BIOLOGICAL AND CLINICAL IMPLICATIONS OF OBESITY GENOMICS IN ANCESTRALLY DIVERSE POPULATIONS

Daeeun Kim

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Epidemiology in the Gillings School of Global Public Health.

Chapel Hill
2023

Approved by:

Christy L. Avery

Penny Gordon-Larsen

Mariaelisa Graff

Kari E. North

Laura M. Raffield

ABSTRACT

Daeeun Kim: Biological and Clinical Implications of Obesity Genomics in Ancestrally Diverse Populations
(Under the direction of Kari E. North)

Obesity, a major risk factor for numerous health outcomes, particularly cardiovascular diseases (CVD), is a highly polygenic trait. Thousands of obesity-associated genetic loci have been identified, facilitating more accurate risk prediction through polygenic risk scores (PRS). Nonetheless, significant research gaps in obesity genomics exist, notably regarding two key aspects: (1) Heterogeneities in PRS prediction across different PRS estimation methods, self-reported race/ethnicity, and different individual-level contexts, and (2) Heterogeneities in shared genetic underpinnings between obesity and dyslipidemia, a major contributor to CVD risk. This dissertation had two specific aims that addressed these research gaps as follows: to characterize the prediction performance of PRS for obesity traits across different PRS estimation methods and diverse settings, including self-reported race/ethnicity, demographic factors, lifestyle factors, and comorbidities (Aim 1); and to identify shared genetic underpinnings in obesity and lipid traits that increased the risk of obesity but were protective for dyslipidemia, as a means to understand why not all obese populations have high risk of CVD (Aim 2). To achieve these goals, we leveraged data from the Population Architecture Using Genomics and Epidemiology (PAGE) study.

Our findings reveal notable differences in PRS prediction across different PRS estimation methods, self-reported racial/ethnic groups, age, gender, smoking status, hypertension, and type 2 diabetes. We also identified 966 genomic regions (among a total of 2,495 partitioned genomic regions) with shared genetic signals between obesity-related traits and lipid traits, with 16 genomic regions of these loci exhibiting counterintuitive directions (associated with increased body mass index (BMI) but decreased dyslipidemia). In PAGE, we observed significant associations of the PRS constructed from variants within these counterintuitive BMI-HDL bivariate loci with lower levels of CVD risk factors. These results enhance our understanding of the heterogeneous underpinnings of obesity susceptibility.

This work is dedicated to my parents, my lovely kids Sohyun and Soye, and my beloved wife Sookyung.

ACKNOWLEDGMENTS

I would like to express my deep appreciation for those who have supported and helped me throughout my PhD journey. First and foremost, I want to express my immense gratitude to my advisor and dissertation chair, Dr. Kari North, for her unwavering support, dedication, and encouragement during my entire doctoral program. I could never have imagined having such a supportive advisor. Without your guidance, encouragement, and support, I would not have made it through this program. It has been my great fortune and privilege to learn and grow under your leadership and mentorship.

I would like to extend my sincere gratitude to all my committee members, Drs. Christy Avery, Penny Gordon-Larsen, Misa Graff, and Laura Raffield, for their incredibly valuable support, feedback, and contributions to my dissertation. Your expertise, keen insights, and thoughtful guidance have been instrumental in my research.

I express my gratitude to the faculty, postdoctoral researchers, and peers in the CVD-Genetic Epidemiology Lab. The lab's invaluable support, diverse activities, and engaging seminars enriched my doctoral experience. Special thanks to Misa, Heather, and Kris for generously dedicating their time and commitment to guiding trainees like myself. I also wish to recognize all my peers in the lab, with special thanks to Victoria and Alice for their great help with this work.

I am grateful to the CVD Epidemiology group, the Epidemiology department, and the instructors of the epidemiology courses for helping me grow as an epidemiologist. I also express

my thanks to Caroline, Valerie, and Bretney for their administrative support. Thanks to fellow department peers, especially Jack and Oumar, for our coffee times during the pandemic.

I express deep gratitude to all coauthors, collaborators, and participants of the PAGE, GIANT consortium, CCHC, and SLS studies. As a PhD student, it has been a privilege to conduct research with such valuable data. I am truly grateful for the financial support during my doctoral program from the Fulbright scholarship, AHA predoctoral Fellowship, and the PAGE 3B grant, in which I participated as a GRA.

I sincerely thank Dr. Joohon Sung, my MPH advisor in Korea, who introduced me to this fascinating field of genetic epidemiology. I also thank the Mahanaim community at the First Korean Baptist Church of Raleigh for being a great support for my family.

I am deeply grateful to my parents and in-laws for their unwavering support throughout this journey. You understood my decision to study in the United States and always prayed for me, supported me, and encouraged me. My adorable kids, Sohyun and Soye, have been my inspiration. And, most importantly, to my beloved wife, Sookyung, none of this would have been possible without her sacrifices and dedication. Above all, I thank God, who always walks beside me every step of this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AAAGC	African Ancestry Anthropometry Genetic Consortium
AFR	African population
ALT	Alanine transaminase
ARIC	Atherosclerosis Risk in Communities Study
BF%	Body fat percentage
BMI	Body mass index
CAD	Coronary artery diseases
CARDIA	Coronary Artery Risk Development in Young Adults
CDC	Centers for Disease Control and Prevention
CHD	Coronary heart disease
CNS	Central nervous system
CT	Computerized tomography
CVD	Cardiovascular disease
DBP	Diastolic blood pressure
DEXA	Dual X-ray absorptiometry
EAS	East Asian population
ER	Endoplasmic reticulum
EUR	European population
FA	Favorable adiposity
FFA	Free fatty acid
GARNET	The Genomics and Randomized Trials Networks
GECCO	The Genetics and Epidemiology of Colorectal Cancer Consortium
GIANT consortium	The Genetic Investigation for Anthropometric Traits consortium
GRS	Genetic risk score
GWAS	Genome-wide association study
HC	Hip circumference
HCHS/SOL	Hispanic Community Health Study/Study of Latinos
HDL	High-density lipoprotein

HIPFX	The Hip Fracture GWAS
HIS	Hispanic/Latino population
HISLA Consortium	Hispanic/Latino Anthropometry Consortium
HM3	HapMap phase 3
IL-6	Interleukin-6
IRB	Internal Review Board
LDL	Low-density lipoprotein
LLS	The Long Life Study
MAF	Minor allele frequency
MARNW	Metabolically at-risk normal weight
MARO	Metabolically at-risk obesity
MCSC	Mount Sinai Medical Center
MEC	Multiethnic Cohort
MEC-AABC	Multiethnic Cohort substudy of breast cancer in African American
MEC-AAPC	Multiethnic Cohort substudy of prostate cancer in African American
MEC-HIBC	Multiethnic Cohort substudy of breast cancer in Native Hawaiian
MEC-JABC	Multiethnic Cohort substudy of breast cancer in Japanese American
MEC-JAPC	Multiethnic Cohort substudy of prostate cancer in Japanese American
MEC-LABC	Multiethnic Cohort substudy of breast cancer in Hispanic/Latinos
MEC-LAPC	Multiethnic Cohort substudy of prostate cancer in Hispanic/Latinos
MEC-Sigma	Multiethnic Cohort-the Slim Initiative in Genomic Medicine for the Americas
MEGA	Multi-Ethnic Genotyping Array
MESA	Multi-Ethnic Study of Atherosclerosis
MHNW	Metabolically healthy normal weight
MHO	Metabolically healthy obesity
MI	Myocardial infarction

MOPMAP	The Modification of PM-Medicate Arrhythmogenesis in Population study
MR	Mendelian randomization
MRI	Magnetic resonance imaging
NHLBI	The National Heart, Lung, and Blood Institute
Ob/DysL(+) loci	Genomic loci associated with higher obesity risk and higher dyslipidemia risk
Ob/DysL(–) loci	Genomic loci associated with higher obesity risk and lower dyslipidemia risk
P+T	Pruning and thresholding method
PAGE	Population Architecture Using Genomics and Epidemiology
PGS	Polygenic score
PRS	Polygenic risk scores
PRS-BMI	Polygenic risk scores for BMI
PRS-WHRadj.BMI	Polygenic risk scores for WHRadj.BMI
SAS	South Asian population
SAT	Subcutaneous adipose tissue
SBP	Systolic blood pressure
SES	Socioeconomic status
SIGMA	The Slim Initiative in Genomic Medicine for Americas
SNP	Single nucleotide polymorphism
T2D	Type 2 diabetes
TC	Total cholesterol
TG	Triglycerides
UFA	Unfavorable adiposity
UKB	UK Biobank
VAT	Visceral adipose tissue
WC	Waist circumference
WHI	Women's Health Initiative
WHI-SHARe	The Women's Health Initiative-SNP Health Association Resource

WHMS	The Women's Health Initiative Memory Study
WHO	World Health Organization
WHRadj.BMI	Body mass index-adjusted waist-hip ratio

CHAPTER 1: SPECIFIC AIMS

A. Rationale

Obesity is an enormous global public health burden. The prevalence of obesity has tripled since 1975.[1] In 2016, more than 2.5 billion adults were overweight or obese[1], and 340 million children and adolescents aged between 5 and 19 were overweight or obese.[2] Since obesity is a major risk factor for numerous health outcomes, including cardiometabolic diseases[3], the rapid increase in the global obesity burden requires immediate public health action and a better understanding of obesity pathogenicity to prevent it.

Genetic epidemiology of obesity will improve our understanding of the pathogenesis of disease and may help develop novel and effective interventions and prevention strategies. Although the current obesogenic environment has been a critical component of secular trends of increasing obesity, inter-individual variability in response to external environmental factors for obesity is largely driven by genetic underpinnings.[4] Indeed, the heritability of obesity ranges from 40% to 70%[5], and studies on obesity genomics have identified not only genes causing monogenic forms of obesity but also hundreds of obesity-associated genomic loci that primarily contribute to common polygenic obesity.[4]

Obesity genomic studies are particularly beneficial for public health because they will enable early risk prediction and targeted interventions for obesity[4] by leveraging individuals' genetic risk information. Obesity risk prediction by genetic information – which is available from birth in theory – is particularly important since obesity can begin in earlier life, and it is

difficult to reverse obesity in older children or adults.[6] Moreover, as the early intervention or prevention efforts for obesity are relatively low risk and high benefit, suboptimal risk prediction (i.e., high false positive rate) is still allowable.[4] In addition, the prediction of individuals' genetic risk for obesity will enable effective intervention strategies targeted to high-risk subgroups of individuals, enabling precision prevention. For instance, prior studies revealed that individuals' aggregate genetic risk based on the central nervous system (CNS)-associated genetic variants showed different patterns of relationships with obesity and eating behaviors compared to the aggregate genetic risk based on the non-CNS-associated genetic variants.[7,8]

Genetic epidemiologic studies of obesity can also elucidate a variety of biological mechanisms causing obesity and linking obesity to subsequent health outcomes, including cardiovascular disease (CVD), Type 2 diabetes (T2D), or cancer. Understanding of the underlying genetics and biological mechanisms can provide novel insights into the heterogeneous relationships between developing obesity and downstream complications and reveal novel drug targets for obesity and its complications.[9,10] Even at present, different obesity-causing monogenic mutations and their revealed pathways are being used in target discovery and development. For instance, in case of leptin-deficient obesity due to the mutations in the *LEP* gene, recombinant leptin is administered to treat this specific type of obesity, whereas in case of monogenic obesity related to *LEPR, PCSK1* and *POMC* deficiency can be treated with an *MC4R* agonist.[4] Likewise, novel findings on the genetic underpinnings of the biological pathways to various CVDs may provide novel insight into drug targets for weight loss but also for downstream diseases like CVD. A variety of potential biological pathways linking excess adiposity and cardiometabolic disorders – e.g., dyslipidemia, diabetes, and coronary artery disease (CAD) – have been suggested.[10,11] Genomic studies have also identified the

heterogeneous associations between some adiposity-increasing alleles and the risk of metabolic disorders.[12-18] In this regard, findings from genetic epidemiological studies that reveal underlying biological mechanisms will be leveraged for novel prevention or therapeutic targets.

Nevertheless, there are several important research gaps in obesity genomic studies, and this dissertation will focus on two important research gaps. First, there is a lack of understanding of the potential heterogeneities in the prediction performance of obesity PRS in various settings. New PRS estimation methods have been developed, but they have not been thoroughly evaluated in diverse populations. Also, potential racial/ethnic differences in prediction performance have not been fully vetted. Furthermore, various individual-level contexts, such as demographic, lifestyle, and comorbidity status, may affect the prediction performance of PRS, yet these contextual factors are understudied. Second, although each obesity-associated variant is expected to have a unique influence on obesity and cardiometabolic complications, the potential roles of obesity-associated variants in cardiometabolic disorders have not been thoroughly characterized. Indeed, we have a very limited understanding of the actual causal genes and variants that underlie obesity-associated genetic variants identified from genome-wide association studies (GWAS). Identifying the genes underlying obesity will improve our understanding of the pathophysiological pathways causing obesity and subsequent health outcomes, in particular among diverse ancestral groups where shorter haplotypes limit the variants and candidate genes to be brought forward for functional studies.

In summary, the current dissertation will address these two major research gaps in the genetic epidemiology of obesity. First, our work will characterize the context-specific performance of obesity PRS across populations. An understanding of these heterogeneities is critical as PRS moves into the clinical domain. Second, we will consider the heterogeneous

association between obesity and its cardiometabolic complications so that we can better understand the molecular mechanisms of obesity and possibly reveal molecular subtypes of obesity. In particular, the heterogeneous relationships between obesity-associated variants and dyslipidemia will be prioritized, as these relationships have been understudied compared to other cardiometabolic traits such as T2D.

Therefore, the current dissertation has two aims.

B. Aim 1

Aim 1. Characterize and evaluate the utility of trans-ancestry obesity PRS in the ancestrally diverse PAGE study.

1a. Construct the trans-ancestry and ancestry-specific PRS for overall obesity (PRS-BMI) and central obesity (BMI-adjusted WHR; PRS-WHRadj.BMI) based on the latest trans-ancestry GWAS of obesity traits in the GIANT consortium.

1b. Characterize and evaluate the predictive performance of obesity in PAGE study by different PRS estimation methods – i.e., Pruning and Thresholding (P+T) [19], PRS-CS(x)[20,21] – and by subgroups defined by self-reported race/ethnicity, sex, age groups, smoking status, physical activity status, T2D status, and hypertension status.

C. Aim 2

Aim 2. Identify genetically correlated loci that jointly influence obesity and dyslipidemia in heterogeneous directions and investigate the potential pathophysiological implications of these heterogeneous pleiotropic loci in ancestrally diverse populations.

2a. Identify genetic loci associated with both obesity and dyslipidemia risk using large-scale publicly available UK Biobank (UKB) GWAS of BMI and lipid traits. Specifically, identify genomic loci associated with higher obesity risk and lower dyslipidemia risk (Ob/DysL(–) loci) and genomic loci associated with higher obesity risk and higher dyslipidemia risk (Ob/DysL(+) loci).

2b. Investigate the biological implications of identified Ob/DysL(–) loci and Ob/DysL(+) loci discovered from UKB using Ob/DysL(–) loci- and Ob/DysL(+) loci-based PRS. Specifically, we will prioritize potential causal genes underlying Ob/DysL(–) loci and investigate the unique association patterns of the two subtypes of obesity PRS with cardiometabolic profile and CVD events in the PAGE study.

D. Hypotheses

(Aim 1) There will be heterogeneities in the prediction performance of obesity PRS by PRS estimation methods, by race/ethnicity, and by various individual-level contextual variables.

(Aim 2) Local genetic correlation analysis will identify novel genomic loci that influence both adiposity and lipid traits in heterogeneous ways. Investigation of the potential biological implications for the identified loci will help better understand the heterogeneous biological pathways related to obesity, dyslipidemia, and CVD.

E. Public Health Impact

The proposed research will fill two critical research gaps in genetic epidemiological studies of obesity. The current research will contribute to a better understanding of the pathogenesis of obesity, heterogeneity in obesity prediction across contexts, and genomic regions

that influence the risk of obesity and that also have an important impact on CVD risk factors, in this case, dyslipidemia.

CHAPTER 2: BACKGROUND AND SIGNIFICANCE

In this section, I will broadly review the current knowledge on the epidemiology and genomics of obesity to explain the background and highlight the significance of the aims of the dissertation. To accomplish this, I will divide this literature review into four parts – 1) general epidemiology of obesity, 2) obesity genetics, 3) PRS for obesity, and 4) heterogeneities in obesity complications. In the first section, I will describe definitions and measures of obesity, the burden of disease, risk factors for obesity, and biological mechanisms of body weight control. In the second section, I will summarize the current understanding of obesity genetics – heritability of obesity, monogenic obesity, and polygenic obesity. In the third section, I will explain PRS in general and PRS for obesity. Lastly, I will introduce the relationship between obesity and various cardiometabolic consequences – biological mechanisms, the heterogeneous nature of obesity consequences, and its underlying genomics.

A. Epidemiology of Obesity

A.1. Definition

Obesity is defined as excessive body fat and is usually measured by body mass index (BMI) – body weight (kg) divided by the square of height (m).[2] Although body fat is not directly measured by BMI, it serves as an easy measure and repeatable estimate.[22,23] According to the guidelines from the US Centers for Disease Control and Prevention (CDC), obesity status among adults can be categorized as normal ($18.5\ kg/m^2 \leq BMI < 25\ kg/m^2$), overweight ($BMI \geq 25\ kg/m^2$), obesity ($BMI \geq 30\ kg/m^2$), and severe obesity ($BMI \geq 40\ kg/m^2$). [2,24]

Table 2.1. Classification of adult obesity based on body mass index (BMI) measures

Classification	**BMI (kg/m^2)**
Underweight	< 18.5
Normal	≥ 18.5 and < 25.0
Overweight	≥ 25.0 and < 30
Obese:	
Class I	≥ 30.0 and < 35.0
Class II	≥ 35.0 and < 40.0
Class III	≥ 40.0

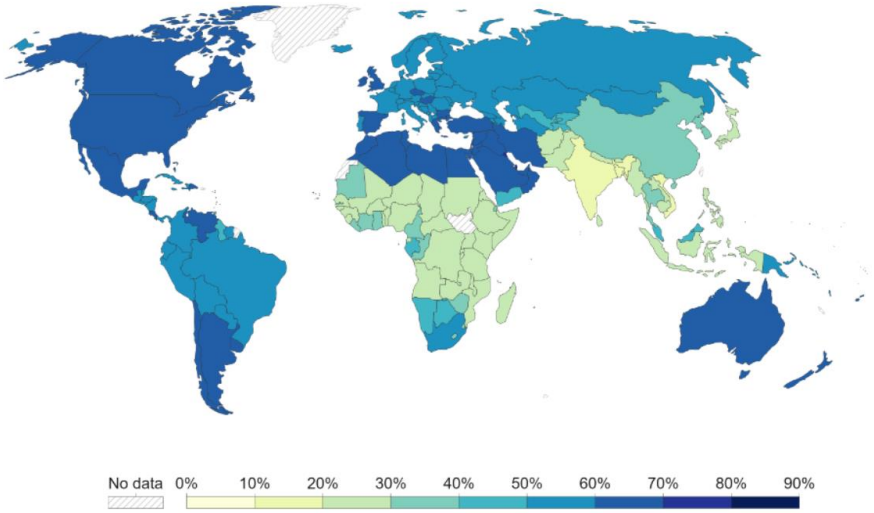
Source: CDC[24]

A.2. Burden of disease

For adults, the global prevalence of overweight and obesity was 26% and 13%, respectively, in 2016 (**Figure 2.1-A**), and it has almost tripled since 1975.[2] The increase in overweight/obesity prevalence was more dramatic among individuals aged 5-19 years, escalating from 4% in 1975 to 18% in 2016 (**Figure 2.1-B**).[2] Overweight/obesity-related deaths have exceeded the deaths related to underweight.[2] About 4 million deaths were attributable to high BMI (including people without obesity) in 2015, and more than two-thirds of the high BMI-related deaths were caused by CVD.[25] In the US, obesity-related medical expenditures were estimated at about $173 billion in 2019.[26]

In the US, NHANES 2021 reported that the prevalence of adult obesity was 41.9% in 2017 – 2020. Along with the global trend, obesity prevalence in the US has increased from 30.5% in 1999 - 2000 to 41.9% in 2017 – 2020 (c.f., from 4.6% to 9.2% for severe obesity).[27] A more serious problem is that obesity disproportionately impacts people from historically marginalized populations. The obesity prevalence among Non-Hispanic Black adults, Hispanic adults, non-Hispanic White adults, and non-Hispanic Asian adults was 49.9%, 45.6%, 41.4%, and 16.1%, respectively.[28]

A. Prevalence of overweight or obesity in 2016 among adults



B. Prevalence of overweight or obesity in 2016 among children and adolescents (5- 19 years old)

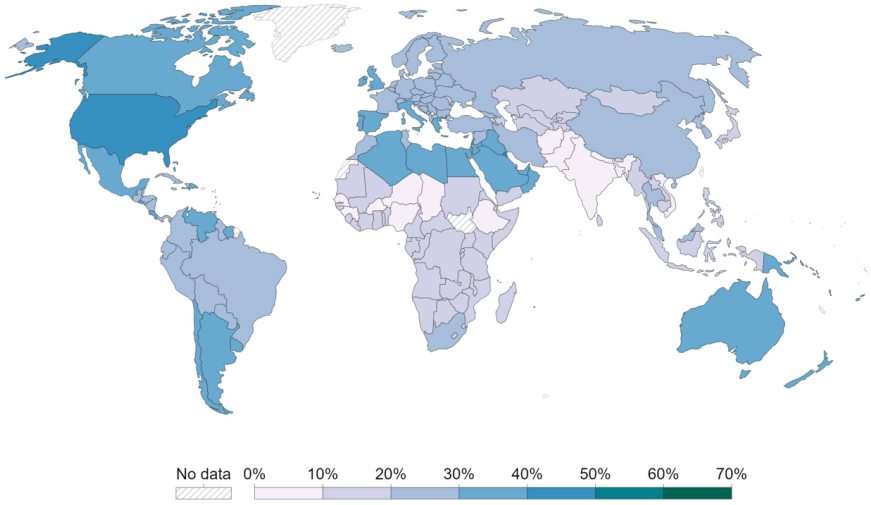


Figure 2.1. The global prevalence of overweight and obesity among adults (panel A) and children (panel B)**.** (A) Overweight or obesity among adults is defined as a BMI greater than or equal to 25. The global prevalence of overweight or obesity was 39% in 2016. At the highest (in most high-income countries), the prevalence was over 60%, and at the lowest end (e.g., South Asia and Sub-Saharan Africa), the prevalence was about 25%. (B) Overweight or obesity among children and adolescents (aged 5 - 19 years) is defined as weight-for-height greater than one standard deviation from the median of the World Health Organization (WHO) Child Growth Standards. The global prevalence was 18%, and the prevalence was greater than 10% in most areas. (Graphic adapted from Hannah Ritchie and Max Roser (2017) - "Obesity." Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/obesity' [Online Resource]; Data source: WHO, Global Health Observatory (2022))

A.3. Measures of obesity

Due to its convenience, BMI is the most commonly used proxy measure of body fat and obesity – far less expensive and intrusive than other accurate body fat measures like magnetic resonance imaging (MRI) or computerized tomography (CT). However, there are several limitations to using BMI for measuring body fat. First, since BMI cannot capture individuals' body composition, it is impossible to differentiate fat mass, which is more closely related to obesity complications, from lean mass (muscle and skeletal mass), and thus, variabilities in fatness and metabolic risk profiles within the same BMI may exist.[2,29] In addition, BMI cannot provide any information on the distribution of body fat, which is another important factor for obesity pathogenicity in addition to overall fatness. Ectopic/visceral fat deposition is more likely to lead to obesity-associated cardiometabolic disorder than subcutaneous fat deposition do[30], but the distribution of body fat was not captured by BMI. For these reasons, though it is not solely because of the limitation of BMI, obesity defined by BMI demonstrated heterogeneous cardiometabolic consequences. For instance, about 30% of people with obesity (BMI $\geq$ 30) have a normal metabolic risk profile, whereas about 23% of people with a normal BMI range have an abnormal metabolic risk profile.[31]

To complement the second limitation of BMI (inability to measure fat distribution) and to capture the fat accumulation in the abdominal area (as a proxy of visceral fat deposition), additional anthropometric measures, which are still relatively convenient to measure, such as waist circumference (or BMI-adjusted waist circumference) and waist-hip ratio (WHR; or BMI-adjusted WHR) have been used.

Dual X-ray absorptiometry (DEXA) has been accepted as a gold standard non-invasive measure of body composition, especially for fat-free mass, fat mass, and bone mineral density,

for specific regions of the body – e.g., arms, legs, and truncal region.[32] Estimates of body fat percentage by DEXA are highly accurate and reproducible, so they have been used as reference measures.[32] Also, since the exposure to radiation by X-ray is extremely low, it is considered safe for children (but not for pregnant women). However, it demonstrated a limited performance in differentiating visceral from subcutaneous fat.[32]

CT scan and MRI are known to be the most precise techniques for measuring regional (at the tissue-organ level) and whole-body adiposity.[32] These methods can differentiate visceral adipose tissue (VAT) from subcutaneous adipose tissue (SAT). MRI does not require radiation exposure, whereas CT entails exposure to radiation.[32] Both methods are much more expensive than DEXA or anthropometric measures.[32]

A.4. Risk factors

Obesity is a complex and multifactorial disease occurring when there is more energy intake than energy expenditure over an extended duration.[33] Surplus energy is generally converted to body fat and stored in adipose tissue.[33] When this happens for prolonged periods of time, it results in increased adipose tissue volume and mass. Any factors influencing energy metabolism, including dietary factors, physical activity, sedentariness, sleep, genetics, and socio-economic factors, are some typical examples of obesity risk factors.[34,35] It has been highlighted that the recent rapid increase in the global obesity burden is attributable to obesogenic behaviors and environment[36] that can be characterized as energy-dense food and physical inactivity[37-39] in the context of underlying genetic vulnerabilities.

An increase in energy intake via changes in dietary patterns and a decrease in energy expenditure via a modern sedentary lifestyle are two major promoting risk factors for obesity. First, physical inactivity has been repeatedly and consistently associated with obesity. As an

example, a recent study considered the longitudinal relationship between the daily hours spent watching TV and obesity incidence and reported that children and adolescents who watched TV more than 5 hours/day had 4.6 times the odds of being overweight among those who watched less than 2 hours/day.[40] Second, changes in dietary patterns have led to increases in daily calorie consumption. The changes included an increased intake of high-fat and carbohydrate foods and soft drinks and a low intake of fruits and vegetables. These changes were partly attributable to increased portion sizes, energy contents per serving, and lower food prices (summarized in [41]).

In addition to the changes in physical activity status and dietary pattern, socioeconomic status (SES), smoking status, and sleep duration have also been consistently associated with the risk of obesity. Lower parental SES, especially for parental education[42], maternal smoking during pregnancy[43], and a shorter sleep duration[44] were linked to an increased risk of obesity during childhood.

