0.293	-0.0091±0.0015	4.0E-09	442,391
rs60804348	12	114,419,506	*RBM19*	intergenic	A	G	0.068	-0.0189±0.0028	9.2E-12	442,391
12:117595440_CACT_C	12	117,595,440	*FBXO21*	intron	I	D	0.778	-0.0128±0.0017	2.3E-14	442,391
rs117923369	12	120,840,074	*MSI1*	regulatory	T	C	0.051	-0.0245±0.0032	2.0E-14	442,391
rs10846920	12	122,943,231	*ZCCHC8*	intergenic	T	C	0.740	0.0107±0.0016	1.4E-11	442,391
rs2229840	12	124,826,462	*NCOR2*	missense	T	C	0.160	0.0122±0.0018	2.5E-11	535,885
rs35221880	12	133,301,500	*ANKLE2*	upstream	T	C	0.269	0.0095±0.0016	1.7E-09	442,391
rs9579775	13	20,616,557	*ZMYM2*	intron	A	C	0.864	-0.0125±0.0021	4.1E-09	442,391

rsid	chr	pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs61944841	13	27,049,616	*CDK8*	intergenic	A	G	0.415	0.0118±0.0015	7.9E-16	442,391
rs9512696	13	28,012,527	*MTIF3*	intron	A	G	0.338	-0.0107±0.0015	3.5E-13	442,391
rs1933437	13	28,624,294	*FLT3*	missense	A	G	0.625	-0.0115±0.0014	9.1E-17	532,966
13:31010888_CAA_C	13	31,010,888	*LINC01058*	intergenic	I	D	0.356	-0.0112±0.0015	1.2E-13	442,391
rs7332115	13	33,147,548	*PDS5B*	intergenic	T	G	0.628	0.0117±0.0014	5.6E-17	530,616
rs75983170	13	40,788,838	*LINC00548*	intergenic	T	G	0.665	-0.0091±0.0015	1.8E-09	442,391
13:50426772_TTTTA_T	13	50,426,772	*CTAGE10P*	intergenic	D	I	0.307	0.0101±0.0015	3.5E-11	442,391
rs4942930	13	51,173,280	*DLEU1*	intron	T	C	0.539	0.009±0.0014	1.4E-10	442,391
rs12429545	13	54,102,206	*LINC00558*	intergenic	A	G	0.129	0.0178±0.002	2.2E-18	512,061
rs2121058	13	58,627,256	*LINC02338*	intergenic	T	C	0.772	0.0174±0.0017	1.7E-25	442,391
rs34500367	13	58,686,522	*LINC02338*	intergenic	A	T	0.967	0.0423±0.0062	8.6E-12	442,391
rs4055791	13	59,266,053	*LINC00374*	intergenic	T	C	0.418	-0.014±0.0014	3.8E-23	442,391
rs1949204	13	65,475,834	*LINC01052*	intergenic	T	G	0.239	-0.0102±0.0016	4.2E-10	442,391
rs59041875	13	66,203,733	*LINC01052*	intergenic	D	I	0.563	-0.0101±0.0015	5.9E-12	442,391
rs2010695	13	67,424,416	*PCDH9-AS2*	intron	T	C	0.347	-0.0088±0.0015	2.1E-09	442,391
rs1360371	13	78,455,230	*EDNRB-AS1*	intergenic	T	C	0.225	-0.0129±0.0017	1.3E-14	442,391
rs1441264	13	79,580,919	*LINC00331*	intergenic	A	G	0.592	0.0117±0.0015	7.9E-16	442,391
rs61975177	13	96,309,730	*DZIP1*	intron	A	C	0.278	0.0092±0.0016	3.4E-09	442,391
rs55911231	13	96,983,940	*HS6ST3*	intron	T	C	0.411	0.0113±0.0014	1.9E-15	442,391
rs7333559	13	100,546,450	*LOC101927437*	3' UTR	A	G	0.790	-0.0114±0.0017	3.2E-11	442,391
rs9559013	13	108,000,884	*FAM155A*	intron	A	G	0.138	0.0125±0.002	6.6E-10	442,391
rs750598	13	111,028,978	*COL4A2*	intron	A	G	0.338	0.0087±0.0015	4.0E-09	442,391
rs9560114	13	112,187,882	*LINC02337*	intergenic	A	T	0.735	0.0117±0.0016	2.3E-13	442,391
14:25931404_GTC_G	14	25,931,404	*STXBP6*	intergenic	D	I	0.679	0.0159±0.0015	1.9E-26	442,391
rs11624548	14	29,737,466	*PRKD1*	intron	T	G	0.643	0.012±0.0015	2.1E-16	442,391
rs61980001	14	30,430,683	*PRKD1*	intron	T	C	0.038	0.0235±0.0037	3.0E-10	442,391
14:33298731_CA_C	14	33,298,731	*AKAP6*	intron	I	D	0.546	-0.0113±0.0014	1.1E-15	442,391
rs555928282	14	35,568,947	*LOC101927178*	non-coding transcript exon variant	I	D	0.177	-0.0127±0.0018	5.8E-12	442,391
rs872281	14	40,834,177	*LINC02315*	intergenic	T	C	0.177	-0.011±0.0018	2.4E-09	442,391
rs12889085	14	42,885,336	*LRFN5*	intron	A	G	0.575	-0.0085±0.0014	2.2E-09	442,391
rs5808362	14	47,321,109	*MDGA2*	intron	D	I	0.557	-0.0126±0.0014	1.3E-18	442,391
rs2065999	14	54,656,551	*CDKN3*	regulatory	T	C	0.618	0.0091±0.0014	2.3E-10	442,391
14:56116712_TA_T	14	56,116,712	*KTN1*	intron	D	I	0.371	-0.0103±0.0015	2.6E-12	442,391
rs148410947	14	66,498,900	*FUT8*	intergenic	T	C	0.222	0.0118±0.0017	2.5E-12	442,391
rs35230100	14	70,348,141	*SMOC1*	intron	I	D	0.433	-0.0098±0.0015	1.7E-11	442,391
rs763388	14	73,322,566	*DPF3*	intron NMD transcript	T	C	0.615	0.0111±0.0014	1.1E-14	442,391
rs10142359	14	73,884,540	*NUMB*	intron	A	G	0.520	-0.0089±0.0014	1.5E-10	442,391
rs7141420	14	79,899,454	*NRXN3*	intron	T	C	0.516	0.0128±0.0014	1.1E-19	442,391

rsid	chr	pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs57984207	14	80,641,446	*DIO2*	intergenic	C	G	0.896	-0.0141±0.0023	7.7E-10	442,391
14:91499131_GA_G	14	91,499,131	*RPS6KA5*	intron	D	I	0.603	0.0146±0.0015	9.4E-24	442,391
rs7153027	14	92,427,222	*TRIP11*	intergenic	A	C	0.576	-0.0109±0.0014	1.2E-15	526,523
rs10498635	14	93,103,309	*RIN3*	intron	T	C	0.183	0.0133±0.0017	1.4E-14	534,021
rs1015032	14	93,903,994	*UNC79*	intron	A	G	0.562	0.0139±0.0014	4.9E-23	442,391
rs28929474	14	94,844,947	*SERPINA1*	missense	T	C	0.020	-0.0415±0.0048	3.9E-18	535,885
rs4082793	14	99,700,080	*BCL11B*	intron	T	C	0.427	-0.0091±0.0014	1.4E-10	442,391
rs7161194	14	101,529,005	*MIR377*	upstream	A	G	0.336	0.0146±0.0015	3.9E-21	442,391
rs8022504	14	102,740,275	*MOK*	intron	A	C	0.915	-0.0192±0.0025	1.9E-14	442,391
rs3803286	14	103,246,470	*TRAF3*	intron	A	G	0.333	0.0122±0.0015	1.9E-16	442,391
rs1136165	14	103,988,180	*CKB*	synonymous	T	G	0.628	0.0141±0.0015	1.8E-22	442,391
rs1818917	15	23,941,678	*NDN*	intergenic	T	C	0.512	-0.0084±0.0014	2.3E-09	442,391
rs4779541	15	31,851,884	*OTUD7A*	intron	T	C	0.534	-0.0083±0.0014	3.1E-09	442,391
rs2178004	15	41,991,315	*MGA*	missense	A	T	0.184	0.0122±0.0017	1.9E-12	524,520
rs2453533	15	45,641,225	*GATM*	intergenic	A	C	0.371	0.0082±0.0014	4.0E-09	535,885
rs12439798	15	46,584,787	*LOC105370802*	intergenic	T	G	0.433	0.0111±0.0014	3.0E-15	442,391
rs113875061	15	48,930,234	*FBN1*	intron	I	D	0.125	0.0132±0.0022	4.7E-09	442,391
rs6493534	15	52,379,736	*MAPK6*	intron	T	C	0.421	-0.0091±0.0014	1.8E-10	442,391
15:53492263_AT_A	15	53,492,263	*LINC02490*	intergenic	D	I	0.059	-0.0221±0.003	1.5E-13	442,391
rs147384209	15	55,454,511	*LOC105370829*	intergenic	T	C	0.024	-0.0277±0.0046	2.0E-09	442,391
rs8040114	15	59,957,585	*BNIP2*	non-coding transcript exon variant	T	C	0.615	-0.0088±0.0014	8.0E-10	442,391
rs890156	15	60,913,340	*RORA-AS1*	intron	A	T	0.448	0.0083±0.0014	3.8E-09	442,391
rs72749772	15	62,367,013	*C2CD4A*	downstream	T	C	0.981	0.032±0.0052	6.8E-10	442,391
rs56187480	15	63,789,479	*USP3*	intergenic	A	G	0.345	-0.0095±0.0015	8.9E-11	442,391
15:64898757_CA_C	15	64,898,757	*ZNF609*	intron	I	D	0.084	0.0188±0.0031	1.2E-09	442,391
rs144910307	15	66,506,435	*MEGF11*	intron	A	G	0.733	0.0099±0.0016	4.8E-10	442,391
rs11631381	15	67,339,185	*LOC102723493*	intron	A	G	0.488	-0.0117±0.0014	7.0E-17	442,391
rs35874463	15	67,457,698	*SMAD3*	missense	A	G	0.942	0.0216±0.003	4.2E-13	442,391
rs4776970	15	68,080,886	*MAP2K5*	intron	A	T	0.644	0.0164±0.0015	2.4E-29	442,391
rs2241423	15	68,086,838	*MAP2K5*	intron	A	G	0.223	-0.0172±0.0016	7.0E-27	532,966
15:73049542_CT_C	15	73,049,542	*ADPGK*	non-coding transcript exon variant	I	D	0.345	-0.0097±0.0016	7.3E-10	442,391
rs74022954	15	73,657,063	*HCN4*	intron	T	C	0.073	-0.0169±0.0027	3.2E-10	442,391
rs28588430	15	74,223,430	*LOXL1*	intron	C	G	0.492	0.0159±0.0014	4.9E-30	442,391
rs572865226	15	74,380,752	*GOLGA6A*	intergenic	T	G	0.925	-0.0252±0.0031	6.3E-16	442,391
rs35032603	15	74,753,425	*UBL7*	5' UTR	T	C	0.963	-0.0224±0.0037	1.2E-09	442,391
rs11636613	15	77,335,902	*TSPAN3*	upstream	A	G	0.321	-0.0135±0.0015	1.4E-20	522,343
rs11854808	15	78,011,104	*LINGO1*	intron	A	G	0.283	-0.0098±0.0016	5.3E-10	442,391
rs1443658	15	79,386,366	*RASGRF1*	upstream	A	G	0.580	0.0118±0.0014	1.0E-16	442,391

rsid	chr	pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs2759301	15	80,994,288	*ABHD17C*	intron	A	G	0.449	0.0106±0.0014	4.6E-14	442,391
rs301853	15	84,078,467	*SH3GL3*	intron	A	T	0.834	-0.0111±0.0019	4.8E-09	442,391
rs11259936	15	84,580,582	*ADAMTSL3*	intron	A	C	0.479	0.0274±0.0013	1.1E-92	535,885
rs150961	15	85,086,552	*UBE2Q2P1*	intron	A	G	0.710	0.0135±0.0016	3.4E-18	442,391
rs11633157	15	89,398,939	*ACAN*	synonymous	T	C	0.019	0.0438±0.0054	6.8E-16	442,391
rs3817428	15	89,415,247	*ACAN*	missense	C	G	0.733	-0.0233±0.0015	3.5E-53	529,789
rs62020775	15	89,960,286	*MIR9-3HG*	regulatory	A	T	0.142	-0.0135±0.002	2.1E-11	442,391
rs7173947	15	95,270,467	*LOC440311*	intergenic	T	C	0.643	-0.013±0.0014	3.1E-20	514,980
rs289482	15	98,232,714	*LINC00923*	intergenic	T	C	0.435	-0.0087±0.0014	8.0E-10	442,391
rs2715423	15	99,511,873	*PGPEP1L*	missense	A	G	0.285	-0.0115±0.0015	6.5E-15	534,313
rs2581348	15	100,514,063	*ADAMTS17*	3' UTR	T	C	0.648	-0.0098±0.0015	1.9E-11	442,391
rs72755233	15	100,692,953	*ADAMTS17*	missense	A	G	0.113	0.0338±0.0021	1.1E-57	535,885
rs28583508	15	100,800,049	*ADAMTS17*	intron	T	G	0.661	0.012±0.0015	7.5E-16	442,391
rs7201895	16	407,723	*AXIN1*	intergenic	A	G	0.355	-0.0093±0.0015	2.5E-10	442,391
rs34762152	16	1,634,385	*IFT140*	missense,NMD transcript	T	C	0.064	0.0164±0.0027	1.7E-09	535,885
rs75510884	16	2,161,796	*PKD1*	synonymous	A	G	0.098	-0.0144±0.0024	9.2E-10	442,391
rs3794701	16	3,730,578	*TRAP1*	intron	A	G	0.480	-0.0102±0.0014	3.3E-13	442,391
rs879620	16	4,015,729	*ADCY9*	3' UTR	T	C	0.615	0.0151±0.0014	7.5E-26	442,391
rs4786924	16	6,772,948	*RBFOX1*	intron	A	C	0.206	-0.0102±0.0017	3.9E-09	442,391
rs8062638	16	9,720,130	*GRIN2A*	intergenic	A	C	0.306	-0.0094±0.0015	5.0E-10	442,391
rs6497676	16	10,180,223	*GRIN2A*	intron	A	G	0.629	-0.0097±0.0014	2.0E-11	442,391
rs12444979	16	19,933,600	*GPRC5B*	intergenic	T	C	0.142	-0.0301±0.0019	6.4E-56	535,885
rs868554	16	20,050,466	*GPR139*	intron NMD transcript	C	G	0.760	-0.0143±0.0016	3.6E-18	442,391
16:20259934_AT_A	16	20,259,934	*GP2*	intergenic	D	I	0.130	-0.0123±0.0021	8.8E-09	442,391
rs9652589	16	20,370,816	*PDILT*	missense	T	C	0.514	-0.0105±0.0013	5.0E-15	532,966
rs9922288	16	24,550,930	*RBBP6*	5' UTR	A	G	0.236	0.0169±0.0017	3.1E-24	442,391
rs78457529	16	24,950,880	*ARHGAP17*	missense	T	C	0.012	-0.0464±0.0061	3.3E-14	535,885
16:28556564_CT_C	16	28,556,564	*NUPR1*	downstream	I	D	0.630	-0.0122±0.0014	2.9E-17	442,391
rs3814883	16	29,994,922	*TAOK2*	synonymous	T	C	0.483	0.0175±0.0014	4.4E-36	442,391
rs111739641	16	30,510,984	*ITGAL*	intron	D	I	0.498	0.0116±0.0014	1.4E-16	442,391
rs12716979	16	31,011,821	*STX1B*	intron	T	G	0.625	0.0144±0.0014	3.1E-23	442,391
rs8050390	16	31,539,386	*AHSP*	intron	A	G	0.360	0.0104±0.0015	6.0E-13	442,391
rs6500208	16	49,011,249	*CBLN1*	intergenic	A	G	0.207	0.0103±0.0017	2.6E-09	442,391
rs4533281	16	49,725,798	*ZNF423*	intron	A	G	0.755	0.0106±0.0016	7.3E-11	442,391
rs12927850	16	50,044,403	*CNEP1R1*	intergenic	A	G	0.373	0.0095±0.0015	1.3E-10	442,391
rs76488452	16	53,756,885	*FTO*	intron	A	G	0.948	-0.0118±0.0031	1.7E-04	442,391
rs1421085	16	53,800,954	*FTO*	intron	T	C	0.596	-0.0544±0.0014	2.1E-346	535,885
rs1861867	16	53,848,561	*FTO*	intron	A	G	0.400	-0.0399±0.0015	6.8E-160	442,391
rs11373	16	56,545,175	*BBS2*	missense	T	C	0.827	-0.011±0.0018	4.2E-10	535,885