A.5. Biological mechanisms of body weight controls

Recent research has suggested that body weight is maintained at a set point across the life course, which is maintained through the equilibrium of caloric intake and energy expenditure, involving genetic and biological factors, environmental factors, and behavioral factors[45]. Indeed, body weight is actively defended through homeostatic regulation involving the interplay of the cognitive and executive brain functions (controlling hedonic processes) and the metabolic brain functions (metabolic processes) in response to internal and external disturbances.[46] The brain receives and processes external and internal cues and coordinates adaptive behavioral, autonomic, and endocrine responses essential for maintaining body energy balance.[46] Berthoud et al. (2017) summarized the processes into the categories of monitoring nutrients, regulation of appetite and food consumption by the nervous system, and regulation of energy expenditure by

the nervous system.[47] The brain senses the nutrients in the external environment through classical senses (visual, olfactory, auditory, and oral taste) and the nutrients absorbed into the blood through vagal sensory nerves and gastrointestinal (GI)-derived hormones. By combining external and internal information, the brain undergoes metabolic adaptations and engages in suitable behavioral responses.[47] The hypothalamus serves as a hub for regulating appetite, especially where AGRP/NPY and POMC/CART neurons interpret internal and external signals, stimulating or suppressing appetite and influencing ingestive behavior.[46,47] The brain can also be involved in pathways related to energy expenditure, including resting metabolism, thermogenesis, and physical activity.[47]

B. Genetics of obesity

Although the current obesogenic environment has been a critical component of secular trends of increasing obesity, inter-individual variability in response to external environmental factors for obesity is largely driven by genetic underpinnings.[4] Indeed, the heritability of obesity was estimated to range from 40% to 70%.[5] Genomic studies have identified not only genes causing monogenic forms of obesity but also thousands of obesity-associated genomic loci that primarily contribute to common polygenic forms of obesity.[4] In this section, I will summarize the heritability of obesity and the current evidence on monogenic and polygenic forms of obesity. However, although sometimes obesity is classified into two different categories (i.e., monogenic obesity and polygenic obesity), all obesity shares similar underlying biology.[4] In particular, the central nervous system plays an important role in both monogenic and polygenic obesity.[4] **Figure 2.2** shows the spectrum of key characteristics of monogenic and polygenic manifestation of obesity – i.e., monogenic forms of obesity are characterized by high overall genetic contribution, with a single mutation in one gene, with large genetic effects by the small number

of variants, rare, high penetrance, and less environmental influence; polygenic forms of obesity are characterized with modest overall genetic contributions, with numerous variants in or near multiple genes, small effect by every single variant, common, low penetrance, and environmental influence.[4]

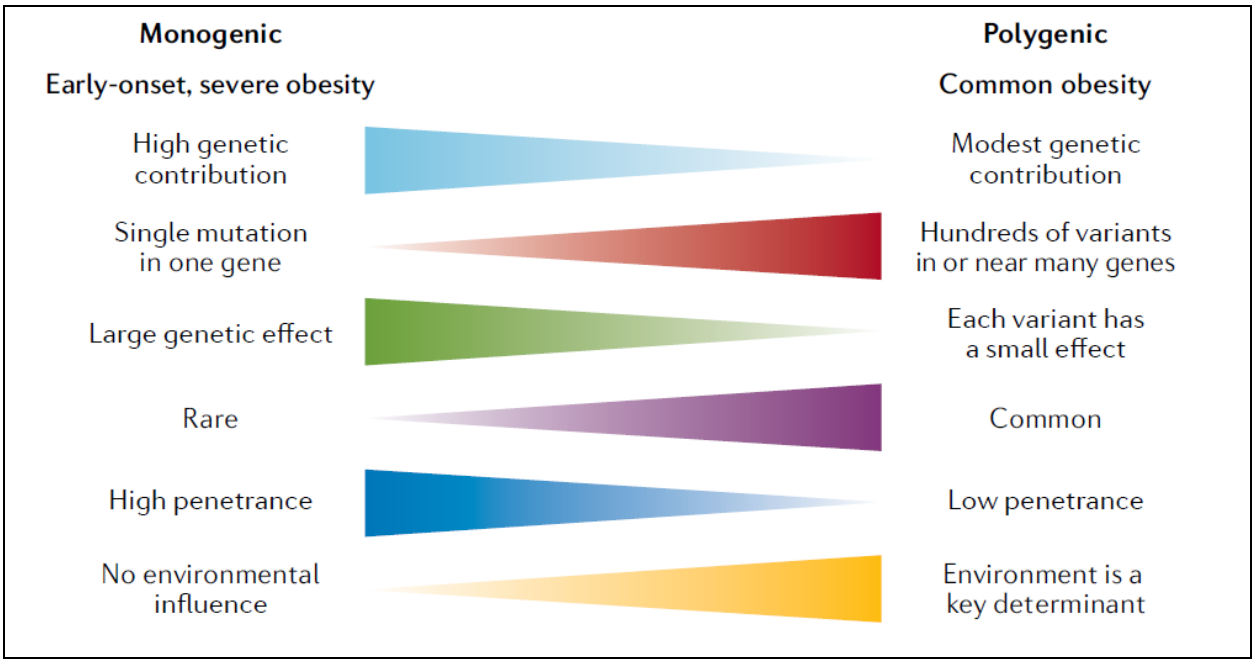


Figure 2.2. Key features of monogenic and polygenic forms of obesity**.** Monogenic forms of obesity are characterized by high overall genetic contribution, with a single mutation in one gene, with large genetic effects by the small number of variants, rare, high penetrance, and less environmental influence. Polygenic forms of obesity are characterized by modest overall genetic contributions, with numerous variants in or near multiple genes, small effect by every single variant, common, low penetrance, and environmental influence (Adapted from Loos RJF, Yeo GSH. The genetics of obesity: from discovery to biology. *Nat Rev Genet.* 2022;23(2):120-133.[4])

B.1. Heritability of obesity

Heritability is defined as the proportion of total variation in a given trait that is explained by genetic variation within a population. Heritability estimates have been used to assess if there are any genetic contributions and, if so, the amount of overall genetic contributions to a given trait. There have been studies to estimate the heritability of various obesity-related measures, including BMI, which were summarized in previous review articles[5,48]. Overall, the heritability

of obesity ranges between 40% and 50% after adjusting for age and sex.[49] However, heritability estimates are population-specific and display wide heterogeneities across study populations, study designs, and sample sizes.[49] For example, twin studies tend to have higher estimates than family studies, and studies of individuals with obesity tend to have higher heritability estimates than studies of individuals with normal weight (**Figure 2.3**).[49,50] Other than BMI, the heritability of fat mass and body fat percentage have been estimated between 40% and 50%, which is comparable to overall BMI estimates.[49,51] Visceral fat measures reveal higher heritability estimates than other regional fat depots, including the upper/lower body fat, subcutaneous adiposity, hepatic adiposity, and other ectopic fat depots (**Figure 2.3**).[49] Taken together, the body of literature reveals obesity as a highly heritable trait, supporting the study of discovery genetics.

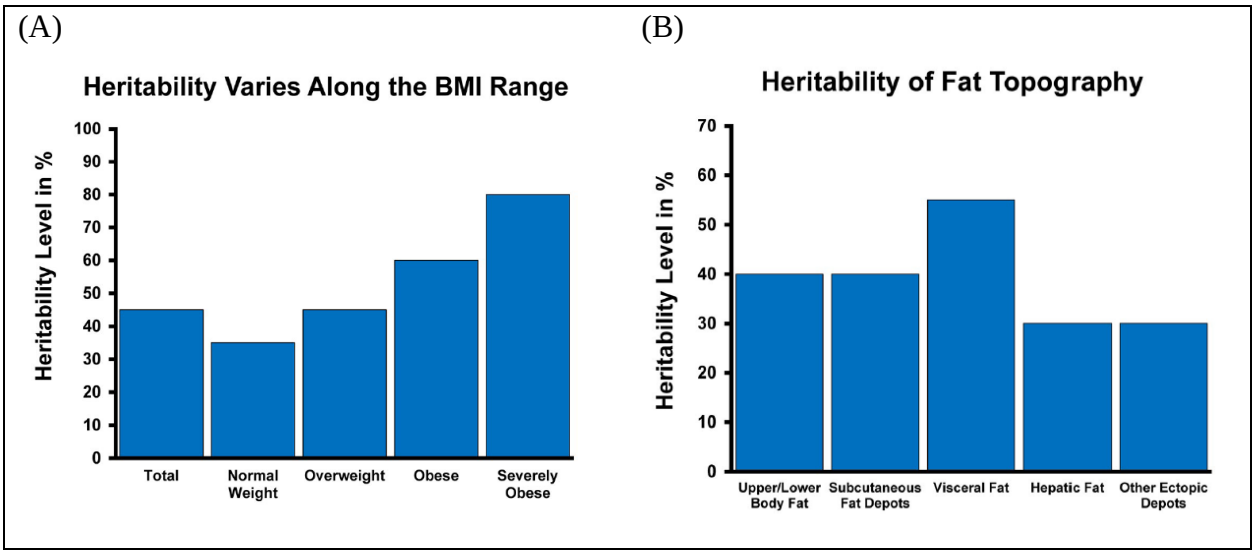


Figure 2.3. Heritability estimates of BMI by obesity classes (A) and by regional fat depot (B). (A) The heritability estimates of BMI increase linearly across obesity classes, from normal weight (~30%) to severe obesity (~80%). (B) Heritability estimates also vary by fat topography, but the evidence related to visceral fat, hepatic fat, and other ectopic fat are based on smaller studies in comparison to upper/lower fat and subcutaneous fat. (Adapted from Bouchard C. Genetics of Obesity: What We Have Learned Over Decades of Research. *Obesity (Silver Spring).* 2021;29(5):802-820.[49])

B.2. Monogenic obesity

Monogenic (non-syndromic) obesity is rare in the population, and mutations in genes (e.g., *LEP, LEPR, MC4R, POMC, or PCSK1*) within the essential energy metabolism pathways

have been identified, for example, the leptin/melanocortin pathway.[4,52,53] The proportion of severe early-onset obesity attributable to monogenic forms of obesity is estimated as less than 5% (but it can vary across different populations)[49], and it was predicted that 12,800 individuals with obesity in the U.S. are MC4R pathway-deficient due to mutation in *POMC*, *PSCK1*, and *LEPR* genes[54]. It was known that some classic intervention strategies for common forms of obesity – e.g., lifestyle modification or bariatric surgery – are not effective for those individuals who have monogenic obesity.[55] Despite its rarity and unique characteristics, studies of monogenic obesity have provided critical insights into the underlying biological mechanisms for developing obesity.[56] Biological mechanisms for some forms of monogenic obesity for leptin and the melanocortin receptor 4 genes are described below.

LEP and LEPR Leptin, an adipokine released by the white adipose tissue, has a crucial function in energy metabolisms along with its receptor (leptin receptor). In an energy surplus condition, secretion of leptin normally leads to decreases in food intake and increases in energy expenditure (e.g., through thermogenesis), resulting in weight loss.[57] This leptin-related negative feedback loop is imperative to maintain energy homeostasis and body weight.[57] Leptin-deficient mice showed high food intake and low energy expenditure and ended up developing severe obesity.[58] Mutations in *LEP* and *LEPR* lead to the development of severe obesity (reviewed in [59]). Patients with severe obesity due to congenital leptin deficiency (though it is rare) could be treated with external leptin administration.[60] Multiple mutations in *LEP* or *LEPR* - p.L72S, p.N103K, p.R105W, p.H118L, p.S141C, p.W121X, c.104_106delTCA, c.135del3bp, c.398delG c.481_482delCT, c.163C>T, and p.P316T, have been extensively studied. [59]

MC4R Obesity caused by mutations in the *MC4R* gene is the most well-described form of monogenic obesity.[61] The *MC4R* is part of the melanocortin system that regulates body weight

and energy homeostasis by modulating appetite and eating or reward-related behaviors (reviewed in [62]). The original functions of *MC4R* involve energy homeostasis - regulating energy intake and expenditure – by interacting with the brain rewarding system.[63,62] Previous studies reported that *MC4R* knockout mice showed obesity, hyperphagia, hyperglycemia, hyperinsulinemia,[64] and reduced cholecystokinin satiety response[65].

B.3. Polygenic obesity

Although monogenic obesity is accompanied by severe and early-onset forms, the most common form of obesity is polygenic obesity.[66] Common polygenic obesity is characterized by numerous common genetic factors with small effect size, and their interplay with external factors (behaviors or environment) affect the risk of developing obesity.[66] Genetic underpinnings of common polygenic obesity have been revealed through investigations of genome-wide single nucleotide polymorphisms (SNP), also known as a genome-wide association study (GWAS). In general, GWAS aims to scan the whole genome and detect common SNPs (minor allele frequency (MAF) > 5% or 1%) – rather than rare SNPs (MAF < 1%) – associated with an obesity-related trait. Since the first identification of the fat-mass and obesity-associated gene (*FTO*) as associated in 2007[67,68], more than 1,000 obesity-associated genetic variants with small effect sizes have been discovered through GWAS of BMI.[4]

In a study of more than 300,000 individuals, 97 genome-wide significant ($p < 5\times10^{-8}$) SNPs, 2,346 SNPs with $p < 5\times10^{-3}$, and about 1.3M of HapMap 3 variants accounted for 2.7%, 6.6%, and 21.6% of BMI variance, respectively.[69] In other studies, the proportion of BMI variance explained by millions of genome-wide common SNPs – SNP heritability – was estimated from 23 to 25%.[70,71] As illustrated above, while each common BMI-associated SNP

has a small effect on BMI, cumulative effects of the common SNPs explained the total BMI variance substantially.

Since 2007, more than 60 GWAS of obesity-related traits – including BMI, WHR, obesity classes, or regional fat measures – have identified more than 1,000 obesity-associated genetic variants.[4] A list of studies that reported at least one genome-wide significant ($p < 5E-8$) obesity-associated variant to NHGRI GWAS catalog (as of 11/29/2022) is shown in the appendix table (**Supplementary Table 1**). Anthropometric measures such as BMI (as a measure of overall obesity) and WHR (as a measure of central obesity) are widely used as obesity-related phenotypes. Of note, there have been two contradicting viewpoints on the GWAS of obesity; one supported the advantages of GWAS of anthropometric traits like BMI since it enabled the large sample sizes[72], and the other maintained the need for GWAS of more refined obesity phenotypes[73].

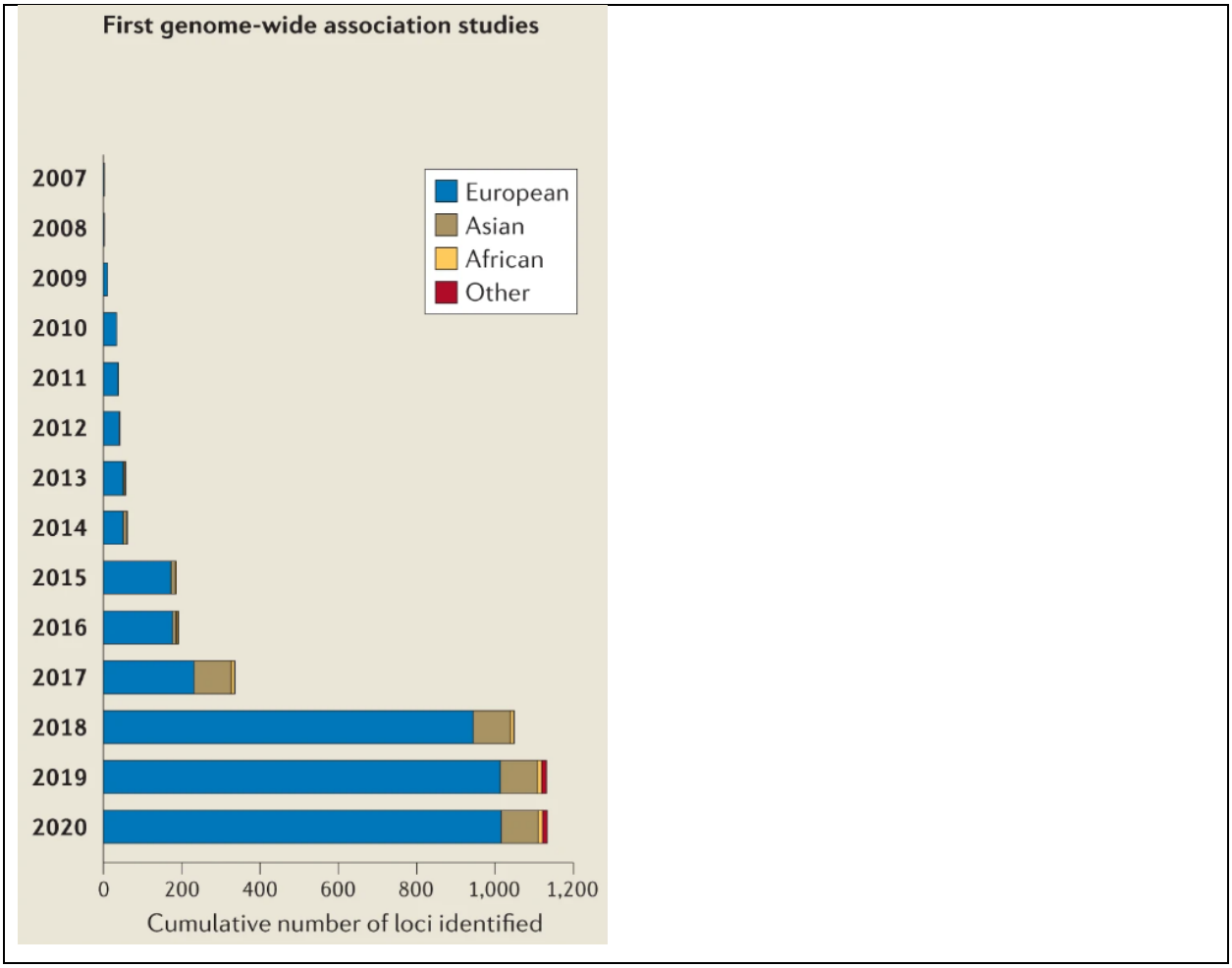


Figure 2.4. Cumulative number of obesity-associated loci identified (2007 – 2020). More than 1,000 obesity-related loci have been identified cumulatively from GWAS since 2007. Yet, most of the identified loci were discovered from populations of European ancestry. (Adapted from Loos RJF, Yeo GSH. The genetics of obesity: from discovery to biology. *Nat Rev Genet.* 2022;23(2):120-133 [4].)

GWAS of overall obesity. Several large-scale meta-analyses of BMI have been conducted to elucidate the genetic architecture of overall fatness. As introduced earlier, the Genetic Investigation for Anthropometric Traits (GIANT) consortium meta-analyzed the association results for BMI in 339,224 individuals (mostly (~95%) of European descent) from 125 studies (82 with GWAS results and 43 with results from Metabochip) and identified 97 genome-wide significant loci (56 of novel associations).[69] A more recent study combined the summary statistics from the previous GIANT BMI GWAS and GWAS of BMI in the UK Biobank (sample

N ~ 450,000) and meta-analyzed them (a total sample N ~ 700,000).[74] The study identified 941 approximately independent BMI-associated SNPs, including 751 novel loci (at a more stringent genome-wide significant $p < 1 \times 10^{-8}$). [74]

GWAS of BMI have revealed genes involved in novel pathways and provided crucial biological implications for obesity etiology that had not been discovered from the studies on the monogenic forms of obesity.[75] Functional analyses – e.g., enrichment analysis – conducted for the BMI-associated loci identified by GWAS revealed a large proportion of genes involved in CNS-related processes. Specifically, among the 31 significantly enriched tissues, 27 were parts of the CNS, including the hypothalamus and pituitary gland (appetite-related), hippocampus, and limbic system (learning, cognition, emotion, and memory-related).[69] These findings describe a critical relationship among the brain, behaviors, and energy balance.[69,76] Unlike monogenic obesity mutations, a great number of BMI-associated variants are located in regions of the genome that are non-coding or regulatory.[76]

While most GWAS of BMI focused on common variants (MAF > 5%), there have been some studies focusing on rare or low-frequency coding variants. Turcot et al. (2018) conducted association analyses to identify rare or low-frequency (MAF < 5%) coding SNPs associated with BMI using an exome array (number of variants ~ 246,328).[77] The study meta-analyzed summary results of more than 700,000 individuals from 125 studies and discovered 14 coding variants in 13 genes, a part of which was newly implicated in obesity biology. As expected, discovered rare variants demonstrated greater effect sizes (~ 10 times) than common variants did.[77] In addition, a recent study analyzed the whole exome sequencing data from 645,626 individuals (428,719 of European descent from the UK Biobank, 121,061 of European descent from the MyCode Community Health Initiative cohort, and 95,846 of admixed population from the Mexico City

Prospective study) and identified 16 BMI-associated genes.[78] The presence of rare nonsynonymous variants in the genes was linked to BMI, which includes five brain-expressed G protein-coupled receptors (*CALCR, MC4R, GIPR, GPR151*, and *GPR75*).[78] Among the identified genes, *GPR75* (1.8kg/m^2 lower BMI among carriers on average) was further investigated using knock-out mice models.[78] *Gpr75* knock-out mice displayed less weight gain in a high-fat diet model compared to the wild-type mice.[78] Functional analyses from the studies on coding variants also highlighted the importance of CNS-related pathways in overall adiposity.[77,78]

In addition to BMI, BF% has been studied as an estimate of overall adiposity in much smaller sample sizes than for BMI. As an example, a GWAS of BF% with more than 100,000 individuals identified 12 genetic loci (8 previously reported for BMI and BF% and four novel associations).[15] A group of loci among the BF%-associated loci were more strongly associated with BMI in comparison to BF%, or vice versa, suggesting some distinct genetic effects for BF% and BMI.[15]

GWAS of central obesity. In addition to overall obesity (primarily measured by BMI), many studies have attempted to identify genetic loci specifically associated with fat distribution, especially central obesity. BMI-adjusted WHR (WHRadj.BMI) is a well-known anthropometric proxy measure for body fat distribution and central obesity. The GIANT consortium conducted meta-analyses of GWAS results for WHRadj.BMI in up to 224,459 individuals (142,762 individuals from 57 cohorts with GWAS data and 81,697 individuals from 44 cohorts genotyped on the Metabochip) and identified 49 (33 novel) loci associated with WHRadj.BMI.[79] The gene-set enrichment analyses suggested that fat distribution is closely related to adipose tissue biology (adipogenesis, angiogenesis, and transcriptional regulation) and insulin resistance.[79] Also, a substantial portion of the WHRadj.BMI-associated loci (20 of 49 loci) demonstrated significant

differences by sex – most of the sexually dimorphic loci had a stronger influence among females.[79] Then, a follow-up large-scale study combined the GIANT GWAS results and the UK Biobank GWAS of WHRadj.BMI (a total of more than 690,000 individuals) and reported 463 genome-wide significant associations (spanning 346 loci).[80] As noted previously, a large proportion of the associations (in Pulit et al., 105 associations were dimorphic) were sexually dimorphic, and females tended to have more associated variants.[80] Furthermore, a study with ExomeChip (i.e., only included variants in protein-coding regions) in 344,369 individuals identified a total of 56 coding variants significantly associated with WHRadjBMI, and 31 of them were associated specifically with WHRadjBMI and not with BMI.[81] Some studies conducted GWAS on more accurate regional fat distribution – e.g., VAT and SAT – measured by CT, MRI, DEXA, and BIA[82-85], and they suggested several novel variants and related pathways. Also, tissue enrichment analyses revealed that unlike BMI, body fat ratio-associated genes were not enriched in CNS tissue gene sets and showed different patterns of enrichment.[85]

C. **Genetics of obesity in diverse populations**

One major limitation of genetic studies of obesity (and human genetics in general) is the underrepresentation of diverse populations. A majority of genomic studies (including the genetic studies of obesity) have been conducted in individuals of European descent – more than 80% of study participants included in the NHGRI GWAS catalog are of European ancestry (as of 2016).[86] Among non-European ancestry groups that are present, East Asians are the most widely studied. The proportion of Asian populations in the GWAS catalog increased from 3% in 2009 to 14% in 2016, and 64% of those were of East Asian ancestry specifically.[86] In terms of obesity specifically, large-scale GWAS for obesity-related traits identified several novel BMI-associated loci.[87,88] Several large genomic consortia for ancestrally diverse populations have been published

recently, especially for African ancestry and admixed ancestry (Hispanic/Latino) populations – e.g., African Ancestry Anthropometry Genetic Consortium (AAAGC), Population Architecture Using Genomics and Epidemiology (PAGE), and Hispanic/Latino Anthropometry(HISLA) Consortium.[86]

AAAGC conducted large-scale genome-wide meta-analyses and replication analyses in up to 52,895 individuals of African ancestry for BMI and up to 23,095 individuals for WHRadj.BMI.[89] The study reported ten genome-wide significant associations for BMI (three novel associations (*IRX4/IRX2*, *INTS10/LPL*, and *MCL1*) and four genome-wide significant associations for WHRadj.BMI (three novel associations near *TCF7L2/HABP2*, *SSX2IP*, and *PDE3B*). When combined with European GWAS, there were additional novel loci (*SPRYD7/DLEU2, CASC8,* and *ZDHHC1/HSD11B2*) for WHRadj.BMI. In addition, when the GWAS results for African ancestry were added to those for European ancestry, the fine-mapping analyses yielded more tractable credible sets (containing ≤ 20 variants) than for the European ancestry results only.[89] The study findings highlight the need for increased ancestry-diverse obesity genetics studies.[86]

The PAGE study was developed to conduct genetic epidemiological studies in ancestrally diverse populations in the US. The PAGE study was drawn from several existing population-based cohort studies and hospital-based biobank data – Hispanic Community Health Study/Study of Latinos (HCHS/SOL), Women's Health Initiative (WHI), Multiethnic Cohort (MEC), Coronary Artery Risk Development in Young Adults (CARDIA), Multi-Ethnic Study of Atherosclerosis (MESA), Atherosclerosis Risk in Communities Study (ARIC), and the Icahn School of Medicine at Mount Sinai BioMe biobank (www.pagestudy.org). Gong et al.(2018) conducted a trans-ethnic GWAS for BMI in more than 102,000 European American, African

American, Hispanic/Latino, Asian American, and Native Hawaiian populations.[90] Individuals were genotyped on ~200,000 SNPs on the Illumina Metabochip and imputed to the 1000 Genome Projects Phase 1.[90] The study replicated 15 of 21 known BMI loci available for the Metabochi and discovered two new loci (at the Metabochip-wide significance level $p < 2.5E-7$).[90] Recently, the PAGE investigators, along with other academic collaborators, designed Multi-Ethnic Genotyping Array (MEGA) to improve the coverage of non-European genetic variation. Using the genetic data in 49,839 non-European individuals genotyped on the MEGA, the PAGE investigators conducted a GWAS of 26 clinical and behavioral traits, including BMI and WHRadj.BMI.[91] Two novel loci were identified for BMI and WHRadj.BMI (one for each trait). The identified BMI SNP was more common in African ancestry populations (minor allele frequency (MAF) of 0.08) and in Hispanic/Latino populations (MAF of 0.01), in comparison to Native American ancestry populations (MAF of 0.001), Asian ancestry populations (absent), and primarily European ancestry (in 1000 Genome reference) population (absent).[86,91]

The HISLA consortium was formed to address the paucity of genomic studies of obesity in Hispanic/Latino populations.[86] The consortium included more than 23 studies as well as two consortia (the Slim Initiative in Genomic Medicine for the Americas (SIGMA) consortium and the Consortium for the Analysis of the Diversity and Evolution of Latin America).[86] Fernandez-Rhodes et al. (2022) conducted a GWAS of anthropometric traits in HISLA (59,771 for stage 1 discovery meta-analysis and 10,538 for stage 2 replication meta-analysis) and identified one novel BMI loci (*PAX3*) and two novel signals in established loci for BMI (rs17361324 in *ADCY5* and rs148899910 in *ILRUN*).[92] When combined with AAAGC and the GIANT consortia, three novel BMI one novel WHRadj.BMI loci were identified, and three novel signals were established loci for BMI and two for WHRadj.BMI.[92] The study also found that trans-ancestral

meta-analysis demonstrated a small-to-moderate influence of residual population stratification on the SNPs' estimated effect sizes.[92] The findings of the study provided additional insights into the genetic underpinnings of obesity-related traits and highlighted the importance of including diverse populations.[92]

D. Polygenic risk scores for obesity

D.1. Genetic risk prediction for complex traits using polygenic risk scores

For the past decade, GWAS of complex traits has identified numerous associated genetic variants, especially as a form of SNP. Results of the GWAS have revealed that many complex traits have a polygenic nature, which is influenced by thousands of SNPs with small effect sizes.[93,94] In order to measure individuals' genetic predisposition to polygenic traits, PRS (also known as polygenic score (PGS) or genetic risk scores (GRS)), as an aggregate genetic risk measure, was suggested. Generally, PRS is defined as a weighted sum of the number of risk alleles from a set of selected SNPs.[95] Thus, information on the risk allele of a certain SNP and its effect size (i.e., weight) is required to construct PRS, and the information is inferred from the results of GWAS analyses.

$PRS_i = \sum_{j=1}^{m} x_{ij}\widehat{\beta_j}$ [96]

PRS_i: PRS for ith individual
x_{ij}: the genotype for ith individual and jth SNP (0, 1, or 2)
$\widehat{\beta_j}$: the estimated effect size of jth SNP (from GWAS summary statistics)
m: the number of SNPs selected for PRS construction

With the increased availability of large GWAS summary statistics and individual genotype data in many cohort studies and biobanks, the number of publications on PRS has

rapidly increased[97], and evaluating the potential utilities of PRS has become an actively studied area. Potential utilities of PRS include risk prediction and stratification, disease subtyping, individualized intervention, and dissecting disease biology.[98]

One major utility of PRS is disease risk prediction and risk stratification (i.e., identifying those most at risk).[96] PRS can be used as another risk factor of a certain health outcome in addition to the existing risk factors, and adding PRS to the existing risk prediction model can improve the accuracy of risk prediction.[95] For instance, the predictive accuracy for coronary heart disease was improved by the addition of PRS to the existing Framingham risk score and the ACC/AHA13 scores.[99] Despite the current low predictive accuracy, the upper limit of PRS's accuracy, in theory, is determined by the SNP heritability – the proportion of phenotypic variance explained by SNPs in GWAS – of a given trait.[95,97] In terms of risk stratification, studies showed that PRS can identify a greater number of high-risk groups whose risk is comparable to rare monogenic mutation. If there are population-based screening and preventive measures, implementation of PRS is particularly of interest and could benefit public health.[95] One unique feature of PRS compared to other risk factors is that genetic risk can be available at birth and is not influenced by other environmental factors (but PRS can vary by PRS estimation methods or GWAS summary statistics).

In addition, PRS can be used in subtyping diseases. A previous study on PRS for T1D highlighted the utility of T1D PRS in discriminating T1D from T2D. [100] Similarly, a study on breast cancer also suggested the potential of PRS in disease subtyping (estrogen-receptor-positive or – negative) by developing subtype-specific PRS.[101]

Apart from the abovementioned clinical utilities, PRS can help elucidate underlying disease biology. Since obesity is a major risk factor or predictor of numerous health outcomes,

PRS for obesity, as a genetic instrumental variable, can be utilized to assess the causal relationship between obesity and correlated health outcomes.[102]

D.2. PRS estimation methods

Recently, numerous PRS estimation methods have been developed and assessed.[20,21,103-107] In this proposal, the two major PRS estimation methods are described. One is the Pruning and Thresholding (P+T) method, which is the most commonly and widely used method. The other method is PRS-CS[20] (and PRS-CSx[21] as an extension of PRS-CS), which has been reported as one of the best methods in many previous studies.[97]

P+T is considered as a basic method and has been widely used for many traits[103], so it has been used as a benchmark method in many PRS method developing studies. P+T sets a certain p-value threshold to filter in SNPs with significant effects on the trait and utilizes the LD clumping process with a certain LD r^2 threshold to remove the correlated SNPs.[108,109] It uses the effect sizes from GWAS as weights for PRS.[110] Multiple p-values thresholds are applied in a tuning population, and a p-value with the highest accuracy will be chosen.[109] The underlying assumption of the P+T method is that selected SNPs are not correlated with each other, and they independently and additively affect the trait of interest.[109]

For better effect size estimation and prediction accuracy, PRS-CS has been developed based on the Bayesian framework, which considers all genome-wide markers simultaneously to calculate each variant's posterior effect size.[20,109,111] It applies one hyper-parameter, the global shrinkage parameter, and a continuous shrinkage prior to the effect sizes of the variants.[111] For the global shrinkage parameter, in PRS-CS, it was optimized through grid-search (partial Bayesian approach), whereas, in PRS-CS-auto, it was learned from GWAS summary statistics through a fully Bayesian approach and placed with a half Cauchy prior.[20,97] An independent

gamma-gamma prior was assigned to the local shrinkage parameter.[20] PRS-CSx is an extension of PRS-CS, enabling the incorporation of the multiple GWAS summary statistics from different populations, and it showed better prediction performance for the ancestrally diverse populations.[21] Since PRS-CS utilizes external LD reference (e.g., 1000G), if there are systemic differences in LD structure between the GWAS populations and the reference panel, the predictive performance is expected to be reduced.[97]

The P+T method implicitly makes a sparsity assumption that only a certain proportion of SNPs has non-zero effect sizes, and the rest of the SNPs have exactly zero effect sizes so that a sparse set of SNPs affects the trait of interest.[21] On the contrary, PRS-CS makes a polygenic assumption that all SNPs have non-zero effects on the trait of interest.[21] As illustrated above, each PRS method has distinct assumptions for the genetic architecture – e.g., distributions and effect sizes of casual variants[103 109] and relies on different algorithms to compute the effect estimates[97], so the best performing PRS methods can vary depending on the actual genetic architecture of the trait of interest or study settings. A proper selection of the PRS method is important for prediction accuracy because the prediction accuracy could be reduced by the imprecise effect size estimation for each SNP.[111,112]

D.3. Current status of polygenic risk prediction for obesity

As described above, obesity is a major contributing factor to various cardiometabolic disorders, and the rapid increase in obesity prevalence is a significant public health threat. Obesity can begin in earlier life, and it has a long-term influence on cardiovascular health in later in life [113]. Also, it is difficult to reverse obesity in older children or adults.[6] In this regard, it is crucial to predict the risk of obesity before its onset and to implement effective preventive strategies for those with high obesity risk.[114]

Despite robust associations between GWAS loci and obesity traits, early studies generating a GRS for obesity using only known SNPs performed poorly– i.e., the proportion of variance in BMI explained by GWAS variants ranged from 0.34% to 2.70% (reviewed in [114]). In contrast, recent studies have demonstrated considerable improvements in the predictive performance of obesity-related PRS. For example, Khera and colleagues in 2019 constructed PRS for BMI (PRS-BMI), including more than 2 million variants, and reported a strong correlation between BMI and PRS-BMI.[70] Individuals within the top 10% of the PRS-BMI had 2.9kg/m^2 higher BMI than those within the lowest 90% of the PRS-BMI. Moreover, the OR for severe obesity was 4.2 (top 10% of PRS-BMI vs. 90% PRS-BMI).[70] In the same study, the correlation between BMI and PRS-BMI was 0.29.[70] Another prospective study demonstrated that although BMI at a specific time point (rather than a PRS-BMI) tended to be a better predictor of future BMI, PRS-BMI displayed significant additional explanatory capacity in the prediction model.[115] In summary, the prediction performance of obesity has been improved by an increased GWAS sample size and by novel PRS estimation methods, and the obesity PRS provides additional explanatory capacity to the existing prediction models with traditional risk factors.

Obesity, defined by BMI, often tends to be treated as a uniform condition; however, obesity is heterogeneous in many ways – e.g., monogenic vs. polygenic, severity (severe vs. mild), age of onset (early onset vs. late onset), and cardiometabolic complications (obesity with complications vs. obesity without complications).[114] (See more in next chapter) Thus, in order to precisely predict the risk of different forms of obesity and its subsequent complications and to implement targeted prevention strategies, genetic underpinnings of these various aspects of obesity should be thoroughly investigated as well. There have been attempts to assess the genetic risk prediction for different conditions of obesity. For example, a previous study reported that the

discriminative accuracy of obesity PRS increased with obesity severity (from obese class 1 to obese class 2).[116] Also, another study investigated the correlation between obesity PRS and weight in different age groups and the distribution of obesity PRS in different weight categories (i.e., underweight, normal, overweight, obese, and severely obese).[70]

Previous studies on obesity PRS highlighted the differences in predictive accuracy between different ancestry groups – i.e., better performance among European ancestry and limited performance among non-European, especially African ancestry[115,117] – possibly due to different genetic architecture, LD structure, allele frequencies (different tagging SNPs). This is a critical component of my research and is addressed in section F. Research gaps.

Early identification of high-risk groups for obesity at a young age could be transformative, as many downstream diseases result from obesity, including CVD, T2D, cancers, etc.[114] However, many non-invasive prevention strategies like lifestyle changes are not effective in the long term, and it is possible that information on high risk by genetics may not promote preventive behaviors effectively.[114,118,119] Nonetheless, the identification of genetic risk factors for obesity has the potential to revolutionize the drug market with the development of targeted therapies.

E. **Obesity and cardiometabolic consequences**

In this section, I will describe the relationships between obesity and cardiometabolic health outcomes and the genetic underpinnings of these relationships. First, I will summarize the current literature on the causal roles of obesity in the pathogenesis of various CVDs, as reported in MR studies. Then, I will introduce some important potential biological pathways from obesity to CVD. In the following section, I will address the heterogeneous influence of obesity on CVD

despite the overall close link between obesity and CVD risk. Lastly, I will describe how genetic studies have helped us better understand the heterogeneous impact of obesity on CVD risk and summarize the findings to date.

E.1. MR Studies of Obesity and CVD.

Obesity is a major risk factor for cardiometabolic risk factors and CVD. From epidemiological studies, it has been well-established that excess body weight is closely related to CVD[120] and its risk factors, including elevated blood pressure[121], diabetes[122], and high blood cholesterol level[123]. However, despite the strong and consistent close associations between obesity and CVD or CVD risk factors, the causal relationships have been less certain, partly due to the limitations of observational studies. Mendelian randomization (MR) provides an opportunity for causal inference using genetic instruments that, by definition, are non-confounded. Several MR studies have leveraged the results of GWAS of obesity to elucidate the causal relationship between obesity and CVD or CVD risk factors. As expected, the results of the MR studies were mostly supportive of the causal roles of obesity in most CVD risk factors (e.g., T2D, fasting insulin, systolic blood pressure (SBP), diastolic blood pressure (DBP), triglycerides (TG), and high-density lipoprotein (HDL) cholesterol). (Summarized in[10]) Also, MR studies, in general, have provided supportive evidence of the causal roles of obesity in CVD events but with some divergent evidence gradients by different types of CVD – i.e., strong evidence for aortic valve stenosis, heart failure, deep vein thrombosis, hypertension, atrial fibrillation, and peripheral artery disease and low-level evidence for subarachnoid hemorrhage, abdominal aneurysm, intracerebral hemorrhage, ischemic stroke, transient ischemic attack, and stroke.[124]

E.2. Biological mechanisms for metabolic consequences of excess adiposity

As described above, it is widely accepted that obesity plays a causal role in various CVD risk factors and CVD outcomes. Multiple biological mechanisms have been posited to explain the pathway from excess adiposity to CVD outcomes. The biological mechanisms include (not an exhaustive list): 1) systemic inflammation, 2) neuroendocrine factors, 3) endothelial dysfunction, 4) hemostatic factors, and 5) ectopic fat deposition (documented in [125]). Following is a summary of each category.

A typical feature of obesity is increased systemic inflammation, which can partly explain the link between obesity and CVD. Several potential mechanisms have been proposed to explain the relationship between obesity and inflammation (reviewed in [125]). First, due to adipocyte hypoxia, adipose tissue may induce inflammatory responses, including increased secretion of IL-6, leptin, and TNF-α.[126] Second, because of hypoxia or surplus nutrients in obesity, unfolded proteins can accumulate in the endoplasmic reticulum (ER) (so-called ER stress), which leads to stimulation of the inflammatory response through NF-κB-IκB kinase and JNK-AP1 pathways.[127] Third, intensified level of systemic inflammation in obesity might be due to the release of free fatty acid (FFA) by lipolysis[128], and the increase in FFAs induces a lipotoxic state, oxidative stress to the ER, and further pro-inflammatory response through TLR2/4 and JNK signaling pathways.[125,128]

Second, as a part of the endocrine system, adipose tissue secretes numerous adipokines such as leptin, adiponectin, interleukin-6 (IL-6), and TNF-α, and these adipokines play essential roles in energy homeostasis. Excess fat deposition in ectopic sites can lead to dysregulation of the adipokine profile, and this may cause an atherosclerotic response and subsequent CVD.[129,130] People with obesity tend to have elevated leptin levels as a manifestation of hyperleptinemia or leptin resistance. Elevated leptin levels lead to increased CRP levels and oxidative stress in

vascular endothelial cells, which in turn leads to atherogenic responses.[131] Also, leptin is involved in the renal sodium secretion pathway (by regulating urinary excretion of nitric oxide metabolites) so that it can influence vascular tone, blood pressure levels, and atherogenic response via dysregulated blood pressure.[132]

Third, endothelial dysfunction – i.e., an early marker of atherosclerotic disease – could be a link between obesity and CVD. Among people with obesity, individuals become insulin resistant and central wave reflection is impaired[133], and endothelium-dependent vasodilation was significantly dysregulated.[134] Elevated FFA among people with obesity is known to play a significant role in developing endothelial dysfunctions through insulin resistance, inflammation, and oxidative stress (reviewed in [135]).

Fourth, thrombosis/blood coagulation is also a potential mediating factor for the association between obesity and CVD. Chronic inflammation and impaired fibrinolysis (i.e., degradation of the fibrin clot by plasmin) are two major pathways through which obesity is closely associated with elevated thrombosis.[136] In addition, adipokines and microRNA are modifying factors for the association between obesity and thrombosis.[136]

Lastly, ectopic fat deposition, especially for epicardial adipose tissue, could contribute to the development of CAD via elevated levels of pro-inflammatory cytokines directly secreted from the epicardial adipose tissue (as a paracrine system).[125] Also, excessive fat depots in other ectopic sites – e.g., abdominal, heart, and liver – lead to increased circulating blood volume and pro-inflammatory cytokines, which may result in a higher stroke volume, cardiac wall stress, and myocardial injury.[137]

The above-described biological mechanisms linking obesity and CVD risk factors are only the most prominent pathways. There will be many more pathways to be revealed as

epidemiological and biological studies accumulate more relevant evidence. It should also be noted that these pathways can be bidirectional; for example, there is also excellent evidence that increased inflammation causes obesity.[138]

E.3. Heterogeneity in cardiometabolic consequences of obesity

Despite the close association between obesity and CVD (and its risk factors), there is substantial heterogeneity in the cardiometabolic consequences of obesity. So-called metabolically healthy obesity (MHO) and metabolically at-risk normal weight (MARNW) are two examples of such heterogeneity. A NHANES study recently reported that 31.7% of people with obesity were metabolically healthy, whereas 23.5% of people with normal weight were metabolically unhealthy.[31] There have been more than 30 different definitions of MHO used in different studies[139], but most commonly, MHO has been defined as having ≤ 1 component among the following abnormal metabolic profiles[140] – elevated blood pressure, TG levels, fasting glucose levels, and low HDL cholesterol levels (or ≤ 2 components when including high waist circumference as another factor).[141]

Many studies have suggested that the risks of developing obesity-related diseases vary among those who are MHO, MARNW, MHNW and MARO.[142] Most studies suggest that MHO is more likely to develop cardiometabolic disorders (e.g., CVD, cerebrovascular disease, hypertension, insulin resistance, and T2D) than the metabolically healthy normal weight (MHNW) group but less likely than metabolically at-risk obesity (MARO) group (summarized in [142]). Some studies have suggested that MHO is a mere transitional state from MHNW to MARO.

Although some demographic or lifestyle factors – e.g., younger age, female, non-Hispanic black race/ethnicity, relatively lower BMI and waist circumference, and healthier lifestyle among MHO groups compared to MARO groups – can explain these phenotypes in part[31,142], adjusting

for these known factors does not remove differences among MHO, MARO, MARNW, and MHNW.[142] This suggests that there are biological mechanisms explaining the heterogeneous consequences of obesity. Indeed, as described in the previous section, there are numerous biological pathways from excess adiposity to cardiometabolic consequences, and the heterogeneous biological responses to excess adiposity can influence the inter-individual variabilities in obesity-associated complications.

E.4. Genetic investigations for heterogenous cardiometabolic consequences of obesity

BMI-associated genetic variants and genes could play unique roles in different biological mechanisms. As recent large-scale genomic studies (e.g., GWAS) have contributed to better biological understandings of disease pathogenesis by characterizing the disease-associated genetic variants, characterizing pleiotropic SNPs with obesity-increasing and lipid-lowering effects, and vice versa, they could provide important biological and mechanistic implications on the observed MHO or MARNW phenotypes. To be specific, although there has not been a GWAS study specifically on the MHO or MARNW phenotypes, some previous studies have identified genetic variants demonstrating counter-intuitive associations with adiposity and cardiometabolic profiles – i.e., an allele of a given SNP is associated with increased adiposity but with 'favorable' or 'protective' cardiometabolic profile (e.g., lower T2D risk, TG levels, glucose levels, or blood pressure levels).[142] Identifying the pathways that underlie these shared counterintuitive genetic effects[142] may provide insights into the manifestation of the MHO phenotype. The following section summarizes the previous genomic studies which have addressed the metabolic heterogeneities of obesity.

One notable example of an obesity-increasing allele that is also protective against cardiometabolic disease is rs2943650-C of the *IRS1* locus. This variant was identified as a BF%-

associated variant (rs2943650-T), but the BF%-decreasing allele of the variant was associated with an increased risk of an abnormal metabolic profile – e.g., insulin resistance, dyslipidemia, diabetes, and CAD.[13] Further study showed that the BF%-increasing allele of rs2943650 was associated with higher SAT but not associated with VAT. Thus, it can be inferred that the genetic variant can lead to an increase in BF% due to fat accumulation in SAT.[13] Additional GWAS on BF% identified other variants (rs6738627 in *GRB14*, rs3761445 in *PLA2G6*, and rs6857 in *TOMM40*) whose BF%-increasing allele were also associated with protective cardiometabolic profile.[15] One variant (rs6738627), similarly for rs2943650 near *IRS1*, was thought to play a role in influencing insulin sensitivity via the regulation of body fat distribution. Two other variants (rs3761445 and rs6857) from that study may be involved in different pathways to impact higher BF% but more protective cardiometabolic profile than through body fat distribution.[15]

Also, in Scott et al. (2014), an insulin resistance genetic score was derived from 10 fasting insulin-associated variants demonstrating an association with lower HDL and higher TG based on the findings from a previous study[143]. The insulin resistance score was associated with decreased BMI and gluteofemoral fat mass and with increased ALT and γ-glutamyl transferase.[17] These findings suggest an independent role of insulin resistance and fat distribution, not mediated through BMI, in developing T2D.[17]

Several other studies have been conducted to explicitly identify or characterize variants associated with both risk of obesity (e.g., BMI) and risk of cardiometabolic disorders (e.g., T2D) but in counterintuitive directions – i.e., adiposity-increasing variants associated with protective cardiometabolic profiles.[12,16-18,144,145] These studies reported several SNPs that were associated with both obesity and protective cardiometabolic profiles, and those SNPs were listed in **Supplementary Table 2**.

First, among 19 previously identified fasting insulin-associated variants[143,146], Yaghootkar et al. (2014) grouped a cluster of 11 variants and showed that the 11 variants-based genetic risk score was associated with lipodystrophy-like metabolic profiles (i.e., increased fat accumulation in the visceral area compared to subcutaneous area, higher risk of T2D, hypertension, and CAD, but lower BMI).[145] This finding was replicated in UK Biobank data (N = 164,609), building on the evidence that some genetic variants are associated with higher overall adiposity but with protective metabolic profiles, possibly through the capacity of body fat accumulation.[18]

Lotta et al. (2017) utilized GWAS summary statistics for insulin resistance-related traits including fasting insulin, HDL cholesterol, and TG) and identified 53 insulin resistance loci (43 of them were novel) by aligning the risk alleles from the three GWAS results for higher fasting insulin, higher TG, and lower HDL.[14] GRS based on these 53 aligned variants was associated with an increased risk of T2D and CHD, lower BMI and BF%, and higher WHR, and it also supported the hypothesis that a limited subcutaneous fat storage capacity can lead to insulin resistance.[14]

Ji et al. (2018) identified 14 variants (7 novel variants) that showed a "favorable adiposity" pattern among the 33 significant associations from BF% GWAS and a multivariate GWAS for a group of metabolic traits (body fat percentage (BF%), HDL cholesterol, adiponectin, sex hormone-binding globulin, TG, fasting insulin, and alanine transaminase (ALT)).[144] Martin et al. (2021) implemented a similar approach with an increased number of samples in a multivariate GWAS and identified 254 variants showing significant associations from both BF% GWAS and the multivariate GWAS. Then, the 254 variants were grouped into 36 favorable adiposity (FA) variants and 38 unfavorable adiposity (UFA) variants using a k-means clustering approach.[16]

Lastly, a recent study by Huang et al. (2021) conducted pair-wise cross-phenotype meta-analyses for pairs between 3 adiposity traits (BMI, WHR, and BF%) and eight cardiometabolic

traits (HDL, low-density lipoprotein (LDL), TG, fasting insulin, fasting glucose, SBP, CAD, and T2D) using 11 publicly available GWAS summary statistics to identify variants associated with higher adiposity but with protective cardiometabolic profile.[12] Follow-up analyses suggested potential pathways linking the adiposity-increasing variants and protective cardiometabolic profiles such as fat distribution, adipocyte function, insulin-glucose signaling, energy expenditure, fatty acid oxidation, browning of white adipose tissue, and inflammation.[12]

To sum up, though it is still unclear whether the counter-intuitive associations (between obesity variants and protective cardiometabolic profile) are driven by horizontal pleiotropy (i.e., genetic variants influencing both obesity and a cardiometabolic trait through different mechanisms) or by a specific type of protective adiposity (i.e., genetic variants leading to protective cardiometabolic profile through protective adiposity), the evidence for genes underlying these processes is rapidly accumulating.

F. Research gaps

This dissertation will focus on two major research gaps in the genetic epidemiology of obesity.

First, there has been no thorough investigation of the performance of polygenic risk prediction for obesity in various settings. In terms of PRS modeling or estimation methods, new PRS estimation methods have been developed, but they have not been thoroughly evaluated in diverse populations. Also, despite recent efforts to include more non-European populations in genomic research, the number of studies and the sample sizes of non-European population-based studies are substantially smaller than for European-based studies. Although many genomic findings are shared across populations, population-specific effects have been noted[147]; thus, the lack of diversity in genomic research hampers the identification of population-specific disease-

causing variants.[86] For instance, if some crucial variants in a specific population have low frequency or are not detectable in European populations, those variants are likely to be missed from discovery analysis.[91] Furthermore, although various demographic (age and sex)[148], lifestyle (e.g., smoking status)[149-151], and comorbid conditions (e.g., T2D and hypertension; possibly through medication, physical activity, and dietary habits) are known to modify the genetic effects on obesity-related traits, the performance of PRS across these settings has not been thoroughly investigated. Most studies have applied a single PRS, assuming that the prediction performance is the same for all individuals and populations. A lack of consideration of heterogeneities in prediction performance may limit the clinical impact of obesity PRS – e.g., risk group identification or targeted prevention efforts.

Second, although each obesity-associated variant is expected to have a unique influence on obesity and cardiometabolic complications, very little is known about the pleiotropic effects of obesity-associated variants on downstream cardiometabolic disorders. As a part of the effort to address this research gap, some recent studies (described in the previous section) have focused on pleiotropic obesity loci (i.e., genetic loci influencing both obesity and another trait), especially counter-intuitive associations with cardiometabolic profiles. By identifying bivariate (obesity and cardiometabolic trait) alleles with heterogeneous directions, thousands of obesity variants can be classified into subcategories by their potential roles in downstream cardiometabolic disease. Several different approaches (e.g., multivariate GWAS or using a novel composite trait to represent ‘favorable adiposity’ or ‘lipodystrophy-like trait’) have been used to identify bivariate loci. However, an emerging genomic analysis tool, a local genetic correlation approach (more will be described in the research plan section), has not been widely implemented despite its potential to discover novel bivariate loci. In addition, the previously identified

pleiotropic loci have not been validated in diverse populations. As with other genomic research, these loci were discovered in European ancestry populations, and it is unknown whether the identified bivariate loci show comparable influences on obesity and cardiometabolic traits in different ancestries. Therefore, it is necessary to further identify the bivariate loci for obesity and cardiometabolic traits, in particular, for lipid profiles, as the obesity-lipid bivariate connection has been understudied when compared to T2D or glycemic traits, even in European ancestry populations.

The proposed aims will fill the above-mentioned research gaps, leading to an improved understanding of the heterogeneous impact of polygenic risk prediction for obesity and pleiotropic obesity loci among diverse populations.

CHAPTER 3: RESEARCH PLAN

A. Overview

In this section, I will describe the study populations, variables (phenotype traits of interest, genetic data, and covariates), and an analysis plan for the proposed aims. In **Aim 1**, I will construct the obesity PRS using the latest and largest trans-ancestry GWAS of obesity-related traits from the GIANT consortium (N ~ 2 million) (1a) and evaluate and characterize the prediction performance among ancestrally diverse populations of PAGE study (1b). In **Aim 2**, I will identify the genetically correlated genomic loci between obesity and dyslipidemia in opposing directions by local genetic correlation analysis (2a) and investigate the potential influence of the correlated loci on obesity, dyslipidemia, and downstream CVD outcome (2b).

B. Study populations

B.1. Population Architecture using Genetics and Epidemiology: The PAGE study

The PAGE consortium was launched in 2008 along with NHGRI's effort to expand the ancestral diversity in genomic studies.[91,152] In this dissertation, all participants with relevant genetic and phenotypic information from PAGE participating cohort studies will be included. The PAGE cohort studies include the ARIC, CARDIA, HCHS/SOL, WHI, MEC, and Icahn School of Medicine at Mount Sinai BioMe biobank. Based on self-identified racial/ethnic groups, participants were classified as Hispanic/Latino, non-Hispanic Black, Asian American, Native American, Native Hawaiian, and non-Hispanic White. A total of 88,402 participants will be analyzed for BMI – the most available trait in this dissertation. The distribution of participants

whose genetic and phenotypic information is available is presented in **Table 3.1**. Participants in the PAGE study will be a target population for Aim 1 and a validating population for Aim 2. Following are brief descriptions of the PAGE-participating cohort studies.

ARIC, funded by the National Heart, Lung, and Blood Institute (NHLBI), is an ongoing community-based prospective cohort study primarily aiming to investigate the etiology of atherosclerosis and its clinical outcomes.[153] A random sample of 15,792 adults aged 45 – 64 years at baseline was initially recruited between 1987 and 1989 (approximately 4,000 participants for each of four communities in the U.S. – Forsyth County, NC; Jackson, MS; Washington County, MD; Minneapolis, MN).[153] Participants have received standardized examinations on their demographic, social, and health status approximately every five years.

BioMe, funded by The Charles Bronfman Institute for Personalized Medicine, is an electronic medical record-linked biobank whose participants were based on consented and volunteered patients in the Mount Sinai Medical Center (MSMC) (among over 70,000 inpatients and 800,000 outpatients annually).[154] The MSMC serves racially/ethnically diverse communities of the upper Manhattan area, which includes Central Harlem (predominantly non-Hispanic Black), East Harlem (predominantly Hispanic/Latino), and Upper East Side (predominantly non-Hispanic White). There have been more than 57,843 participants (21% Non-Hispanic Black, 34% Hispanic/Latino, 31% Non-Hispanic White, and 14% of other ancestry groups) enrolled in BioMe since 2007 (as of Feb 2021). Among them, a total of 32,344 participants have been genotyped (as of Feb 2021) so that they can be investigated in genomic studies (https://icahn.mssm.edu/research/ipm/programs/biome-biobank/facts).

CARDIA, funded by NHLBI, is a community-based prospective cohort study aiming to investigate the influencing factors for the development of coronary heart disease and its risk

factors during young adulthood.[155] Initial recruitment was done in 1985 – 1986, and a total of 5,116 Non-Hispanic Black (52%) and Non-Hispanic White (48%), aged 18 – 30 years, participated from four urban communities – 1,179 from Birmingham, AL; 1,109 from Chicago, IL; 1,402 from Minneapolis, MN; and 1,426 from Oakland, CA.[155] In the recruiting step, participants were selected for the cohort to be balanced in age (> or ≤ 24 years), educational level (> or ≤ 12 years), sex, and race/ethnicity.[155] After the initial examination, participants were asked to respond to the subsequent assessments in 1987 – 1988 (Year 2), 1990 – 1991 (Year 5), 1992 – 1993 (Year 7), 1995 – 1996 (Year 10), 2000 – 2001 (Year 15), 2005 – 2006 (Year 20), 2010 – 2011 (Year 25), and 2015 – 2016 (Year 35) (and currently Year 40 exam is ongoing as of Dec 2022). Data collection included the potential influencing factors for coronary heart disease – e.g., blood pressure, glucose levels, blood cholesterol levels, anthropometric traits, lifestyle factors, and family history.

HCHS/SOL, funded by NHLBI and other institutes, is a community-based prospective cohort study of Hispanic/Latino populations in the U.S. aiming to determine the role of acculturation in the prevalence and incidence of diseases and to identify influencing factors for the health of Hispanic/Latino populations. A total of more than 16,000 participants who were self-identified as Hispanic/Latinos and aged 18 – 74 years were recruited between 2008 and 2011 from four study sites – Bronx, NY; Chicago, IL; Miami, FL; and San Diego, CA. The study was designed to enroll 4,000 participants (2,500 aged 45 – 74 years and 1,500 aged 18 – 44 years) in each study site and to have at least 2,000 participants in each of the four groups of origin – Cuban, Puerto Rican, Mexican, or Central/South American.[156] The participants received extensive baseline examinations on psych-social and clinical factors during 2008 – 2011. A

follow-up assessment for the cohort was done during 2015 – 2017; the third exam is in progress now, and annual follow-up interviews via phone calls are ongoing.

MEC, funded by the National Cancer Institute, is a prospective cohort study to investigate lifestyle and genetic risk factors for cancer in the U.S.[157] A total of 215,251 adults aged 45 – 75 years at baseline were recruited between 1993 and 1996 from Hawaii and L.A. County, CA.[157] Ethnic distributions of the participants were 16.3% of Non-Hispanic Black, 22.0% of Hispanic/Latino, 26.4% of Japanese American, 6.5% of Native Hawaiian, 22.9% of Non-Hispanic White, and 5.8% of other ethnic groups.[157] During 2001 – 2006, a prospective biospecimen collection (i.e., biospecimen collected before the onset of disease; blood, urine, mouthwash, saliva, or viable lymphocytes) was done for a subset of participants (75,928 as of April 2019) (https://www.uhcancercenter.org/for-researchers/mec-cohort-composition). In this dissertation, eight ancillary studies will be included – the Slim Initiative in Genomic Medicine for the Americas (MEC-Sigma) (a type 2 diabetes study in Hispanic/Latino adults); MEC-AAPC, MEC-JAPC, and MEC-LAPC (studies of prostate cancer in Non-Hispanic Black, Japanese American, and Hispanic/Latino men, respectively); MEC-AABC, MEC-JABC, MEC-LABC, and MEC-HIBC (studies of breast cancer in Non-Hispanic Black, Japanese American, Hispanic/Latino women, and Native Hawaiian women, respectively).

WHI, funded by NHLBI, is a prospective cohort study to investigate the health of postmenopausal women in the U.S., especially for preventing CVD, breast cancer, colon cancer, and osteoporotic fractures in women aged 50 – 79 years.[158] A total of 161,808 participants were recruited between 1993 and 1998 at 40 clinical centers across the U.S. There are two different parts in WHI – one is the WHI Clinical Trial (~64,500), a randomized clinical trial of hormone therapy, dietary intervention, and calcium/vitamin D supplements, and the other is WHI

Observational Study (~100,000), investigating incidence, risk factors, and potential interventions for CVD, cancer, and osteoporotic fractures.[158] Followings are ancillary studies that will be included in our analyses – the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO); the Modification of PM-Mediate Arrhythmogenesis in Population study (MOPMAP); the Genomics and Randomized Trials Networks (GARNET); the Hip Fracture GWAS (HIPFX); the Long Life Study (LLS); the Women's Health Initiative Memory Study (WHIMS); and the Women's Health Initiative-SNP Health Association Resource (WHI-SHARe),

Table 3.1. Sample sizes of PAGE participants in this proposal by study and self-report race/ethnicity for BMI (upper) and WHRadjBMI (lower)

	European	African	Hispanic	East Asian	Native Hawaiian	American Indian	Other	Total
BMI								
ARIC	9,233	2,811	0	0	0	0	0	12,044
BioME	1,970	5,938	8,059	716	0	51	945	17,679
CARDIA	1,652	889	0	0	0	0	0	2,541
MEC	0	6,980	6,355	5,817	3,414	0	0	22,566
SOL	0	0	7,237	0	0	0	0	7,237
WHI	12,578	8,663	4,177	356	0	493	69	26,336
Total	**25,433**	**25,281**	**25,828**	**6,889**	**3,414**	**544**	**1,014**	**88,403**
WHRadjBMI								
ARIC	9,228	2,811	0	0	0	0	0	12,039
CARDIA	1,651	888	0	0	0	0	0	2,539
MEC	0	4,089	4,376	4,875	2,643	0	0	15,983
SOL	0	0	7,221	0	0	0	0	7,221
WHI	12,529	8,628	4,160	355	0	491	69	26,232
Total	**23,408**	**16,416**	**15,757**	**5,230**	**2,643**	**491**	**69**	**64,014**

B.2. The GIANT consortium

The GIANT consortium aims to discover genetic determinants contributing to body size and shape (measured via height, BMI, and WHR). Hundreds of studies have participated in the consortium, and meta-analyses of study-specific GWAS results have identified thousands of anthropometric trait-associated genetic loci. The number of participating studies has been expanded to improve the power to detect novel genetic loci. The most recent results included a

total of about 5.4 million participants for height[159] and about 2 million participants for BMI (under review for publication) from multiple self-reported racial/ethnic groups. Meta-analysis of height GWAS has the largest available set, and its sample size summary is shown as follows (**Figure 3.1**).