rsid	chr	pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs73586863	16	67,011,228	*CES3*	downstream	A	G	0.960	-0.0226±0.0036	2.4E-10	442,391
16:67554712_GA_G	16	67,554,712	*LOC100505942*	intron	D	I	0.078	0.0185±0.0026	2.8E-12	442,391
rs2307022	16	68,381,978	*PRMT7*	intron	A	G	0.331	0.0133±0.0015	3.1E-19	442,391
rs5011579	16	69,187,318	*UTP4*	intron	C	G	0.285	-0.0111±0.0015	7.7E-13	442,391
rs1364063	16	69,588,572	*MIR1538*	TF binding site	T	C	0.589	0.0128±0.0014	5.8E-21	535,885
16:70252464_CA_C	16	70,252,464	*SMG1P7*	downstream	I	D	0.494	-0.0123±0.0014	3.3E-18	442,391
rs117512594	16	71,685,344	*PHLPP2*	intron	A	G	0.915	-0.0156±0.0025	8.2E-10	442,391
rs811054	16	72,251,132	*PMFBP1*	intergenic	T	C	0.537	0.0094±0.0014	2.0E-11	442,391
rs62053191	16	72,993,706	*ZFHX3*	synonymous	A	G	0.084	-0.0148±0.0025	4.0E-09	442,391
rs4493083	16	75,294,589	*BCAR1*	intron	T	C	0.507	-0.0109±0.0014	9.1E-15	442,391
rs4074541	16	78,345,435	*WWOX*	intron	A	C	0.659	-0.009±0.0015	8.2E-10	442,391
rs12599980	16	81,587,084	*CMIP*	intron	T	C	0.408	-0.01±0.0014	1.8E-12	442,391
rs9936936	16	81,644,087	*CMIP*	non-coding transcript exon variant	T	C	0.530	0.0065±0.0014	2.9E-06	442,391
rs2303232	16	84,808,969	*USP10*	intron	C	G	0.488	0.0096±0.0014	3.0E-11	442,391
rs1552657	16	86,424,697	*LINC00917*	intergenic	A	G	0.444	-0.0087±0.0014	6.4E-10	442,391
rs76520574	16	88,321,027	*LINC02182*	intron	T	C	0.041	-0.0239±0.0036	2.5E-11	442,391
rs62068640	16	89,496,727	*ANKRD11*	intron	T	C	0.442	-0.0086±0.0014	2.5E-09	442,391
rs57207396	17	1,630,208	*WDR81*	missense	T	C	0.261	0.0093±0.0015	9.8E-10	531,750
rs758959824	17	1,823,490	*RTN4RL1*	intergenic	D	I	0.156	-0.0147±0.0019	3.7E-14	442,391
rs10852932	17	2,143,460	*SMG6*	5' UTR	T	G	0.351	0.0148±0.0014	6.9E-26	534,313
17:4175556_GTATA_G	17	4,175,556	*UBE2G1*	3' UTR	D	I	0.021	-0.0302±0.0049	9.6E-10	442,391
rs78378222	17	7,571,752	*TP53*	3' UTR	T	G	0.988	-0.0606±0.0064	3.9E-21	442,391
rs57828851	17	7,732,351	*DNAH2*	intron	A	T	0.581	-0.0094±0.0014	3.6E-11	442,391
rs10684399	17	15,945,474	*NCOR1*	intron	I	D	0.558	0.0087±0.0014	6.3E-10	442,391
rs2605136	17	18,136,750	*LLGL1*	intron	T	C	0.355	0.009±0.0015	7.1E-10	442,391
rs1320251	17	21,264,396	*KCNJ12*	regulatory	T	C	0.454	-0.0123±0.0014	1.6E-18	442,391
rs113000401	17	27,530,458	*MYO18A*	non-coding transcript exon variant	T	C	0.049	-0.0199±0.0033	1.1E-09	442,391
rs536851066	17	30,251,481	*SUZ12*	intergenic	I	D	0.089	-0.016±0.0027	2.2E-09	442,391
rs11868042	17	32,250,375	*ASIC2*	intron	A	T	0.183	0.0116±0.0018	2.7E-10	442,391
rs2306593	17	34,866,546	*MYO19*	intron	T	C	0.488	-0.0107±0.0014	1.6E-14	442,391
rs12450937	17	35,494,511	*ACACA*	intron	T	C	0.357	0.0089±0.0015	9.7E-10	442,391
rs349792	17	39,261,560	*KRTAP4-8*	upstream	A	T	0.691	-0.0111±0.0015	2.2E-13	442,391
17:42325043_CTGGCAGAGTTAGCTCTCTGACCCTT_C	17	42,325,043	*SLC4A1*	downstream	I	D	0.975	-0.0315±0.0046	6.2E-12	442,391
rs188710438	17	43,573,649	*PLEKHM1*	intergenic	T	C	0.226	0.014±0.0017	3.5E-17	442,391
rs138464472	17	44,130,160	*KANSL1*	intron	D	I	0.286	0.0145±0.0016	2.5E-20	442,391
rs56755526	17	44,793,627	*NSF*	intron	I	D	0.743	-0.0127±0.0016	1.2E-14	442,391

rsid	chr	pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
17:45481842_TAA_T	17	45,481,842	*EFCAB13*	intron	D	I	0.080	-0.0205±0.0026	5.7E-15	442,391
rs376610477	17	46,292,292	*SKAP1*	intron	D	I	0.649	0.0146±0.0016	2.1E-19	442,391
rs12603429	17	46,836,576	*TTLL6*	downstream	C	G	0.456	0.0111±0.0014	2.2E-15	442,391
rs11079849	17	47,090,785	*IGF2BP1*	intron	T	C	0.328	-0.0129±0.0015	5.3E-18	442,391
17:47382240_GTA_G	17	47,382,240	*ZNF652*	intron	D	I	0.394	-0.017±0.0015	2.6E-30	442,391
rs28412876	17	47,454,515	*LOC102724596*	intron	T	G	0.376	0.0156±0.0014	6.1E-27	442,391
rs74929176	17	54,905,494	*C17orf67*	intron	T	C	0.227	-0.0111±0.0017	3.4E-11	442,391
rs72845886	17	61,666,687	*DCAF7*	3' UTR	T	C	0.058	-0.0173±0.0029	1.3E-09	535,885
rs236531	17	68,217,860	*KCNJ2*	intergenic	T	C	0.416	0.0089±0.0014	2.5E-10	481,528
rs2019877	17	71,759,960	*LINC00469*	intron	T	C	0.436	-0.0083±0.0014	6.3E-09	442,391
rs7218677	17	76,146,520	*C17orf99*	intron	A	T	0.688	0.0092±0.0015	2.7E-09	442,391
rs12941365	17	76,897,733	*CEP295NL,TIMP2*	intron	T	G	0.412	0.0096±0.0014	1.2E-11	442,391
rs2587507	17	77,790,135	*LINC01977*	intergenic	T	C	0.494	0.0093±0.0013	3.8E-12	528,831
rs11150745	17	78,757,626	*RPTOR*	intron	A	G	0.681	0.0138±0.0015	4.4E-20	442,391
rs3935190	17	79,084,367	*BALAP2*	intron	A	G	0.536	-0.0114±0.0014	5.1E-16	442,391
rs10664154	17	79,375,625	*BAHCC1*	non-coding transcript exon variant	D	I	0.344	0.0095±0.0015	1.5E-10	442,391
rs4625783	17	80,086,395	*CCDC57*	missense	T	C	0.407	0.01±0.0014	4.9E-13	512,821
rs6506537	18	794,273	*YES1*	intron	T	C	0.754	-0.01±0.0016	8.7E-10	442,391
rs60764613	18	1,839,911	*LINC00470*	intergenic	T	G	0.145	0.0121±0.002	1.2E-09	442,391
rs5822974	18	7,548,483	*PTPRM*	intergenic	D	I	0.812	0.0111±0.0018	7.3E-10	442,391
rs17498752	18	9,254,785	*ANKRD12*	missense	T	C	0.087	-0.0153±0.0024	1.4E-10	535,885
rs1788808	18	21,090,023	*C18orf8*	intron	A	G	0.504	0.0101±0.0014	4.8E-13	442,391
rs34866209	18	30,458,628	*CCDC178*	intron	D	I	0.222	0.01±0.0017	2.4E-09	442,391
rs1941706	18	31,223,776	*ASXL3*	intron	A	G	0.537	-0.0105±0.0014	8.1E-14	442,391
rs1443641	18	32,690,756	*MAPRE2*	intron	A	T	0.542	-0.008±0.0014	1.5E-08	442,391
rs559231	18	39,644,247	*PIK3C3*	intron	T	G	0.393	0.0086±0.0014	1.9E-09	442,391
rs9956387	18	44,773,382	*SKOR2*	missense	A	T	0.497	0.0083±0.0014	3.0E-09	442,391
rs7239114	18	45,921,214	*ZBTB7C*	intron	A	G	0.542	0.0082±0.0014	6.0E-09	442,391
rs7243172	18	52,472,235	*RAB27B*	intron	T	G	0.542	-0.009±0.0014	1.5E-10	442,391
rs11659764	18	53,335,512	*LINC01415*	upstream	A	T	0.053	-0.0255±0.0031	4.5E-16	442,391
rs200519957	18	56,878,075	*GRP*	intergenic	D	I	0.835	0.0132±0.0019	2.7E-12	442,391
rs8086105	18	57,685,001	*PMAIP1*	upstream	T	C	0.745	0.0145±0.0016	4.1E-19	442,391
rs17782313	18	57,851,097	*MC4R*	intergenic	T	C	0.767	-0.0417±0.0016	1.0E-151	532,966
rs201873971	18	57,863,192	*MC4R*	upstream	T	C	0.879	-0.0426±0.0026	1.8E-62	442,391
rs12457883	18	57,870,593	*MC4R*	intergenic	A	T	0.930	0.0268±0.0027	1.6E-22	442,391
rs2331933	18	57,955,945	*MC4R*	intergenic	T	G	0.650	-0.0204±0.0015	1.9E-43	442,391
rs111389401	18	58,420,595	*MC4R*	intron	T	C	0.019	-0.0475±0.0053	1.6E-19	442,391
rs12454712	18	60,845,884	*BCL2*	intron	T	C	0.624	-0.0106±0.0014	1.9E-13	442,391
rs8089514	18	69,224,478	*LINC01541*	intron	A	T	0.368	0.0087±0.0015	2.9E-09	442,391

rsid	chr	pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs11150911	18	73,498,528	*LINC01898*	intergenic	A	C	0.276	0.0097±0.0016	6.8E-10	442,391
rs3957285	19	1,891,992	*ABHD17A*	intergenic	A	G	0.544	0.0107±0.0014	3.3E-14	442,391
19:2230029_CTT_C	19	2,230,029	*DOT1L*	3' UTR	I	D	0.942	-0.0215±0.003	9.7E-13	442,391
rs7351050	19	4,044,579	*ZBTB7A*	3' UTR	A	G	0.162	-0.0129±0.0019	1.6E-11	442,391
rs243342	19	4,406,160	*CHAF1A*	intron	A	G	0.545	0.0105±0.0014	5.6E-14	442,391
rs11878235	19	7,976,698	*MAP2K7*	intron	A	G	0.595	-0.0099±0.0014	4.6E-12	442,391
rs62621197	19	8,670,147	*ADAMTS10*	missense	T	C	0.037	0.0414±0.0038	5.7E-27	442,391
19:8675220_AACAC_A	19	8,675,220	*ADAMTS10*	intron	D	I	0.628	-0.0079±0.0015	7.2E-08	442,391
rs35915221	19	9,498,618	*ZNF559-ZNF177*	intergenic	I	D	0.586	0.0115±0.0014	6.7E-16	442,391
rs2878723	19	10,011,375	*OLFM2*	intron	T	C	0.689	-0.0097±0.0015	2.2E-10	442,391
rs1078264	19	12,963,143	*MAST1*	splice region,synonymous	T	C	0.732	-0.0109±0.0015	5.6E-13	532,966
rs531173354	19	13,946,454	*LOC284454*	non-coding transcript exon variant	I	D	0.334	0.0101±0.0015	1.6E-11	442,391
rs7246865	19	17,219,105	*MYO9B*	intron	A	G	0.259	-0.0117±0.0016	3.1E-13	442,391
rs199590561	19	18,381,169	*KIAA1683*	intron	I	D	0.269	0.0107±0.0016	4.1E-11	442,391
rs113230003	19	18,460,956	*PGPEP1*	intron	A	G	0.260	-0.012±0.0016	6.6E-14	442,391
rs2051815	19	18,835,115	*CRTC1*	intron	A	G	0.324	-0.0123±0.0015	2.3E-16	442,391
rs1064395	19	19,361,735	*NCAN*	3' UTR	A	G	0.160	-0.0207±0.0018	1.5E-29	530,616
rs75270598	19	19,864,849	*LINC00663*	downstream	T	C	0.154	-0.0133±0.0019	7.6E-12	442,391
rs8102137	19	30,296,853	*CCNE1*	regulatory	T	C	0.670	-0.0135±0.0014	4.6E-21	530,616
rs73019624	19	30,709,251	*ZNF536*	downstream	A	G	0.207	-0.0115±0.0017	3.3E-11	442,391
rs73026723	19	31,017,177	*ZNF536*	intron	T	C	0.155	-0.0173±0.0019	2.6E-19	442,391
rs29941	19	34,309,532	*KCTD15*	downstream	A	G	0.327	-0.0135±0.0014	7.4E-21	512,061
rs2075650	19	45,395,619	*TOMM40*	intron	A	G	0.854	0.0127±0.0019	2.5E-11	535,885
rs1800437	19	46,181,392	*GIPR*	missense	C	G	0.195	-0.0202±0.0017	1.0E-32	528,184
rs3810291	19	47,569,003	*ZC3H4*	3' UTR	A	G	0.676	0.0231±0.0014	3.6E-58	531,792
rs200876443	19	50,363,585	*PTOV1*	3' UTR	I	D	0.964	-0.0377±0.0038	1.3E-23	442,391
rs760129606	19	51,812,546	*IGLON5*	upstream	D	I	0.462	0.0105±0.0014	1.6E-13	442,391
rs6132308	20	2,102,598	*STK35*	intron	T	C	0.578	0.0083±0.0014	5.2E-09	442,391
rs633284	20	2,904,143	*PTPRA*	intron	A	T	0.484	0.0094±0.0014	3.3E-11	442,391
rs2145270	20	6,621,685	*LINC01713*	intergenic	T	C	0.621	0.0179±0.0014	2.0E-38	535,885
rs1578407	20	17,196,608	*PCSK2*	intergenic	T	C	0.728	0.01±0.0016	1.9E-10	442,391
rs742698	20	20,083,949	*CFAP61*	intron	T	C	0.301	0.0096±0.0015	3.2E-10	442,391
rs4815025	20	21,142,523	*KIZ*	non-coding transcript exon variant	C	G	0.331	0.0094±0.0014	6.4E-11	526,353
rs116948922	20	25,534,854	*NINL*	intron	T	C	0.033	-0.0235±0.004	3.0E-09	442,391
rs375341392	20	30,986,406	*ASXL1*	intron	D	I	0.182	0.011±0.0018	2.2E-09	442,391
20:32563820_GTATATA_G	20	32,563,820	*RALY-AS1*	intergenic	D	I	0.601	0.01±0.0014	2.3E-12	442,391
rs7268343	20	42,754,240	*JPH2*	intron	A	G	0.122	0.0141±0.0021	4.7E-11	442,391