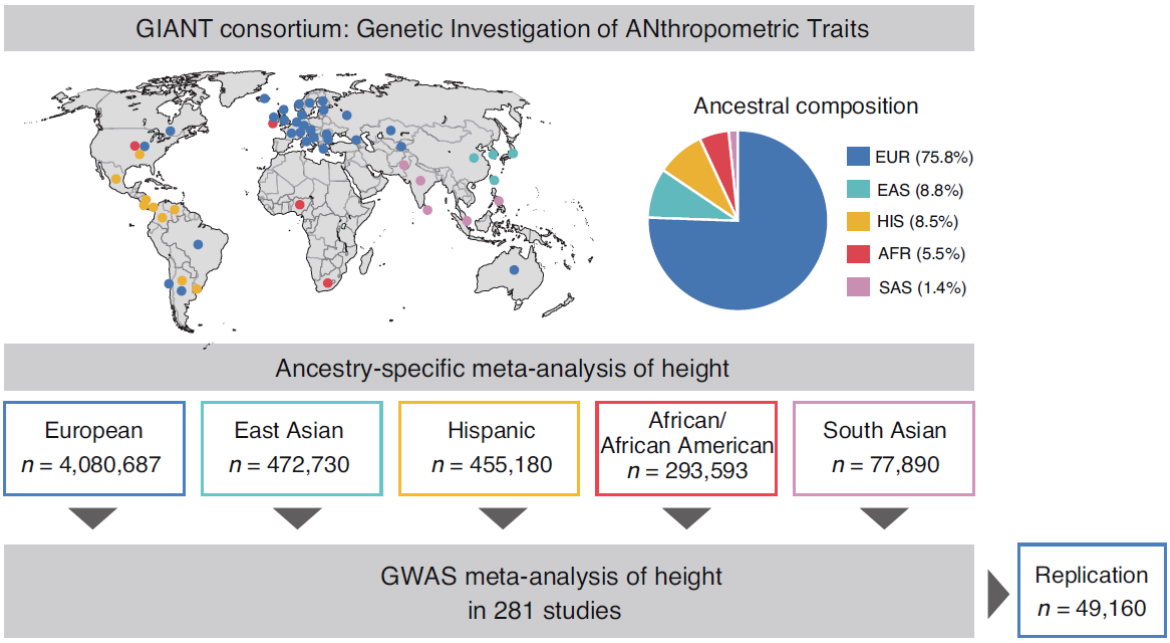


Figure 3.1. Geographical mapping and ancestries composition of 281 studies meta-analyzed in the latest GIANT Height GWAS. In the latest publication from the GIANT consortium, the GWAS meta-analysis of height consists of about 4M Europeans, 472K East Asians, 455K Hispanics, 293K Africans, and 78K South Asians. (Adapted from Yengo L, Vedantam S, Marouli E, et al. A saturated map of common genetic variants associated with human height. *Nature.* 2022;610(7933):704-712) [159]

For this dissertation, I will utilize the meta-analysis results for two obesity-related anthropometric traits (BMI and WHR) from the GIANT consortium to construct an overall obesity and a central obesity PRS (PRS-BMI and PRS-WHRadjBMI). To maintain the independence of the target population (PAGE study) from the base samples, the meta-analysis results were obtained after removing GWAS from PAGE participating studies. Following is a brief description of the meta-analysis conducted by the GIANT consortium. First, all individual studies were quality-controlled using the EasyQC[160] software and checked for the total number

of variants included, the total number of variants not in the reference panels, imputation quality scores, genomic inflation factor, and phenotype transformation. Variants with imputation quality score > 0.3, Hardy-Weinberg Equilibrium p-value > 1E-8, and minor allele count > 5 from each individual study were included in the analysis.[159] Then, meta-analyses were conducted by each ancestry group (EUR, EAS, HIS, AFR, and SAS) using RAREMETAL[161] to account for multi-allelic variants. Then, a fixed-effect meta-analysis of five ancestry groups was conducted to get the association results for trans-ancestry GWAS summary statistics.[159] I will use this trans-ancestry and ancestry-specific GWAS summary statistics to construct genome-wide obesity PRS.

B.3. UK Biobank

UKB is a large-scale prospective cohort study of more than 500,000 people from the United Kingdom with the primary aim of improving the prevention, diagnosis, and treatment of various diseases' onset later in life.[162] Participants aged 40 – 69 were recruited between 2006 – 2010.[162]. Participants' phenotypic and genotypic information, including questionnaires, physical and blood measures, genome-wide genotyping data, imaging data, and health outcomes, has been collected.[162] The UKB is available for paid access to researchers. In 2018, the Pan-ancestry GWAS of UK Biobank (https://pan.ukbb.broadinstitute.org/docs/study-design) presented a multi-ancestry GWAS of 7,221 phenotypes, including anthropometric and obesity-related measures. **Table 3.2** shows the number of samples and traits included in the projects. GWAS analysis was conducted using SAIGE[163] to implement a linear mixed model – with a kinship matrix as a random effect and covariates as fixed effects. Continuous traits were rank-based inverse normalized within each ancestry group, and covariates included in GWAS were age, sex, age*sex, age^2, age^2*sex, and the first 10 PCs (https://github.com/atgu/ukbb_pan_ancestry/wiki/QC). I will utilize the publicly available

GWAS summary statistics (available from: http://www.nealelab.is/uk-biobank/) for BMI and lipid traits (HDL, LDL, and TG) as a discovery sample for local genetic correlation analysis for the first step of aim 2.

Table 3.2. Sample sizes in the base UKB GWAS for aim two by ancestry groups

Population	BMI	HDL	LDL	TG
African ancestry	6545	5754	6200	6211
Admixed American ancestry	971	854	938	937
Central/South Asian ancestry	8646	7688	8404	8415
East Asian ancestry	2693	2342	2568	2570
European ancestry	419163	367021	400223	400639
Middle Eastern ancestry	1572	1364	1498	1499

Source: https://github.com/atgu/ukbb_pan_ancestry/wiki

C. Measurement of variables

Individual-level genetic and phenotypic data from PAGE are described in the following sections.

C.1. Genetic information

In the original PAGE study, a total of 53,426 non-European ancestry (African ancestry, Hispanic/Latino, East Asian, Native Hawaiian, and American Indian participants) samples from different participating studies were genotyped on the MEGA at the Center for Inherited Disease Research.[91] The MEGA was collaboratively designed by the PAGE II investigators, Illumina, and the Consortium on Asthma among African-ancestry Populations in the Americas to better capture the genetic diversity among populations of non-European ancestry.[164] The content of the MEGA was determined after considering some backbone content – e.g., Infinium HumanCore BeadChip, African Diaspora Consortium Power Chip, enhanced cross-population tagging content, diverse exonic content, tagging SNPs identified in published GWAS, SNPs documented

in UCSC browser track, and all clinically significant SNPs – and additional hand curated custom content suggested by PAGE investigators – e.g., regulatory variants with differential function in laboratory studies, enhanced coverage of tag SNPs for candidate genes or regions, expanded coverage of exonic regions for candidate genes or regions, comprehensive fine-mapping coverage for GWAS catalog reports, and clinically significant SNPs associated with traits of interest.[164]

In addition to the MEGA genotyping platform, some participants from ARIC, BioMe, CARDIA, MEC, and WHI were genotyped separately on Illumina or Affymetrix arrays by each study or ancillary study.

The number of samples included in this proposal (especially for the analysis of BMI) by study, self-reported race/ethnicity, and genotyping platform is shown in **Table 3.3**. A total of 38,971 samples that will be included in the current analysis were genotyped on MEGA, and the remaining 49,632 samples were genotyped on the non-MEGA array.

Table 3.3. Number of participants in PAGE genotyped on MEGA and non-MEGA array by study and by ancestry

Study	Race/ethnicity	MEGA	non-MEGA (Illumina or Affymetrix)
ARIC	European	0	9233
	African	0	2811
BioMe	European	0	1970
	African	4192	1746
	Hispanic/Latino	4294	3765
	East Asian	716	0
	American Indian	51	0
	Other	920	25
CARDIA	European	0	1652
	African	0	889
MEC	African	4467	2513
	Hispanic/Latino	24	6331
	East Asian	2972	2845
	Native Hawaiian	3106	308
HCHS/SOL	Hispanic/Latino	7237	0
WHI	European	0	12578
	African	6102	2761
	Hispanic/Latino	4106	71
	East Asian	291	65
	American Indian	493	0
	Other	0	69

The following table summarizes the genotyping platform, QC criteria, imputation methods, and reference panel that each study and ancillary study implemented.

Table 3.4. Summary of the non-MEGA genotype and quality control information in the PAGE

Study	Ancillary Study	Genotyping Platform	Sample Call Rate	HWE threshold	Imputation	Reference Panel
ARIC		Affymetrix GeneChip SNP Array 6.0	90%	$p>10^{-6}$	IMPUTE version 2.3.2	1000 Genome phase 3 v 5
BioMe		Affymetrix GeneChip SNP Array 6.0 and Illumina OmniExpressExome Array	95%	$p>5x10^{-5}$	IMPUTE version 2.3.2	1000 Genome phase 3 v 5
CARDIA		Affymetrix GeneChip SNP Array 6.0	95%	$p>10^{-6}$	IMPUTE version 2.3.2	1000 Genome phase 3 v 5
MEC	JAPC	Illumina Human660W_Quad_v1 Array	95%	NA	MACH	HapMap Phase 2
	LAPC	Illumina Human660W_Quad_v1 Array	95%	NA	MACH	HapMap Phase 2
	AAPC	Illumina Human1M-Duo Array	95%	NA	MACH	HapMap Phase 2
	LA T2D 2.5M	Illumina HumanOmni2.5-4v1_B Array	95%	NA	IMPUTE version 2.2.0	1000 Genomes Phase I integrated variant set
	AABC	Illumina Human1M-Duo Array	95%	NA	MACH	HapMap Phase 2
	LABC	Illumina Human660W_Quad_v1 Array	95%	NA	MACH	HapMap Phase 2
	JABC	Illumina Human660W_Quad_v1 Array	95%	NA	MACH	HapMap Phase 2
	HIBC	Illumina Human660W_Quad_v1 Array	95%	NA	MACH	HapMap Phase 2
WHI	GARNET	Illumina Human Omni1-Quad v1-0 B	98%	p>10-6	IMPUTE version 2.3.2	1001 Genome phase 3 v 5
	GECCO	Illumina 610 and Cytochip 370K	98%	p>10-6	IMPUTE version 2.3.2	1002 Genome phase 3 v 5
	HIPFX	Illumina 50K and 610K	98%	p>10-6	IMPUTE version 2.3.2	1003 Genome phase 3 v 5
	MOPMAP	Affymetrix Gene Titan, Axiom Genome-Wide, Human CEU I Array Plate	90%	p>10-6	IMPUTE version 2.3.2	1004 Genome phase 3 v 5
	WHIMS	Human OmniExpress Exome-8v1_B Genome-Wide Human	98%	p>10-6	IMPUTE version 2.3.2	1005 Genome phase 3 v 5
	LLS	Human OmniExpress Exome-8v1_A Genome-Wide Human	98%	p>10-6	IMPUTE version 2.3.2	1006 Genome phase 3 v 5
	WHI-SHARe	Affymetrix Gene Chip SNP Array 6.0	98%	p>10-6	IMPUTE version 2.3.2	1007 Genome phase 3 v 5
*MEGA		Infinium Expanded Multi-Ethnic Genotyping Array	98%	p>10-6	IMPUTE version 2.3.2	1000 Genome phase 3 v 5

Human genome build 37 and dbSNP version 150 were used for all cases.

C.2. Phenotypic information

Anthropometric measures

BMI will be used as a proxy measure of overall adiposity. BMI was derived from weight and height measured at baseline visit (at the time of enrollment) for ARIC, CARDIA, HCHS/SOL, and WHI. For 140 WHI participants who were missing in height or weight at baseline, height and/or weight measures at 1-year or 3-year follow-up substituted the missing baseline measures.[165] In MEC and BioMe biobank, height, and weight measures were self-reported, and this self-reported baseline height and weight measures were used to generate BMI at baseline.

WHR will be used as a continuous proxy measure of central adiposity, and it was derived from waist circumstance (WC) and hip circumference (HC) measures in the PAGE study. As with other anthropometric traits, WC and HC were measured during baseline visits. WC was measured using a tape measure at the natural waist level in a horizontal plane, rounded to the nearest 0.5 cm.[166] Self-reported WC and HC measures were collected in MEC.[91] BioMe did not collect WC or HC measures.

Cardiovascular disease risk factors

Lipid trait. HDL-C, TC, and TG levels were measured from fasting blood, and the Friedewald Equation was used to calculate LDL-C levels from other lipid measures. If measured TG levels were greater than 400mg/dL, LDL-C levels were not calculated. In addition, following previous studies, medication status was adjusted by adding a constant (**Table 3.5**).[167,168] The largest constant was applied if more than one medication was reported. Those who had not fasted

for 8 hours or were pregnant at measure were excluded from the harmonized phenotype database. Natural-log transformation was applied to TG levels after adjusting for medication.

Table 3.5. Constants used for medication adjustment of lipid levels in the PAGE study.

Medication	Constants (mg/dL)			
	HDL	LDL	TC*	TG
Statins	-2.3	49.9	52.1	18.4
Fibrates	-5.9	40.1	46.1	57.1
Bile acid sequestrants	-1.9	40.5	0	0
Niacin	-9.9	24.7	34.6	89.4
Cholesterol absorption inhibitors	0	40.5	40.5	0

Source: [168]

*TC: Total Cholesterol

Glycemic traits. Fasting blood glucose levels and insulin levels were measured at baseline visits using standard assays after 8 hours of fasting. HbA1c levels were measured during follow-up visits for all cohort studies except for HCHS/SOL. Participants without diabetes (normoglycemia) were defined as having fasting glucose < 5.6 mmol/L or HbA1c < 38 mmol/mol and aged over 40. I will exclude those under 40 years old were glucose < 5.6 mmol/L or HbA1c < 38 mmol/L from the analysis. Participants with pre-diabetes were defined as having glucose ≥ 5.6 mmol/L or HbA1c ≥ 38 mmol/mol. Lastly, participants with diabetes were defined based on ADA criteria (by medication, report diagnosis, fasting glucose ≥ 7 mmol/L or HbA1c ≥ 48 mmol), or random glucose > 11.11 mmol/L, and aged ≥ 21 years at the time of diagnosis (to avoid potential misclassification between T1D and T2D).

Blood pressure was measured using a standardized protocol. Participants were considered hypertensive using the following criteria (if met at least one criterion): 1) SBP ≥ 140 mmHg, 2) DBP ≥ 90 mmHg, 3) any antihypertensive medication reported, or 4) ICD-9 codes 401. x or ICD-10 codes I10.x - I15.x. [91]

Cardiovascular diseases

Some of the PAGE participating cohorts have information (prevalence, incidence, or death) on cardiovascular diseases. ARIC, MEC, and WHI ascertained the prevalence or incidence of myocardial infarction (MI), coronary heart disease (CHD), or stroke.

In ARIC, information on CHD events, including hospitalization and deaths, was collected through annual follow-up interviews and community surveillance.[169] Definitions of CHD events included acute hospitalized MI, definite fatal CHD, MI diagnosed by ECG, and revascularization.[169]

In MEC, as described in previous studies[170], CHD cases and controls from several nested case-control substudies in MEC will be used in the current dissertation. CHD cases were ascertained through the participants' medical records from the California Hospital Discharge Data (1990 - 2012) and the Centers for Medicare and Medicaid Services claim files (outpatients) (1999 - 2011), which were linked to MEC study - c.f., some participants from Hawaii (76.6% of Japanese American) were not available for hospital discharge data. Case definitions for CHD were based on ICD-9 codes (DX 410 - 414) for ischemic heart disease as the principal or first diagnosis code and the principal or first procedure code. Also, if a primary cause of death is MI (ICD-9 DX410, ICD-10 I21) or other CHD (ICD-9 DX411-414, ICD-10 I20, I22-25), these individuals were included as cases. Both prevalent (~20%; ascertained at baseline) and incident (~80%; ascertained during follow-up) CHD cases were ascertained.[169] Controls were selected among those without a history of heart attack or angina from the baseline questionnaire or all follow-up questions.

In WHI, CHD events were identified through a self-reported questionnaire and adjudicated by physicians after reviewing the chart within 3 months.[171] CHD cases were defined as individuals who had a history of MI (self-reported) or a revascularization procedure at baseline and/or manifested a definitive MI, went through a revascularization procedure, or died from CHD during follow-up.[171]

Lifestyle factors

Smoking status and physical activity will be considered as lifestyle factors, and they were measured differently across different cohorts. Smoking status was summarized into a variable classifying participants into never-smokers, former smokers, and current smokers. For physical activity, a binary variable was created to classify the participants into two groups – the bottom 20th percentile, by sex, for each cohort as the sedentary group and the rest (top 80th percentile per sex and cohort) as the non-sedentary group.

D. Statistical analyses

D.1. Aim 1. Characterize and evaluate the utility of trans-ancestry obesity PRS in the ancestrally diverse PAGE study

In aim 1, I will construct PRS-BMI and PRS-WHRadjBMI by using different genome-wide PRS estimation methods (P+T and PRS-CS(x)), and I will compare the prediction performance among PRS by different methods in the PAGE study. Then, I will characterize the prediction performance in various conditions in the PAGE study.

D.1.1. Construction of obesity PRS

I will construct genome-wide polygenic risk scores using three different methods described in the previous section – P+T, PRS-CS, and PRS-CSx. We will use the effect size estimates for variants from the trans-ancestry or ancestry-specific GIANT GWAS results for BMI and WHRadjBMI. (In PRS-CS and PRS-CSx, the estimated effects will be adjusted using the Bayesian approach.) The estimated effect sizes will be the genome-wide inputs for PRS calculation. As obesity is a highly polygenic trait, the use of genome-wide variants, instead of limiting the variants with statistical significance, would better capture the polygenic nature of obesity.[172] Indeed, previous literature demonstrated a better predicting performance when using genome-wide polygenic scores than using variants with genome-wide significance.[70] Although the PRS calculation step has the basic framework in common as described in Chapter 2, section D.1 (i.e., PRS for an individual=$\sum \beta_i SNP_i$, where SNP_i stands for the individual's dosage for the *i*th SNP and β_i is the estimated association between *i*th SNP and BMI from the GWAS), each PRS estimation method takes different approach when selecting SNPs (for PRS-CS and PRS-CSx, HapMap phase 3 variants will be used; for P+T, only independent index SNPs of each locus will be included) or deciding SNPs' effect size (PRS-CS and PRS-CSx adjust the SNPs' effect size based on Bayesian approach; P+T method uses the raw effect sizes from the base GWAS).

P + T

P+T method filters in only significantly associated SNPs based on a predefined p-value cut-off value (i.e., thresholding) and select the best (i.e., most significantly associated with BMI or WHRadjBMI) independent SNP in a given locus (i.e., clumping) based on a base GWAS

summary statistics. Independence between SNPs is usually decided by the LD R^2 between a pair of SNPs, and the cut-off criterion for the independence can vary by studies.

Before clumping and thresholding the SNPs, it should be decided which reference panel will be used to calculate LD R^2 between two SNPs. For each ancestry-specific GWAS result, I will use the matched population group from 1000 Genome reference population - i.e., EUR, AFR, AMR, EAS, SAS – to get the LD structures. However, there is no reference for the trans-ancestry GIANT GWAS; thus, I will construct a trans-ancestry reference population (called an "ALL" population) by combining randomly selected ancestry-specific 1000 Genome reference populations proportional to the distribution of different populations in GIANT GWAS. The number of participants in each ancestry in the base GIANT (without PAGE participants) is shown in the following table (the 4th column), and the number of randomly selected 1000 Genome populations for each ancestry is also shown in the table (the 6th column) - I will include the maximum number of EUR population, and other ancestry groups will be proportional to the number of EUR population.

Table 3.6. Distribution of continental ancestry in the reference population to be generated based on the 1000 Genome Phase 3 populations

Ancestry	**Total**	**Excluded PAGE samples**	**Remaining samples**	**Proportion**	**The new trans-ancestry reference population**
European	1,595,348	39,358	1,555,990	79.19%	489 (maximum available)
Hispanic	58,160	28,498	29,662	1.51%	9
East Asian	263,383	4,266	259,117	13.19%	81
African	114,335	27,030	87,305	4.44%	27
South Asian	44,704	11,906	32,798	1.67%	10
Total	2,075,930	111,058	1,964,872	100%	617

In addition, before clumping, the base GIANT GWAS was additionally cleaned by excluding variants with missing beta, sample size less than ⅓ of maximum sample size, minor allele frequency less than 0.001, or minor allele count less than 5 will be excluded from the base GIANT GWAS results before clumping.

In the clumping step, several parameters will be specified to conduct clumping and thresholding – especially for LD R^2 cut-off criterion (0.1, 0.2, and 0.5), LD window sizes (250kb or 500kb), and significant p-value thresholds (5E-2, 5E-3, 5E-5, 5E-7, and 5E-9), and based on combinations of criteria, there will be different sets of SNPs filtered in for constructing PRS. To find the best-performing combination, I will randomly divide the target PAGE samples into two independent sets by sex, study, and race/ethnicity stratum; one is the tuning sample, and the other is the testing sample (N ~ 44,000 for BMI and ~32,000 for WHRadjBMI in each tuning and testing set).

In practice, LD clumping will be conducted using the '--clump' command from PLINK software for all possible combinations of LD R^2 criteria (0.1, 0.2, and 0.5), LD window sizes (250kb and 500kb), and populations (ALL, EUR, AFR, AMR, and EAS). Once LD clumping is done, I will additionally filter the variant with different p-value thresholds (5E-2, 5E-3, 5E-5, 5E-7, and 5E-9).

For P+T, the raw effect estimates for SNP on BMI or WHRadjBMI from the GIANT GWAS will be used as PRS weights. PRS for PAGE individuals will be calculated using the '--score' function in PLINK software. The PAGE 1000 Genome imputed genetic data will be filtered using an imputation quality score, and variants with an imputation score less than 0.4 will be removed from the score calculation.

PRS-CS

PRS-CS reweights the effects estimates for a given SNP from the base GWAS results using the Bayesian approach. I will apply PRS-CS[20] to GIANT GWAS of BMI and WHRadjBMI after excluding variants with low reliability - missing effects estimates, low sample size (sample < ⅓ of maximum sample size), and rare variants (minor allele frequency < 0.001 or minor allele count < 5). PRS-CS uses genome-wide HapMap phase3 (HM3) variants (N ~ 1.3M) and requires an external LD reference panel. Since I will use trans-ancestry base GWAS in this proposal, I will utilize the trans-ancestry LD reference panel generated previously. PRS-CS needs information on the sample size of the GWAS and the 'phi' parameter (a global shrinkage parameter). The sample size will be specified as the 90th percentile of the sample size distribution across the variants in the GIANT GWAS. For the global shrinkage parameter, I will try different phi parameters (auto option, 1, 0.01, 0.0001, and 0.000001) and test the prediction performance of PRS for each phi parameter in tuning sample (except for 'auto' option since it does not need additional tuning sample) to decide the best-performing phi parameter. The tuning and testing samples will be the same sets that will be used in the P+T method.

PRS will be calculated with the weights estimated from PRS-CS for available HM3 variants in PAGE samples using the '--score' function in PLINK software. Variants with low imputation quality (imputation quality score < 0.4) will be removed from the PAGE genetic data before the PRS calculation. For PRS-CS, I will only construct trans-ancestry PRS-BMI and PRS-WHRadjBMI and ancestry-specific PRS will be estimated based on ancestry-specific GWAS using PRS-CSx (which will be described in the next section).

PRS-CSx

PRS-CSx[21] is known to have advantages for studies with heterogeneous population groups. I will apply the PRS-CSx method using ancestry-specific GIANT GWAS summary statistics (for EUR, AFR, HIS, and EAS) with the same inclusion/exclusion criteria as for PRS-CS – i.e., missing effects estimates, low sample size (sample < ⅓ of maximum sample size), or rare variants (minor allele frequency < 0.001 or minor allele count < 5). As in PRS-CS, the global shrinkage parameter and sample size information for each ancestry group should be provided. In this dissertation, I will try 1, 0.01, 0.0001, 0.000001, and auto as global shrinkage parameters and the 90th percentile of the sample size distribution as sample size parameters. I will use an ancestry-specific 1000 Genome LD reference panel (EUR, AFR, AMR, EAS, and SAS), which is provided by the authors. PRS-CSx will estimate ancestry-specific variants' weights – i.e., ω_{EUR}, ω_{AFR}, ω_{AMR}, ω_{EAS}, and ω_{SAS} – as well as an inverse-variance weighted meta-analysis of ancestry-specific weights (ω_{META}).

Individuals' ancestry-specific PRS – score-EUR, score-AFR, score-AMR, score-EAS, and score-SAS – in the PAGE population will be calculated using the '--score' function in PLINK software. As in other methods, variants with low imputation quality (imputation quality score < 0.4) will be removed from the PAGE genetic data before the PRS calculation. Then, in the tuning sample, the following linear regression model will be fitted, and the beta for each ancestry will be estimated.

$$\text{BMI (or WHRadjBMI)} \sim \beta_{EUR} \cdot \text{score-EUR} + \beta_{AFR} \cdot \text{score-AFR} + \beta_{AMR} \cdot \text{score-AMR} + \beta_{EAS} \cdot \text{score-EAS} + \beta_{SAS} \cdot \text{score-SAS}$$

Subsequently, these beta estimates for ancestry-specific scores will be applied to the testing sample, and the prediction performance will be evaluated.

Additionally, an inverse-variance weighted meta-analysis of ancestry-specific weights (META) will be calculated from the PRS-CSx, and these 'META' weights will be applied to the testing sample (without the tuning step).

D.1.2. Evaluation of prediction performance in PAGE study

The prediction performance of PRS-BMI in the PAGE study will be evaluated by R^2 (the proportion of variance in an outcome variable explained by PRS) from the linear regression models. The outcome variable will be inverse-normalized residuals of BMI after adjusting out other covariates, and PRS-BMI will be used as an explanatory variable. Likewise, the prediction performance of PRS-WHRadjBMI in the PAGE study will be evaluated by R^2 values from the linear regression models with inverse normalized residuals of WHRadjBMI after adjusting out other covariates as an outcome and PRS-WHRadjBMI as an explanatory variable.

In each analysis stratum, residual generation models are as follows for each sex.

[BMI] BMI ~ age + self-reported race/ethnicity + study + genotype platform + PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + PC7 + PC8 + PC9 + PC10 … (1)

[WHRadjBMI] WHR ~ BMI + age + self-reported race/ethnicity + study + genotype platform + PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + PC7 + PC8 + PC9 + PC10 … (2)

These sex-specific residuals will be inverse normalized. PRS will be standardized as a mean of 0 and a standard deviation of 1 for each analysis stratum. Then, R^2 will be estimated from the following linear regression models.

[BMI] Inverse normalized residuals from (1) ~ standardized PRS-BMI

[WHR] Inverse normalized residuals from (2) ~ standardized PRS-WHRadjBMI

D.1.3. Characterization of obesity PRS in PAGE study

To characterize the prediction performance of obesity PRS in the PAGE study, I will first stratify the PAGE sample by several different variables and compare the prediction performance across these strata. I will select variables known to be associated with obesity and potentially influencing the prediction accuracy of obesity PRS. Those variables include sex and age group as demographic variables, smoking status and physical activity as lifestyle factors, and T2D status and hypertension status as cardiometabolic comorbidities. For the age group, I will divide the participants into under and over 50, assuming 50 is an approximate age for menopause. For smoking status, I will classify the participants into never-smokers, former smokers, and current smokers. For physical activity, I will dichotomize the physical activity status into low (sedentary) and high (non-sedentary) groups. For T2D status, participants will be classified as a non-diabetic, prediabetic, and diabetic group. For hypertension, I will dichotomize a group with normal BP levels and a group with hypertension. To maximize the available sample size for the stratified analysis, I will use PRS constructed by PRS-CS(auto) since the PRS-CS(auto) can be applied directly to the testing sample with no need for an additional tuning sample.

D.2. Aim 2. Investigation of genetically correlated loci that jointly influence obesity and dyslipidemia

In Aim 2, I will identify genetic loci that are simultaneously associated with obesity (BMI) and dyslipidemia (HDL, LDL, and TG) and classify these bivariate loci into two different categories based on the direction of the local genetic correlation coefficients - one is Ob/DysL(+) loci (significant local genetic correlation with (+) sign) and the other is Ob/DysL(–) loci

(significant local genetic correlation with (–) sign). Then, I will investigate the potential influence of these loci on obesity, dyslipidemia, and other subsequent CVD-related factors in the PAGE study by testing the associations with Ob/DysL(+) loci- Ob/DysL(–) loci-based obesity PRS.

D.2.1. Identification of the genetically correlated loci that jointly influence obesity and dyslipidemia

Ob/DysL(+) and Ob/DysL(–) loci will be identified by local genetic correlation analysis using a pair of UKB GWAS summary statistics for obesity (BMI) and lipid traits (HDL, LDL, and TG). Local genetic correlation analyses will be conducted using the *LAVA* R package. A total of 3 obesity-lipid trait pairs (BMI-HDL, BMI-LDL, and BMI-TG) will be analyzed separately.

Here is a brief summary of the local genetic correlation approach implemented in this proposal.[173] LAVA, like other local genetic correlation estimation tools, was developed to estimate the locus-level genetic correlation between two phenotypes. LAVA first estimates the local genetic signal (measured by local heritability (h^2)) as follows.[173]

$$Y_p = X\alpha_p + \epsilon_p$$

Y_p : Standardized phenotype vector
X: genotype matrix with K_{snp} $SNPs$ (standardized)
α_p: vector of joint SNP effects (accounting for LD)
ϵ_p : vector of normally distributed residuals with variance η_p^2

$\hat{\alpha}_p = (X^T X)^{-1} X^T \boldsymbol{Y_p}$, if the local SNP LD matrix is denoted as $S = cor(X)$ and the vector of estimated marginal SNP effects are denoted as $\hat{\beta}_p$ (not accounting for LD), $\hat{\alpha}_p = S^{-1}\hat{\beta}_p$. That is, if marginal SNP effects are obtained from GWAS summary statistics, we can

estimate the joint SNP effects ($\hat{\alpha}_p$) using a reference population's LD structure.[173] Using the estimated joint SNP effects, local residual phenotypic variance (η_p^2) and t the proportion of phenotypic variance explained by the SNPs within the locus (local h^2) can be estimated.[173] Then, it estimates bivariate local genetic correlations. The local genetic effects (G) can be defined as $G = X\alpha$ (α is a K (number of SNPs in the locus) by P (number of phenotypes) matrix of joint SNP effects). The realized covariance matrix of G is denoted as follows (Ω).[173]

$$\Omega = \begin{pmatrix} \omega_p^2 & \omega_{qp} \\ \omega_{pq} & \omega_p^2 \end{pmatrix}$$

ω_p^2 : local genetic variance of G_p for phenotype p
ω_{pq} : local genetic covariance of G_p and G_q for phenotype p and q

Then, the local r_g can be calculated by $\rho_{pq} = \frac{\omega_{pq}}{\sqrt{\omega_p^2 \omega_q^2}}$, and ρ_{pq}^2 will be considered as the proportion of variance in the local genetic effects G_p explained by G_q.[173] Since G is not actually observed, Ω should be estimated using the Method of Moments, not computed directly.[173] The significance of the correlation will be determined using simulation-based p-values.[173] This local genetic correlation analysis will be especially useful for situations where some signals appear in opposing directions at different regions and nullify each other at a global level – i.e., the absence of global genetic correlation despite the presence of local genetic correlation in opposing directions, whereas global genetic correlation captures only the average genetic correlation across the whole genome and sometimes cannot differentiate the null genetic correlation.[173]

LAVA utilizes pre-partitioned genomic regions to get a local genetic correlation estimate for each locus. I will use 2,495 pre-partitioned genome that has been provided by the developers of LAVA (https://github.com/cadeleeuw/lava-partitioning). These partitioned genomic blocks

were generated based on the 1000 Genome European reference population on build hg19/GRCh37 to get approximately LD-independent genomic blocks across the whole genome.

As described earlier, LAVA first performs the univariate test to filter in the loci where a significant local genetic influence (measured by local heritability (h2)) on adiposity or lipid traits is estimated. It will exclude the loci without any significant local heritability for either of the two traits from the following bivariate analysis (correlation analysis). Then, local genetic correlation coefficients between a pair of obesity traits and lipid traits will be estimated among the significant univariate loci.

I will define the bivariate loci as follows. Bivariate loci are genomic regions showing significant local heritability estimates (Bonferroni-corrected $p < 0.00002$ (=0.05/2,495); call it as "univariate loci") and local genetic correlation coefficients (Bonferroni-corrected $p < 0.05$ / number of tested loci (univariate loci) for each obesity-lipid pair). I will classify the bivariate loci into two different groups based on their directions of association with dyslipidemia risk. In other words, if a given bivariate locus shows positive local genetic correlation coefficients between obesity and dyslipidemia (i.e., $rg < 0$ for BMI-HDL, $rg > 0$ for BMI-LDL and BMI-TG pairs), the locus will be classified as Ob/DysL(+) locus whereas if the bivariate locus shows negative local genetic correlation coefficients (i.e., $rg > 0$ for BMI-HDL, $rg < 0$ for BMI-LDL and BMI-TG), the locus will be classified as Ob/DysL(–) locus.

Table 3.7. Classification of Ob/DysL(–) and Ob/DysL(+) loci based on local heritability analysis and local genetic correlation analysis

	Ob/DysL(–)	**Ob/DysL(+)**
Step 1. Local heritability (h2)	$p < 0.00002$ (= 0.05/2495)	$p < 0.00002$ (= 0.05/2495)
Step 2. Local genetic correlation (rg)	$p < 0.05$ / N tested loci $rg > 0$ for BMI-HDL $rg < 0$ for BMI-LDL, BMI-TG	$p < 0.05$ / N tested loci $rg < 0$ for HDL-BMI $rg > 0$ for LDL-BMI, TG-BMI

After identifying the Ob/DysL(–) loci and Ob/DysL(+) loci, I will assess if the identified loci are previously reported (**Supplementary Table 2**) or novel. Since previous studies were conducted at the variant level, not the locus level, I will consider certain loci as replicated loci if the known variants were within the identified loci.

D.2.2. Prioritization of genes underlying counter-intuitive Ob/DysL(–) loci

To investigate the biological implications of the identified BMI-lipid bivariate loci and to prioritize potential causal genes underlying these counter-intuitive loci - Ob/DysL(–), I will conduct TWAS-FUSION[174] and identify potential genes whose genetically predicted expression levels were associated with the BMI or lipid traits. I will integrate each GWAS summary result (BMI, HDL, LDL, and TG) with reference gene expression levels in Whole Blood samples from the Cardiovascular Risk in Young Finns Study (YFS)[175] and adipose tissue from Metabolic Syndrome in Men Study (METSIM).[176] Then, I will filter the genes located within the bivariate loci (based on the start and the end position of the genes) and identify the overlapping genes from the BMI and corresponding lipid trait. I will also examine directional consistency by comparing TWAS Z scores for BMI and the corresponding lipid trait. For example, I will verify if a gene within BMI-HDL Ob/DysL(–) loci had the same direction of effect in the TWAS Z-score for both BMI and HDL. Based on the known roles of the overlapped genes (reported in public databases (e.g.,) PubMed or Genecards), I will infer potential pathways simultaneously influencing BMI and lipid traits.

D.2.3. Potential influence of the bivariate loci on obesity, dyslipidemia, and CVD-related factors in PAGE study

I will investigate the potential influence of Ob/DysL(+) loci or Ob/DysL(–) loci on obesity, dyslipidemia, and CVD-related factors compared to that of overall obesity loci in the

PAGE study. To do this, I will derive the PRS for obesity based on Ob/DysL(+)-loci or Ob/DysL(–) loci (PRS -Ob/DysL(+) or PRS-Ob/DysL(–), respectively) and test the association between the PRS-Ob/DysL(+) or PRS-Ob/DysL(–) and obesity traits (BMI and obesity status), lipid traits (HDL, LDL, TG, and dyslipidemia status), and other CVD-related factors (glycemic traits, blood pressure tratis, and CVDs). I hypothesize that PRS-Ob/DysL(–) will be associated with protective dyslipidemia and CVD risk profile but positively associated with obesity, whereas PRS-Ob/DysL(+) will be associated with adverse dyslipidemia and CVD risk profile and positively associated with obesity risk (as expected for overall PRS-BMI).

To construct PRS-Ob/DysL(+) and PRS-Ob/DysL(–), I will utilize publicly available PRS weights for BMI prepared and provided by ExPRSweb[177] (https://exprsweb.sph.umich.edu/). The PRS weights for BMI were estimated using PRS-CS ($N_{variants}$ = 1,113,832; Pearson correlation between PRS and BMI in testing sample = 0.321[177]) methods based on UKB GWAS summary statistics for BMI. I will restrict the genetic variants to those located in the Ob/DysL(–) bivariate loci and Ob/DysL(+) bivariate loci for the PRS-Ob/DysL(–) and PRS-Ob/DysL(+), respectively, and apply the weights to our target population, the PAGE study. The association will be tested in the available subset (available for lipid traits and CVD-related factors) of the PAGE study. The linear regression model to be tested is as follows.

[Quantitatvie measures]

Outcome trait ~ PRS-Ob/DysL(+) or PRS-Ob/DysL(–) + age + sex + study + self-reported race/ethnicity + genotype platform + PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + PC7 + PC8 + PC9 + PC10

[Categorical measures]

log(odds of being case) ~ PRS-Ob/DysL(+) or PRS-Ob/DysL(–) + age + sex + study + self-reported race/ethnicity + genotype platform + PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + PC7 + PC8 + PC9 + PC10

CHAPTER 4: MANUSCRIPT 1: CHARACTERIZING POLYGENIC RISK SCORES FOR OBESITY TRAITS ACROSS DIVERSE POPULATIONS AND SETTINGS

A. Overview

Obesity, a major driver of the population burden of cardiovascular disease (CVD), is a highly heritable trait. Thousands of obesity-associated genetic loci have been identified, enabling the construction of obesity polygenic risk scores (PRS) for risk prediction. However, current PRS are largely developed and tested based on genetic studies of non-Hispanic White populations, and thus far, prediction performance among ancestrally diverse populations has been poor. In addition, little is known about the potential heterogeneities in the prediction performance of obesity PRS across different contexts defined by demographic, lifestyle, and comorbid factors. In this regard, we aimed to characterize the performance of PRS for obesity traits (body mass index (BMI) and BMI-adjusted waist-to-hip ratio (WHRadjBMI)) in the diverse *Population Architecture using Genomics and Epidemiology* (PAGE) study.

Using the latest GWAS of BMI (80% of non-Hispanic White, 13% of East Asians, 4% of non-Hispanic Black, 1.5% of Hispanics and South Asians) and WHRadjBMI (84% of non-Hispanic White, 12% of East Asians, 0.6% of non-Hispanic Black, 0.8% of Hispanics and 2.7% of South Asians) from the GIANT consortium, we applied scores derived using the pruning and thresholding (P+T) method and PRS-CS method, a Bayesian approach using genome-wide SNPs (HapMap Phase 3 variants), to the PAGE participants and evaluated the prediction performance [variance explained by PRS (R^2) of the regression models]. We also investigated stratum-specific prediction performance of PRS for obesity traits by demographic factors (age group ($>$ or ≤ 50

years) and sex(females and males), lifestyle factors (smoking status (current smokers, former smokers, and never-smokers) and physical activity status (sedentary and non-sedentary), and comorbid status [T2D status (T2D, prediabetes, and normoglycemic) and hypertension status (hypertensive and normotensive)].

Prediction performance was improved by applying PRS-CS methods compared to P+T methods across all self-reported racial/ethnic groups (R^2 from 6.6% to 9.0% for PRS for BMI and from 2.9% to 4.6% for PRS for WHRadjBMI). However, we observed substantial differences in the prediction performance of PRS across self-reported race/ethnicity groups, especially between non-Hispanic White (R^2 of 14.0% for BMI and R^2 of 7.1% for WHRadjBMI) and non-Hispanic Black (R^2 of 7.1% for BMI and R^2 of 2.7% for WHRadjBMI) populations. Heterogeneities in the prediction performance of PRS-BMI and PRS-WHRadjBMI by different stratifying variables were also noted – i.e., age group, sex, smoking status, T2D status, and hypertension status for PRS-BMI and sex, T2D status, and hypertension status for PRS-WHRadjBMI.

Our results reinforce the need for more large-scale GWAS of obesity-related traits among diverse race/ethnic groups to improve the prediction performance of PRS for obesity-related traits. In addition, the current findings demonstrate that beyond the heterogeneities in race/ethnicity performance, other contextual factors have a measurable impact on prediction performance and, therefore, must be evaluated prior to the application of PRS in the clinical setting.

B. Introduction

Obesity has been associated with a wide swath of cardiometabolic disorders[3], as well as other disorders, with the rapid increase in obesity prevalence a significant public health threat.[1,2] Obesity often begins in early life, and it has a long-term influence on cardiometabolic health later in life.[113] Also, it is difficult to reverse obesity, once prevalent, in older children or adults.[6] Since obesity is highly heritable – heritability estimates ranged from 40% to 70%[4], it may be useful to identify individuals with a high genetic predisposition to obesity before its onset and to focus prevention efforts among those at the highest genetic risk of obesity.[114] Indeed, early identification of high-risk groups for obesity at a young age could be transformative, as many downstream diseases result from obesity, including cardiovascular diseases, cancers, etc.[114]

With the large-scale GWAS of obesity traits and novel polygenic risk scores (PRS) estimation methods, risk prediction using PRS has substantially improved. A previous study constructed PRS for BMI (PRS-BMI), including more than 2 million variants, and explained 8.41% of the variance in BMI.[70] The study suggested that PRS-BMI be implemented to identify high-risk individuals at birth for targeted and cost-effective prevention strategies.[70]

However, a majority of genomic studies have been conducted in individuals of European populations, and ancestral diversity has been lacking.[147,86] Although many genomic findings are shared across populations, population-specific effects, distinct patterns of linkage disequilibrium, and heterogeneity in SNP effect size across ancestries[147] limit the generalizability of genetic risk prediction across populations.[91] Also, despite the recent advancement in PRS estimation methods, the potential benefit of recent advances in diverse racial/ethnic populations is unclear. This makes it difficult for ancestrally diverse populations to benefit from genomic research and precision medicine[86], which may exacerbate the already evident obesity health disparities among

populations[147,178]. In addition, although various demographic (age and sex)[148], lifestyle (e.g., smoking status)[149-151], and comorbid conditions (e.g., T2D and hypertension; possibly through medication, physical activity, and dietary habits) are known to modify the genetic effects on obesity-related traits, the performance of PRS across these settings has not been thoroughly investigated. Most studies have applied a single PRS, assuming that the prediction performance is the same for all individuals and populations. A lack of consideration of heterogeneities in prediction performance may limit the clinical impact of PRS – e.g., risk group identification or targeted prevention efforts. In this regard, we aimed to evaluate the prediction performance of obesity PRS across different PRS estimation methods and race/ethnicity and characterize the prediction performance of obesity PRS across multiple demographic, lifestyle, and obesity comorbidity contexts.

C. Methods

C.1 Study Population

C.1.1 Genetic Investigation of ANthropometric Traits (GIANT) consortium

The GIANT consortium aims to discover genetic determinants contributing to body size and shape (measured via height, BMI, and WHR). Hundreds of studies have participated in the consortium, and meta-analyses of study-specific GWAS results have identified thousands of anthropometric trait-associated genetic loci. The number of participating studies has been expanded to improve the power to detect novel genetic loci. The most recent results included a total of about 5.4 million participants for height[159], about 2 million participants for BMI (manuscript in preparation), and about 1 million participants for WHR (adjusted for BMI) from multiple self-reported racial/ethnic groups. We utilized the latest GWAS of BMI and

WHRadjBMI from the GIANT consortium as the base GWAS of the PRS. Since PAGE participating studies were part of the GIANT consortium, we excluded PAGE participating studies from the discovery GWAS results to maintain sample independence[179] between the base GWAS and target population of the PRS analysis. A total of 1.95 million participants (79.5% of non-Hispanic White, 12.9% of East Asian, 4.5% of non-Hispanic Black, 1.5% of Hispanic/Latino, and 1.7% of South Asian) were included in the GWAS of BMI (excluding PAGE studies), and a total of 1.08 million participants (84.3% of non-Hispanic White, 11.7% of East Asian, 0.6% of non-Hispanic Black, 0.8% of Hispanic/Latino, and 2.7% of South Asian) were included in the GWAS of WHRadjBMI.

C.1.2 Population Architecture using Genetics and Epidemiology: The PAGE study

The PAGE consortium was established in 2008 as a part of NHGRI's initiative to expand ancestral diversity in genomic research.[91,152] All individuals with relevant genetic and phenotypic data from PAGE participating studies were included in the current study. The PAGE participating studies include the Atherosclerosis Risk in Communities (ARIC), Coronary Artery Risk Development in Young Adults Study (CARDIA), Hispanic Community Health Study / Study of Latinos (HCHS/SOL), Women's Health Initiative (WHI), Multiethnic Cohort Study (MEC), and Icahn School of Medicine at Mount Sinai BioMe biobank. Participants were categorized into different self-reported racial/ethnic groups, such as non-Hispanic White, non-Hispanic Black, Hispanic/Latino, Asian, Native Hawaiian, and Native American. Additional descriptions of these populations can be found in the previous literature.[91,152]

C.2 Measurement

C.2.1 Genetic Information

A total of 38,940 and 26,329 participants included in the PRS-BMI and PRS-WHRadjBMI analysis, respectively, were genotyped on the MEGA at the Center for Inherited Disease Research[91 164], and the remaining 49,405 and 37,615 were genotyped on the non-MEGA (Illumina or Affymetrix) arrays (**Table 4.1, 4.2**). The 1000 Genome imputed genetic data were filtered using imputation quality score, removing variants with imputation scores below 0.4.

C.2.2 Phenotype Information

BMI and WHRadjBMI were used as surrogate continuous measures of overall and central obesity, respectively. Blood glucose levels and insulin levels were measured after an 8-hour fast during the baseline visit. We categorized individuals' diabetes status according to the American Diabetes Association (ADA) criteria. Blood pressure was measured following a standardized procedure. Participants were classified hypertensive if they met at least one of the following criteria using the following criteria: 1) Systolic blood pressure (SBP) ≥ 140 mmHg, 2) Diastolic blood pressure (DBP) ≥ 90 mmHg, 3) reported use of any antihypertensive medication, or 4) ICD-9 codes 401. x or ICD-10 codes I10.x - I15.x.[91] Past and current smoking status and physical activity status were considered as lifestyle factors, but the measurement varied among different cohorts, as detailed in the supplement. Detailed descriptions of the phenotype information are provided in the **Supplementary Information.**

C.3 Statistical Analysis

C.3.1 Construction of PRS for obesity-related traits

We constructed genome-wide PRS using two different methods – P+T and PRS-CS(x). All PRS estimation methods have the following PRS calculation formula in common.

$$\text{PRS for an individual} = \sum \beta_i SNP_i$$

(where SNP_i stands for the individual's dosage for the *i*th SNP and β_i is the estimated association between *i*th SNP and BMI or WHRadj.BMI from the GWAS)

There were differences between the P+T method and PRS-CS (and PRS-CSx) in terms of the SNP list and the weights assigned to the SNPs used for PRS calculation. P+T method used a set of independent SNPs within a given locus after LD clumping based on a certain LD R^2 threshold and significance threshold, whereas PRS-CS and PRS-CSx used a pre-defined list of SNPs – e.g., HapMap3 variants – variants regardless of the variants' significance in base GWAS. While P+T adopted the raw effect estimates from the base GWAS for variants' weight in PRS calculation, PRS-CS and PRS-CSx reweighted the variants' effect estimates using a Bayesian approach. We used trans-ancestry and ancestry-specific (i.e., non-Hispanic black, East Asian, non-Hispanic white, Hispanic, and South Asian) GIANT GWAS results for BMI and WHRadj.BMI (self-reported non-Hispanic White, non-Hispanic Black, Hispanic/Latino, East Asian, and South Asian populations) as the discovery GWAS for PRS calculation. PAGE-specific studies were used for training and testing the PRS. Detailed descriptions on how each PRS estimation method was applied are described in the **Supplementary Information**.

C.3.2 Evaluation of prediction performance of obesity PRS in the PAGE study

We randomly divided the target PAGE samples into two independent sets by sex, study, and race/ethnicity stratum; one was the tuning sample, and the other was the testing sample (N ~ 44,000 for BMI and ~32,000 for WHRadjBMI in each tuning and testing set). We tuned the parameters for P+T (LD R^2, LD window size, and p-value threshold), PRS-CS (global shrinkage parameters), and PRS-CSx (global shrinkage parameters and the weights for ancestry-specific scores) in the tuning sample and evaluated the prediction performance of different PRS methods in the testing sample with the best-performing parameters for each method.

The prediction performance of obesity PRS (PRS-BMI and PRS-WHRadjBMI) in the PAGE study was evaluated by R^2 (the proportion of variance in an outcome variable explained by PRS) from the linear regression models. The outcome variable was the residuals of BMI or WHR after adjusting out other covariates, and the explanatory variable was PRS-BMI or PRS-WHRadjBMI. In each analysis stratum, residual generation models were as follows for each sex.

[BMI] *BMI ~ age + self-reported race/ethnicity + study + genotype platform + PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + PC7 + PC8 + PC9 + PC10 … (1)*

[WHRadj.BMI] *WHR ~ BMI + age + self-reported race/ethnicity + study + genotype platform + PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + PC7 + PC8 + PC9 + PC10 … (2)*

These sex-specific residuals were inverse normalized. PRS was standardized as a mean of 0 and a standard deviation of 1 for each analysis stratum. Then, R^2 was estimated from the following linear regression models.

[BMI] Inverse normalized residuals from (1) ~ standardized PRS-BMI

[WHR] Inverse normalized residuals from (2) ~ standardized PRS-WHRadj.BMI

C.3.3 Characterization of obesity PRS in the PAGE study

To characterize the prediction performance of obesity PRS in the PAGE study, we stratified the PAGE sample by different factors hypothesized to modify SNP effect size and thus PRS performance and compared the prediction performance across strata. We selected variables known to be associated with obesity and potentially influencing the prediction accuracy of obesity PRS. Those variables included sex and age group ($\leq$ 50 years and > 50 years), smoking status (current smokers, former smokers, and never smokers) and physical activity (sedentary and non-sedentary) as lifestyle factors, and T2D status (T2D, prediabetes, and normoglycemic) and hypertension status (hypertensive and normotensive) as cardiometabolic comorbidities of obesity. For the age group, we divided the participants into under and over 50, defining age 50 as a mid-life point and also the point at which the majority of women have achieved menopause. For smoking status, we classified the participants into never-smokers, former smokers, and current smokers. For physical activity, we dichotomized the physical activity status into low (sedentary) and high (non-sedentary) groups. For T2D status, participants were classified into normal glucose tolerance, prediabetes, and diabetes groups. For hypertension, we dichotomized into a group with hypertension and those without hypertension. To maximize the available sample size for the stratified analysis, we used PRS constructed by PRS-CS(auto) since the PRS-CS(auto) method does not require an additional tuning sample and can be applied directly to the testing sample.

D. Results

A total of 88,345 individuals and 63,944 individuals were included in the analyses of BMI and WHR, respectively (**Tables 4.3, 4.4).** Mean (SD) ages were 54.9 (11.4) years for the BMI set and 56.0 (10.7) years for the WHR set. The proportion of female participants was 68.4%

for the BMI set and 73.3% for the WHR set. The mean (SD) BMI for the BMI set was 28.6 (6.12) kg/m^2, and the mean (SD) WHR for the WHR set was 0.87 (0.09).

D.1 Prediction Performance by PRS Estimation Methods

In the PRS-BMI analyses, two PRS-CS methods outperformed the P+T methods across all race/ethnicity groups (**Figure 4.1A** and **Table 4.5**). For example, in race/ethnicity-pooled results, the R^2 was 36% higher using the PRS-CS(tuned phi) in comparison to the P+T method [6.60% → 8.95%]. Two PRS-CSx-META methods also demonstrated relatively high performance compared to the two PRS-CS methods, except for the race/ethnicity-pooled analysis. In addition, while the PRS-CSx (phi: auto) method demonstrated comparatively high performance among non-Hispanic White, Asians, and Native Hawaiians, it showed a lower performance (similar to that of P+T methods) among non-Hispanic Black and Hispanic/Latino and even the lowest performance among race/ethnicity-pooled results and American Indians. The PRS-CSx (phi: tuned) demonstrated consistently low performance across all race/ethnicity groups. In American Indian-specific analyses, the performance by estimation methods could not be evaluated due to small sample sizes.

Based on the performance of PRS-CS (phi: auto) – one of the best-performing PRS across all racial/ethnic groups, the prediction performance was highest among non-Hispanic White (R^2 of 14.0%), as expected, given the overrepresentation (~80%) of non-Hispanic White in the GIANT GWAS results that served as the base GWAS. The prediction performance was ranked in following order: Hispanic/Latino, Asian, Native Hawaiian, non-Hispanic Black, and American Indian. Of note, in the base GIANT GWAS, the proportion of Hispanic/Latino populations was only 1.5%, yet the prediction performance among Hispanic/Latino was comparable to that among non-Hispanic White. Among the three self-identified racial/ethnic

groups with comparable sample sizes in PAGE (non-Hispanic White, non-Hispanic Black, and Hispanic/Latino), the non-Hispanic White R^2 was almost twice as high as the non-Hispanic Black results.

In the PRS-WHRadjBMI analyses, the two PRS-CS methods consistently outperformed the P+T methods across all race/ethnicity groups (**Figure 4.1B** and **Table 4.6**). In race/ethnicity-pooled results, the R^2 was 61% higher using the PRS-CS(tuned phi) method when compared to the P+T method [2.87% → 4.64%]. Regarding the PRS-CSx methods, the PRS-CSx-META (phi: tuned) demonstrated relatively high prediction performance compared to the P+T methods and PRS-CSx (non-META) methods, however, the PRS-CSx-META (phi: auto) did not perform well compared to PRS-CSx-META (phi: tuned), and the performance of PRS-CSx-META (phi: auto) was even lower than that of the P+T methods for all racial/ethnic groups except for non-Hispanic White and Asians. The prediction performance of PRS-CSx (phi:auto) was far lower than other methods among race/ethnicity-pooled results, Native Hawaiians, and American Indians; however, it exhibited comparable performance among other race/ethnicity groups. The PRS-CSx (phi:tuned) method displayed consistently the lowest performance across all ancestry groups.

The performance of PRS-CS (phi: auto) for WHRadjBMI was the highest among non-Hispanic White (R^2 of 7.13%) as expected by the sample size distribution in the base GWAS (~84.3% of non-Hispanic White), and it was followed by Hispanic/Latino, Native Hawaiian, non-Hispanic Black, and Asian. Of note, the precision of the estimates for American Indian was so low (i.e., a wider confidence interval than other racial/ethnic groups). The prediction performance of WHRadjBMI was about or less than a half of that of BMI (8.9% for BMI and 4.4% for WHRadjbMI).

D.2 Prediction Performance of PRS-BMI and PRS-WHRadjBMI in Different Strata

We assessed the prediction performance of PRS-BMI and PRS-WHRadjBMI in different strata by demographic variables (age and sex), lifestyle variables (smoking status and physical activity status), and comorbid CVD stata (Hypertension and T2D status) (**Table 4.7-4.8** and **Figure 4.2**). We detected significant differences in prediction performance by age group, sex, smoking status, hypertension status, and T2D status for PRS-BMI and by sex and hypertension and T2D status for PRS-WHRadjBMI. Specifically, the prediction performance of PRS-BMI was higher among > 50 years group (vs. ≤ 50 years), females (vs. males), former smokers or non-smokers (vs. current smokers), hypertensive group (vs. normotensive group), and prediabetes group (vs. T2D control groups or diabetes group). In addition, the prediction performance of PRS-WHRadjBMI was higher among females (vs. males), the normotensive group (vs. hypertensive group), and the T2D control group or prediabetes group (vs. diabetes group).

E. Discussion

In this study, the Bayesian PRS estimation methods performed better than the P+T methods across all racial/ethnic groups in PAGE participants. However, we still noted a substantially poorer performance among non-Hispanic Black populations compared to non-Hispanic White populations. The performance among other populations groups fell between that of non-Hispanic White and non-Hispanic Black. We also observed distinct patterns of PRS performance across demographic, lifestyle, and CVD risk factor-informed strata, further supporting the importance of context when applying PRS to populations with often unique patterns of gene-environment interaction. As precision medicine advances and PRS are applied clinically, we must have an understanding of the accuracy of our prediction tools in each

population. Overall, our study demonstrated that PRS prediction performance varied substantially by PRS estimation methods, racial/ethnic groups, and various individual-level contexts. Therefore, we strongly advocate for the consideration of performance-influencing factors before applying PRS in clinical settings.

Improved prediction accuracy of PRS by Bayesian approaches – using the larger number of genome-wide SNPs and reweighting the SNPs' effect sizes by Bayesian regression – when compared to estimates using the P+T method has been reported many times[70,20,107,180], including previous studies of BMI.[20] We confirmed that the improved performance by PRS estimation methods was applicable to all race/ethnicity groups in PAGE participants. It has also been argued that PRS-CS methods perform better for highly polygenic traits, influenced by numerous genome-wide SNPs with small impact rather than by a small set of significant SNPs with larger effect sizes,[105] and our results for these obesity-related traits are no exception. We also tested variations of the PRS-CS methods using ancestry-specific GWAS results (e.g., PRS-CSx-META); however, we did not find strong evidence of improvement in prediction accuracy, and, in some cases, we observed under-performance of PRS-CSx methods compared to PRS-CS method (which is based on one large trans-ancestry GWAS results). The discrepancies between our results and the original PRS-CSx report[21] might be due to the relatively small sample sizes for non-European-specific discovery GWAS results. Since, unlike the other methods, PRS-CS (auto) (where global shrinkage parameter is automatically acquired from data)[20] did not require a separate tuning data set, and its prediction performance was better in most comparisons, we implemented the PRS-CS (auto) method for all subsequent analyses. Approaches to improve PRS performance in diverse populations is a topic of major importance currently, our data suggest there is still room for substantial improvement in prediction accuracy.

We observed substantial differences in PRS performance across self-identified race/ethnic subpopulations in the PAGE study, especially between non-Hispanic White populations and non-Hispanic Black populations. The poor performance of PRS in self-identified non-European populations is well described,[147,181,182] and likely driven by continental ancestral differences, for example, differences in allele frequencies and linkage disequilibrium, and an underrepresentation of diversity in discovery GWAS. However, cultural and environmental population differences by race/ethnicity may also have influenced distinct environmental and demographic factors.[181,183-185] Unfortunately, due to historical limitations of the data collection and characterization, we are unable to distinguish true ancestry effects from race/ethnicity effects. As described, the predominance of non-Hispanic White populations in the discovery GWAS is one likely explanation for such differences (79.2% and 84.3% non-Hispanic White populations for BMI and WHRadjBMI GWAS, respectively), as others have also observed.[181] This discrepancy highlights the need for larger GWAS studies in diverse populations, especially in African and African American populations.[181,186]

PRS-WHRadjBMI displayed lower predictive performance when compared to PRS-BMI (about half R^2 compared to PRS-BMI). Such findings are unsurprising, given the smaller sample sizes of the discovery GWAS for waist traits (about 2 million in GWAS of BMI vs. 1 million in GWAS of WHR) (*unpublished; manuscript in preparation*). Other factors likely influencing these differences include the more error-prone measurement of WHR (reviewed in [187]). On the other hand, there may be real differences in the genetic architecture of BMI and WHR, particularly as demonstrated in the recent large-scale GWAS of overall adiposity and central adiposity (or fat distribution)[4] and the generally lower heritability estimates for waist measures compared to BMI (67.8% for BMI and 37.3% for waist circumference) [188]. Thus, further studies

are required to understand the lower performance of PRS-WHRadjBMI and explore ways to improve the performance. Unlike BMI, WHR measurements are less prevalent in clinical settings (e.g., EHR database) yet they have strong associations with various CVD and CVD risk factors. Therefore, improving prediction accuracy of PRS-WHRadjBMI is particularly important, as it could serve as a robust genetic instrumental variable of central obesity or as an important genetic predictor for CVD risk in clinical settings.

We observed the heterogeneities in obesity PRS prediction performance across different contexts, such as demographic factors, lifestyle factors, and comorbidities, implicating the importance of accounting for gene-environmental interactions in obesity genetic risk prediction. Previous literature has also demonstrated an important impact of gene-environmental interactions on PRS performance.[189,190] Below we describe each contextual factor in detail, highlighting literature that may support the heterogeneous results observed.