rsid	chr	pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs2425847	20	44,907,927	*CDH22*	intron	A	G	0.591	-0.009±0.0014	2.5E-10	442,391
rs112852122	20	47,498,117	*ARFGEF2*	intergenic	A	G	0.159	-0.0112±0.0019	8.0E-09	442,391
rs5014389	20	48,597,280	*SNAI1*	upstream	A	T	0.104	0.0148±0.0023	1.0E-10	442,391
rs6096548	20	50,347,558	*ATP9A*	intron	A	T	0.257	0.0094±0.0016	4.1E-09	442,391
rs6096886	20	50,951,298	*ZFP64*	intron	A	G	0.810	0.0183±0.0018	8.8E-25	442,391
rs2207895	20	54,387,375	*CBLN4*	intergenic	A	G	0.190	-0.0113±0.0018	2.6E-10	442,391
rs61734651	20	61,451,332	*COL9A3*	missense,splice region	T	C	0.071	-0.0246±0.0028	4.0E-18	442,391
rs150998792	20	62,130,403	*EEF1A2*	splice region,intron	T	G	0.050	-0.0203±0.0033	9.8E-10	442,391
rs7264802	20	62,692,440	*TCEA2*	intron	A	G	0.252	-0.0119±0.0016	2.3E-13	442,391
rs2823991	21	18,082,592	*MIR99AHG*	regulatory	T	G	0.740	-0.0107±0.0016	3.0E-11	442,391
rs2830585	21	28,305,212	*ADAMTS5*	missense	T	C	0.160	0.0124±0.0018	1.4E-11	532,966
rs13047416	21	40,309,436	*LOC400867*	intron	C	G	0.623	0.0102±0.0014	2.0E-12	442,391
rs394608	21	46,581,798	*ADARB1*	intron	T	C	0.462	-0.01±0.0014	1.6E-12	442,391
rs4680	22	19,951,271	*COMT*	missense,NMD transcript	A	G	0.517	0.01±0.0013	1.3E-13	531,750
rs35359276	22	20,769,186	*ZNF74*	downstream	I	D	0.293	0.0093±0.0016	3.7E-09	442,391
rs878718	22	31,552,702	*MIR3928*	upstream	A	G	0.436	0.0102±0.0014	5.9E-13	442,391
rs470072	22	32,263,131	*DEPDC5*	intron	T	C	0.459	-0.0103±0.0014	7.1E-14	510,887
rs75425504	22	40,684,279	*TNRC6B*	intron	D	I	0.650	0.0176±0.0015	6.2E-33	442,391
rs139497	22	41,640,098	*RANGAP1*	upstream	T	C	0.688	-0.0128±0.0015	4.5E-17	442,391
rs5751239	22	42,592,239	*TCF20*	intron	T	C	0.522	0.0085±0.0014	1.6E-09	442,391
rs9614666	22	45,829,560	*RIBC2*	downstream	A	G	0.837	0.0144±0.0019	3.3E-14	442,391
rs9617050	22	50,708,731	*MAPK11*	5' UTR	A	G	0.519	-0.0083±0.0014	7.8E-09	442,391
rs1379871	23	31,854,782	*DMD*	intron	C	G	0.343	0.0097±0.0012	3.0E-15	442,391
rs35473057	23	53,601,990	*HUWE1*	intron	A	C	0.466	-0.0137±0.0012	7.9E-32	442,391
rs41303733	23	67,652,748	*OPHN1*	missense	T	C	0.085	-0.0128±0.0021	7.1E-10	442,391
rs11539157	23	68,381,264	*PJA1*	missense	A	C	0.238	-0.0181±0.0014	3.0E-40	442,391
rs5922922	23	83,467,844	*JDP2*	intergenic	A	G	0.147	0.0096±0.0016	4.7E-09	442,391
rs6524628	23	85,558,345	*DACH2*	intron	A	G	0.414	-0.0081±0.0012	5.5E-12	442,391
rs5920845	23	99,966,059	*SYTL4*	intron	T	C	0.332	0.0085±0.0012	6.6E-12	442,391
rs12833777	23	100,785,073	*ARMCX4*	intron NMD transcript	A	T	0.555	-0.0069±0.0012	2.9E-09	442,391
rs12560103	23	101,383,199	*TCEAL2*	downstream	T	C	0.641	0.0094±0.0012	9.9E-15	442,391
rs6567865	23	109,701,864	*RGAG1*	downstream	A	C	0.399	0.0088±0.0012	8.6E-14	442,391
rs138073470	23	117,843,442	*IL13RA1*	intergenic	A	G	0.831	-0.0157±0.0016	1.0E-23	442,391
rs747153900	23	131,290,808	*FRMD7*	intergenic	A	G	0.228	-0.012±0.0014	1.5E-17	442,391
rs1190742	23	136,075,730	*GPR101*	non-coding transcript exon variant	C	G	0.502	-0.0077±0.0012	3.4E-11	442,391
rs750212356	23	137,050,402	*ZIC3*	intergenic	I	D	0.154	-0.0149±0.0017	3.7E-18	442,391
Women-only										
rs55656112	1	172,189,764	*DNM3*	intron	A	C	0.788	0.0209±0.0033	1.2E-10	240,552
rs925422	4	60,254,101	*LINC02429*	intergenic	T	G	0.259	0.019±0.0031	6.0E-10	240,552

rsid	chr	pos	Nearest gene	Type	EA	OA	EAF	Beta±SE	*p*	N
rs2667360	4	140,793,531	*MAML3*	intron	A	T	0.694	0.0175±0.0029	2.0E-09	240,552
rs11424823	5	171,284,642	*FBXW11*	downstream	D	I	0.380	0.0169±0.0029	3.6E-09	240,552
rs41271299	6	19,839,415	*ID4*	intron	T	C	0.052	-0.0378±0.006	4.1E-10	240,552
15:59278382_TTG_T	15	59,278,382	*RNF111*	intron	D	I	0.751	0.0202±0.0032	4.8E-10	240,552
rs727702	15	70,006,236	*PCAT29*	intergenic	A	G	0.117	0.0244±0.0042	4.9E-09	240,552
17:56542985_CA_C	17	56,542,985	*HSF5*	intron	I	D	0.273	0.0189±0.0032	3.3E-09	240,552
Men-only										
rs60745892	1	66393106	*PDE4B*	intron	D	I	0.658	0.0192±0.0031	4.0E-10	201,839
rs823094	1	205689807	*NUCKS1*	intron	T	G	0.574	0.0192±0.0029	6.2E-11	201,839
rs111501168	2	113993983	*PAX8-AS1*	non-coding transcript exon variant	T	C	0.839	-0.0246±0.004	8.6E-10	201,839
rs77169818	18	74980601	*GALR1*	missense	A	T	0.957	0.0422±0.0066	1.7E-10	235,388
rs4804524	19	10867448	*DNM2*	intron	A	G	0.612	-0.0179±0.003	2.4E-09	201,839
rs5934507	23	8917206	*FAM9B*	intergenic	A	G	0.732	-0.0152±0.0023	6.1E-11	201,839

Supplementary Table 3.7: 119 variants for FFMI which are solely driven by the association with height

rsid	chr	Position	Nearest gene	EA	OA	EAF	beta FFMI observed	Beta FFMI expected	p diff	p FFMI	p FFM	p hght
rs9970807	1	56965664	*PLPP3*	T	C	0.092	0.014	0.026	4.9E-02	1.8E-09	7.6E-02	2.2E-06
rs12089815	1	91189933	*BARHL2*	A	G	0.548	-0.009	-0.016	5.2E-02	5.2E-10	2.0E-02	1.7E-07
rs143792309	1	92505309	*EPHX4*	T	C	0.064	0.017	0.040	2.1E-03	1.2E-09	2.2E-01	5.6E-11
rs984222	1	119503843	*TBX15*	C	G	0.386	0.014	0.027	4.2E-04	2.3E-24	8.8E-05	1.7E-16
rs10802069	1	119517357	*TBX15*	T	C	0.387	0.014	0.027	2.7E-04	8.3E-24	7.6E-05	2.1E-16
rs1194610	1	154296076	*AQP10*	T	C	0.765	-0.010	-0.014	2.6E-01	2.3E-09	1.1E-02	3.2E-05
rs35208023	1	156415673	*C1orf61*	T	C	0.669	0.012	0.018	9.5E-02	5.9E-15	4.0E-03	2.3E-08
rs10919515	1	170869758	*MROH9*	A	T	0.095	-0.020	-0.037	1.1E-02	4.1E-14	1.7E-02	1.1E-10
rs4916229	1	171443368	*PRRC2C*	C	G	0.904	-0.019	-0.028	1.1E-01	3.8E-15	1.2E-03	3.9E-07
rs12079987	1	221715102	*DUSP10*	A	G	0.024	-0.032	-0.057	2.5E-02	2.7E-12	1.2E-02	1.9E-08
rs300789	2	89910	*FAM110C*	A	G	0.798	-0.010	-0.027	1.3E-04	3.5E-09	4.4E-01	8.3E-13
rs404108	2	33301389	*LTBP1*	A	G	0.765	-0.010	-0.025	1.8E-04	2.1E-09	3.4E-01	1.8E-11
rs11690012	2	43542480	*THADA*	C	G	0.23	-0.010	-0.017	6.9E-02	5.9E-09	5.7E-03	2.4E-06
rs6760379	2	55279536	*RTN4*	T	C	0.693	0.010	0.018	3.3E-02	8.3E-11	8.4E-03	2.5E-07
rs12476772	2	60203917	*MIR4432HG*	A	C	0.331	0.010	0.019	1.0E-02	1.9E-11	1.1E-02	2.5E-08
rs34355307	2	86800506	*RNF103-CHMP3*	A	G	0.61	-0.011	-0.019	2.6E-02	8.5E-15	2.3E-04	5.6E-09
rs10489971	2	102527827	*MAP4K4*	T	C	0.307	-0.009	-0.014	2.1E-01	2.1E-09	2.2E-03	4.4E-05
rs10497820	2	199612796	*SATB2*	A	C	0.341	0.011	0.017	1.0E-01	4.0E-13	1.5E-04	1.8E-06
rs2108485	2	242021742	*SNED1*	T	G	0.155	-0.014	-0.028	3.9E-03	3.3E-15	1.1E-02	1.3E-11
rs9852062	3	45373442	*LARS2*	A	T	0.556	-0.010	-0.016	6.6E-02	1.4E-11	3.2E-03	3.1E-06
rs72906474	3	47817007	*SMARCC1*	T	G	0.584	-0.012	-0.025	8.5E-04	2.9E-18	8.8E-04	7.1E-14
rs141434554	3	52208046	*POC1A*	D	I	0.78	-0.013	-0.030	8.8E-05	1.5E-13	8.6E-02	3.8E-14
rs68093214	3	70657785	*FOXP1*	T	C	0.751	0.011	0.016	2.3E-01	5.2E-12	1.2E-03	2.2E-05
rs2399213	3	107395302	*BBX*	T	C	0.819	-0.014	-0.021	1.2E-01	1.2E-14	1.9E-04	1.5E-07
rs17619973	3	114417675	*ZBTB20*	A	G	0.925	0.016	0.041	2.0E-04	2.2E-09	2.0E-01	2.8E-11
rs9881641	3	117602500	*LINC02024*	T	C	0.687	-0.009	-0.018	1.4E-02	1.9E-09	7.3E-02	2.2E-07
rs17282078	3	124481760	*ITGB5*	A	T	0.131	-0.015	-0.025	4.0E-02	1.5E-12	7.7E-03	1.1E-07
rs76594121	3	128189391	*DNAJB8-AS1*	T	G	0.953	0.024	0.043	2.4E-02	1.1E-12	1.8E-02	9.3E-09
rs9827728	3	133971558	*RYK*	A	T	0.88	-0.015	-0.034	5.8E-04	3.1E-12	1.1E-01	2.7E-12
rs2051559	4	3298800	*RGS12*	T	C	0.867	-0.015	-0.029	5.2E-03	1.1E-12	5.8E-02	9.7E-10
rs6833878	4	10005555	*SLC2A9*	A	T	0.732	-0.010	-0.021	3.7E-03	1.3E-09	1.4E-01	3.6E-09
rs3209570	4	38699657	*KLF3*	A	G	0.399	-0.010	-0.014	1.8E-01	2.8E-11	1.2E-03	5.6E-06
rs781659	4	57779851	*REST*	A	G	0.573	-0.009	-0.021	8.6E-04	6.4E-11	3.6E-02	1.4E-10
rs141374503	4	73179445	*ADAMTS3*	T	C	0.004	0.075	0.158	2.6E-03	2.6E-12	2.4E-02	2.9E-09
rs1054627	4	88732692	*IBSP*	A	G	0.313	0.010	0.015	1.9E-01	1.0E-11	4.8E-03	2.5E-05
rs201081507	4	102681041	*BANK1*	A	G	0.943	-0.022	-0.050	6.7E-04	1.4E-11	2.3E-02	4.2E-11
4:103981698_GT_G	4	103981698	*SLC9B2*	D	I	0.404	-0.010	-0.019	2.1E-02	4.7E-13	1.0E-03	4.0E-08
rs75544266	4	104584997	*TACR3*	T	C	0.056	-0.022	-0.037	4.1E-02	1.5E-12	3.8E-03	3.2E-08

rsid	chr	Position	Nearest gene	EA	OA	EAF	beta FFMI observed	Beta FFMI expected	p diff	p FFMI	p FFM	p hght
rs7688177	4	120760872	*LINC01365*	T	C	0.77	-0.011	-0.023	7.6E-03	1.5E-11	1.4E-02	6.9E-09
rs7666785	4	144060464	*LOC105377623*	A	G	0.4	0.009	0.015	1.2E-01	9.6E-12	1.1E-02	1.2E-06
rs11933087	4	145722862	*HHIP*	A	T	0.629	0.012	0.014	5.7E-01	1.1E-16	8.1E-05	3.5E-06
rs35016840	4	151239774	*LRBA*	T	C	0.641	-0.010	-0.016	9.0E-02	3.0E-11	7.1E-03	2.1E-06
rs459193	5	55806751	*C5orf67*	A	G	0.254	-0.013	-0.024	1.1E-02	7.3E-18	1.7E-03	1.5E-11
rs55963623	5	64555615	*ADAMTS6*	T	C	0.456	0.013	0.020	4.9E-02	1.3E-20	8.7E-05	1.5E-11
rs62360508	5	65177737	*ERBIN*	T	C	0.913	0.015	0.035	2.3E-03	4.9E-10	2.2E-02	3.5E-10
rs10942491	5	86382726	*MIR4280*	C	G	0.553	-0.010	-0.014	1.9E-01	1.3E-11	5.1E-04	1.5E-05
rs35843836	5	88798726	*MEF2C-AS1*	A	T	0.633	-0.015	-0.029	7.1E-05	8.2E-24	1.5E-04	8.5E-20
rs149457	5	107438057	*FBXL17*	T	C	0.17	-0.019	-0.031	1.0E-02	2.3E-24	8.1E-05	8.9E-13
rs1479585	5	140976528	*DIAPH1*	C	G	0.607	0.011	0.019	1.9E-02	6.7E-15	4.5E-04	4.7E-08
rs9394951	6	43350753	*ZNF318*	T	C	0.567	-0.009	-0.023	9.8E-05	3.7E-11	1.7E-01	3.1E-12
rs78856780	6	56939238	*ZNF451*	A	G	0.856	-0.012	-0.028	1.2E-03	1.2E-09	6.6E-02	7.0E-09
rs4719730	7	929123	*GET4*	T	C	0.175	-0.011	-0.017	2.1E-01	1.9E-09	1.6E-03	1.4E-05
rs4721089	7	1872921	*MAD1L1*	T	C	0.784	0.011	0.017	1.3E-01	2.3E-10	3.0E-04	1.6E-05
rs62441156	7	5347446	*TNRC18*	A	G	0.773	-0.010	-0.025	4.0E-04	3.0E-09	8.6E-02	1.3E-10
rs12669977	7	13132400	*ARL4A*	T	G	0.502	0.008	0.015	4.0E-02	3.1E-09	2.8E-01	1.3E-06
rs17141862	7	19770756	*TMEM196*	T	C	0.137	-0.014	-0.031	9.1E-04	6.4E-12	4.0E-02	1.9E-12
rs28423374	7	20579647	*ABCB5*	T	C	0.736	-0.011	-0.017	1.7E-01	6.8E-13	1.0E-03	7.8E-06
rs2067087	7	27241660	*HOTTIP*	C	G	0.715	0.011	0.022	4.9E-03	1.2E-12	1.3E-02	4.4E-10
rs7384844	7	28025308	*JAZF1*	A	T	0.556	0.009	0.016	3.5E-02	8.1E-10	9.0E-03	2.2E-05
rs554794335	7	32931486	*KBTBD2*	D	I	0.689	0.010	0.020	6.9E-03	4.0E-10	7.1E-02	6.4E-08
rs329270	7	35075619	*DPY19L1*	A	G	0.515	0.010	0.022	6.9E-04	5.0E-13	2.6E-02	2.2E-12
rs1207728	7	96638021	*DLX6-AS1*	C	G	0.225	0.010	0.020	2.1E-02	2.1E-09	6.8E-02	1.3E-07
rs28566086	7	99025490	*ATP5J2-PTCD1, PTCD1*	C	G	0.114	-0.018	-0.036	1.1E-03	1.1E-15	4.4E-03	3.6E-14
rs12669944	7	100446365	*SLC12A9-AS1*	A	G	0.35	-0.008	-0.016	2.3E-02	2.9E-07	1.3E-01	1.8E-07
rs10500039	7	114351267	*FOXP2*	T	C	0.578	-0.011	-0.012	7.9E-01	1.1E-14	8.0E-05	4.6E-05
rs7010322	8	13196821	*DLC1*	T	C	0.36	-0.009	-0.017	2.4E-02	7.4E-10	2.7E-02	1.0E-06
rs4082204	8	38328902	*FGFR1*	A	G	0.402	0.010	0.014	2.1E-01	1.0E-11	4.3E-03	5.9E-06
rs7838717	8	145504343	*BOP1*	T	C	0.365	-0.010	-0.017	6.7E-02	3.6E-11	2.0E-02	1.5E-07
rs10960276	9	11819686	*TYRP1*	A	C	0.356	-0.009	-0.016	1.0E-01	1.8E-10	1.2E-02	1.9E-06
rs4741546	9	15846112	*CCDC171*	T	C	0.396	-0.008	-0.013	1.7E-01	1.0E-08	1.4E-02	3.1E-05
rs544957562	9	33978015	*UBAP2*	A	T	0.845	0.016	0.022	2.5E-01	1.1E-15	3.5E-04	2.7E-06
rs7874181	9	85858511	*FRMD3*	T	G	0.847	-0.012	-0.018	2.5E-01	6.4E-10	2.0E-03	7.3E-06
rs35870355	9	92039730	*SEMA4D*	A	T	0.235	0.010	0.019	3.5E-02	7.3E-10	2.8E-02	1.2E-05
rs2672813	9	98781179	*ERCC6L2,LINC00092*	A	G	0.44	-0.010	-0.024	1.2E-04	4.8E-13	2.0E-01	8.4E-13
rs3838333	9	117787666	*TNC*	I	D	0.429	0.009	0.017	1.7E-02	2.7E-10	5.6E-02	3.2E-08

rsid	chr	Position	Nearest gene	EA	OA	EAF	beta FFMI observed	Beta FFMI expected	p diff	p FFMI	p FFM	p hght
rs80031633	9	120687550	*TLR4*	A	T	0.457	-0.009	-0.014	1.5E-01	2.8E-10	3.2E-03	4.9E-05
rs2267958	9	131015279	*DNM1*	A	G	0.512	-0.009	-0.020	1.6E-03	7.2E-10	4.1E-02	7.3E-10
rs2435381	10	43678796	*CSGALNACT2*	T	C	0.277	-0.010	-0.020	1.3E-02	3.5E-10	3.7E-02	4.2E-09
rs3824780	10	105384826	*SH3PXD2A*	T	C	0.356	0.009	0.023	2.6E-04	1.2E-10	8.0E-02	7.3E-12
rs117118217	10	131783328	*EBF3*	C	G	0.018	0.033	0.057	8.6E-02	2.9E-09	8.6E-03	4.3E-05
rs11824377	11	11787253	*MIR8070*	A	G	0.441	0.009	0.013	2.0E-01	4.8E-10	2.8E-04	2.3E-05
rs900144	11	13294268	*ARNTL*	T	C	0.568	0.010	0.014	2.7E-01	2.3E-12	1.0E-04	6.8E-06
rs72632979	11	16615883	*SOX6*	A	G	0.829	-0.013	-0.021	8.2E-02	1.3E-11	3.1E-02	1.3E-06
rs610476	11	86184102	*ME3*	A	G	0.41	-0.009	-0.014	1.6E-01	1.9E-09	2.1E-03	2.0E-05
rs1064939	11	118396331	*LOC101929089*	A	T	0.977	0.034	0.072	2.0E-03	1.6E-12	3.3E-02	2.7E-10
rs11218510	11	121922587	*MIR100HG*	A	G	0.401	-0.009	-0.016	7.5E-02	2.3E-10	9.4E-03	4.0E-06
rs10893502	11	126254018	*ST3GAL4*	T	C	0.408	0.008	0.021	4.2E-04	1.0E-08	3.0E-01	3.4E-10
rs7936928	11	130279168	*ADAMTS8*	T	C	0.397	0.009	0.018	1.1E-02	1.1E-09	4.4E-02	8.7E-09
rs145878042	12	48143315	*RAPGEF3*	A	G	0.989	-0.055	-0.121	6.3E-05	9.8E-18	1.1E-02	3.0E-16
12:51589293_TA_T	12	51589293	*POU6F1*	D	I	0.784	-0.008	-0.019	1.1E-02	3.2E-06	2.5E-01	2.9E-06
rs730560	12	57938565	*DCTN2*	A	G	0.582	0.010	0.015	2.4E-01	1.9E-12	3.8E-04	1.5E-05
rs1154752	12	68089328	*LINC02421*	T	C	0.606	-0.009	-0.015	8.2E-02	2.2E-09	3.7E-02	2.8E-06
rs490872	12	69645864	*CPSF6*	A	G	0.274	0.011	0.017	1.4E-01	2.5E-12	2.9E-03	3.3E-07
13:31010888_CAA_C	13	31010888	*LINC01058*	I	D	0.356	-0.011	-0.015	3.9E-01	1.2E-13	4.9E-04	3.9E-05
rs9560114	13	112187882	*LINC02337*	A	T	0.735	0.012	0.021	2.4E-02	2.3E-13	2.8E-04	1.1E-08
rs10142359	14	73884540	*NUMB*	A	G	0.52	-0.009	-0.022	1.3E-04	1.5E-10	2.2E-01	1.7E-12
15:73049542_CT_C	15	73049542	*ADPGK*	I	D	0.345	-0.010	-0.015	2.2E-01	7.3E-10	2.5E-03	2.8E-05
rs35032603	15	74753425	*UBL7*	T	C	0.963	-0.022	-0.053	1.5E-03	1.2E-09	1.6E-01	1.0E-09
rs2715423	15	99511873	*PGPEP1L*	A	G	0.285	-0.012	-0.021	1.1E-02	6.5E-15	3.4E-02	7.5E-10
rs3794701	16	3730578	*TRAP1*	A	G	0.48	-0.010	-0.019	1.4E-02	3.3E-13	1.1E-03	1.1E-08
rs8050390	16	31539386	*AHSP*	A	G	0.36	0.010	0.014	3.9E-01	6.0E-13	1.7E-04	2.8E-05
16:67554712_GA_G	16	67554712	*LOC100505942*	D	I	0.078	0.019	0.036	7.0E-03	2.8E-12	2.1E-02	4.1E-09
rs5011579	16	69187318	*UTP4*	C	G	0.285	-0.011	-0.022	4.9E-03	7.7E-13	3.6E-02	2.9E-11
rs2303232	16	84808969	*USP10*	C	G	0.488	0.010	0.020	7.4E-03	3.0E-11	8.0E-03	3.9E-10
rs57828851	17	7732351	*DNAH2*	A	T	0.581	-0.009	-0.022	5.4E-04	3.6E-11	2.8E-02	1.1E-11
rs2605136	17	18136750	*LLGL1*	T	C	0.355	0.009	0.020	2.7E-03	7.1E-10	6.8E-02	3.5E-09
rs113000401	17	27530458	*MYO18A*	T	C	0.049	-0.020	-0.043	5.3E-03	1.1E-09	7.8E-02	1.0E-08
rs138464472	17	44130160	*KANSL1*	D	I	0.286	0.015	0.030	1.5E-04	2.5E-20	5.4E-04	4.6E-15
rs376610477	17	46292292	*SKAP1*	D	I	0.649	0.015	0.022	7.1E-02	2.1E-19	2.0E-04	1.0E-08
rs7218677	17	76146520	*C17orf99*	A	T	0.688	0.009	0.014	2.5E-01	2.7E-09	2.1E-02	3.6E-05
rs12941365	17	76897733	*CEP295NL-TIMP2*	T	G	0.412	0.010	0.017	3.5E-02	1.2E-11	5.9E-03	1.3E-07

rsid	chr	Position	Nearest gene	EA	OA	EAF	beta FFMI observed	Beta FFMI expected	p diff	p FFMI	p FFM	p hght
rs4625783	17	80086395	*CCDC57*	T	C	0.407	0.010	0.024	1.3E-04	4.9E-13	3.0E-01	7.4E-12
rs5822974	18	7548483	*PTPRM*	D	I	0.812	0.011	0.026	1.6E-03	7.3E-10	5.8E-02	1.7E-09
19:2230029_CTT_C	19	2230029	*DOT1L*	I	D	0.942	-0.022	-0.032	1.7E-01	9.7E-13	9.1E-04	4.7E-07
rs75270598	19	19864849	*LINC00663*	T	C	0.154	-0.013	-0.021	1.0E-01	7.6E-12	4.3E-03	1.1E-05
rs2425847	20	44907927	*CDH22*	A	G	0.591	-0.009	-0.020	1.5E-03	2.5E-10	4.7E-02	8.1E-09
rs6096548	20	50347558	*ATP9A*	A	T	0.257	0.009	0.020	6.8E-03	4.1E-09	2.9E-02	6.1E-07
rs150998792	20	62130403	*EEF1A2*	T	G	0.05	-0.020	-0.041	1.5E-02	9.8E-10	2.7E-02	2.5E-06
rs7264802	20	62692440	*TCEA2*	A	G	0.252	-0.012	-0.019	9.6E-02	2.3E-13	3.9E-03	4.3E-07
rs35359276	22	20769186	*ZNF74*	I	D	0.293	0.009	0.020	1.2E-02	3.7E-09	2.1E-02	2.3E-07

Supplementary Table 3.8: Low-frequency and rare coding variants reaching genome-wide significance for BF%

MarkerName	Chr	Pos	EA	OA	Annotation	*Gene*	EAF	MAF	Beta	SE	*P* value	N	Analysis
Sex-combined analysis													
rs9429157	1	45,808,863	A	G	nonsynonymous	*TOE1*	0.041	0.041	-0.02	0.004	4.08E-10	550,180	Exome-Sex-combined
rs148330006	1	86,048,526	C	G	nonsynonymous	*CYR61*	0.992	0.008	-0.05	0.008	4.06E-12	511,386	Exome-Sex-combined
rs56228576	1	150,530,548	C	G	missense_variant	*ADAMTSL4*	0.969	0.031	0.03	0.004	1.06E-14	442,278	GWAS-Sex-combined
rs72704117	1	155,175,089	T	C	nonsynonymous	*THBS3*	0.023	0.023	-0.03	0.005	1.82E-12	550,180	Exome-Sex-combined
rs41314549	1	155,290,231	T	C	nonsynonymous	*FDPS*	0.972	0.028	0.03	0.004	5.23E-10	550,180	Exome-Sex-combined
rs61741479	1	173,915,909	C	G	nonsynonymous	*RC3H1*	0.031	0.031	-0.03	0.004	9.21E-11	528,666	Exome-Sex-combined
rs34246968	1	177,929,501	C	G	missense_variant	*SEC16B*	0.969	0.031	-0.03	0.004	2.02E-09	442,278	GWAS-Sex-combined
rs1229984	4	100,239,319	T	C	nonsynonymous	*ADH1B*	0.025	0.025	-0.03	0.005	4.59E-11	472,606	Exome-Sex-combined
rs61749613	5	82,815,170	A	G	nonsynonymous	*VCAN*	0.959	0.041	0.03	0.004	5.45E-17	550,180	Exome-Sex-combined
rs114285050	5	145,895,394	A	G	stopgain	*GPR151*	0.008	0.008	-0.05	0.008	5.52E-10	550,180	Exome-Sex-combined
rs11538263	6	31,601,735	A	G	nonsynonymous	*PRRC2A*	0.044	0.044	-0.02	0.004	1.04E-11	529,263	Exome-Sex-combined
rs2228265	6	35,253,974	T	C	nonsynonymous	*ZNF76*	0.021	0.021	0.03	0.005	2.62E-09	544,113	Exome-Sex-combined
rs33959228	6	35,259,397	T	C	nonsynonymous	*ZNF76*	0.022	0.022	0.04	0.005	4.52E-14	550,180	Exome-Sex-combined
rs3087653	9	15,459,821	A	C	nonsynonymous	*SNAPC3*	0.952	0.048	0.02	0.003	8.00E-14	518,951	Exome-Sex-combined
rs150090666	11	14,865,399	T	C	stop_gained	*PDE3B*	0.001	0.001	0.15	0.025	2.56E-09	442,278	GWAS-Sex-combined
rs3730071	12	49,168,798	A	C	nonsynonymous	*ADCY6*	0.030	0.030	-0.02	0.004	2.99E-09	548,316	Exome-Sex-combined
rs11551274	12	49,224,108	C	G	missense_variant	*DDX23*	0.031	0.031	-0.03	0.004	3.44E-09	442,278	GWAS-Sex-combined
rs61754230	12	72,179,446	T	C	nonsynonymous	*RAB21*	0.020	0.020	0.03	0.005	1.93E-10	550,180	Exome-Sex-combined
rs145350287	12	120,907,309	A	T	nonsynonymous	*SRSF9*	0.040	0.040	-0.03	0.004	2.16E-17	513,150	Exome-Sex-combined
rs527456159	12	122,685,176	A	G	missense_variant	*LRRC43*	0.044	0.044	-0.02	0.004	1.82E-10	442,278	GWAS-Sex-combined
rs72681869	14	50,655,357	C	G	nonsynonymous	*SOS2*	0.011	0.011	-0.06	0.007	1.29E-15	523,167	Exome-Sex-combined
rs72683923	14	50,735,947	T	C	synonymous_variant	*L2HGDH*	0.980	0.020	0.04	0.005	5.10E-11	442,278	GWAS-Sex-combined
rs12595158	15	62,316,035	T	C	missense_variant	*VPS13C*	0.025	0.025	-0.03	0.005	3.81E-09	525,103	GWAS-Sex-combined
rs117133016	16	2,816,627	C	G	missense_variant	*SRRM2*	0.010	0.010	0.05	0.007	1.38E-11	442,278	GWAS-Sex-combined

rs78457529	16	24,950,880	T	C	nonsynonymous	*ARHGAP17*	0.012	0.012	0.04	0.006	1.49E-09	550,180	Exome-Sex-combined
rs2229616	18	58,039,276	T	C	missense_variant	*MC4R*	0.020	0.020	-0.07	0.005	1.02E-40	522,124	GWAS-Sex-combined
rs62621197	19	8,670,147	T	C	missense_variant	*ADAMTS10*	0.037	0.037	-0.02	0.004	1.67E-09	442,278	GWAS-Sex-combined
rs112693590	19	46,274,553	A	G	missense_variant	*DMPK*	0.049	0.049	-0.02	0.004	9.23E-13	463,359	GWAS-Sex-combined
rs6050446	20	25,195,509	A	G	nonsynonymous	*ENTPD6*	0.033	0.033	-0.03	0.004	3.60E-13	547,742	Exome-Sex-combined
rs41278126	20	40,079,655	C	G	nonsynonymous	*CHD6*	0.044	0.044	0.02	0.004	3.64E-09	544,084	Exome-Sex-combined
Women analysis													
rs74580294	12	122,622,795	A	G	nonsynonymous	*MLXIP*	0.950	0.050	0.04	0.006	5.42E-11	266,853	Exome-Women

Supplementary Table 3.9: Low-frequency and rare coding variants reaching genome-wide significance for FFMI

rsid	MarkerName	EA	OA	EAF	Gene	Type	Beta	SE	p-value	N
Sex-combined										
rs61744853	chr1:35453651	a	c	0.991	*ZMYM6*	missense	-0.044	0.007	4.1E-10	535,885
rs75770915	chr1:39799912	t	c	0.046	*MACF1*	missense	-0.025	0.003	2.2E-15	535,885
rs1746842	chr1:39878815	a	g	0.955	*KIAA0754*	missense	0.026	0.003	1.1E-15	514,980
rs141845046	chr1:154987704	t	c	0.023	*ZBTB7B*	missense	0.034	0.005	8.8E-14	535,885
rs11465205	chr1:155160941	t	c	0.031	*MUC1*	missense	0.026	0.004	4.0E-11	517,010
rs72704117	chr1:155175089	t	c	0.023	*THBS3*	missense	0.050	0.005	8.9E-29	535,885
rs41314549	chr1:155290231	t	c	0.972	*FDPS*	missense	-0.040	0.004	3.8E-23	535,885
rs76381440	chr1:156146546	t	c	0.021	*SEMA4A*	missense	0.040	0.005	1.0E-17	535,885
rs34246968	chr1:177929501	c	g	0.969	*SEC16B*	missense	-0.031	0.004	1.4E-14	442,391
rs2293072	chr2:220045454	a	g	0.039	*FAM134A*	missense	-0.025	0.004	4.9E-12	442,391
rs41291734	chr3:50513613	t	c	0.035	*CACNA2D2*	missense	0.024	0.004	2.5E-11	535,885
rs11722554	chr4:5016883	a	g	0.038	*CYTL1*	missense	0.026	0.004	1.4E-13	522,343
rs141374503	chr4:73179445	t	c	0.004	*ADAMTS3*	missense	0.075	0.011	2.6E-12	535,885
rs61749613	chr5:82815170	a	g	0.959	*VCAN*	missense	-0.038	0.003	8.4E-30	535,885
rs78727187	chr5:127668685	t	g	0.006	*FBN2*	missense	-0.057	0.009	5.8E-11	535,885
rs41290587	chr5:151045923	t	c	0.007	*SPARC*	missense	0.051	0.008	3.0E-10	535,885
rs78247455	chr5:176722005	a	g	0.026	*NSD1*	missense	0.033	0.004	7.7E-15	535,885
rs34672415	chr6:34839644	a	g	0.016	*UHRF1BP1*	missense	0.047	0.005	2.4E-18	535,885
rs7761870	chr6:35423886	t	c	0.016	*FANCE*	missense	0.045	0.005	2.8E-17	535,885
rs2766597	chr6:35765043	a	g	0.986	*CLPS*	missense	-0.036	0.006	1.6E-09	531,750
rs41271629	chr6:86257229	t	g	0.971	*SNX14*	missense	0.025	0.004	3.0E-10	512,183
rs62621812	chr7:127015083	a	g	0.022	*ZNF800*	missense	0.046	0.005	5.5E-22	535,885
rs34804482	chr11:27389739	t	c	0.974	*LGR4*	missense	0.035	0.004	1.7E-16	535,885
rs145878042	chr12:48143315	a	g	0.989	*RAPGEF3*	missense	-0.055	0.006	9.8E-18	534,021
rs1126930	chr12:49399132	c	g	0.036	*PRKAG1*	missense	0.031	0.004	4.7E-18	529,789
rs78607331	chr12:57648644	t	c	0.045	*R3HDM2*	missense	0.020	0.003	7.5E-10	495,549
rs28929474	chr14:94844947	t	c	0.020	*SERPINA1*	missense	-0.042	0.005	3.9E-18	535,885
rs3784635	chr15:62254989	t	c	0.976	*VPS13C*	missense	0.027	0.004	9.0E-10	532,966
rs12595158	chr15:62316035	t	c	0.021	*VPS13C*	missense	-0.029	0.005	7.5E-10	535,885
rs28559926	chr15:89400043	c	g	0.039	*ACAN*	missense	0.031	0.004	9.8E-17	442,391
rs28407189	chr15:89400680	a	g	0.971	*ACAN*	missense	-0.034	0.004	7.4E-18	535,885
rs78457529	chr16:24950880	t	c	0.012	*ARHGAP17*	missense	-0.046	0.006	3.3E-14	535,885
rs16957289	chr16:67325711	t	c	0.043	*KCTD19*	missense	0.021	0.003	5.8E-10	514,980
rs9922085	chr16:67397580	c	g	0.044	*LRRC36*	missense	0.020	0.003	1.1E-09	529,789
rs8052655	chr16:67409180	a	g	0.044	*LRRC36*	missense	0.020	0.003	1.4E-09	530,616
rs16957415	chr16:67418957	a	g	0.958	*LRRC36*	missense	-0.020	0.003	1.3E-09	535,885
rs5030980	chr16:67516945	t	c	0.042	*AGRP*	missense	0.021	0.003	4.3E-10	535,885
rs79051270	chr16:69153251	t	g	0.040	*CHTF8*	missense	-0.021	0.004	1.7E-09	442,391
rs149615348	chr16:84900645	a	g	0.007	*CRISPLD2*	missense	0.049	0.008	5.6E-10	535,885

rs5036	chr17:42338945	t	c	0.978	*SLC4A1*	missense	-0.030	0.005	1.0E-10	522,343
rs2229616	chr18:58039276	t	c	0.020	*MC4R*	missense	-0.065	0.005	3.1E-39	442,391
rs62621197	chr19:8670147	t	c	0.037	*ADAMTS10*	missense	0.041	0.004	5.7E-27	442,391
rs35384424	chr19:46145661	t	c	0.044	*C19orf83*	missense	-0.021	0.004	4.3E-09	442,391
Men-only										
rs77169818	chr18:74980601	a	t	0.954	*GALR1*	missense	0.042	0.007	1.7E-10	235,388