First, the lower prediction performance of PRS-BMI among current smokers (vs. never smokers or former smokers) and among T2D cases (vs. prediabetes) was likely driven by well-known changes or fluctuations in body weight due to smoking or T2D pathogenicity.[191,192] It is well known that cigarette smoking is inversely associated with BMI, possibly due to reduced appetite[193] and that smoking cessation is related to weight gain.[194,195] Thus, genetic predisposition to increased BMI may be partly suppressed by current cigarette smoking, resulting in the reduced prediction performance of PRS-BMI. In addition, patients with T2D are likely to experience fluctuations in weight – either from weight loss related to a healthier lifestyle[196] or weight loss related to T2D pathogenicity during the course of the disease[197] or medication,[198] thereby limiting prediction performance. For PRS-WHRadjBMI, we also observed lower

prediction performance in the T2D case group compared to the normoglycemic T2D control group or the prediabetes group, but not by smoking status.

In terms of age, previous studies have demonstrated both higher and lower prediction performance of PRS-BMI as populations age.[199,115] Our results align more closely with studies that report a larger proportion of variation in BMI explained by PRS-BMI at older ages when compared to younger ages (e.g., [115]). Differences in lifestyle factors by strata may explain these apparent age effects. Indeed, we observed a higher proportion of current smokers among the ≤ 50 years group, which may have also influenced prediction performance differences across ages. In addition, the discovery and PAGE populations were similarly middle-aged, thus it is unsurprising that the mid-to-older aged strata had improved prediction performance. We observed no differences in prediction performance between the two age groups for PRS-WHRadjBMI, perhaps due to smaller sample sizes or possible confounding by sex.

We speculate that sex differences in PRS performance were driven by differences in the sample size of the discovery GWAS, with more females than males, much like in our study strata. Differences may also be driven by demographic confounders of sex differences, as the females were older and less likely to smoke, both of which had demonstrated impact on our PRS performance [(mean age was higher among females – 55.8 (SD: 11.0) years – when compared to males -52.9 (SD: 12.0) years-) and the proportion of current smokers was lower among females (13.6%) than among males (21.1%)]. PRS-WHRadjBMI also performed better among females [5.27% (95% CL: 4.87 – 5.66) in females vs. 2.04% (95% CL: 1.62 – 2.47) in males]. This finding is consistent with the previous literature on sex-specific genetic effects for waist traits (i.e., higher heritability among females)[79].

Differences in the prediction performance of obesity-related PRS by hypertension and T2D status may be driven by demographic and CVD risk factor differences between groups. For example, we observed that groups with higher mean BMI and more adverse metabolic health (the >50 years group when stratified by age group, the hypertensive group when stratified by hypertension status, and the prediabetes group when stratified by T2D status) displayed higher prediction performance of PRS-BMI. This finding is in line with a recent study that demonstrated a stronger genetic predisposition to obesity in the context of obesogenic environments.[189] However, unlike in the PRS-BMI, we observed higher prediction performance among normotensive individuals (vs. hypertensive individuals) for PRS-WHRadjBMI.

All taken together, we suggest four overarching reasons for heterogeneous effects across contexts– 1) differences in sample characteristics between the discovery GWAS populations and target populations (e.g., age and sex distribution in base GWAS); 2) differences in genetic architecture or biological mechanisms between subgroups (e.g., sexual dimorphism in WHR or altered biological mechanisms due to aging or comorbidity); 3) potential biological influences by external exposures (e.g., smoking or medications); and 4) strong lifestyle modifications (e.g., intentional weight loss or physical activity). Thus, it can be inferred that any individual-level factors that are related to the above categories may have a potential to influence the prediction performance of obesity PRS. In addition, combinations of these factors (e.g., young female smokers vs. older male non-smokers) could have an impact that is greater than each factor individually. Furthermore, although the current study is limited by only available variables which is not necessarily relevant for predicting obesity risk at birth (e.g., by age group or by smoking status), the current results suggested that other important context-related variables, which can be

determined at birth (e.g., Socio-economic status of household), should be accounted for when predicting obesity risk.

This study has notable strengths. First, the total sample size of the PAGE study was extensive, enabling a comprehensive characterization of the PRS-BMI and PRS-WHRadjBMI. The distribution of self-identified race/ethnicity in the PAGE study was balanced, ensuring that the race/ethnicity-pooled results were not biased toward a certain racial/ethnic group. Also, the phenotypes of the PAGE participants were extensively measured, allowing for the thorough characterization of the PRS-BMI and PRS-WHRadjBMI in various contexts.

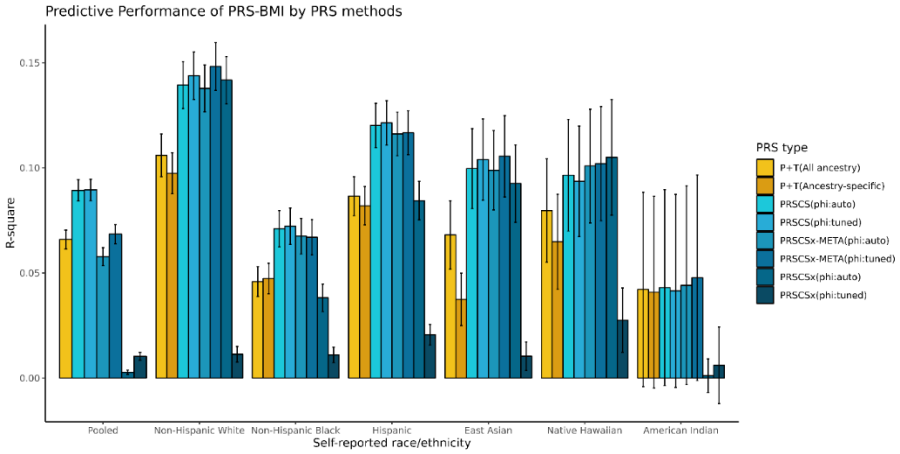
The current study also had limitations. First, since the discovery GWAS was predominantly from non-Hispanic White study populations, the performance of stratum-specific prediction may have been unduly influenced by an already poor performance of PRS across self-identified race/ethnicity. In addition, we implemented broad race/ethnicity categories, which likely encompassed individuals with a variety of genetic ancestries and thus varying prediction performance within self-reported race/ethnicity stratum.[200] Furthermore, since the phenotype measures we used in the current analyses were not from multiple time points, our inference on the differences in prediction performance over time and across the life course is limited.

Our findings illustrated an improvement in obesity PRS prediction performance by Bayesian estimation methods regardless of racial/ethnic groups. However, the race/ethnicity-specific results demonstrated a decreased PRS prediction performance for populations external to the base GWAS populations (i.e., mostly EUR so far). Also, the current results revealed an importance of contextual heterogeneities for PRS performance by demographic, lifestyle, and comorbidity status. All such heterogeneities limit the potential clinical use of PRS for obesity.

Therefore, the current results reinforce the need for the evaluation of context specific influences before the application of PRS in the clinical settings.

F. Main Findings and Figures

A. Prediction performance of PRS-BMI by PRS methods and self-reported race/ethnicity



B. Prediction performance of PRS-WHRadjBMI by PRS methods and self-reported race/ethnicity

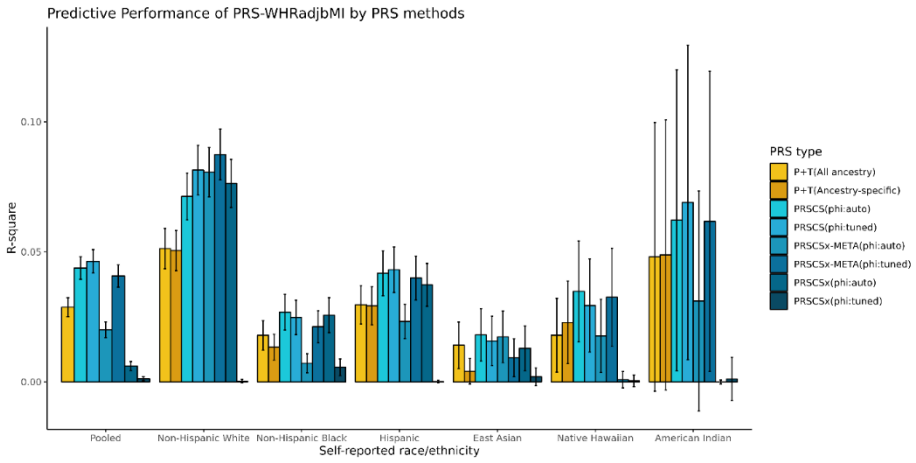


Figure 4.1 Prediction performance (R2) of PRS-BMI (A) and PRS-WHRadjBMI (B) by different PRS methods and by self-reported race/ethnicity groups in the PAGE study. We evaluated two categories of PRS estimation methods, including multiple specification options for each method. R^2 represented a proportion of variance in BMI (A) or WHR (B) explained by PRS after adjusting for age, sex, study, genotype panels, self-reported race/ethnicity, and ten genetic principal components. BMI was also accounted for in the models with PRS-WHRadjBMI. The PRS-CS and PRS-CSx-META outperformed P+T methods for both BMI and WHRadjBMI across all self-reported race/ethnicity groups. We also observed substantial differences in performance between non-Hispanic white and non-Hispanic black.

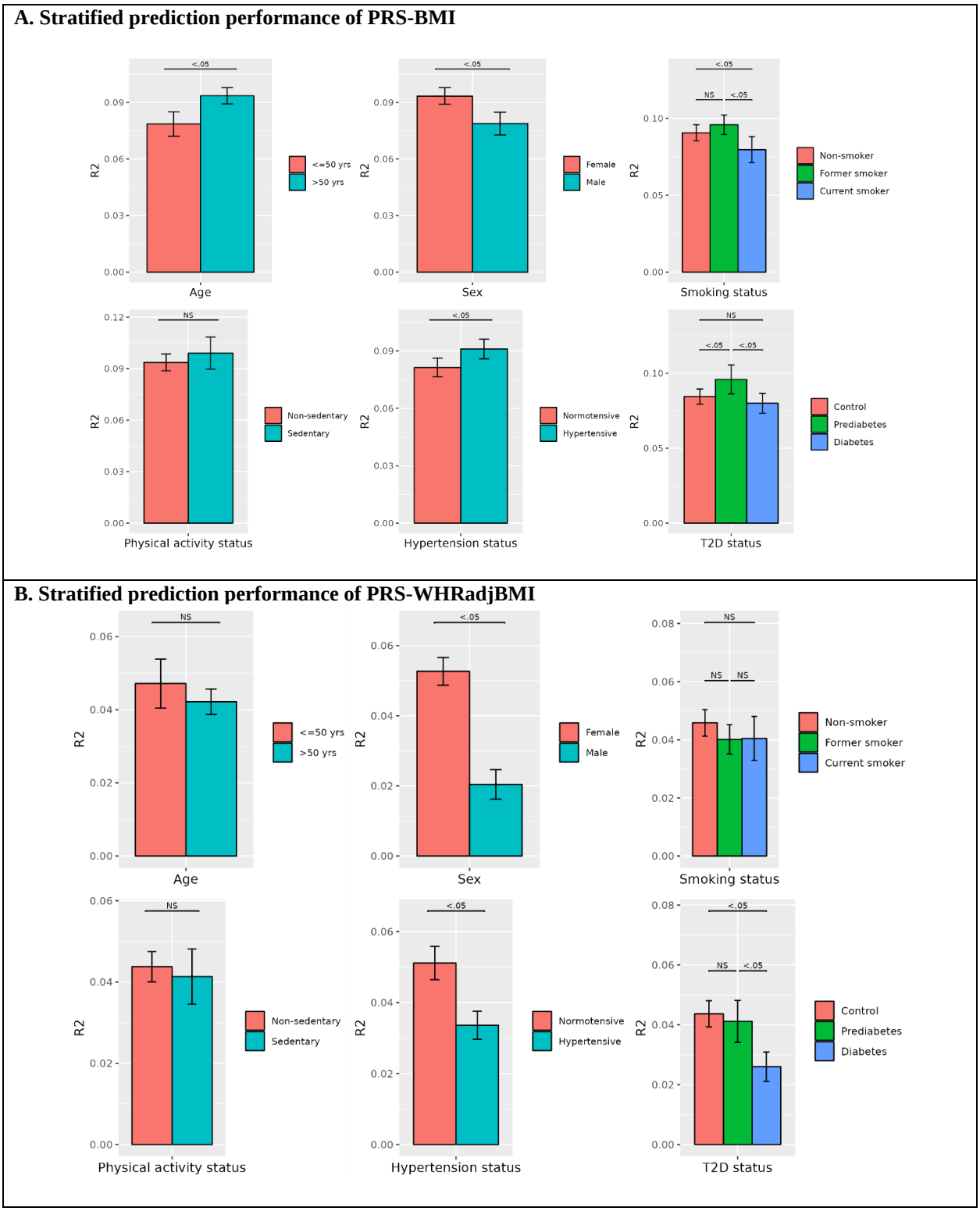


Figure 4.2 Stratified prediction performance of PRS-BMI (A) and PRS-WHRadjBMI (B)**.** We stratified participants by demographic (age and sex), lifestyle (smoking status and physical activity status), and comorbid condition (hypertension status and T2D status) variables and assessed prediction performance in each stratum. We observed significant differences in prediction performance between age groups, sex, smoking status, hypertension status, and T2D status for PRS-BMI (A) and between sex, hypertension status, and T2D status for PRS-WHRadjBMI (B).

G. Supplement

G.1 Supplemental Methods

G.1.1 Phenotype Information

BMI was used as a surrogate measure of overall fatness. BMI was calculated using weight and height measured at the initial visit (at the time of enrollment) for participants in ARIC, BioMe biobank, CARDIA, HCHS/SOL, and WHI. In cases where height or weight data were absent at the baseline (for WHI), measurements at 1-year or 3-year follow-ups were used.[165] In the MEC study, participants reported height and weight information, which was used to generate BMI at baseline.

WHR was used as a surrogate measure of fat distribution, and it was calculated using waist circumstance (WC) and hip circumference (HC) measures. These measurements, along with other anthropometric traits, were collected during participants' baseline visits. WC was measured using a tape measure at the natural waist level, with a precision of up to 0.5 cm.[166] In the case of MEC, participants self-reported their WC and HC measures.[91] BioMe did not collect WC or HC measures.

Glycemic traits. Blood glucose levels and insulin levels were measured after an 8-hour fast during the baseline visit. HbA1c levels were assessed during follow-up visits for all cohort studies except for HCHS/SOL. Participants without diabetes (normoglycemia) were identified if their fasting glucose level was < 5.6 mmol/L or HbA1c level was < 38 mmol/mol and they were over 40 years old. If individuals were under 40 years old, fasting glucose level < 5.6 mmol/L or HbA1c level < 38 mmol/L, these participants were excluded from the analysis. Participants with pre-diabetes were defined as having glucose levels ≥ 5.6 mmol/L or HbA1c levels≥ 38

mmol/mol. Lastly, individuals with diabetes were identified based on ADA criteria (including medication use, reported diagnosis, fasting glucose ≥ 7 mmol/L, or HbA1c ≥ 48 mmol/mol) or if their random glucose level was > 11.11 mmol/L, and they were aged ≥ 21 years at the time of diagnosis to prevent potential misclassification between T1D and T2D.

Blood pressure was measured following a standardized procedure. Participants were classified hypertensive if they met at least one of the following criteria using the following criteria: 1) SBP ≥ 140 mmHg, 2) DBP ≥ 90 mmHg, 3) reported use of any antihypertensive medication, or 4) ICD-9 codes 401. x or ICD-10 codes I10.x - I15.x. [91]

Cardiovascular diseases. A part of PAGE participating cohorts gathered data on CVD, including their prevalence, incidence, or related deaths. ARIC, MEC, and WHI specifically collected information on the occurrence of myocardial infarction (MI) and stroke, as well as the deaths resulting from MI or stroke. Further details regarding how CVD status was ascertained for each study are shown in the following section.

Lifestyle factors. Past and current smoking status and physical activity status were taken into account as lifestyle factors, but the measurement varied among different cohorts. Smoking status was categorized into never-smokers, former smokers, and current smokers. In terms of physical activity, a binary variable was set to classify the participants into two groups – the bottom 20th percentile for each cohort as the sedentary group and the rest (top 80th percentile cohort) as the non-sedentary group.

G.1.2 CVD Ascertainment by PAGE-participating studies

In ARIC, information on the CHD events including hospitalization and deaths were collected through annual follow-up interviews and community surveillance.[169] Definitions of CHD events included acute hospitalized MI, definite fatal CHD, MI diagnosed by ECG, and revascularization.[169]

In MEC, As described in previous studies[170], CHD cases and controls from several nested case-control substudies in MEC were used in the current analysis. CHD cases were ascertained through the participants' medical record from the California Hospital Discharge Data (1990 - 2012) and the Centers for Medicare and Medicaid Services claim files (outpatients) (1999 - 2011), which were linked to MEC study - c.f., some participants from Hawaii (76.6% of Japanese American) were not available for hospital discharge data. Case definitions for CHD were ICD-9 codes (DX 410 - 414) for ischemic heart disease as the principal or first diagnosis code and the principal or first procedure code. Also, if a primary cause of death is MI (ICD-9 DX410, ICD-10 I21) or other CHD (ICD-9 DX411-414, ICD-10 I20, I22-25), these individuals were included as cases. Both prevalent (~20%; ascertained at baseline) and incident (~80%; ascertained during follow-up) CHD cases were ascertained.[169] Controls were selected among those without history of heart attack or angina from the questionnaire at baseline or all follow-up questions.

In WHI, CHD events were identified through self-reported questionnaire and adjudicated by physicians after reviewing the chart within 3 months.[171] CHD cases were defined as individuals who had a history of MI (self-reported) or a revascularization procedure at baseline, and/or manifested a definitive MI, went through a revascularization procedure, or died from CHD during follow-up. [171]

G.1.3 PRS Estimation

P + T. For each ancestry-specific GWAS result, we used the matched population group from the 1000 Genome reference populations - i.e., EUR, AFR, AMR, EAS, SAS – to get the LD structures for clumping. Regarding the trans-ancestry GWAS results, since there are no corresponding reference populations, we constructed a trans-ancestry reference population (called an "ALL" population) by combining randomly selected ancestry-specific 1000 Genome reference populations proportional to the distribution of different populations in GIANT GWAS. Before clumping, the base GIANT GWAS was additionally cleaned by excluding variants with missing beta, sample size less than ⅓ of maximum sample size, minor allele frequency less than 0.001, or minor allele count less than 5. The parameters specified for the clumping step were as follows – LD R^2 cut-off criterion (0.1, 0.2, and 0.5), LD window sizes (250kb or 500 kb), and significant p-value thresholds (5E-2, 5E-3, 5E-5, 5E-7, and 5E-9). We derived sets of filtered SNPs by each combination of parameters and constructed PRS from each set of SNPs. The best-performing specification was selected from a tuning sample, and the PRS with the best-performing specification was tested in a separate testing set. PRS for the PAGE participants were calculated using the *'--score'* function in PLINK software.

PRS-CS. PRS-CS reweights the effects estimates for a set of SNPs (HapMap3 variants were most commonly used) from the base GWAS results using the Bayesian approach. We applied PRS-CS[20] to GIANT GWAS of BMI and WHRadjBMI after excluding variants with low reliability - missing effects estimates, low sample size (sample < ⅓ of maximum sample size), and rare variants (minor allele frequency < 0.001 or minor allele count < 5). Since PRS-CS requires an external LD reference panel to calculate the weights assigned to each SNP, we used a trans-ancestry LD reference panel derived from the 1000 Genome Project. The sample size was specified as the 90th percentile of the sample size distribution of the variants in the GIANT

GWAS. Different global shrinkage parameters (1, 0.01, 0.0001, and 0.000001) were used in the weight estimation step, and the best-performing parameter from a tuning set was selected to be tested in a testing set. An additional 'auto' option for the global shrinkage parameter was tested in a testing set (c.f., the 'auto' option did not require a tuning sample). Individuals' PRS was be calculated with the weights estimated from PRS-CS for available HM3 variants in PAGE samples using *'--score'* function in PLINK software.

PRS-CSx. Since PRS-CSx[21] is known to have advantages for studies with heterogeneous population groups, we implemented PRS-CSx[21] to the PAGE populations using the ancestry-specific GIANT GWAS results (EUR, AFR, HIS, and EAS) after applying the same variant QC criteria as for PRS-CS – i.e., missing effects estimates, low sample size (sample < ⅓ of maximum sample size), or rare variants (minor allele frequency < 0.001 or minor allele count < 5). As in PRS-CS, we applied different global shrinkage parameters and specified the sample size information for each ancestry group as the 90th percentile of the sample size distribution of the variants in ancestry-specific GWAS. For the external LD reference, we used an ancestry-specific 1000 Genome LD reference panel (EUR, AFR, AMR, EAS, and SAS) provided by the authors.

PRS-CSx estimated ancestry-specific weights assigned to the HM3 variants – i.e., ω_{EUR}, ω_{AFR}, ω_{AMR}, ω_{EAS}, and ω_{SAS} – as well as in a verse-variance weighted meta-analysis of ancestry-specific weights (ω_{META}). Then, individuals' ancestry-specific PRS – score-EUR, score-AFR, score-AMR, score-EAS, and score-SAS – in the PAGE population was calculated using the '--score' function in PLINK software. In the tuning set, the following linear regression model was fitted, and the beta for each ancestry was estimated.

$$\text{BMI (or WHRadjBMI)} \sim \beta_{EUR} \cdot \text{score-EUR} + \beta_{AFR} \cdot \text{score-AFR} + \beta_{AMR} \cdot \text{score-AMR} + \beta_{EAS} \cdot \text{score-EAS} + \beta_{SAS} \cdot \text{score-SAS}$$

Subsequently, these beta estimates for ancestry-specific scores were applied to the testing sample, and the prediction performance was evaluated. Additionally, an inverse-variance weighted meta-analysis of ancestry-specific weights (META) was calculated from the PRS-CSx, and these 'META' weights was applied to the testing sample (without the tuning step).

G.2 Supplemental Tables and Figures

Supplemental tables (Table 4.1 – Table 4.14) are below.

Table 4.1. The number of participants in the current analysis genotyped on MEGA and non-MEGA array by study and by ancestry (PRS-BMI set)

Study	**Race/ethnicity**	**MEGA**	**non-MEGA (Illumina or Affymetrix)**
ARIC	European	0	9233
	African	0	2811
BioMe	European	0	1970
	African	4188	1744
	Hispanic/Latino	4293	3764
	East Asian	716	0
	American Indian	51	0
	Other	920	25
CARDIA	European	0	1652
	African	0	889
MEC	African	4465	2513
	Hispanic/Latino	24	6330
	East Asian	2972	2845
	Native Hawaiian	3105	308
HCHS/SOL	Hispanic/Latino	7234	0
WHI	European	0	12563
	African	6092	2553
	Hispanic/Latino	4098	71
	East Asian	291	65
	American Indian	491	0
	Other	0	69

Table 4.2. The number of participants in the current analysis genotyped on MEGA and non-MEGA array by study and by ancestry (PRS-WRHadjBMI set)

Study	Race/ethnicity	MEGA	non-MEGA (Illumina or Affymetrix)
ARIC	European	0	9228
	African	0	2811
CARDIA	European	0	1651
	African	0	888
MEC	African	3058	1031
	Hispanic/Latino	21	4355
	East Asian	2662	2213
	Native Hawaiian	2455	188
HCHS/SOL	Hispanic/Latino	7218	0
WHI	European	0	12500
	African	6063	2546
	Hispanic/Latino	4076	70
	East Asian	289	65
	American Indian	487	0
	Other	0	69

Table 4.3. Distribution of variables (BMI analysis set)

		Total (N=88345)	Non-Hispanic White (N=25418)	Non-Hispanic Black (N=25255)	Hispanic (N=25814)	Asian (N=6889)	Native Hawaiian (N=3413)	American Indian (N=542)	other (N=1014)
Age		54.9 (11.4)	57.3 (11.0)	54.8 (11.2)	52.3 (12.2)	57.2 (9.51)	54.0 (7.07)	58.4 (7.62)	46.9 (14.2)
Sex									
	Male	27898 (31.6%)	6200 (24.4%)	6994 (27.7%)	9528 (36.9%)	3295 (47.8%)	1377 (40.3%)	22 (4.1%)	482 (47.5%)
	Female	60447 (68.4%)	19218 (75.6%)	18261 (72.3%)	16286 (63.1%)	3594 (52.2%)	2036 (59.7%)	520 (95.9%)	532 (52.5%)
Study									
	ARIC	12044 (13.6%)	9233 (36.3%)	2811 (11.1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
	BioME	17671 (20.0%)	1970 (7.8%)	5932 (23.5%)	8057 (31.2%)	716 (10.4%)	0 (0%)	51 (9.4%)	945 (93.2%)
	CARDIA	2541 (2.9%)	1652 (6.5%)	889 (3.5%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
	MEC	22562 (25.5%)	0 (0%)	6978 (27.6%)	6354 (24.6%)	5817 (84.4%)	3413 (100%)	0 (0%)	0 (0%)
	SOL	7234 (8.2%)	0 (0%)	0 (0%)	7234 (28.0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
	WHI	26293 (29.8%)	12563 (49.4%)	8645 (34.2%)	4169 (16.2%)	356 (5.2%)	0 (0%)	491 (90.6%)	69 (6.8%)
BMI		28.6 (6.12)	27.7 (5.60)	30.0 (6.72)	29.2 (5.91)	25.1 (4.08)	29.1 (6.09)	30.1 (6.27)	27.1 (5.88)
HDL		50.7 (16.0)	51.0 (16.5)	54.4 (16.2)	48.6 (14.6)	49.6 (17.2)	41.3 (14.8)	52.4 (13.1)	50.6 (18.7)
	Missing	42846 (48.5%)	14090 (55.4%)	12113 (48.0%)	9108 (35.3%)	4900 (71.1%)	1882 (55.1%)	40 (7.4%)	713 (70.3%)
LDL		136 (40.6)	134 (38.6)	140 (43.5)	133 (39.7)	141 (38.7)	144 (36.7)	140 (39.3)	122 (40.4)
	Missing	43737 (49.5%)	14261 (56.1%)	12470 (49.4%)	9399 (36.4%)	4946 (71.8%)	1889 (55.3%)	53 (9.8%)	719 (70.9%)
TG		132 (80.3)	127 (78.0)	111 (64.1)	150 (87.9)	139 (82.0)	128 (74.7)	158 (85.3)	153 (95.9)
	Missing	42798 (48.4%)	14078 (55.4%)	12222 (48.4%)	8980 (34.8%)	4899 (71.1%)	1882 (55.1%)	45 (8.3%)	692 (68.2%)
DBP		78.4 (12.4)	76.2 (11.5)	82.5 (12.7)	77.6 (12.3)	77.9 (12.0)	NA (NA)	79.1 (10.8)	79.4 (12.6)
	Missing	25400 (28.8%)	527 (2.1%)	7659 (30.3%)	7629 (29.6%)	5930 (86.1%)	3413 (100%)	9 (1.7%)	233 (23.0%)
SBP		130 (21.6)	127 (20.4)	136 (22.4)	129 (21.2)	126 (21.4)	NA (NA)	131 (19.9)	130 (22.8)

Table 4.3. Distribution of variables (BMI analysis set)

	Total (N=88345)	Non-Hispanic White (N=25418)	Non-Hispanic Black (N=25255)	Hispanic (N=25814)	Asian (N=6889)	Native Hawaiian (N=3413)	American Indian (N=542)	other (N=1014)
Missing	25429 (28.8%)	533 (2.1%)	7662 (30.3%)	7636 (29.6%)	5930 (86.1%)	3413 (100%)	9 (1.7%)	246 (24.3%)
Hypertension								
No(1)	43491 (49.2%)	15394 (60.6%)	9729 (38.5%)	14344 (55.6%)	2168 (31.5%)	943 (27.6%)	263 (48.5%)	650 (64.1%)
Yes (2)	43503 (49.2%)	9301 (36.6%)	15359 (60.8%)	11266 (43.6%)	4685 (68.0%)	2258 (66.2%)	270 (49.8%)	364 (35.9%)
Missing	1351 (1.5%)	723 (2.8%)	167 (0.7%)	204 (0.8%)	36 (0.5%)	212 (6.2%)	9 (1.7%)	0 (0%)
MI								
No	80257 (90.8%)	22768 (89.6%)	22822 (90.4%)	23924 (92.7%)	6273 (91.1%)	3046 (89.2%)	504 (93.0%)	920 (90.7%)
Yes	8088 (9.2%)	2650 (10.4%)	2433 (9.6%)	1890 (7.3%)	616 (8.9%)	367 (10.8%)	38 (7.0%)	94 (9.3%)
Stroke								
No	80250 (90.8%)	23655 (93.1%)	22484 (89.0%)	23955 (92.8%)	5750 (83.5%)	2931 (85.9%)	508 (93.7%)	967 (95.4%)
Yes	8095 (9.2%)	1763 (6.9%)	2771 (11.0%)	1859 (7.2%)	1139 (16.5%)	482 (14.1%)	34 (6.3%)	47 (4.6%)
Fasting Glucose	5.44 (1.25)	5.46 (1.00)	5.47 (1.51)	5.48 (1.26)	5.12 (1.16)	4.99 (1.11)	5.88 (1.85)	5.67 (1.37)
Missing	40424 (45.8%)	7514 (29.6%)	12312 (48.8%)	13152 (50.9%)	4601 (66.8%)	1798 (52.7%)	70 (12.9%)	977 (96.4%)
Fasting Insulin	10.5 (14.2)	9.70 (7.89)	12.0 (23.6)	11.0 (9.55)	6.68 (5.75)	8.41 (7.40)	11.8 (11.5)	11.7 (6.53)
Missing	40645 (46.0%)	7394 (29.1%)	12460 (49.3%)	13221 (51.2%)	4736 (68.7%)	1793 (52.5%)	65 (12.0%)	976 (96.3%)
HOMA-IR	2.57 (2.30)	2.42 (2.04)	2.82 (2.57)	2.73 (2.37)	1.56 (1.48)	1.95 (2.11)	3.21 (3.44)	3.08 (2.19)
Missing	41479 (47.0%)	7817 (30.8%)	12710 (50.3%)	13313 (51.6%)	4772 (69.3%)	1815 (53.2%)	75 (13.8%)	977 (96.4%)
HbA1c	40.8 (13.3)	37.8 (9.99)	46.8 (18.0)	40.5 (12.3)	42.5 (12.9)	NA (NA)	51.0 (25.5)	48.4 (16.9)
Missing	66231 (75.0%)	16304 (64.1%)	20896 (82.7%)	17564 (68.0%)	6749 (98.0%)	3413 (100%)	528 (97.4%)	777 (76.6%)
T2D Status								
T2D	23387 (26.5%)	3531 (13.9%)	7882 (31.2%)	7566 (29.3%)	2666 (38.7%)	1324 (38.8%)	150 (27.7%)	268 (26.4%)