Chapter 4

Supplementary Table 4.1: Loci reaching genome-wide significance in meta-GWAS for unadjusted fat and lean mass compartments

Locus	Chr:pos	Sentinel variant	EA	OA	EAF	Nearest gene	Phenotype	Beta±SE	p-value	hetpval	N	Lead variant for trait (if other than sentinel variant)
1	1:11125729	rs2039841	t	c	0.30	*EXOSC10*	Leg fat	0.0742±0.0131	1.5E-08	5.7E-01	14,736	
2	1:41544279	rs2885697	g	t	0.33	*SCMH1*	Lean arm	0.0534±0.0094	1.3E-08	2.1E-01	25,429	
3	1:103549862	rs4391683	a	g	0.59	*COL11A1*	Android lean	-0.0506±0.0091	2.2E-08	5.6E-01	25,419	
							Trunk lean	-0.0509±0.009	1.8E-08	9.5E-01	25,425	
4	2:610603	rs2867131	t	c	0.17	*TMEM18*	Android lean	-0.0675±0.0118	1.2E-08	2.1E-03	25,419	rs2683992
							Trunk lean	-0.0676±0.0118	1.1E-08	2.0E-03	25,425	rs2683992
							Appendicular lean	-0.0746±0.0118	2.7E-10	3.3E-05	25,428	
							Leg lean	-0.0781±0.0118	3.8E-11	1.0E-05	25,426	
							Total lean	-0.0745±0.0118	3.0E-10	1.5E-04	25,427	
							Gynoid lean	-0.0725±0.0127	1.2E-08	5.5E-05	25,418	rs7563362
5	2:165543199	rs3769869	g	a	0.20	*COBLL1*	Leg fat	0.0606±0.0111	4.8E-08	4.7E-02	25,426	
6	2:227637763	rs4675093	a	t	0.93	*IRS1*	Android lean	0.0958±0.0175	4.3E-08	4.6E-01	25,419	rs13401886
							Total lean	0.0955±0.0175	4.5E-08	4.2E-01	25,427	rs13401886
							Trunk lean	0.1002±0.0175	9.6E-09	8.8E-01	14,736	rs13401886
							Gynoid lean	0.0961±0.0175	3.9E-08	5.3E-01	25,418	
7	3:12365308	rs13083375	t	g	0.12	*PPARG*	Arm fat	0.0864±0.0138	3.6E-10	1.5E-01	25,424	
8	3:65024223	rs371476659	c	g	0.98	*ADAMTS9-AS2*	Visceral fat	-0.2729±0.0496	3.8E-08	2.5E-01	14,736	
9	3:129131263	rs4688812	a	g	0.99	*EFCAB12*	Trunk lean	0.355±0.0639	2.8E-08	6.3E-03	14,736	
10	3:172168507	rs519384	t	a	0.71	*GHSR*	Appendicular lean	-0.06±0.0098	9.6E-10	6.0E-01	25,428	
							Leg lean	-0.0616±0.0098	3.6E-10	6.3E-01	25,426	
							Total lean	-0.0658±0.0098	2.1E-11	4.5E-01	25,427	
							Trunk lean	-0.0645±0.0098	5.2E-11	2.6E-01	25,425	
							Gynoid lean	-0.0592±0.0098	1.8E-09	6.2E-01	25,418	rs12638147
							Android lean	-0.0631±0.0098	1.4E-10	2.6E-01	25,419	rs6765792
11	4:17998426	rs2061456	c	a	0.74	*LCORL*	Android lean	-0.0657±0.0101	8.8E-11	8.2E-01	25,419	
							Appendicular lean	-0.0583±0.0101	8.1E-09	1.9E-01	25,428	

Locus	Chr:pos	Sentinel variant	EA	OA	EAF	Nearest gene	Phenotype	Beta±SE	p-value	hetpval	N	Lead variant for trait (if other than sentinel variant)
							Gynoid lean	-0.0625±0.0101	6.8E-10	5.7E-01	25,418	
							Leg lean	-0.0593±0.0101	4.5E-09	2.7E-01	25,426	
							Total lean	-0.0702±0.0101	3.9E-12	3.6E-01	25,427	
							Trunk lean	-0.0762±0.0101	4.8E-14	7.1E-01	25,425	
12	4:56683846	rs7679852	t	c	0.66	*LOC644145*	Appendicular lean	0.0514±0.0093	3.6E-08	2.1E-02	25,428	
							Lean arm	0.048±0.0093	2.8E-07	8.0E-03	25,429	rs11722734
13	4:89737457	rs7694958	t	c	0.47	*FAM13A*	Leg fat	-0.0499±0.0089	2.0E-08	5.1E-01	25,426	
14	4:146164089	rs12504656	t	g	0.33	*OTUD4*	Gynoid lean	0.0542±0.0096	1.5E-08	7.6E-01	25,418	
15	6:26233387	rs10946808	a	g	0.72	*HIST1H1D*	Total lean	0.0561±0.0098	1.1E-08	8.9E-01	25,427	
							Trunk lean	0.0563±0.0098	9.5E-09	6.9E-01	25,425	
16	6:27884127	rs148298110	a	t	0.93	*OR2B2*	Leg fat	0.1029±0.0175	4.3E-09	9.2E-01	25,426	
							Peripheral fat	0.1001±0.0175	1.1E-08	9.2E-01	25,424	
17	6:31270735	rs6908171	c	a	0.41	*HLA-C*	Gynoid lean	0.0716±0.0119	1.7E-09	6.2E-01	14,736	
							Leg lean	0.0656±0.0119	3.1E-08	6.4E-01	14,736	
18	6:34161299	rs56348792	t	c	0.13	*GRM4*	Android lean	0.0763±0.0132	7.5E-09	4.2E-01	25,419	
							Appendicular lean	0.0774±0.0132	4.4E-09	8.2E-01	25,428	
							Leg lean	0.0762±0.0132	7.5E-09	7.1E-01	25,426	
							Total lean	0.0829±0.0132	3.2E-10	8.3E-01	25,427	
							Trunk lean	0.0787±0.0132	2.5E-09	6.9E-01	25,425	rs10807137
19	6:43757896	rs998584	c	a	0.52	*VEGFA*	Leg fat	0.0494±0.009	3.4E-08	4.1E-02	25,426	
20	6:108945370	rs2022464	c	a	0.71	*FOXO3*	Gynoid lean	0.0537±0.0097	3.6E-08	7.8E-01	25,418	rs2764264
							Android lean	0.0608±0.0097	4.4E-10	1.3E-01	25,419	
							Leg lean	0.0537±0.0097	3.3E-08	6.6E-01	25,426	
							Total lean	0.0577±0.0097	3.0E-09	4.6E-01	25,427	
							Trunk lean	0.0547±0.0097	2.0E-08	2.1E-01	25,425	
21	6:127036174	rs853965	t	c	0.50	*MIR588*	Appendicular lean	-0.0511±0.0089	8.1E-09	2.1E-01	25,428	rs6569469
							Gynoid lean	-0.0554±0.0089	4.4E-10	2.7E-01	25,418	
							Leg lean	-0.0514±0.0089	6.8E-09	2.1E-01	25,426	
							Total lean	-0.0593±0.0089	2.1E-11	1.0E-01	25,427	
							Trunk lean	-0.0619±0.0089	2.9E-12	8.4E-02	25,425	
							Android lean	0.0603±0.0089	1.1E-11	2.8E-01	25,419	rs853971

Locus	Chr:pos	Sentinel variant	EA	OA	EAF	Nearest gene	Phenotype	Beta±SE	p-value	hetpval	N	Lead variant for trait (if other than sentinel variant)
							Gynoid fat	0.1191±0.0205	6.6E-09	9.1E-02	25,417	rs577721086
	6:127454893	rs72959041	g	a	0.95	*RSPO3*	Leg fat	0.139±0.0205	1.2E-11	2.3E-01	25,426	
							Peripheral fat	0.1242±0.0205	1.4E-09	1.5E-01	25,424	
							Gynoid lean	0.0582±0.0101	7.4E-09	4.4E-01	25,418	rs6926186
							Total lean	0.0687±0.0101	8.4E-12	4.8E-01	25,427	rs6926186
							Lean arm	0.0722±0.0101	7.2E-13	5.1E-01	25,429	rs7744830
22	6:130374461	rs7740107	t	a	0.26	*L3MBTL3*	Android lean	0.0601±0.0101	2.4E-09	6.4E-01	25,419	
							Trunk lean	0.0611±0.0101	1.3E-09	6.9E-01	25,425	
							Appendicular lean	0.0675±0.0101	1.8E-11	3.8E-01	25,428	rs12661188
							Leg lean	0.0624±0.0101	5.4E-10	3.6E-01	25,426	rs9402214
							Leg lean	0.0626±0.0109	1.0E-08	6.3E-01	25,426	rs2190507
							Android lean	0.0695±0.011	2.2E-10	4.7E-02	25,419	
							Total lean	0.0652±0.0109	2.5E-09	4.0E-01	25,427	
23	7:92236829	rs4272	g	a	0.21	*CDK6*	Trunk lean	0.0696±0.0109	2.0E-10	1.3E-01	25,425	
							Appendicular lean	0.0613±0.0109	2.0E-08	6.4E-01	25,428	rs42038
							Gynoid lean	0.071±0.011	9.2E-11	6.0E-01	25,418	rs42038
							Total lean	0.0544±0.0091	2.7E-09	3.1E-01	25,427	rs6951150
24	7:150681914	rs6951150	c	t	0.63	*NOS3*	Appendicular lean	0.0583±0.0091	1.7E-10	4.8E-01	25,428	
							Leg lean	0.0606±0.0091	3.4E-11	4.7E-01	25,426	
25	8:57259146	rs78329725	a	g	0.74	*SDR16C5*	Trunk lean	-0.058±0.0102	1.1E-08	3.3E-01	25,425	
							Appendicular lean	0.0543±0.0091	2.9E-09	3.1E-01	25,428	rs1466190
26	8:78086732	rs1377246	g	c	0.62	*PEX2*	Leg lean	0.0549±0.0091	1.8E-09	3.0E-01	25,426	
							Total lean	0.0544±0.0091	2.6E-09	3.8E-01	25,427	
27	9:103385293	rs200914342	a	g	0.18	*CAVIN4*	Trunk lean	-0.1062±0.0188	1.7E-08	1.9E-01	14,736	
							Gynoid fat	0.0555±0.0101	4.5E-08	2.9E-01	25,417	
28	10:118613931	rs9421247	t	c	0.26	*ENO4*	Leg fat	0.0581±0.0101	9.7E-09	1.3E-01	25,426	
							Peripheral fat	0.0582±0.0101	9.3E-09	2.2E-01	25,424	
							Appendicular lean	0.0608±0.0108	2.0E-08	7.9E-01	25,428	
							Leg lean	0.0608±0.0108	2.1E-08	6.7E-01	25,426	
29	11:27703188	rs6484320	a	t	0.79	*BDNF-AS*	Total lean	0.0624±0.0108	8.5E-09	6.6E-01	25,427	
							Gynoid lean	0.0565±0.0109	1.9E-07	3.4E-01	25,418	rs12273363
30	12:4384844	rs76895963	g	t	0.02	*CCND2-AS1*	Android lean	0.2324±0.0348	2.3E-11	5.1E-01	25,419	

Locus	Chr:pos	Sentinel variant	EA	OA	EAF	Nearest gene	Phenotype	Beta±SE	p-value	hetpval	N	Lead variant for trait (if other than sentinel variant)
							Appendicular lean	0.2396±0.0347	5.1E-12	1.6E-01	25,428	
							Lean arm	0.199±0.0347	1.0E-08	2.2E-01	25,429	
							Gynoid lean	0.2607±0.0348	6.3E-14	2.2E-01	25,418	
							Leg lean	0.2412±0.0347	3.7E-12	1.9E-01	25,426	
							Total lean	0.2521±0.0347	3.6E-13	1.4E-01	25,427	
							Trunk lean	0.2347±0.0347	1.4E-11	1.6E-01	25,425	
31	12:46707618	rs1586905	a	t	0.61	*SLC38A2*	Trunk lean	0.0665±0.012	3.3E-08	1.5E-01	14,736	
							Gynoid lean	0.0677±0.0121	2.3E-08	5.9E-01	14,736	
32	12:50261809	rs3205718	t	c	0.38	*FAIM2*	Leg lean	0.0553±0.0093	2.3E-09	1.4E-01	25,426	
33	12:104001177	rs11111676	a	t	0.86	*STAB2*	Lean arm	0.1177±0.0214	3.7E-08	1.1E-01	14,736	
							Total lean	0.1192±0.0214	2.4E-08	2.3E-02	14,736	
34	12:124409502	rs7133378	a	g	0.33	*DNAH10*	Leg fat	0.0607±0.0095	1.4E-10	6.3E-02	25,426	
35	13:51088356	rs1638703	g	c	0.74	*DLEU1*	Appendicular lean	0.0674±0.0102	3.4E-11	8.5E-01	25,428	
							Leg lean	0.07±0.0102	5.5E-12	8.3E-01	25,426	
							Total lean	0.0597±0.0102	4.2E-09	8.9E-01	25,427	
36	16:29923510	rs4548895	a	g	0.61	*KCTD13*	Appendicular lean	0.0531±0.0092	8.1E-09	6.9E-01	25,428	
							Lean arm	0.051±0.0092	3.1E-08	2.0E-01	25,429	
							Leg lean	0.0513±0.0092	2.5E-08	8.0E-01	25,426	rs9921544
37	16:53845487	rs11642841	a	c	0.40	*FTO*	Subcutaneous android fat	0.0619±0.009	7.1E-12	2.5E-01	25,351	rs56094641
							Visceral fat	0.0599±0.009	3.4E-11	6.5E-02	25,347	rs56094641
							Arm fat	0.0676±0.009	6.1E-14	5.8E-01	25,424	rs55872725
							Total fat	0.0681±0.009	4.1E-14	4.8E-01	25,429	rs55872725
							Leg lean	0.0582±0.009	1.0E-10	4.6E-01	25,426	rs55872725
							Fat and	0.0669±0.009	1.2E-13	2.0E-01	25,423	rs62033400
							Peripheral fat	0.0586±0.009	8.1E-11	7.1E-01	25,424	rs62033400
							Trunk fat	0.069±0.009	1.9E-14	2.4E-01	25,430	rs62033400
							Gynoid fat	0.0615±0.009	9.8E-12	6.3E-01	25,417	
							Leg fat	0.0535±0.009	3.0E-09	6.6E-01	25,426	
							Appendicular lean	0.0573±0.009	1.9E-10	4.0E-01	25,428	
							Total lean	0.0541±0.009	1.8E-09	2.9E-01	25,427	

Locus	Chr:pos	Sentinel variant	EA	OA	EAF	Nearest gene	Phenotype	Beta±SE	p-value	hetpval	N	Lead variant for trait (if other than sentinel variant)
38	17:29211667	rs7223535	g	a	0.73	*ATAD5*	Appendicular lean	0.081±0.0099	3.6E-16	1.8E-01	25,428	
							Lean arm	0.0674±0.0099	1.2E-11	1.2E-01	25,429	
							Gynoid lean	0.0612±0.01	7.9E-10	1.7E-01	25,418	
							Leg lean	0.0814±0.0099	2.5E-16	1.7E-01	25,426	
							Total lean	0.076±0.0099	2.0E-14	2.1E-01	25,427	
							Android lean	0.0585±0.01	5.0E-09	7.7E-01	25,419	rs113934718
							Trunk lean	0.0636±0.01	2.0E-10	4.8E-01	25,425	rs113934718
39	17:48628160	rs62621401	a	g	0.98	*SPATA20*	Arm fat	0.2224±0.0364	1.0E-09	5.4E-01	25,424	
40	17:61998879	rs2040347	a	g	0.33	*GH1*	Appendicular lean	-0.0522±0.0094	2.9E-08	4.3E-01	25,428	rs2854152
							Leg lean	-0.0528±0.0094	2.0E-08	4.7E-01	25,426	rs2854152
							Android lean	-0.0538±0.0094	1.2E-08	5.5E-01	25,419	
							Total lean	-0.0597±0.0094	2.2E-10	5.6E-01	25,427	
							Trunk lean	-0.0615±0.0094	6.4E-11	6.9E-01	25,425	
							Gynoid lean	-0.0599±0.0094	2.1E-10	6.7E-01	25,418	rs1051684
41	18:57913965	rs35693910	a	g	0.27	*MC4R*	Leg lean	0.072±0.0101	1.1E-12	5.4E-01	25,426	rs538656
							Android lean	0.0648±0.0101	1.6E-10	5.2E-01	25,419	
							Appendicular lean	0.0747±0.0101	1.5E-13	4.9E-01	25,428	
							Lean arm	0.0694±0.0101	7.2E-12	5.6E-01	25,429	
							Gynoid lean	0.0678±0.0101	2.1E-11	4.4E-01	25,418	
							Total lean	0.0763±0.0101	4.4E-14	4.3E-01	25,427	
							Trunk lean	0.0729±0.0101	5.7E-13	2.7E-01	25,425	
42	19:437256	rs4605291	t	c	0.99	*SHC2*	Trunk lean	0.2309±0.0406	1.3E-08	3.7E-07	25,425	
43	19:49650319	rs6509418	g	a	0.60	*PPFLA3*	Lean arm	0.0512±0.0091	2.1E-08	4.8E-01	25,429	
							Total lean	0.0498±0.0091	4.8E-08	4.3E-01	25,427	
44	20:32263624	rs3213180	g	c	0.90	*E2F1*	Trunk lean	0.0873±0.0149	5.0E-09	5.6E-01	25,425	
45	20:34025756	rs143384	a	g	0.59	*GDF5*	Android lean	-0.0531±0.009	4.2E-09	6.1E-01	25,419	
							Appendicular lean	-0.0664±0.009	1.7E-13	4.6E-01	25,428	
							Gynoid lean	-0.0738±0.009	2.8E-16	3.6E-01	25,418	
							Leg lean	-0.0704±0.009	5.5E-15	3.3E-01	25,426	
							Total lean	-0.0607±0.009	1.7E-11	5.8E-01	25,427	
							Trunk lean	-0.0497±0.009	3.5E-08	7.9E-01	25,425	

Analysis plans

Analysis plan 1: Exome chip analyses for BF% and FFMI

This protocol was modified from the GIANTexome Analysis Plan – 22/April/2014 – prepared by Ruth Loos and her colleagues for the GIANT Consortium ExomeChip effort.

1. Introduction

AIM: This analysis plan aims to coordinate collection of summary statistics data of ExomeChip analyses for body fat percentage and related traits:

- Fat percentage (FatPCT) in %
- Fat free mass-index (FFMi) in kg per m^2
- Circulating leptin levels in ng/ml
- Plasma adiponectin levels in μg/ml

Note: Fat percentage can be measured via dual-energy x-ray absorptiometry (DEXA) and with a portable bioimpedance monitor only. If you are using DEXA, we would like you to use the "Lean mass" variable (rather than the "fat free mass" variable, as this includes bone mineral content).
Many of you have run single variant analyses before. For this effort, we will use either RAREMETALWORKER or RvTests because:

1. We can perform multi-variant tests such as gene burden tests centrally using summary stats you will generate in your individual studies.
2. We can perform conditional analyses centrally because the software generates LD matrices for all variants within 1Mb windows.
3. We can correct for population stratification and cryptic relatedness in the best possible way.

All of this can be done with just a few commands in RAREMETALWORKER or RvTests and this analysis plan describes how to do it. The only part that requires special attention is the generation of a *.vcf file containing your genotypes and the steps required to ensure all alleles are on the correct strand AND that you have labelled the correct alleles as reference and non-reference. These steps are described below.

Please contact us if you have any question:

Ruth Loos (ruth.loos@mssm.edu), Claudia Schurmann (claudia.schurmann@mssm.edu)

2. General methodology

Each individual study will perform data quality control (QC) and analysis and provide summary results for meta-analysis. Results files will be deposited to a central repository, where QC/data cleaning and meta-analysis will be performed within each trait working group.

OUTCOMES

- Fat percentage (FatPCT) in %
- Fat free mass-index (FFMi) in kg per m2
- Circulating leptin levels in ng/ml
- Circulating leptin levels in ng/ml adjusted for body mass index (BMI)
- Circulating leptin levels in ng/ml adjusted for FatPCT
- Plasma adiponectin levels in µg/ml
- Plasma adiponectin levels in µg/ml adjusted for body mass index (BMI)
- Plasma adiponectin levels in µg/ml adjusted for FatPCT

EXCLUSIONS

- Individuals < 18 years
- Pregnant women

STRATIFYING ANALYSES BY

- Race/ethnicity
- Disease status; i.e. analyze cases and controls separately, if the disease status influences the outcome.

ANALYSES TO PERFORM

- ALL individuals
- MEN only
- WOMEN only

NOTES

- With longitudinal data, please use an examination that maximizes sample size, restricting to adult subjects (≥ 18 years of age). All subjects included in analysis should utilize data obtained from the same examination.
- For studies with "unrelated" individuals (i.e. not family-based), we are specifically asking you to NOT exclude people who you find might be closely/cryptically related. Instead

family/cryptic relatedness will be handled by the genomic relationship matrix created and used in the RAREMETALWORKER and RvTests part of the protocol (described below).

IMPORTANT:

Throughout the text we refer to principal components (PCs), which are used to control for population stratification and other confounders. Here, we don't distinguish between PCs generated using software like smartPCA/EIGENSOFT or components calculated using multidimensional scaling (MDS) as implemented in PLINK. For your analyses, you should include either PCs or MDS components as covariates in the models. Principal components in studies with related individuals to be calculated in founders.

3. Preparing data for analysis – Files and software to download

A. FILES TO DOWNLOAD

Folder

/data/RELEVANT_FILES/CHR24_26/CHR24-26.txt

/data/RELEVANT_FILES/FORCE_ALLELES/FORCEALLELES_V1.0.gz

/data/RELEVANT_FILES/FILES_FOR_PC1_PC2_SCATTERPLOTS/Files_for_scatter_plots.rar

For studies from the CHARGE Consortium only (and only if variants have not yet been flipped to the "+" strand):

/data/RELEVANT_FILES/FORWARD_TO_POSTIVE_FLIPPING_LIST_CHARGE/SNP_list_to_be_flipped_KL_TW.txt

For studies that used the "Exome-chip QC SOP.v5.pdf"

/data/RELEVANT_FILES/UK_CHR_UPDATES/UK_chr_updates.txt

/data/RELEVANT_FILES/UK_CHR_UPDATES/UK_pos_updates.txt

/data/RELEVANT_FILES/FLIPPING_23SNPs /SNP_list_to_be_flipped_Exome-chip_QC_SOP_studies_only_23SNPs.txt

B. SOFTWARE TO BE DOWNLOADED

B.1. SOFTWARE TO PREPARE GENOTYPE FILES

B.1.1. Download PLINK

Download PLINK from

http://pngu.mgh.harvard.edu/~purcell/plink/download.shtml#download

B.1.2. Download PLINK/SEQ

Download PLINK/SEQ from https://atgu.mgh.harvard.edu/plinkseq/download.shtml

Download the LOCDB, REFDB and SEQDB files for the hg19 reference genome (http://atgu.mgh.harvard.edu/plinkseq/resources.shtml) and save them in a resources directory: /path/to/hg19.

The guidelines about how to use PLINK/SEQ can be found at https://atgu.mgh.harvard.edu/plinkseq/.

B.1.3. Download tabix

You can download tabix at http://sourceforge.net/projects/samtools/files/.

B.1.4. Download checkVCF

Download the necessary resource files (checkVCF-20140116.tar.gz) and scripts at https://github.com/zhanxw/checkVCF.

B.1.5. Download Python

Download Python from https://www.python.org/downloads/

B.2. ANALYSIS SOFTWARE

Analysts are free to choose which software (RvTests or RAREMETALWORKER) they use. The most notable differences to users are the interface, in particular the way covariates are organized. RvTests organizes covariates in a similar way to PLINK, where covariates are provided in a separate file. RAREMETALWORKER was derived from Merlin, so it supports inputs more similar to Merlin, in that covariates are stored in PED file, and you also need a DAT file to describe the covariates. Both tools have a very simple interface to use, and the input for one tool can be easily converted to the other.

THE SOFTWARE VERSIONS LISTED BELOW ARE THE LATEST RELEASES AS OF TODAY. HOWEVER, IF AT THE TIME OF ANALYSES A MORE RECENT VERSION IS AVAILABLE, PLEASE USE THAT ONE.

B.2.1. Download RvTests

The source codes (version 20141006) files can be found at https://github.com/zhanxw/rvtests/releases/. For binary executable files, download from https://github.com/zhanxw/rvtests/releases/download/v1.8.4/rvtests-20141006.tar.gz and then unpack the downloaded file:

tar xvzf rvtests-20140416.tar.gz

rvtests and vcf2kinship can be found under rvtests/executable/.

For source code, unpack the download zipped source codes or go to your cloned github repository, then type "make". Upon successful compilation, rvtests and vcf2kinship can be found under rvtests/executable/.

B.2.2. Download RAREMETALWORKER

To download the package with source code (version 4.13.5), please go to the following:
http://genome.sph.umich.edu/wiki/RAREMETAL_DOWNLOAD_%26_BUILD#Where_to_Download
Then use the following command to decompress:
tar xvzf raremetal.4.13.5.tgz
Finally, go to
./raremetal_4.13.5/raremetal/raremetalworker/src/
and use the following command to compile:
make all

4. Preparing data for analysis -- Genotypes

Please determine if data from your study has been called by the CHARGE Exome Chip centralized effort led by Megan Grove at the University of Texas Health Science Center at Houston, Texas for the HumanExome V1.0 BeadChips, for which the best practices and joint calling procedures will apply (Grove *et al.* PloSOne 8(7) :e68095).
For all other studies we will assume that your genotype data has been called with the Illumina GenomeStudio GENCALL and subsequently recalled using zCALL, applying the additional sample QC and variant QC as described in "Exome-chip QC SOP.v5.pdf".
Both Grove *et al.* PLoS ONE 2013.pdf and Exome-chip QC SOP.v5.pdf can be found on the SFTP site: /data/RELEVANT_FILES/QC/.

Please, report your calling and QC procedures together with ExomeChip version in the tracking form ExomeFatPercentage_STUDY_PHENO_DATE.xls

NOTE: You may have prepared your genotype files before for other analyses or for other consortia, but PLEASE DOUBLE-CHECK, whether the procedures followed before are consistent with the ones described below.

A. QUALITY CONTROL

I. VARIANT EXCLUSIONS

Variant exclusions have been applied by the central calling effort (Grove *et al.* PloSOne 8(7): e68095) and/or are described by "Exome-chip QC SOP.v5.pdf" procedures.

Please, remove all poorly genotyped and called variants that did not pass QC criteria, but DO NOT remove the monomorphic variants at this stage. [Monomorphic variants will be removed from the actual analyses at a later stage].

In addition, please exclude variants with:

- Allele frequency strongly associated with plate assignment
- Call discordance of overlapping SNPs with previous data, such as GWAS or custom chips

II. ONLY FOR STUDIES THAT USED THE EXOME-CHIP QC SOP.V5 PROTOCOL - UPDATE POSITIONS TO MATCH ILLUMINA MANIFEST

As part of the Exome-chip QC SOP.v5.pdf, many studies updated the chr and positions of some SNPs with uncertain mappings (N=91). To align the positions across all studies, we suggest that the Illumina manifest positions are used. The two commands below will update these SNPs.

```
plink --noweb --bfile QCD_UK_PLINK_FILE --update-map UK_chr_updates.txt --update-chr \
--make-bed –out QCD_UK_PLINK_FILE_chr_updates
plink --noweb --bfile QCD_UK_PLINK_FILE_chr_updates --update-map UK_pos_updates.txt \
--make-bed --out QCD_PLINK_FILE
```

III. SAMPLE EXCLUSIONS

Best practices for sample exclusions are described in Grove *et al.* PloSOne 8(7): e68095) and "Exome-chip QC SOP.v5.pdf", for which exclusions include:

- Sample call rate < 98% (after joint or zCAll calling)
- Outlying heterozygosity rate based on global heterozygosity for >1% SNPs and <1% SNPs
- Deviation from HWE P<e-06

- Gender mismatch
- Discordance with existing genotype information (e.g. GWAS, custom arrays, fingerprint)
- Population IBS/MDS/PCA clustering outliers (described below)

NOTE: for studies with "unrelated" individuals (i.e. not family-based), we are specifically asking you to NOT exclude people who you find might be closely/cryptically related. Instead family/cryptic relatedness will be handled by the genomic relationship matrix created and used in the RAREMETALWORKER and RvTests as part of the protocol (described below).

B. LIMIT THE STUDY TO THE AUTOSOMAL AND X CHROMOSOMES

Only autosomal and X chromosomes are included in the analyses; variants on Y, XY and MT are not included. Therefore, remove variants on chromosomes 24, 25 & 26. If you genotyped the Illumina ExomeChip v1.0 you can use the CHR24-26.txt file, located on the SFTP site: /data/RELEVANT_FILES/CHR24_26/. Otherwise, create your own CHR24-26.txt file that contains all variants on chromosome 24, 25 & 26 (one variant per line). Please, contact us if creating such files is difficult.

Use the following command:

```
plink --noweb --bfile QCD_PLINK_FILE --exclude CHR24-26.txt \
--make-bed --out QCD_PLINK_FILE_CHR1_23
```

where

QCD_PLINK_FILE: the plink binary file

CHR24-26.txt: a list of variants on Y (24), XY (25) and MT (26) chromosomes to be removed from the plink binary file

QCD_PLINK_FILE_CHR1_23: the binary output file that contains variants on autosomal and X chromosomes

C. ALLELE ALIGNMENT

Please note that this is the only slightly fiddly section of the document – we apologize, but it is important that we get all alleles aligned to the correct strand, and that for each variant the correct allele is labelled as "reference" – two slightly different things! Again, all of this will be relevant to

other consortia as well, so once you have done it, running again for other phenotypes will be very straightforward. Prepare genotype files separately by ancestry/ethnicity.

I. ONLY FOR STUDIES THAT HAVE NOT BEEN CALLED AS PART OF THE CHARGE CONSORTIUM - FLIP 23 SNPS TO "+" STRAND

Studies that followed the Exome-chip QC SOP.v5.pdf protocol and have not been called as part of the CHARGE consortium have to flip 23 SNPs to the + strand.

```
plink --noweb --bfile QCD_PLINK_FILE_CHR1_23 \
--flip SNP_list_to_be_flipped_Exome-chip_QC_SOP_studies_only_23SNPs.txt \
--make-bed --out QCD_PLINK_FILE_CHR1_23_STRANDALIGNED
```

where

QCD_PLINK_FILE_CHR1_23: the plink binary file

SNP_list_to_be_flipped_Exome-chip_QC_SOP_studies_only_23SNPs.txt: a simple text file of the 23 SNPIDs which are to be strand-flipped -one SNP per line.

QCD_ PLINK_FILE_CHR1_23_STRANDALIGNED: the binary output file aligned to the + strand. This file can then go into the forcealleles command below.

II. STUDIES FROM THE CHARGE CONSORTIUM ONLY– FLIP ALLELES TO "+" STRAND

For studies that were genotyped and QC'd as part of the CHARGE consortium, we would require that they flip SNPs to be aligned to the + strand. This can be done using the command below for plink.

NOTE: Before using the command below, make sure you have not already flipped these (for a previous analysis).

```
plink --noweb --bfile QCD_PLINK_FILE_CHR1_23 \
--flip SNP_list_to_be_flipped_KL_TW.txt --make-bed \
--out QCD_PLINK_FILE_CHR1_23_STRANDALIGNED
```

where

QCD_PLINK_FILE_CHR1_23: the plink binary file

SNP_LIST_TO_BE_FLIPPED_KL_TW.txt: a simple text file of SNPIDs which are to be strand-flipped - one SNP per line.

QCD_ PLINK_FILE_CHR1_23_STRANDALIGNED: the binary output file aligned to the + strand. This file can then go into the forcealleles command below.

III. INDELS

If you chose to include INDELs, please make sure to recode them using I/D only.

IV. CREATE UNIFORM LIST OF REFERENCE ALLELES USING A FORCEALLELES COMMAND IN PLINK

ALL STUDIES (CHARGE studies included), please, run this as default, even if you think your reference alleles are all coded correctly. In this step, we are making sure that for each variant the correct allele is labeled as "reference" allele.

Please, force the alleles indicated as the desired reference allele prior to generating another VCF file using the following plink command where the FORCEALLELES_V1.0 file (located on the SFTP site: /RELEVANT_FILES/FORCE_ALLELES/) consists of two columns listing "exomechip SNP ID" and "desired reference allele".