Table 4.3. Distribution of variables (BMI analysis set)

	Total (N=88345)	**Non-Hispanic White (N=25418)**	**Non-Hispanic Black (N=25255)**	**Hispanic (N=25814)**	**Asian (N=6889)**	**Native Hawaiian (N=3413)**	**American Indian (N=542)**	**other (N=1014)**
Pre-diabetes	12682 (14.4%)	5863 (23.1%)	2613 (10.3%)	3756 (14.6%)	222 (3.2%)	108 (3.2%)	75 (13.8%)	45 (4.4%)
T2D controls	42640 (48.3%)	14276 (56.2%)	12067 (47.8%)	10104 (39.1%)	3592 (52.1%)	1981 (58.0%)	300 (55.4%)	320 (31.6%)
Other controls	9636 (10.9%)	1748 (6.9%)	2693 (10.7%)	4388 (17.0%)	409 (5.9%)	0 (0%)	17 (3.1%)	381 (37.6%)
Smoking status								
Non-smoker	42387 (48.0%)	11484 (45.2%)	11286 (44.7%)	13767 (53.3%)	3558 (51.6%)	1448 (42.4%)	267 (49.3%)	577 (56.9%)
Former smoker	29811 (33.7%)	9512 (37.4%)	8586 (34.0%)	7329 (28.4%)	2655 (38.5%)	1327 (38.9%)	200 (36.9%)	202 (19.9%)
Current smoker	14138 (16.0%)	4025 (15.8%)	4809 (19.0%)	3938 (15.3%)	578 (8.4%)	618 (18.1%)	57 (10.5%)	113 (11.1%)
Missing	2009 (2.3%)	397 (1.6%)	574 (2.3%)	780 (3.0%)	98 (1.4%)	20 (0.6%)	18 (3.3%)	122 (12.0%)
Physical activity status								
Physically active	49338 (55.8%)	15368 (60.5%)	12632 (50.0%)	12691 (49.2%)	5271 (76.5%)	2978 (87.3%)	352 (64.9%)	46 (4.5%)
Physically non-active	14146 (16.0%)	4270 (16.8%)	4407 (17.5%)	4239 (16.4%)	754 (10.9%)	331 (9.7%)	125 (23.1%)	20 (2.0%)
Missing	24861 (28.1%)	5780 (22.7%)	8216 (32.5%)	8884 (34.4%)	864 (12.5%)	104 (3.0%)	65 (12.0%)	948 (93.5%)

Table 4.4. Distribution of variables (WHRadjBMI analysis set)

		Total (N=63944)	Non-Hispanic White (N=23379)	Non-Hispanic Black (N=16397)	Hispanic (N=15740)	Asian (N=5229)	Native Hawaiian (N=2643)	American Indian (N=487)	other (N=69)
Age		56.0 (10.7)	57.2 (11.1)	56.3 (10.0)	53.2 (11.5)	58.8 (7.19)	53.8 (6.98)	59.5 (5.78)	61.0 (5.18)
Sex									
	Male	17047 (26.7%)	5122 (21.9%)	3022 (18.4%)	5276 (33.5%)	2554 (48.8%)	1073 (40.6%)	0 (0%)	0 (0%)
	Female	46897 (73.3%)	18257 (78.1%)	13375 (81.6%)	10464 (66.5%)	2675 (51.2%)	1570 (59.4%)	487 (100%)	69 (100%)
Study									
	ARIC	12039 (18.8%)	9228 (39.5%)	2811 (17.1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
	CARDIA	2539 (4.0%)	1651 (7.1%)	888 (5.4%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
	MEC	15983 (25.0%)	0 (0%)	4089 (24.9%)	4376 (27.8%)	4875 (93.2%)	2643 (100%)	0 (0%)	0 (0%)
	SOL	7218 (11.3%)	0 (0%)	0 (0%)	7218 (45.9%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
	WHI	26165 (40.9%)	12500 (53.5%)	8609 (52.5%)	4146 (26.3%)	354 (6.8%)	0 (0%)	487 (100%)	69 (100%)
WHR		0.873 (0.0927)	0.860 (0.0948)	0.855 (0.0907)	0.896 (0.0884)	0.903 (0.0806)	0.905 (0.0853)	0.839 (0.0779)	0.862 (0.0907)
BMI		28.5 (5.89)	27.7 (5.57)	30.0 (6.46)	29.0 (5.63)	25.2 (3.95)	28.8 (5.86)	30.2 (6.27)	30.1 (5.52)
HDL		50.7 (15.5)	50.7 (16.2)	54.5 (15.7)	48.7 (14.0)	48.3 (16.6)	41.5 (14.9)	52.0 (12.7)	NA (NA)
	Missing	26177 (40.9%)	12747 (54.5%)	5953 (36.3%)	2544 (16.2%)	3541 (67.7%)	1312 (49.6%)	11 (2.3%)	69 (100%)
LDL		137 (39.7)	134 (38.4)	143 (42.6)	133 (38.2)	144 (36.3)	144 (36.6)	140 (38.7)	NA (NA)
	Missing	26915 (42.1%)	12901 (55.2%)	6234 (38.0%)	2791 (17.7%)	3578 (68.4%)	1319 (49.9%)	23 (4.7%)	69 (100%)
TG		129 (77.1)	126 (77.5)	106 (58.2)	146 (84.2)	139 (79.5)	128 (73.9)	160 (84.1)	NA (NA)
	Missing	26471 (41.4%)	12778 (54.7%)	6155 (37.5%)	2594 (16.5%)	3547 (67.8%)	1312 (49.6%)	16 (3.3%)	69 (100%)
DBP		77.6 (12.0)	75.8 (11.3)	82.3 (12.2)	76.2 (11.7)	80.9 (11.5)	NA (NA)	79.0 (10.7)	81.9 (11.5)
	Missing	16217 (25.4%)	6 (0.0%)	4093 (25.0%)	4600 (29.2%)	4875 (93.2%)	2643 (100%)	0 (0%)	0 (0%)
SBP		129 (21.1)	126 (20.4)	135 (21.9)	126 (20.0)	134 (22.2)	NA (NA)	131 (19.7)	140 (23.6)

Table 4.4. Distribution of variables (WHRadjBMI analysis set)

	Total (N=63944)	Non-Hispanic White (N=23379)	Non-Hispanic Black (N=16397)	Hispanic (N=15740)	Asian (N=5229)	Native Hawaiian (N=2643)	American Indian (N=487)	other (N=69)
Missing	16211 (25.4%)	7 (0.0%)	4090 (24.9%)	4596 (29.2%)	4875 (93.2%)	2643 (100%)	0 (0%)	0 (0%)
Hypertension								
No(1)	32147 (50.3%)	14175 (60.6%)	6203 (37.8%)	9388 (59.6%)	1356 (25.9%)	768 (29.1%)	230 (47.2%)	27 (39.1%)
Yes (2)	30520 (47.7%)	8487 (36.3%)	10030 (61.2%)	6148 (39.1%)	3849 (73.6%)	1716 (64.9%)	248 (50.9%)	42 (60.9%)
Missing	1277 (2.0%)	717 (3.1%)	164 (1.0%)	204 (1.3%)	24 (0.5%)	159 (6.0%)	9 (1.8%)	0 (0%)
MI								
No	57957 (90.6%)	20865 (89.2%)	14792 (90.2%)	14690 (93.3%)	4719 (90.2%)	2380 (90.0%)	453 (93.0%)	58 (84.1%)
Yes	5987 (9.4%)	2514 (10.8%)	1605 (9.8%)	1050 (6.7%)	510 (9.8%)	263 (10.0%)	34 (7.0%)	11 (15.9%)
Stroke								
No	58560 (91.6%)	21762 (93.1%)	14800 (90.3%)	14879 (94.5%)	4303 (82.3%)	2295 (86.8%)	456 (93.6%)	65 (94.2%)
Yes	5384 (8.4%)	1617 (6.9%)	1597 (9.7%)	861 (5.5%)	926 (17.7%)	348 (13.2%)	31 (6.4%)	4 (5.8%)
Fasting Glucose	5.43 (1.22)	5.46 (1.00)	5.47 (1.49)	5.45 (1.19)	5.11 (1.15)	4.95 (1.07)	5.88 (1.85)	5.67 (1.37)
Missing	17507 (27.4%)	5520 (23.6%)	3894 (23.7%)	3715 (23.6%)	3058 (58.5%)	1269 (48.0%)	19 (3.9%)	32 (46.4%)
Fasting Insulin	10.5 (14.4)	9.70 (7.89)	12.0 (23.9)	11.0 (9.46)	6.66 (5.79)	8.21 (7.11)	11.9 (11.5)	11.7 (6.53)
Missing	17753 (27.8%)	5400 (23.1%)	4056 (24.7%)	3793 (24.1%)	3193 (61.1%)	1266 (47.9%)	14 (2.9%)	31 (44.9%)
HOMA-IR	2.57 (2.28)	2.42 (2.04)	2.84 (2.56)	2.73 (2.34)	1.56 (1.47)	1.87 (1.93)	3.22 (3.45)	3.08 (2.19)
Missing	18543 (29.0%)	5821 (24.9%)	4287 (26.1%)	3869 (24.6%)	3226 (61.7%)	1284 (48.6%)	24 (4.9%)	32 (46.4%)
HbA1c	38.6 (10.5)	37.6 (9.91)	45.4 (17.3)	37.0 (4.78)	57.6 (NA)	NA (NA)	55.3 (23.5)	NA (NA)
Missing	46694 (73.0%)	14607 (62.5%)	13834 (84.4%)	9831 (62.5%)	5228 (100.0%)	2643 (100%)	482 (99.0%)	69 (100%)
T2D Status								
T2D	15538 (24.3%)	3238 (13.9%)	4883 (29.8%)	4166 (26.5%)	2117 (40.5%)	963 (36.4%)	143 (29.4%)	28 (40.6%)

Table 4.4. Distribution of variables (WHRadjBMI analysis set)

	Total (N=63944)	**Non-Hispanic White (N=23379)**	**Non-Hispanic Black (N=16397)**	**Hispanic (N=15740)**	**Asian (N=5229)**	**Native Hawaiian (N=2643)**	**American Indian (N=487)**	**other (N=69)**
Pre-diabetes	11838 (18.5%)	5822 (24.9%)	2283 (13.9%)	3371 (21.4%)	188 (3.6%)	91 (3.4%)	74 (15.2%)	9 (13.0%)
T2D controls	32160 (50.3%)	12693 (54.3%)	8358 (51.0%)	6294 (40.0%)	2924 (55.9%)	1589 (60.1%)	270 (55.4%)	32 (46.4%)
Other controls	4408 (6.9%)	1626 (7.0%)	873 (5.3%)	1909 (12.1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Smoking status								
Non-smoker	31025 (48.5%)	10642 (45.5%)	7523 (45.9%)	8817 (56.0%)	2628 (50.3%)	1143 (43.2%)	238 (48.9%)	34 (49.3%)
Former smoker	22345 (34.9%)	8757 (37.5%)	5834 (35.6%)	4423 (28.1%)	2121 (40.6%)	996 (37.7%)	186 (38.2%)	28 (40.6%)
Current smoker	10000 (15.6%)	3858 (16.5%)	2831 (17.3%)	2311 (14.7%)	454 (8.7%)	489 (18.5%)	51 (10.5%)	6 (8.7%)
Missing	574 (0.9%)	122 (0.5%)	209 (1.3%)	189 (1.2%)	26 (0.5%)	15 (0.6%)	12 (2.5%)	1 (1.4%)
Physical activity status								
Physically active	44515 (69.6%)	15329 (65.6%)	10613 (64.7%)	11369 (72.2%)	4486 (85.8%)	2324 (87.9%)	348 (71.5%)	46 (66.7%)
Physically non-active	12750 (19.9%)	4253 (18.2%)	3748 (22.9%)	3730 (23.7%)	629 (12.0%)	245 (9.3%)	125 (25.7%)	20 (29.0%)
Missing	6679 (10.4%)	3797 (16.2%)	2036 (12.4%)	641 (4.1%)	114 (2.2%)	74 (2.8%)	14 (2.9%)	3 (4.3%)

Table 4.5. Prediction Performance of PRS-BMI in PAGE by PRS estimation methods

Ancestry	PRS_method	R^2	SE (R^2)	95% LCL	95% UCL	Beta	SE(beta)	t_value	P-value	N
Pooled	PT(ALL)	0.0660	0.0023	0.0615	0.0704	0.2568	0.0046	55.8457	~0	44170
Pooled	PRSCS-auto	0.0893	0.0026	0.0843	0.0944	0.2988	0.0045	65.8207	~0	44170
Pooled	PRSCS-(Phi-tuned by ancestry)	0.0895	0.0026	0.0845	0.0946	0.2992	0.0045	65.9076	~0	44170
Pooled	PRSCSx-META-auto	0.0578	0.0022	0.0536	0.0620	0.2403	0.0046	52.0420	~0	44170
Pooled	PRSCSx-META-phi.tuned	0.0685	0.0023	0.0639	0.0730	0.2617	0.0046	56.9897	~0	44170
Pooled	PRSCSx-phiauto	0.0028	0.0005	0.0018	0.0038	0.0529	0.0048	11.1303	9.75E-29	44170
Pooled	PRSCSx-phituned	0.0103	0.0010	0.0084	0.0122	0.1016	0.0047	21.4651	1.09E-101	44170
Non-Hispanic White	PT(ALL)	0.1060	0.0052	0.0959	0.1161	0.3255	0.0084	38.8142	1.52E-311	12708
Non-Hispanic White	PT(Anc-specific)	0.0975	0.0050	0.0877	0.1072	0.3121	0.0084	37.0396	2.86E-285	12708
Non-Hispanic White	PRSCS-auto	0.1395	0.0057	0.1283	0.1506	0.3733	0.0082	45.3769	~0	12708
Non-Hispanic White	PRSCS-(Phi-tuned by ancestry)	0.1439	0.0058	0.1326	0.1552	0.3792	0.0082	46.2094	~0	12708
Non-Hispanic White	PRSCSx-META-auto	0.1378	0.0057	0.1267	0.1490	0.3711	0.0082	45.0703	~0	12708
Non-Hispanic White	PRSCSx-META-phi.tuned	0.1483	0.0058	0.1369	0.1597	0.3849	0.0082	47.0275	~0	12708
Non-Hispanic White	PRSCSx-phiauto	0.1418	0.0057	0.1305	0.1530	0.3764	0.0082	45.8145	~0	12708
Non-Hispanic White	PRSCSx-phituned	0.0114	0.0019	0.0077	0.0150	0.1066	0.0088	12.0858	1.91E-33	12708
Non-Hispanic Black	PT(ALL)	0.0459	0.0036	0.0388	0.0530	0.2141	0.0087	24.6386	5.88E-131	12626
Non-Hispanic Black	PT(Anc-specific)	0.0474	0.0037	0.0402	0.0546	0.2176	0.0087	25.0599	2.68E-135	12626
Non-Hispanic Black	PRSCS-auto	0.0711	0.0044	0.0625	0.0797	0.2665	0.0086	31.0842	1.79E-204	12626
Non-Hispanic Black	PRSCS-(Phi-tuned by ancestry)	0.0723	0.0044	0.0636	0.0810	0.2687	0.0086	31.3560	6.74E-208	12626
Non-Hispanic Black	PRSCSx-META-auto	0.0675	0.0043	0.0591	0.0760	0.2598	0.0086	30.2348	6.16E-194	12626
Non-Hispanic Black	PRSCSx-META-phi.tuned	0.0670	0.0043	0.0586	0.0754	0.2588	0.0086	30.1090	2.14E-192	12626
Non-Hispanic Black	PRSCSx-phiauto	0.0382	0.0033	0.0316	0.0447	0.1953	0.0087	22.3843	7.25E-109	12626
Non-Hispanic Black	PRSCSx-phituned	0.0111	0.0019	0.0074	0.0147	0.1051	0.0088	11.8835	2.14E-32	12626
Hispanic	PT(ALL)	0.0866	0.0047	0.0773	0.0959	0.2942	0.0084	34.9795	3.43E-256	12907
Hispanic	PT(Anc-specific)	0.0820	0.0046	0.0729	0.0911	0.2863	0.0084	33.9537	4.06E-242	12907
Hispanic	PRSCS-auto	0.1202	0.0054	0.1096	0.1307	0.3465	0.0083	41.9802	~0	12907
Hispanic	PRSCS-(Phi-tuned by ancestry)	0.1214	0.0054	0.1109	0.1320	0.3483	0.0082	42.2315	~0	12907

Table 4.5. Prediction Performance of PRS-BMI in PAGE by PRS estimation methods

Ancestry	PRS_method	R^2	SE (R^2)	95% LCL	95% UCL	Beta	SE(beta)	t_value	P-value	N
Hispanic	PRSCSx-META-auto	0.1162	0.0053	0.1058	0.1266	0.3407	0.0083	41.1878	~0	12907
Hispanic	PRSCSx-META-phi.tuned	0.1167	0.0053	0.1063	0.1272	0.3416	0.0083	41.3011	~0	12907
Hispanic	PRSCSx-phiauto	0.0844	0.0047	0.0753	0.0936	0.2905	0.0084	34.5003	1.42E-249	12907
Hispanic	PRSCSx-phituned	0.0206	0.0025	0.0158	0.0255	0.1436	0.0087	16.4863	1.91E-60	12907
Asian	PT(ALL)	0.0681	0.0083	0.0519	0.0844	0.2607	0.0164	15.8659	9.11E-55	3445
Asian	PT(Anc-specific)	0.0375	0.0063	0.0250	0.0499	0.1933	0.0167	11.5749	2.00E-30	3445
Asian	PRSCS-auto	0.0997	0.0097	0.0807	0.1187	0.3154	0.0162	19.5248	1.33E-80	3445
Asian	PRSCS-(Phi-tuned by ancestry)	0.1040	0.0098	0.0847	0.1233	0.3221	0.0161	19.9915	3.29E-84	3445
Asian	PRSCSx-META-auto	0.0989	0.0096	0.0800	0.1178	0.3142	0.0162	19.4429	5.61E-80	3445
Asian	PRSCSx-META-phi.tuned	0.1055	0.0099	0.0861	0.1249	0.3245	0.0161	20.1550	1.73E-85	3445
Asian	PRSCSx-phiauto	0.0926	0.0094	0.0742	0.1110	0.3040	0.0162	18.7446	9.97E-75	3445
Asian	PRSCSx-phituned	0.0104	0.0034	0.0037	0.0171	-0.1019	0.0169	-6.0179	1.95E-09	3445
Native Hawaiian	PT(ALL)	0.0797	0.0126	0.0551	0.1043	0.2817	0.0232	12.1488	1.25E-32	1706
Native Hawaiian	PT(Anc-specific)	0.0649	0.0115	0.0424	0.0875	0.2543	0.0234	10.8792	1.07E-26	1706
Native Hawaiian	PRSCS-auto	0.0965	0.0136	0.0699	0.1231	0.3100	0.0230	13.4924	1.71E-39	1706
Native Hawaiian	PRSCS-(Phi-tuned by ancestry)	0.0937	0.0134	0.0674	0.1200	0.3054	0.0230	13.2708	2.54E-38	1706
Native Hawaiian	PRSCSx-META-auto	0.1009	0.0138	0.0738	0.1280	0.3170	0.0229	13.8283	2.68E-41	1706
Native Hawaiian	PRSCSx-META-phi.tuned	0.1021	0.0139	0.0749	0.1292	0.3188	0.0229	13.9167	8.85E-42	1706
Native Hawaiian	PRSCSx-phiauto	0.1050	0.0140	0.0775	0.1325	0.3234	0.0229	14.1417	5.14E-43	1706
Native Hawaiian	PRSCSx-phituned	0.0275	0.0078	0.0122	0.0428	0.1656	0.0238	6.9469	5.29E-12	1706
American Indian	PT(ALL)	0.0422	0.0236	-0.0043	0.0886	0.2002	0.0582	3.4421	0.000669	271
American Indian	PT(Anc-specific)	0.0409	0.0233	-0.0049	0.0867	0.1970	0.0582	3.3858	0.000816	271
American Indian	PRSCS-auto	0.0430	0.0238	-0.0038	0.0899	0.2022	0.0581	3.4778	0.000589	271
American Indian	PRSCS-(Phi-tuned by ancestry)	0.0415	0.0234	-0.0046	0.0876	0.1985	0.0582	3.4128	0.000742	271
American Indian	PRSCSx-META-auto	0.0442	0.0241	-0.0033	0.0916	0.2048	0.0581	3.5252	0.000497	271
American Indian	PRSCSx-META-phi.tuned	0.0477	0.0249	-0.0014	0.0968	0.2128	0.0580	3.6708	0.000292	271
American Indian	PRSCSx-phiauto	0.0011	0.0041	-0.0068	0.0091	0.0330	0.0594	0.5555	0.579	271
American Indian	PRSCSx-phituned	0.0061	0.0093	-0.0122	0.0244	0.0761	0.0592	1.2849	0.2	271

Table 4.6. Prediction Performance of PRS-WHRadjBMI in PAGE by PRS estimation methods

Ancestry	PRS_method	R^2	SE (R^2)	95% LCL	95% UCL	Beta	SE(beta)	t_value	P-value	N
Pooled	PT(ALL)	0.0287	0.0018	0.0251	0.0323	0.1695	0.0055	30.7669	7.08E-205	31993
Pooled	PRSCS-auto	0.0438	0.0022	0.0394	0.0482	0.2093	0.0055	38.2870	1.23E-313	31993
Pooled	PRSCS-(Phi-tuned by ancestry)	0.0464	0.0023	0.0419	0.0509	0.2153	0.0055	39.4481	~0	31993
Pooled	PRSCSx-META-auto	0.0201	0.0016	0.0170	0.0231	0.1416	0.0055	25.5965	4.62E-143	31993
Pooled	PRSCSx-META-phi.tuned	0.0407	0.0022	0.0365	0.0450	0.2018	0.0055	36.8503	3.61E-291	31993
Pooled	PRSCSx-phiauto	0.0061	0.0009	0.0044	0.0078	0.0782	0.0056	14.0230	1.53E-44	31993
Pooled	PRSCSx-phituned	0.0012	0.0004	0.0004	0.0020	0.0347	0.0056	6.2107	5.34E-10	31993
Non-Hispanic White	PT(ALL)	0.0513	0.0040	0.0435	0.0591	0.2264	0.0090	25.1425	6.68E-136	11696
Non-Hispanic White	PT(Anc-specific)	0.0505	0.0039	0.0428	0.0582	0.2247	0.0090	24.9432	7.70E-134	11696
Non-Hispanic White	PRSCS-auto	0.0713	0.0046	0.0623	0.0803	0.2670	0.0089	29.9715	3.22E-190	11696
Non-Hispanic White	PRSCS-(Phi-tuned by ancestry)	0.0815	0.0048	0.0720	0.0910	0.2853	0.0089	32.2084	3.84E-218	11696
Non-Hispanic White	PRSCSx-META-auto	0.0807	0.0048	0.0712	0.0901	0.2839	0.0089	32.0290	7.70E-216	11696
Non-Hispanic White	PRSCSx-META-phi.tuned	0.0874	0.0050	0.0777	0.0972	0.2956	0.0088	33.4722	1.17E-234	11696
Non-Hispanic White	PRSCSx-phiauto	0.0763	0.0047	0.0671	0.0856	0.2762	0.0089	31.0859	6.44E-204	11696
Non-Hispanic White	PRSCSx-phituned	0.0003	0.0003	-0.0003	0.0010	0.0178	0.0092	1.9211	0.0547	11696
Non-Hispanic Black	PT(ALL)	0.0179	0.0029	0.0122	0.0236	0.1338	0.0109	12.2361	3.95E-34	8207
Non-Hispanic Black	PT(Anc-specific)	0.0134	0.0025	0.0084	0.0183	0.1156	0.0110	10.5505	7.39E-26	8207
Non-Hispanic Black	PRSCS-auto	0.0268	0.0035	0.0199	0.0337	0.1637	0.0109	15.0398	1.89E-50	8207
Non-Hispanic Black	PRSCS-(Phi-tuned by ancestry)	0.0248	0.0034	0.0182	0.0314	0.1574	0.0109	14.4466	9.83E-47	8207
Non-Hispanic Black	PRSCSx-META-auto	0.0072	0.0019	0.0035	0.0108	0.0847	0.0110	7.7050	1.46E-14	8207
Non-Hispanic Black	PRSCSx-META-phi.tuned	0.0212	0.0031	0.0151	0.0274	0.1456	0.0109	13.3385	3.61E-40	8207
Non-Hispanic Black	PRSCSx-phiauto	0.0257	0.0034	0.0189	0.0324	0.1601	0.0109	14.7023	2.56E-48	8207
Non-Hispanic Black	PRSCSx-phituned	0.0057	0.0016	0.0024	0.0089	0.0751	0.0110	6.8288	9.17E-12	8207
Hispanic	PT(ALL)	0.0297	0.0038	0.0223	0.0371	0.1722	0.0111	15.5166	1.65E-53	7876
Hispanic	PT(Anc-specific)	0.0293	0.0037	0.0219	0.0366	0.1710	0.0111	15.4078	8.51E-53	7876
Hispanic	PRSCS-auto	0.0418	0.0044	0.0332	0.0505	0.2044	0.0110	18.5394	3.89E-75	7876
Hispanic	PRSCS-(Phi-tuned by ancestry)	0.0432	0.0045	0.0344	0.0519	0.2076	0.0110	18.8445	1.62E-77	7876

Table 4.6. Prediction Performance of PRS-WHRadjBMI in PAGE by PRS estimation methods

Ancestry	PRS_method	R^2	SE (R^2)	95% LCL	95% UCL	Beta	SE(beta)	t_value	P-value	N
Hispanic	PRSCSx-META-auto	0.0232	0.0034	0.0167	0.0298	0.1524	0.0111	13.6892	3.57E-42	7876
Hispanic	PRSCSx-META-phi.tuned	0.0400	0.0043	0.0315	0.0485	0.1998	0.0110	18.1077	7.84E-72	7876
Hispanic	PRSCSx-phiauto	0.0374	0.0042	0.0291	0.0456	0.1932	0.0111	17.4806	3.69E-67	7876
Hispanic	PRSCSx-phituned	0.0001	0.0003	-0.0004	0.0006	0.0114	0.0113	1.0141	0.311	7876
Asian	PT(ALL)	0.0141	0.0046	0.0051	0.0231	0.1186	0.0194	6.1113	1.14E-09	2614
Asian	PT(Anc-specific)	0.0041	0.0025	-0.0008	0.0090	0.0639	0.0195	3.2774	0.00106	2614
Asian	PRSCS-auto	0.0181	0.0052	0.0080	0.0282	0.1343	0.0194	6.9362	5.06E-12	2614
Asian	PRSCS-(Phi-tuned by ancestry)	0.0158	0.0048	0.0063	0.0253	0.1255	0.0194	6.4737	1.14E-10	2614
Asian	PRSCSx-META-auto	0.0173	0.0051	0.0074	0.0272	0.1315	0.0194	6.7877	1.41E-11	2614
Asian	PRSCSx-META-phi.tuned	0.0093	0.0037	0.0020	0.0166	0.0962	0.0194	4.9493	7.92E-07	2614
Asian	PRSCSx-phiauto	0.0130	0.0044	0.0044	0.0216	0.1137	0.0194	5.8560	5.34E-09	2614
Asian	PRSCSx-phituned	0.0020	0.0017	-0.0014	0.0054	0.0447	0.0195	2.2899	0.0221	2614
Native Hawaiian	PT(ALL)	0.0180	0.0072	0.0038	0.0321	0.1337	0.0272	4.9126	1.01E-06	1321
Native Hawaiian	PT(Anc-specific)	0.0229	0.0081	0.0070	0.0388	0.1509	0.0271	5.5594	3.27E-08	1321
Native Hawaiian	PRSCS-auto	0.0348	0.0099	0.0154	0.0542	0.1861	0.0270	6.8989	8.11E-12	1321
Native Hawaiian	PRSCS-(Phi-tuned by ancestry)	0.0294	0.0091	0.0115	0.0473	0.1710	0.0271	6.3210	3.55E-10	1321
Native Hawaiian	PRSCSx-META-auto	0.0177	0.0072	0.0036	0.0318	0.1327	0.0272	4.8770	1.21E-06	1321
Native Hawaiian	PRSCSx-META-phi.tuned	0.0326	0.0096	0.0138	0.0514	0.1800	0.0270	6.6636	3.91E-11	1321
Native Hawaiian	PRSCSx-phiauto	0.0009	0.0016	-0.0023	0.0041	0.0295	0.0274	1.0765	0.282	1321
Native Hawaiian	PRSCSx-phituned	0.0004	0.0011	-0.0018	0.0027	0.0206	0.0275	0.7511	0.453	1321
American Indian	PT(ALL)	0.0481	0.0264	-0.0038	0.1000	0.2182	0.0624	3.4979	0.000558	244
American Indian	PT(Anc-specific)	0.0488	0.0265	-0.0034	0.1011	0.2199	0.0624	3.5254	0.000506	244
American Indian	PRSCS-auto	0.0622	0.0295	0.0040	0.1203	0.2481	0.0619	4.0061	8.22E-05	244
American Indian	PRSCS-(Phi-tuned by ancestry)	0.0690	0.0309	0.0082	0.1298	0.2613	0.0617	4.2355	3.24E-05	244
American Indian	PRSCSx-META-auto	0.0312	0.0216	-0.0114	0.0737	0.1756	0.0629	2.7894	0.0057	244
American Indian	PRSCSx-META-phi.tuned	0.0618	0.0294	0.0038	0.1197	0.2472	0.0619	3.9910	8.72E-05	244
American Indian	PRSCSx-phiauto	0.0000	0.0004	-0.0008	0.0008	-0.0031	0.0639	-0.0478	0.962	244
American Indian	PRSCSx-phituned	0.0011	0.0043	-0.0072	0.0095	0.0335	0.0639	0.5249	0.6	244

Table 4.7. Stratum-specific prediction performance of PRS-BMI by different stratifying variables