NOTE: The FORCEALLELES_V1.0 file only contains the variants that were genotyped on the "Illumina ExomeChip 1.0". If you have genotyped another version of the ExomeChip, you will have to create a bespoke FORCEALLELES files.

```
gunzip FORCEALLELES_V1.0.gz
plink --noweb --bfile QCD_PLINK_FILE_CHR1_23_STRANDALIGNED \
--reference-allele FORCEALLELES_V1.0 --make-bed \
--out QCD_PLINK_FILE_FINAL
```

where

QCD_PLINK_FILE_CHR1_23_STRANDALIGNED is the plink binary file which contains SNPs on autosomal and X chromosomes and is aligned to the + strand.

FORCEALLELES_V1.0 is a simple text file of "exomechip SNP ID" and "desired reference allele" that are to be forced - one SNP per line

QCD_PLINK_FILE_FINAL is the binary output file that contains SNPs on autosomal and X chromosomes, is aligned to the + strand and is coded for "desired reference allele". This file can now be converted to VCF using PLINK/SEQ. See commands below.

D. CREATE YOUR VCF FILE

In this step, you need to make a VCF file using PLINK/SEQ. Make VCF files separately by ancestry/ethnicity.

I. CREATE A NEW PROJECT IN PLINK/SEQ

```
pseq MY_project new-project --resources /path/to/hg19 \
--scratch /path/to/temporary/files
```

II. LOAD EXISTING PLINK FILES

```
pseq MY_project
load-plink --file QCD_PLINK_FILE_FINAL \
--id iid --check-reference
```

III. CREATE YOUR VCF FILE:

```
pseq MY_project
write-vcf > QCD_FILE_FINAL.vcf
```

NOTE: if you want to re-create a project (re-start from i to iii) with the same name as an old project, you have to delete all the output files produced during the creating of the old project.

E. CHECK ORIENTATION OF ALLELES USING CHECKVCF

To check if the above process has worked, i.e. to check if there are any allele or strand flips or differences relative to the reference we need to run the following commands.

I. REMOVE "CHR" PREFIX FROM THE CHROMOSOME MARKERS AND RENAME CHROMOSOME 23 AS "X" IN THE VCF FILE

This step removes "chr" and renames chromosome 23 as X, and it is needed before VCFcheq

```
sed 's/^chr//g' QCD_FILE_FINAL.vcf | sed 's/^23/X/g' > QCD_FILE_FINAL_nochr.vcf
```

II. CHECK THE VCF FILE FOR ALLELE ALIGNMENT.

The checkVCF package includes 3 files: checkVCF.py script, reference genome (hs37d5.fa) in FASTA format and its index file (hs37d5.fa.fai). Please install python and run the following vcf checking scripts on your final, cleaned VCF file. The following command will generate a list of monomorphic sites, sites with reference alleles inconsistent with the

reference genome, sites with invalid genotypes, non-SNP site (e.g. INDELs), and 900 suspicious sites with allele frequencies >1:

```
python checkVCF.py -r hs37d5.fa -o STUDY QCD_FILE_FINAL_nochr.vcf
```

The STUDY.check.ref file will list sites with non-matching reference alleles under the designation example "MismatchRefBase 1:564766:T-C/T", where the base directly following the position is the desired reference allele. If there are allele issues, please resolve based on QCD_PLINK_FILE_FINAL.bim file (please see section C-E), run checkVCF.py again, before moving to the next steps.

Please contact the central team for further help if the "check.ref" file contains variants and you are not sure how to resolve this issue.

F. CREATE FILE WITH ALLELE FREQUENCIES OF ALL CLEAN VARIANTS

Generate a file that lists allele frequencies of ALL variants (including monomorphic SNPs) that passed the variant and sample QC procedures (above). This file will be used to calculate allele frequencies across all studies, and should not depend on phenotypes. Hence, only one file per study is required; i.e. not need to provide this by phenotype. However, as for the analyses, please stratify by ancestry

Use command:

```
pseq MY_project v-freq > STUDY_AFREQ_ANCESTRY.txt
```

Where:

STUDY_AFREQ_ANCESTRY.txt is the output file labeled as follows:

For STUDY use: the same acronym you usually use

For ANCESTRY use: EU: European ancestry; AA: African Ancestry; HL: Hispanic Latino; SA: South Asian; EA: East Asian

e.g. MSSM_AFREQ_HL.txt

NOTE ON DUPLICATED MARKERS: There will be ~800 duplicate markers; most of these are bi-allelic, a few are tri-allelic variants.

For bi-allelic variants, PLINK/SEQ will integrate the genotype information from both duplicate markers into one unique marker, using information from both the original duplicate markers. If genotypes are inconsistent between the original duplicate markers, it will assign a missing value.

For the tri-allelic variants, the allele information (three alleles in one position) of both duplicate makers will be retained during the file conversion. Unfortunately, these markers cannot (yet) be read by RvTests or RAREMETALWORKER (below).

G. MULTIDIMENSIONAL SCALING (MDS) / PRINCIPAL COMPONENTS (PCS) FOR STRATIFICATION

For studies with "unrelated samples" (i.e. not family-based studies), perform multidimensional scaling (MDS) on the basis of the IBS calculated from the LD pruned (r2 < 0.2) MAF 1% markers (use GWAS genotype data, if available). Record the projection of each sample onto the first ten principal components (PCs) for use in downstream analyses to adjust for population structure. Final decisions on whether to exclude samples are likely to be cohort specific, but clear ethnic outliers should be excluded from European-based studies at a minimum. While other methods are available (i.e. STRUCTURE), within PLINK this can be achieved using the command:

```
--cluster --mds-plot 10
```

H. PRINCIPAL COMPONENT ANALYSES AND PLOT

Calculate PC1 and PC2 using common variants (MAF > 1%) and submit a scatterplot of PC1 vs PC2 anchored by 1000G CEU, YRI and CHB/JPT data on all samples included in analyses to allow comparisons of ethnic homogeneity between studies. You can use the PLINK binary files (AIMs.1000G.v2) to anchor your data using the script "Files_for_scatter_plots.rar". The files and relevant scripts are shared at the SFTP site: /data/RELEVANT_FILES/FILES_FOR_PC1_PC2_SCATTERPLOTS.

Provide this plot as PDF file and label them as SCATTER_STUDY_ANCESTRY.pdf; e.g. SCATTER_MSSM_HL.pdf

For ancestry use:

EU: European ancestry;

AA: African Ancestry;

HL: Hispanic Latino;

SAS: South Asian;

EAS: East Asian

5. Analysis with RVTest or RAREMETALWORKER

Rare variant analyses can be carried out either using RvTests or RAREMETALWORKER (both described below). Both can be used for unrelated and related individuals. We ask all studies to use an empirical kinship matrix (aka genomic relationship matrix) for analysis. If you do not already

have a kinship matrix for your study, it can be generated automatically with a specific command in RAREMETALWORKER or RvTests and so requires no extra hands-on time. This will increase the computational time for the analyses, and this time will increase exponentially with sample size. For example, the analyses of a sample of ~10,000 individuals using RAREMETALWORKER took ~22 hours (maximum memory use of 2.3GB) to finish, including generating two kinship matrices (one for autosome and the other for chromosome X), fitting linear mixed models using these two matrices, and deliver single variant analyses results using additive, recessive, and dominant models, and generate three separate sets of covariance matrices, together with the plots in PDF. The same analyses for a sample of around ~2,000 individuals took ~6 hours (max memory use was ~100Mb) using RvTests. Overall, metrics will depend on your server setup, but times and memory use seems comparable for RvTests and RAREMETALWORKER. Thus, if you have a large sample size, you might wish to check your servers' capacity. Again, once made, you can use this matrix for future analyses and collaborations.

Analysts are free to choose which software (RvTests or RAREMETALWORKER) they use. The most notable differences to users are the interface, in particular the way covariates are organized. RvTests organizes covariates in a similar way to PLINK, where covariates are provided in a separate file. RAREMETALWORKER was derived from Merlin, so it supports inputs more similar to Merlin, in that covariates are stored in PED file, and you also need a DAT file to describe the covariates. Both tools have a very simple interface to use, and the input for one tool can be easily converted to the other.

For family-based studies with known pedigree structure, to ensure pedigrees are correctly understood by analysis tools, we recommend that all connecting individuals - including those for whom genotypes are not available - should be dummied into the pedigree file. For example, when analyzing nuclear families where only offspring are genotyped, the pedigree file should include additional rows for the parents. When analyzing more extended pedigrees, additional individuals may need to be included in the pedigree file.

NOTE that we do not ask to perform study-specific multiple-variant analyses, but rather to generate output files that are being used for central multiple-variant meta-analyses (using rareMETALS).

Please remember to

Perform analyses for:

- All individuals
- Men only
- Women only

Stratify analyses by:

- Disease status; i.e. analyze cases and controls separately, in particular if the outcomes (above) are affected by disease status
- Race/ethnicity as appropriate to your sample.

RVTESTS

RvTests software is fast and capable of handling large data set. The input files required (phenotype, covariates) are compatible with PLINK formats. It also calculates Hardy-Weinberg p-value and call rate for quality control purposes. Documentation of RvTests can be found in: https://github.com/zhanxw/rvtests

Fat percentage and ExomeChip - Analysis plan – Dec 8th 2014

16

1. SUMMARY LEVEL STATISTICS FOR META-ANALYSIS OF RARE VARIANT ASSOCIATIONS

The following summary level statistics will be generated for meta-analysis:

a) Allele frequency for each variant
b) Single variant association test statistics, including direction of effect. We use score statistics calculated at each variant site.
c) Covariance matrix for each genetic region. We compute the genotype covariance for all variants in each gene. This matrix reflects linkage disequilibrium in the region.
d) Basic VCF metrics, including reference and alternative alleles, chromosome positions and strand.

The summary statistics can be calculated using the tools RvTests. These require an indexed VCF file (for genotypes) and a PED file (for phenotypes).

2. INDEXING THE VCF FILES

The software works with tabix, and takes indexed bgzip-ed VCF files as input. If your VCF file is not compressed with bgzip, please use the following command to create it (QCD_FILE_FINAL_nochr.vcf.gz):

```
bgzip QCD_FILE_FINAL_nochr.vcf
```

Please index your bgzipp-ed VCF using the following command (QCD_FILE_FINAL_nochr.vcf.gz.tbi):

```
tabix -f -p vcf QCD_FILE_FINAL_nochr.vcf.gz
```

Please make sure there are no entries with -1's or 0's in the chromosome number column.

3. PEDIGREE FILE FORMAT WITH PHENOTYPES

The RvTests tool requires a simple pedigree file that starts with the standard 5 columns (family id, person id, father id, mother id and sex) followed by trait or trait residuals. The trait residuals to use are the transformed phenotypes that were obtained as described above.

4. EXEMPLAR COMMAND

An exemplar command for using RvTests looks like the following (please refer to section 7. for exact commands for your respective data analysis, as trait transformation has to be done beforehand and no covariates are included in the analysis):

```
./vcf2kinship --ped example.ped --out example
./rvtests --inVcf QCD_FILE_FINAL_nochr.vcf.gz --pheno example.ped \
--kinship example.kinship --meta score,cov,dominant,recessive \
--out output_prefix
```

5. SPECIFY VCF FILES

Use --inVcf to specify which VCF to use.

6. SPECIFY PHENOTYPES

Use --mpheno phenoypeColumnNumber or --pheno-name to specify a given phenotype.

Phenotype file is specified by the option --pheno example.ped. The default phenotype column header is "y1". If you want to use alternative columns as phenotype for association analysis (e.g the column with header y2), you may specify the header names using either --mpheno 2 or --pheno-name y2

NOTE: to use --pheno-name, the header line must start with "fid iid" as PLINK requires.

In phenotype file, missing values can be denoted by NA or any non-numeric value. Individuals with missing phenotypes will be automatically dropped from subsequent

association analysis. For each missing phenotype value, a warning will be generated and recorded in the log file.

7. COMMANDS

You need to prepare a phenotype file (example.ped in PLINK format) and a genotype file (QCD_FILE_FINAL_nochr.vcf.gz). You can choose one of the following two scenarios to use RvTests:

a. Analysis of related individuals with known pedigree information in linear mixed models

Generate kinship matrix from pedigree file:

```
./vcf2kinship --ped example.ped --out output --xHemi
```

Generate score statistics and covariance matrix:

```
./rvtests --inVcf QCD_FILE_FINAL_nochr.vcf.gz \
--pheno example.ped --out output --kinship output.kinship \
--xLabel X --meta score,cov,dominant,recessive
```

b. Analysis of unrelated and/or related individuals with empirical kinship in linear mixed models

Generate empirical kinship matrix from VCF file:

```
./vcf2kinship --ped example.ped --inVcf QCD_FILE_FINAL_nochr.vcf.gz \
--bn --out output --xLabel X --xHemi
```

By default, the output file is named as output.kinship, and you can use it with the --kinship option

Generate score statistics and covariance matrix:

```
./rvtests --inVcf QCD_FILE_FINAL_nochr.vcf.gz --pheno example.ped \
--out output --kinship output.kinship --xLabel X \
--meta score,cov,dominant,recessive
```

c. Analyzing sex chromosome

The output of the above scripts (a.- b.) also contains the association results for X linked markers. RvTests recognize male (coded as 1) and female (code as 2) by looking up values in the 5th column of the PED file. Male non-reference genotype in the non-PAR regions of the X chromosome will be coded as 0 and 2 in regression model. Any male sample with

genotypes stored as heterozygous (e.g. 0/1) in this region is skipped from the analysis. The definition of this hemizygote region is the X chromosome except the PAR1 region (X:60001-2699520) or PAR2 region (X:154931044-155270560).

NOTE: The run of vcf2kinship with "--xHemi" option will produce the kinship files for both autosomal region (output.kinship) and X chromosome hemizygote region (output.xHemi.kinship) based on either pedigree information or common exome-chip SNP information. The analysis of the X chromosome hemizygote region requires output.xHemi.kinship file. This file name can be automatically deduced from the parameter of "--kinship" given by the user. For example, if the user specify "--kinship output.kinship", RvTests will look for "output.kinship" as the autosomal kinship file and "output.xHemi.kinship" as the X chromosome hemizygote region kinship file.

Using --out output option, the results are stored in output.MetaScore.assoc (output.MetaRecessive.assoc or output.MetaDominant.assoc) and output.MetaCov.assoc.gz (output.MetaDominantCov.assoc.gz or output.MetaRecessiveCov.assoc.gz). Each output file contains a header with summaries of trait and covariates.

8. OUTPUT FILES

The output.MetaScore.assoc files contain the following columns:

CHROM Chromosome

POS Position

REF Reference Allele

ALT Alternative Allele

N_INFORMATIVE Number of samples in the analysis

AF Allele frequency;

INFORMATIVE_ALT_AC Number of alternative alleles

CALL_RATE Fraction of non-missing genotypes

HWE_PVALUE Hardy-Weinberg Equilibrium P-value.

N_REF, N_HET, N_ALT Number of ref/ref, ref/alt, alt/alt genotypes respectively.

U_STAT, SQRT_V_STAT U statistic and square root of V statistic as in the score test.

ALT_EFFSIZE Estimated effect size using alternative allele.

PVALUE P-value when testing genetic association using score test.

The output.MetaCov.assoc.gz files contain following columns:

CHROM Chromosome

START_POS Position for the first marker in Window

END_POS Position for the last marker in Window

NUM_MARKER Number of markers in Window

MARKER_POS Position for the markers in Window, separated by commas

COV Covariance matrix between test statistics

The Log file is saved in output.log

!! PLEASE; USE THE "FILE NAMING SCHEME" DESCRIBED BELOW BEFORE UPLOADING YOUR DATA !!

RAREMETALWORKER

RAREMETALWORKER provides an alternative to RvTests. This tool handles related individuals using either pedigree information or an empirical kinship matrix estimated from marker data. RAREMETALWORKER also provides an alternative tool for samples of unrelated individuals.

1. KEY FEATURES

Hardy-Weinberg equilibrium p-value and genotype call rate are calculated automatically and included in summary files for later quality control. Some options can automatically compress the results (--zip) to share and allows one to label the significant loci (--labelHits) with gene names in Manhattan plots. Genomic control information will be reported automatically.

RAREMETALWORKER accepts genotypes in VCF or in Merlin pedigree files for input (although we recommend VCF, especially if INDELs are included in the analysis) and (in family samples) should correctly handle variants on the X chromosome.

2. HOW TO EXECUTE

a) To execute, go to /raremetal_4.13.5/raremetal/bin/, and run ./raremetalworker. A summary of available options will appear.

b) For detailed explanations of those options, please refer to the following wiki page: http://genome.sph.umich.edu/wiki/Rare-Metal-Worker#Software_Options

3. INPUT FILES

RAREMETALWORKER reads genotypes from either a PED/DAT file pair or a VCF file. When genotypes are saved in a PED and DAT file, no VCF file is needed. However, when a VCF file is used, the pedigree structure, phenotypes and covariates must be read from a PED/DAT file pair. Sample IDs in the VCF file must match individual IDs (second field) in PED file. To handle unrelated individuals, each individual should have a distinct family ID in the PED file (an easy way to accomplish this is to repeat the individual ID in the family ID column).

We strongly recommend VCF file as input of variant genotypes because VCF has many advantages compared to PED file, for example, INDELs are better defined in VCF but is always ambiguous in PED/DAT files.

a. VCF File

VCF file should be compressed by using bgzip and indexed by using tabix, please refer to section 8 RvTests 2 for commands.

b. PED and DAT Files

Trait values are saved in PED file. Trait descriptions are saved in DAT file.

RAREMETALWORKER can handle MERLIN format or PLINK format (text-only) pedigree files, but the correct label for each field has to be provided in DAT file. For detailed descriptions and examples of Merlin format PED and DAT files or VCF file, refer to the following: http://genome.sph.umich.edu/wiki/Rare-Metal-Worker#Input_Files

In the phenotype file, missing values can be denoted by X as used in MERLIN (http://www.sph.umich.edu/csg/abecasis/merlin/tour/input_files.html). Individuals with missing phenotypes will be automatically dropped from subsequent association analysis.

4. EXAMPLE USAGE

The following examples assume the presence of VCF file.

a. ANALYSIS OF FAMILY SAMPLES (with known pedigree information)

Because genotypes are saved in a VCF file, PED and DAT files are needed for specifying pedigree structure and trait information. This command line fits the model and generates summary statistics.

```
raremetalworker --ped your.ped --dat your.dat \
--vcf QCD_FILE_FINAL_nochr.vcf.gz --kinPedigree \
--recessive --dominant --zip --vcX --separateX --xLabel X
```

b. ANALYSIS USING AN EMPIRICAL KINSHIP MATRIX

This analysis will first calculate an empirical kinship matrix, and then fits the model and generates the summary statistics.

```
raremetalworker --ped your.ped --dat your.dat \
--vcf QCD_FILE_FINAL_nochr.vcf.gz --recessive --dominant \
--zip --kinGeno --kinSave --vcX --separateX --xLabel X
```

NOTE: if you've already generated a kinship based genomic relationship matrix, you can refer point to that with --kinFile.

c. ANALYSIS OF CHROMOSOME X LINKED MARKERS

The output of the above scripts (a.- b.) also contains the association results for X linked markers. In order for RAREMETALWORKER to differentiate between male and female, men should be coded as 1 and women as 2 in the PED file. RAREMETALWORKER takes male genotype to be 0 or 2. This means that whenever an alternative allele is observed in male on chromosome X (non-PAR region), it is considered to have the same effect as a female with homozygous alternative alleles. Any male sample with genotypes stored as heterozygous (e.g. 0/1) in this region is skipped from the analysis. The label of chromosome X can be changed flexibly according to the entry in VCF file (the first column) using --xLabel option. The non-PAR region can also be specified using --xStart and --xEnd options. Then RAREMETALWORKER can automatically differentiate between the PAR and non-PAR region on Chromosome X and decode male genotype accordingly. For example, if the X chromosome in the VCF file is labeled as "X", then adding the following options lets RAREMETALWORKER recognize variants with CHR="X" and 2699520 ≤POS≤154931044 for chromosome X associations.

```
--xLabel X --xStart 2699520 --xEnd 154931044
```

The default values for these options are --xLabel X --xStart 2699520 --xEnd 154931044. The start and end positions are collected from NCBI genome build 37.

For more examples of the usage, please refer to the following:

http://genome.sph.umich.edu/wiki/Rare-Metal-Worker#Example_Command_Lines

For more examples of X-linked marker analysis usage, please refer to the following:

http://genome.sph.umich.edu/wiki/RAREMETALWORKER_X

5. OUTPUT FILES

RAREMETALWORKER will generate seven files for sharing per trait, which should be uploaded to the SFTP site:

- the summary statistics of single variant score tests from the additive model, recessive model and dominant model;
- the linkage disequilibrium (LD) matrices summarising covariance between single marker score statistics from the additive model, recessive model and dominant model;
- a log file with all parameter settings.

These seven files are required for central meta-analysis and QC.

http://genome.sph.umich.edu/wiki/Rare-Metal-Worker#Outputs

1. Summary statistics from single variant tests are saved in the file yourPrefix.traitName.singlevar.score.txt.gz or

yourPrefix.traitName.singlevar.recessive.score.txt.gz or yourPrefix.traitName.singlevar.dominant.score.txt.gz

For detailed explanations, please go to the link above. To change the prefix used to label files use the --prefix option.

The following fields will be included in the output (please double check):

CHROM Chromosome

POS Position

REF Reference Allele

ALT Alternative Allele

N_INFORMATIVE Count of individuals with genotype and phenotype

FOUNDER_AF Allele frequency among founders

ALL_AF Allele frequency across entire sample

INFORMATIVE_ALT_AC Copies of the rare allele among samples with genotype and phenotype

CALL_RATE Fraction of called genotypes

HWE_PVALUE Exact Hardy-Weinberg equilibrium p-value

N_REF Count of reference homozygotes

N_HET Count of heterozygotes

N_ALT Count of alternative allele homozygotes

U_STAT Score statistic numerator

SQRT_V_STAT Score statistic denominator

ALT_EFFSIZE Estimated effect size

PVALUE P-value

This file will also include a header with a summary of trait and covariate distributions (including the mean and standard deviation, for example).

2. LD matrices info is saved in the file yourPrefix.traitName.singlevar.cov.txt.gz (yourPrefix.traitName.recessive.singlevar.cov.txt.gz or yourPrefix.traitName.dominant.singlevar.cov.txt.gz)

CHROM Chromosome

CURRENT_POS Position for the first marker in Window

MARKERS_IN_WINDOW Position for the other markers in window, separated by commas

COV_MATRICES Covariance matrix between test statistics

3. A log file is saved in: yourPrefix.singlevar.log

!! PLEASE; USE THE "FILE NAMING SCHEME" DESCRIBED BELOW BEFORE UPLOADING YOUR DATA !!

9. FILE NAMING SCHEME

Use the following file naming scheme:

STUDY_TRAIT_SEX_STATUS_ANCESTRY_DATE_ANALYST_ ANALYSIS.TXT

STUDY: Name of your study population

TRAIT: FatPCT, FFMi, LEPTIN, LEPTINadjBMI, LEPTINadjFatPCT, ADIPONECTIN, ADIPONECTINadjBMI or ADIPONECTINadjFatPCT

SEX: ALL, MEN, or WOMEN

STATUS: COHORT, CASE or CONTROL (if CASE, specify T2DCASE, or CADCASE, …)

ANCESTRY: EU (European ancestry), AA (African ancestry), HL (Hispanic Latino), EAS (East Asian), SAS (South Asian)

DATA: DDMMYY

ANALYST: Initials of analyst

ANALYSIS: use CLUSTER: for the PC cluster files

When using RvTests use:

MetaCov.assoc.gz

MetaCov.assoc.gz.tbi

MetaScore.assoc

MetaRecessiveCov.assoc.gz

MetaRecessiveCov.assoc.gz.tbi

MetaRecessive.assoc

MetaDominantCov.assoc.gz

MetaDominantCov.assoc.gz.tbi

MetaDominant.assoc

log

When using RAREMETALWORKER use:

singlevar.score.txt.gz

singlevar.cov.txt.gz

recessive.singlevar.score.txt.gz

recessive.singlevar.cov.txt.gz

dominant.singlevar.score.txt.gz

dominant.singlevar.cov.txt.gz

singlevar.log

Please also upload the outputs (7 files) from the checkVCF scripts, using the following conventions:

STUDY.check.ref, STUDY.check.nonSnp etc.

Analysis plan 2: Genome-wide association studies for fat and lean mass compartments

1. Project aims

In this project we aim to identify the genetic determinants of 16 body fat and lean mass compartments measured by DEXA through genome-wide association analyses. Analyses will be run for both sexes combined and stratified by sex. We will analyse the DEXA traits unadjusted and adjusted for total fat mass (fat compartments) and height-squared (lean compartments).
This analysis plan describes the quality control (QC) and transformation of the DEXA phenotypes and the genome-wide association analyses of fat and lean mass compartments. Please contact Claudia Langenberg (Claudia.Langenberg@mrc-epid.cam.ac.uk) and Laura Wittemans (Laura.Wittemans@mrc-epid.cam.ac.uk) in case you have any questions.

2. Phenotype QC and transformation

a) Phenotypes

The 16 traits that will be studied in this project are the following:

- Total fat mass (g) (fat_tot)
- Total lean mass (g) (lea_tot)
- Arms fat mass (g) (fat_arm)
- Arms lean mass (g) (lea_arm)
- Trunk fat mass (g) (fat_tru)
- Trunk lean mass (g) (lea_tru)
- Android fat mass (g) (fat_and)
- Android lean mass (g) (lea_and)
- Gynoid fat mass (g) (fat_gyn)
- Gynoid lean mass (g) (lea_gyn)
- Legs fat mass (g) (fat_leg)
- Legs lean mass (g) (lea_leg)
- Visceral fat mass (g) (fat_vis)
- Subcutaneous fat mass (= android fat mass – visceral fat mass) (g) (fat_scu)
- Appendicular lean mass (= legs lean mass + arms lean mass) (g) (lea_app)
- Peripheral fat mass (= legs fat mass + arms fat mass) (g) (fat_pri)

The first 13 traits are directly generated by the DEXA software, whereas the last 3 can be calculated with the provided formulas.

b) Quality control of phenotypes

DEXA measurements should be omitted if:

- Some tissue is missing, e.g., very tall or broad participants, participants who underwent amputations etc.
- Any other reason for low quality, incomplete or atypical DEXA measurements, such as movement artefacts, pregnant or transgender participants etc.

Participants with prosthetic implants (e.g., hip and knee replacements) should not be excluded.

Please generate histograms of untransformed DEXA traits after QC and provide an overview of your QC procedure of the DEXA data in the info file (see section 4)

If you require assistance with this step, please contact Claudia or Laura.

c) Phenotype transformation

Unadjusted analyses (all phenotypes)

1. Transform phenotypes for men and women separately
2. Natural log transform all DEXA traits
3. Fit a linear model with age and any study-specific covariates as independent variables and DEXA compartment as the dependent variable. Take the residuals
4. Rank-based inverse normal transformation of residuals

Merge transformed phenotypes for men and women back together for sex-combined analyses

Example R code for steps 2-4, for android fat in men:

```
Men<-men[!is.na(men$fat_and),]

log_fat_and<-log(men$fat_and)

Model<-lm(log_fat_and ~ age + study_specific_covariate_1+study_specific_covariate2 + …, data=men)

Residuals<-residuals(Model)

InvNorm_var<-qnorm((rank(Residuals,na.last="keep")-0.5)/sum(!is.na(Residuals)))
```

```
men$invnorm_fat_and<-InvNorm_var
```

Adjusted analyses for fat compartments

Phenotypes: peripheral fat mass, subcutaneous fat mass, visceral fat mass, leg fat mass, gynoid fat mass, android fat mass, trunk fat mass and arm fat mass.

- Transform phenotypes for men and women separately
- Natural log transform all fat DEXA traits
- Fit a linear model with age, total body fat (natural log transformed) and any study-specific covariates as independent variables and DEXA compartment as the dependent variable. Take the residuals
- Rank-based inverse normal transformation of residuals
- Merge transformed phenotypes for men and women back together for sex-combined analyses

Example R code for steps 2-4, for android fat in men:

```
Men<-men[!is.na(men$fat_and) & !is.na(men$fat_tot),]

log_fat_tot<-log(men$fat_tot)

log_fat_and<-log(men$fat_and)

Model<-lm(log_fat_and ~ age + log_fat_tot + study_specific_covariate_1+study_specific_covariate2
+ …, data=men)

Residuals<-residuals(Model)

InvNorm_var<-qnorm((rank(Residuals,na.last="keep")-0.5)/sum(!is.na(Residuals)))

men$invnorm_fat_and_adj<-InvNorm_var
```

Adjusted analyses for lean compartments

Phenotypes: Appendicular lean mass, leg lean mass, android lean mass, trunk lean mass, arm lean mass, total lean mass, gynoid lean mass.

- Transform phenotypes for men and women separately
- Natural log transform all lean DEXA traits
- Fit a linear model with age and height-squared (in cm^2) and any study-specific covariates as independent variables and DEXA compartment as the dependent variable. Take the residuals

- Rank-based inverse normal transformation of residuals
- Merge transformed phenotypes for men and women back together for sex-combined analyses

Example R code for steps 2-4, for appendicular lean mass in men:

```
Men<-men[!is.na(men$app_lea) & !is.na(men$app_lea),]

log_app_lea<-log(men$app_lea)

Model<-lm(log_app_lea ~ age + I(height^2) +
study_specific_covariate_1+study_specific_covariate2 + …, data=men)

Residuals<-residuals(Model)

InvNorm_var<-qnorm((rank(Residuals,na.last="keep")-0.5)/sum(!is.na(Residuals)))

men$invnorm_app_lea_adj<-InvNorm_var
```

3. Genome-wide association analyses

a) Quality control of genotyped data

For QC of the genotyped data, please follow the best practices for QC suggested by the manufacturer of the genotyping chip.

Recommended filters:

Exclude participants if sex mismatch

Exclude variants for which call rate <97%

Exclude variants for which p-value for HWE<10^{-8}

Exclude participants with more than 5% missing variants

Exclude ethnic outliers that do not cluster with their expected ancestry based on a scatter plot of PC1 and PC2 and the 1000G CEU, YRI and CHB/JPT reference panels

Exclude related participants based on IBD>10%

Please provide an overview of the QC strategy you applied for the genotyped data in the info file (see section 4)

b) Imputation

Preferred imputation strategy for **white British** cohorts (aligned to UK biobank):

Impute genotyped data to the reference panel from the Haplotype Reference Consortium and UK10K-1000 Genomes (UK10K-1KG).

Give priority to HRC-imputed variants and fill in with UK10K-1KG-imputed variants (for those which are not present among the HRC imputed variants)

Preferred imputation strategy for **non-British white European** cohorts: Impute genotyped data to reference panel from the Haplotype Reference Consortium

Preferred imputation strategy for **other ancestries**: Impute genotyped data to 1000G reference panel for the matching ancestry.

HRC, UK10K-1KG and 1KG imputation can be done using the Sanger (imputation.sanger.ac.uk/) or Michigan imputation server (imputationserver.sph.umich.edu).

Please get in touch with Claudia or Laura if you have any questions regarding the imputation.

c) Genome-wide association analyses

To run the genome-wide association analyses, we suggest using SNPTEST or BGENIE. Please include at least 4 principal components and other covariates related to the genotyping if necessary for your cohort (e.g., include genotyping chip as covariate is several different types of chip were used for genotyping in your cohort). Please do not include sex as a covariate.

SNPTEST runs on GEN, BGEN and VCF files and will run the traits one by one. Download and more info: https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html

BGENIE only runs on BGEN files and has the advantage that it can analyse a large number of traits in one run, while decompressing and converting the BGEN files only once for all traits. This makes this tool more time-efficient. GWAS results for all traits are stored in one file, which can become very large and therefore more challenging to handle than the results file outputted by SNPTEST. Download and more info: (https://jmarchini.org/bgenie/).

d Quality control of GWAS results

1. Drop variants for which minor allele count (MAC) is less than 10. MAC=2*minor allele frequency*sample size
2. Drop if info score is lower than 0.4 (INFO<0.4)
3. Drop variants with other alleles than A, C, G or T

4. Drop variants for which the absolute value of the effect size or SE is larger than 10

4. Data sharing

Please generate and upload the following figures and files. You will receive the login details to the SFTP in a separate email.

General comment: Given the large number of traits and analyses, it would be very helpful if you could follow the recommended filename structure.

a) File names

Given the large numbers of phenotypes and analyses, could you please follow the following file name structure for all files for sharing:

CohortName_trait_sex_adjustment_date_[type_of_analysis/figure/document].txt/png

CohortName: cohort name

trait: dexa trait using the abbreviations specified under 2.1. (e.g., fat_and)

Sex: combined, men, women

Adjustment: unadj or adj (for total fat or height-adjusted traits).

Date: yymmdd (e.g., 180321 for 21th March 2018)

b) Phenotypes

Histograms of untransformed DEXA traits after QC, for sexes combined. **CohortName_trait_date_hist.png/jpg**

Scatter plot of total mass from DEXA versus weight as measured using scales. **CohortName_weight_dexa_date_scatter.png/jpg**

Table giving for each trait and by sex the sample size, mean, standard deviation, maximum and minimum, and age (mean, SD, maximum and minimum) **Cohort_phenotypes_date_table.txt**

c) GWAS results

Please provide one gzipped file per trait/analysis using the following file name structure:

CohortName_trait_sex_adjustment_date_gwas.txt.gz

Columns to include:

Markername: chr[chromosome number]:[base pair position]:[alleles sorted alphabetically]

E.g., chr1:85421326:AG

Please provide chromosome and position in build genome build GRCh37

Rsid

Effect_allele (as upper case: e.g., G instead of g)

Other_allele (as upper case)

Chr

Bp: basepair position in build genome build GRCh37

EAF: effect allele frequency

INFO: info imputation score

Beta: effect size aligned to effect allele

SE: standard error

P: p-value

d) Cohort info

Please generate a .txt file (**CohortName_data_info.txt**) in which you specify the following characteristics of your cohort and data:

DEXA technology and software

Applied QC strategy for DEXA

Study-specific covariates used in the phenotype transformation

Genotyping chip

QC of genotyped data

Imputation reference panel and software/server

GWAS software + version

Covariates for GWAS: number of principal components and other covariates (if any)