Contextual variables	Subgroups	trait	beta	SE	P_value	t_value	LCL	UCL	R^2	SE(R^2)	LCL (R^2)	UCL (R^2)	N
Age group	<=50 yrs	BMII	2.087	0.041	0	50.643	2.007	2.168	0.079	0.003	0.072	0.085	24234
	>50 yrs	BMI	1.971	0.023	0	87.139	1.926	2.015	0.094	0.002	0.089	0.098	64111
sex	Female	BMI	2.222	0.026	0	86.320	2.171	2.272	0.093	0.002	0.089	0.098	60447
	Male	BMI	1.546	0.029	0	52.806	1.488	1.603	0.079	0.003	0.073	0.085	27898
T2D_status	Prediabetes	BMI	1.998	0.050	0	39.890	1.900	2.096	0.096	0.005	0.086	0.106	12682
	Diebetes	BMI	1.818	0.039	0	46.583	1.741	1.894	0.080	0.003	0.073	0.087	23387
	Control	BMI	1.751	0.026	0	67.334	1.700	1.802	0.084	0.003	0.079	0.090	42640
Hypertension status	Normotensive	BMI	1.797	0.026	0	68.356	1.745	1.848	0.081	0.003	0.076	0.086	43491
	Hypertensive	BMI	2.015	0.029	0	69.485	1.958	2.071	0.091	0.003	0.086	0.096	43503
Smoking status	Former smoker	BMI	2.075	0.034	0	60.774	2.008	2.142	0.096	0.003	0.090	0.102	29811
	Non-smoker	BMI	2.039	0.029	0	70.639	1.982	2.095	0.091	0.003	0.085	0.096	42387
	Current smoker	BMI	1.859	0.049	3.97E-301	38.015	1.763	1.954	0.080	0.004	0.071	0.088	14138
Physical activity status	Non-sedentary	BMI	1.906	0.024	0	77.798	1.858	1.954	0.094	0.002	0.089	0.099	49338
	Sedentary	BMI	2.171	0.052	0	41.956	2.070	2.273	0.099	0.005	0.090	0.108	14146

Table 4.8. Stratum-specific prediction performance of PRS-WHRadjBMI by different stratifying variables

Contextual variables	Subgroups	trait	beta	SE	P_value	t_value	LCL	UCL	R^2	SE(R^2)	LCL (R^2)	UCL (R^2)	N
Age group	<=50 yrs	WHR	0.014	0.001	2.28E-150	26.426	0.013	0.015	0.047	0.003	0.040	0.054	14664
	>50 yrs	WHR	0.015	<0.001	0	47.092	0.014	0.015	0.042	0.002	0.039	0.046	49280
sex	Female	WHR	0.017	<0.001	0	50.850	0.016	0.017	0.053	0.002	0.049	0.057	46897
	Male	WHR	0.008	<0.001	1.46E-76	18.613	0.007	0.008	0.020	0.002	0.016	0.025	17047
T2D_status	Prediabetes	WHR	0.013	0.001	1.34E-106	22.156	0.012	0.014	0.041	0.004	0.034	0.048	11838
	Diebetes	WHR	0.012	0.001	1.68E-93	20.653	0.011	0.013	0.026	0.003	0.021	0.031	15538
	Control	WHR	0.014	<0.001	3.53E-316	38.445	0.013	0.015	0.044	0.002	0.039	0.048	32160
Hypertension status	Normotensive	WHR	0.015	<0.001	0	41.485	0.014	0.015	0.051	0.002	0.046	0.056	32147
	Hypertensive	WHR	0.013	<0.001	1.13E-235	33.072	0.013	0.014	0.034	0.002	0.030	0.038	30520
Smoking status	Former smoker	WHR	0.014	<0.001	9.90E-214	31.546	0.013	0.015	0.040	0.003	0.035	0.045	22345
	Non-smoker	WHR	0.015	<0.001	5.23E-314	38.324	0.014	0.016	0.046	0.002	0.041	0.050	31025
	Current smoker	WHR	0.013	0.001	7.60E-91	20.422	0.012	0.015	0.040	0.004	0.033	0.048	10000
Physical activity status	Non-sedentary	WHR	0.014	<0.001	0	45.395	0.014	0.015	0.044	0.002	0.040	0.047	44515
	Sedentary	WHR	0.014	0.001	1.71E-119	23.492	0.013	0.016	0.041	0.003	0.035	0.048	12750

Table 4.9. Distributions of variables by stratifying variables (BMI set) (1) Age

	AGE		
Cardiometabolic Profile	**<=50 yrs**	**>50 yrs**	
BMI			
Mean (SD)	28.4 (6.62)	28.7 (5.91)	<0.001
Median [Min, Max]	27.2 [12.9, 66.8]	27.8 [10.9, 67.0]	
Sex			
Male	9727 (40.1%)	18171 (28.3%)	<0.001
Female	14507 (59.9%)	45940 (71.7%)	
Age			
Mean (SD)	40.1 (9.71)	60.4 (5.57)	<0.001
Median [Min, Max]	45.0 [18.0, 50.0]	60.0 [51.0, 70.0]	
HDL			
Mean (SD)	50.4 (15.5)	50.9 (16.2)	<0.001
Median [Min, Max]	48.0 [4.00, 142]	49.0 [3.70, 142]	
Missing	8855 (36.5%)	33991 (53.0%)	
LDL			
Mean (SD)	125 (38.9)	142 (40.2)	<0.001
Median [Min, Max]	121 [11.6, 363]	139 [10.2, 372]	
Missing	9075 (37.4%)	34662 (54.1%)	
logTG			
Mean (SD)	4.61 (0.580)	4.79 (0.509)	<0.001
Median [Min, Max]	4.57 [2.08, 6.52]	4.78 [2.48, 6.53]	
Missing	8796 (36.3%)	34002 (53.0%)	
Total Cholesterol			
Mean (SD)	198 (43.5)	220 (44.7)	<0.001
Median [Min, Max]	194 [38.2, 461]	217 [39.1, 483]	
Missing	8765 (36.2%)	33884 (52.9%)	
T2D status			
T2D cases	3604 (14.9%)	19783 (30.9%)	<0.001
Prediabetes	3465 (14.3%)	9217 (14.4%)	
T2D control	7529 (31.1%)	35111 (54.8%)	
Other controls	9636 (39.8%)	0 (0%)	
FBG			
Mean (SD)	5.12 (0.825)	5.55 (1.34)	<0.001
Median [Min, Max]	5.05 [0.160, 14.6]	5.28 [0.160, 14.8]	
Missing	12194 (50.3%)	28230 (44.0%)	
FBI			
Mean (SD)	10.7 (8.32)	10.4 (15.7)	0.003
Median [Min, Max]	8.60 [0.559, 225]	8.00 [0.288, 1580]	
Missing	12182 (50.3%)	28463 (44.4%)	
HOMA-IR			
Mean (SD)	2.51 (2.09)	2.59 (2.37)	<0.001
Median [Min, Max]	1.93 [0.0267, 27.9]	1.92 [0.0169, 29.2]	
Missing	12250 (50.5%)	29229 (45.6%)	
HbA1c			
Mean (SD)	38.8 (12.3)	42.2 (13.9)	<0.001
Median [Min, Max]	35.5 [4.90, 123]	38.8 [14.7, 123]	
Missing	14796 (61.1%)	51435 (80.2%)	
Hypertension Status			
Normotensive	16921 (69.8%)	26570 (41.4%)	<0.001
Hypertenstive	7027 (29.0%)	36476 (56.9%)	
Missing	286 (1.2%)	1065 (1.7%)	
SBP			
Mean (SD)	121 (19.6)	134 (21.4)	<0.001

Table 4.9. Distributions of variables by stratifying variables (BMI set) (1) Age

	AGE		
Cardiometabolic Profile	**<=50 yrs**	**>50 yrs**	
Median [Min, Max]	118 [67.0, 257]	132 [61.0, 247]	
Missing	5704 (23.5%)	19725 (30.8%)	
DBP			
Mean (SD)	75.4 (13.0)	79.7 (11.9)	<0.001
Median [Min, Max]	74.0 [23.0, 150]	80.0 [12.0, 139]	
Missing	5705 (23.5%)	19695 (30.7%)	
MI status			
No	23261 (96.0%)	56996 (88.9%)	<0.001
Yes	973 (4.0%)	7115 (11.1%)	
Stroke status			
No	23410 (96.6%)	56840 (88.7%)	<0.001
Yes	824 (3.4%)	7271 (11.3%)	
Physical status			
Non-sedentary	11082 (45.7%)	38256 (59.7%)	<0.001
Sedantary	2956 (12.2%)	11190 (17.5%)	
Missing	10196 (42.1%)	14665 (22.9%)	
Smoking status			
Never smokers	13038 (53.8%)	29349 (45.8%)	<0.001
Former smokers	5352 (22.1%)	24459 (38.2%)	
Current smokers	5213 (21.5%)	8925 (13.9%)	
Missing	631 (2.6%)	1378 (2.1%)	

Table 4.10. Distributions of variables by stratifying variables (BMI set) (2) Sex

	SEX		
Cardiometabolic Profile	**Male**	**Female**	
BMI			
Mean (SD)	27.7 (4.93)	29.1 (6.55)	<0.001
Median [Min, Max]	27.0 [10.9, 65.0]	28.1 [11.5, 67.0]	
Sex			
Male	27898 (100%)	0 (0%)	<0.001
Female	0 (0%)	60447 (100%)	
Age			
Mean (SD)	52.9 (12.0)	55.8 (11.0)	<0.001
Median [Min, Max]	55.0 [18.0, 70.0]	58.0 [18.0, 70.0]	
HDL			
Mean (SD)	43.9 (14.1)	54.4 (15.8)	<0.001
Median [Min, Max]	42.0 [3.70, 142]	52.0 [4.70, 142]	
Missing	12010 (43.0%)	30836 (51.0%)	
LDL			
Mean (SD)	133 (39.0)	138 (41.3)	<0.001
Median [Min, Max]	131 [11.6, 363]	134 [10.2, 372]	
Missing	12312 (44.1%)	31425 (52.0%)	
logTG			
Mean (SD)	4.78 (0.563)	4.71 (0.527)	<0.001
Median [Min, Max]	4.77 [2.08, 6.53]	4.69 [2.77, 6.52]	
Missing	11865 (42.5%)	30933 (51.2%)	
Total Cholesterol			
Mean (SD)	204 (43.6)	217 (45.7)	<0.001
Median [Min, Max]	201 [38.2, 440]	214 [42.3, 483]	
Missing	11903 (42.7%)	30746 (50.9%)	
T2D status			
T2D cases	8134 (29.2%)	15253 (25.2%)	<0.001
Prediabetes	4618 (16.6%)	8064 (13.3%)	
T2D control	11261 (40.4%)	31379 (51.9%)	
Other controls	3885 (13.9%)	5751 (9.5%)	
FBG			
Mean (SD)	5.40 (1.02)	5.45 (1.32)	<0.001
Median [Min, Max]	5.33 [0.630, 14.8]	5.17 [0.160, 14.7]	
Missing	15498 (55.6%)	24926 (41.2%)	
FBI			
Mean (SD)	10.3 (8.36)	10.5 (15.8)	0.182
Median [Min, Max]	8.00 [0.559, 210]	8.20 [0.288, 1580]	
Missing	15490 (55.5%)	25155 (41.6%)	
HOMA-IR			
Mean (SD)	2.54 (2.19)	2.57 (2.34)	0.182
Median [Min, Max]	1.95 [0.0613, 28.9]	1.91 [0.0169, 29.2]	
Missing	15574 (55.8%)	25905 (42.9%)	
HbA1c			
Mean (SD)	40.6 (13.2)	40.9 (13.4)	0.188
Median [Min, Max]	37.7 [14.7, 123]	37.7 [4.90, 123]	
Missing	18446 (66.1%)	47785 (79.1%)	
Hypertension Status			
Normotensive	13932 (49.9%)	29559 (48.9%)	0.528
Hypertenstive	13848 (49.6%)	29655 (49.1%)	
Missing	118 (0.4%)	1233 (2.0%)	
SBP			
Mean (SD)	129 (20.9)	130 (21.8)	<0.001

Table 4.10. Distributions of variables by stratifying variables (BMI set) (2) Sex

	SEX		
Cardiometabolic Profile	**Male**	**Female**	
Median [Min, Max]	126 [61.0, 247]	128 [72.0, 257]	
Missing	12430 (44.6%)	12999 (21.5%)	
DBP			
Mean (SD)	78.9 (13.0)	78.3 (12.1)	<0.001
Median [Min, Max]	78.0 [12.0, 150]	78.0 [27.0, 139]	
Missing	12410 (44.5%)	12990 (21.5%)	
MI status			
No	24302 (87.1%)	55955 (92.6%)	<0.001
Yes	3596 (12.9%)	4492 (7.4%)	
Stroke status			
No	24863 (89.1%)	55387 (91.6%)	<0.001
Yes	3035 (10.9%)	5060 (8.4%)	
Physical status			
Non-sedentary	14479 (51.9%)	34859 (57.7%)	<0.001
Sedantary	3723 (13.3%)	10423 (17.2%)	
Missing	9696 (34.8%)	15165 (25.1%)	
Smoking status			
Never smokers	9810 (35.2%)	32577 (53.9%)	<0.001
Former smokers	11425 (41.0%)	18386 (30.4%)	
Current smokers	5898 (21.1%)	8240 (13.6%)	
Missing	765 (2.7%)	1244 (2.1%)	

Table 4.11. Distributions of variables by stratifying variables (BMI set) (3) T2D status

Cardiometabolic Profile	T2D Control	Prediabetes	Diabetes	
BMI				
Mean (SD)	27.6 (5.51)	29.3 (5.81)	31.0 (6.44)	<0.001
Median [Min, Max]	26.6 [12.9, 67.0]	28.4 [11.5, 65.4]	29.9 [10.9, 66.8]	
Sex				
Male	11261 (26.4%)	4618 (36.4%)	8134 (34.8%)	<0.001
Female	31379 (73.6%)	8064 (63.6%)	15253 (65.2%)	
Age				
Mean (SD)	58.3 (7.19)	55.3 (8.86)	58.5 (7.49)	<0.001
Median [Min, Max]	59.0 [42.0, 70.0]	56.0 [18.0, 70.0]	59.0 [22.0, 70.0]	
HDL				
Mean (SD)	53.1 (16.9)	49.6 (15.2)	45.9 (14.3)	<0.001
Median [Min, Max]	51.0 [3.70, 142]	47.2 [6.00, 142]	44.0 [4.10, 129]	
Missing	22877 (53.7%)	2090 (16.5%)	14090 (60.2%)	
LDL				
Mean (SD)	138 (40.0)	141 (39.1)	143 (41.6)	<0.001
Median [Min, Max]	136 [10.2, 371]	138 [14.7, 372]	140 [10.5, 363]	
Missing	23234 (54.5%)	2273 (17.9%)	14371 (61.4%)	
logTG				
Mean (SD)	4.68 (0.507)	4.82 (0.498)	4.97 (0.517)	<0.001
Median [Min, Max]	4.65 [2.48, 6.50]	4.80 [3.00, 6.51]	4.94 [2.48, 6.53]	
Missing	22913 (53.7%)	2127 (16.8%)	13996 (59.8%)	
Total Cholesterol				
Mean (SD)	215 (43.8)	218 (43.2)	220 (48.0)	<0.001
Median [Min, Max]	212 [38.2, 483]	215 [48.0, 449]	216 [61.0, 470]	
Missing	22799 (53.5%)	2081 (16.4%)	14011 (59.9%)	
T2D status				
T2D cases	0 (0%)	0 (0%)	23387 (100%)	<0.001
Prediabetes	0 (0%)	12682 (100%)	0 (0%)	
T2D control	42640 (100%)	0 (0%)	0 (0%)	
Other controls				
FBG				
Mean (SD)	4.92 (0.487)	5.76 (0.467)	6.91 (2.22)	<0.001
Median [Min, Max]	5.00 [0.160, 6.99]	5.77 [3.33, 6.95]	6.28 [0.690, 14.8]	
Missing	18666 (43.8%)	802 (6.3%)	15532 (66.4%)	
FBI				
Mean (SD)	7.92 (5.71)	12.6 (8.98)	15.0 (30.2)	0.003
Median [Min, Max]	6.65 [0.288, 147]	10.5 [0.559, 185]	11.1 [1.00, 1580]	
Missing	18950 (44.4%)	936 (7.4%)	15344 (65.6%)	
HOMA-IR				
Mean (SD)	1.75 (1.28)	3.23 (2.26)	4.23 (3.56)	<0.001
Median [Min, Max]	1.45 [0.0169, 26.6]	2.66 [0.104, 28.0]	3.20 [0.120, 29.2]	
Missing	19330 (45.3%)	943 (7.4%)	15773 (67.4%)	
HbA1c				

Table 4.11. Distributions of variables by stratifying variables (BMI set) (3) T2D status

Cardiometabolic Profile	T2D Control	Prediabetes	Diabetes	
Mean (SD)	34.7 (4.21)	38.3 (4.31)	61.0 (20.1)	<0.001
Median [Min, Max]	34.4 [14.7, 107]	38.8 [15.8, 54.1]	55.2 [4.90, 123]	
Missing	35743 (83.8%)	3967 (31.3%)	19493 (83.3%)	
Hypertension Status				
Normotensive	22135 (51.9%)	7291 (57.5%)	5710 (24.4%)	<0.001
Hypertenstive	19560 (45.9%)	5248 (41.4%)	17414 (74.5%)	
Missing	945 (2.2%)	143 (1.1%)	263 (1.1%)	
SBP				
Mean (SD)	130 (21.0)	129 (20.8)	141 (21.4)	<0.001
Median [Min, Max]	128 [72.0, 247]	127 [61.0, 257]	139 [67.0, 243]	
Missing	13060 (30.6%)	914 (7.2%)	10560 (45.2%)	
DBP				
Mean (SD)	78.5 (11.9)	78.3 (12.5)	82.6 (11.8)	<0.001
Median [Min, Max]	78.0 [34.0, 138]	78.0 [12.0, 139]	82.0 [40.0, 130]	
Missing	13045 (30.6%)	915 (7.2%)	10546 (45.1%)	
MI status				
No	39786 (93.3%)	11305 (89.1%)	19614 (83.9%)	<0.001
Yes	2854 (6.7%)	1377 (10.9%)	3773 (16.1%)	
Stroke status				
No	39549 (92.8%)	11811 (93.1%)	19345 (82.7%)	<0.001
Yes	3091 (7.2%)	871 (6.9%)	4042 (17.3%)	
Physical status				
Non-sedentary	25603 (60.0%)	7295 (57.5%)	13018 (55.7%)	<0.001
Sedantary	6350 (14.9%)	2465 (19.4%)	4370 (18.7%)	
Missing	10687 (25.1%)	2922 (23.0%)	5999 (25.7%)	
Smoking status				
Never smokers	20180 (47.3%)	5999 (47.3%)	10202 (43.6%)	<0.001
Former smokers	15199 (35.6%)	4101 (32.3%)	9215 (39.4%)	
Current smokers	6224 (14.6%)	2510 (19.8%)	3485 (14.9%)	
Missing	1037 (2.4%)	72 (0.6%)	485 (2.1%)	

Table 4.12. Distributions of variables by stratifying variables (BMI set) (4) Hypertension status

Cardiometabolic Profile	Hypertension status Normotensive	Hypertensive	
BMI			
Mean (SD)	27.4 (5.54)	29.9 (6.42)	<0.001
Median [Min, Max]	26.6 [12.9, 66.6]	28.8 [10.9, 67.0]	
Sex			
Male	13932 (32.0%)	13848 (31.8%)	0.528
Female	29559 (68.0%)	29655 (68.2%)	
Age			
Mean (SD)	51.3 (13.0)	58.3 (8.32)	<0.001
Median [Min, Max]	54.0 [18.0, 70.0]	59.0 [18.0, 70.0]	
HDL			
Mean (SD)	51.9 (15.9)	49.3 (16.0)	<0.001
Median [Min, Max]	50.0 [3.70, 142]	47.0 [4.00, 142]	
Missing	18349 (42.2%)	23541 (54.1%)	
LDL			
Mean (SD)	131 (39.3)	142 (41.5)	<0.001
Median [Min, Max]	128 [16.9, 372]	139 [10.2, 369]	
Missing	18854 (43.4%)	23920 (55.0%)	
logTG			
Mean (SD)	4.65 (0.540)	4.84 (0.524)	<0.001
Median [Min, Max]	4.62 [2.08, 6.52]	4.82 [2.48, 6.53]	
Missing	18465 (42.5%)	23375 (53.7%)	
Total Cholesterol			
Mean (SD)	207 (43.9)	219 (46.5)	<0.001
Median [Min, Max]	204 [39.1, 483]	216 [42.1, 482]	
Missing	18287 (42.0%)	23407 (53.8%)	
T2D status			
T2D cases	5710 (13.1%)	17414 (40.0%)	<0.001
Prediabetes	7291 (16.8%)	5248 (12.1%)	
T2D control	22135 (50.9%)	19560 (45.0%)	
Other controls	8355 (19.2%)	1281 (2.9%)	
FBG			
Mean (SD)	5.27 (0.963)	5.65 (1.50)	<0.001
Median [Min, Max]	5.16 [0.160, 14.8]	5.33 [0.430, 14.7]	
Missing	17290 (39.8%)	22664 (52.1%)	
FBI			
Mean (SD)	9.48 (8.73)	11.8 (19.1)	<0.001
Median [Min, Max]	7.63 [0.288, 546]	9.07 [0.432, 1580]	
Missing	17550 (40.4%)	22632 (52.0%)	
HOMA-IR			
Mean (SD)	2.27 (1.95)	2.95 (2.64)	<0.001
Median [Min, Max]	1.75 [0.0169, 28.3]	2.20 [0.0596, 29.2]	
Missing	17855 (41.1%)	23131 (53.2%)	
HbA1c			
Mean (SD)	37.8 (9.70)	45.9 (16.7)	<0.001
Median [Min, Max]	35.5 [14.7, 121]	39.9 [4.90, 123]	
Missing	29543 (67.9%)	35427 (81.4%)	
Hypertension Status			
Normotensive	43491 (100%)	0 (0%)	<0.001
Hypertenstive	0 (0%)	43503 (100%)	
Missing			
SBP			
Mean (SD)	117 (13.3)	147 (18.3)	<0.001

Table 4.12. Distributions of variables by stratifying variables (BMI set) (4) Hypertension status

	Hypertension status		
Cardiometabolic Profile	**Normotensive**	**Hypertensive**	
Median [Min, Max]	117 [61.0, 225]	145 [76.0, 257]	
Missing	9296 (21.4%)	15880 (36.5%)	
DBP			
Mean (SD)	71.7 (8.97)	86.9 (10.9)	<0.001
Median [Min, Max]	71.0 [12.0, 130]	87.0 [44.0, 150]	
Missing	9293 (21.4%)	15853 (36.4%)	
MI status			
No	41290 (94.9%)	37680 (86.6%)	<0.001
Yes	2201 (5.1%)	5823 (13.4%)	
Stroke status			
No	42099 (96.8%)	36885 (84.8%)	<0.001
Yes	1392 (3.2%)	6618 (15.2%)	
Physical status			
Non-sedentary	24380 (56.1%)	24597 (56.5%)	<0.001
Sedantary	6679 (15.4%)	7389 (17.0%)	
Missing	12432 (28.6%)	11517 (26.5%)	
Smoking status			
Never smokers	21501 (49.4%)	20198 (46.4%)	<0.001
Former smokers	12794 (29.4%)	16559 (38.1%)	
Current smokers	7775 (17.9%)	6178 (14.2%)	
Missing	1421 (3.3%)	568 (1.3%)	

Table 4.13. Distributions of variables by stratifying variables (BMI set) (5) Smoking status

Cardiometabolic Profile	Smoking status Non-smoker	Former smoker	Current smoker	
BMI				
Mean (SD)	28.7 (6.14)	29.1 (6.09)	27.6 (5.93)	<0.001
Median [Min, Max]	27.6 [10.9, 67.0]	28.1 [13.8, 66.6]	26.6 [13.4, 65.3]	
Sex				
Male	9810 (23.1%)	11425 (38.3%)	5898 (41.7%)	<0.001
Female	32577 (76.9%)	18386 (61.7%)	8240 (58.3%)	
Age				
Mean (SD)	53.8 (12.4)	57.7 (8.98)	52.1 (11.4)	<0.001
Median [Min, Max]	56.0 [18.0, 70.0]	59.0 [18.0, 70.0]	54.0 [18.0, 70.0]	
HDL				
Mean (SD)	52.2 (15.7)	49.9 (16.2)	48.4 (15.9)	<0.001
Median [Min, Max]	50.0 [4.70, 139]	48.0 [3.70, 142]	46.0 [4.00, 142]	
Missing	19779 (46.7%)	15655 (52.5%)	5909 (41.8%)	
LDL				
Mean (SD)	135 (40.3)	139 (40.3)	134 (41.8)	<0.001
Median [Min, Max]	132 [10.2, 372]	137 [13.8, 371]	131 [11.6, 369]	
Missing	20198 (47.7%)	15960 (53.5%)	6066 (42.9%)	
logTG				
Mean (SD)	4.69 (0.542)	4.77 (0.532)	4.76 (0.546)	<0.001
Median [Min, Max]	4.68 [2.77, 6.52]	4.75 [2.48, 6.53]	4.74 [2.08, 6.52]	
Missing	19795 (46.7%)	15637 (52.5%)	5880 (41.6%)	
Total Cholesterol				
Mean (SD)	212 (45.2)	216 (45.1)	209 (46.4)	<0.001
Median [Min, Max]	209 [42.1, 482]	213 [39.1, 464]	206 [38.2, 483]	
Missing	19708 (46.5%)	15583 (52.3%)	5862 (41.5%)	
T2D status				
T2D cases	10202 (24.1%)	9215 (30.9%)	3485 (24.6%)	<0.001
Prediabetes	5999 (14.2%)	4101 (13.8%)	2510 (17.8%)	
T2D control	20180 (47.6%)	15199 (51.0%)	6224 (44.0%)	
Other controls	6006 (14.2%)	1296 (4.3%)	1919 (13.6%)	
FBG				
Mean (SD)	5.40 (1.24)	5.52 (1.32)	5.37 (1.08)	<0.001
Median [Min, Max]	5.17 [0.160, 14.7]	5.28 [0.630, 14.8]	5.23 [0.690, 14.7]	
Missing	18755 (44.2%)	13731 (46.1%)	6358 (45.0%)	
FBI				
Mean (SD)	10.5 (15.6)	10.7 (13.1)	9.91 (12.0)	0.003
Median [Min, Max]	8.26 [0.432, 1580]	8.21 [0.288, 655]	8.00 [0.432, 823]	
Missing	18926 (44.7%)	13760 (46.2%)	6377 (45.1%)	
HOMA-IR				
Mean (SD)	2.55 (2.26)	2.66 (2.44)	2.41 (2.11)	<0.001
Median [Min, Max]	1.93 [0.0169, 29.2]	1.95 [0.0498, 28.4]	1.82 [0.0613, 28.9]	
Missing	19315 (45.6%)	14110 (47.3%)	6463 (45.7%)	
HbA1c				
Mean (SD)	40.4 (13.1)	41.2 (14.0)	40.6 (12.5)	<0.001
Median [Min, Max]	36.6 [14.7, 123]	37.7 [4.90, 123]	37.7 [16.9, 122]	
Missing	31629 (74.6%)	23517 (78.9%)	9231 (65.3%)	
Hypertension Status				
Normotensive	21501 (50.7%)	12794 (42.9%)	7775 (55.0%)	<0.001
Hypertenstive	20198 (47.7%)	16559 (55.5%)	6178 (43.7%)	
Missing	688 (1.6%)	458 (1.5%)	185 (1.3%)	
SBP				
Mean (SD)	130 (21.6)	132 (21.2)	127 (22.2)	<0.001

Table 4.13. Distributions of variables by stratifying variables (BMI set) (5) Smoking status

	Smoking status			
Cardiometabolic Profile	**Non-smoker**	**Former smoker**	**Current smoker**	
Median [Min, Max]	128 [72.0, 244]	130 [61.0, 243]	124 [63.0, 257]	
Missing	10823 (25.5%)	9771 (32.8%)	3717 (26.3%)	
DBP				
Mean (SD)	78.4 (12.3)	79.4 (11.9)	76.6 (13.2)	<0.001
Median [Min, Max]	78.0 [23.0, 150]	79.0 [12.0, 139]	76.0 [22.0, 143]	
Missing	10821 (25.5%)	9760 (32.7%)	3718 (26.3%)	
MI status				
No	39527 (93.3%)	26432 (88.7%)	12458 (88.1%)	<0.001
Yes	2860 (6.7%)	3379 (11.3%)	1680 (11.9%)	
Stroke status				
No	39112 (92.3%)	26584 (89.2%)	12665 (89.6%)	<0.001
Yes	3275 (7.7%)	3227 (10.8%)	1473 (10.4%)	
Physical status				
Non-sedentary	23726 (56.0%)	18164 (60.9%)	7055 (49.9%)	<0.001
Sedantary	6832 (16.1%)	4762 (16.0%)	2395 (16.9%)	
Missing	11829 (27.9%)	6885 (23.1%)	4688 (33.2%)	
Smoking status				
Never smokers	42387 (100%)	0 (0%)	0 (0%)	<0.001
Former smokers	0 (0%)	29811 (100%)	0 (0%)	
Current smokers	0 (0%)	0 (0%)	14138 (100%)	
Missing				

Table 4.14. Distributions of variables by stratifying variables (BMI set) (6) Physical activity status

Cardiometabolic Profile	Physical Activity Non-sedentary	Sedentary	
BMI			
Mean (SD)	28.1 (5.62)	29.8 (6.51)	<0.001
Median [Min, Max]	27.2 [11.5, 66.6]	28.8 [10.9, 66.0]	
Sex			
Male	14479 (29.3%)	3723 (26.3%)	<0.001
Female	34859 (70.7%)	10423 (73.7%)	
Age			
Mean (SD)	56.3 (10.8)	56.3 (10.4)	0.611
Median [Min, Max]	58.0 [18.0, 70.0]	58.0 [18.0, 70.0]	
HDL			
Mean (SD)	50.6 (15.6)	49.7 (14.7)	<0.001
Median [Min, Max]	49.0 [3.70, 142]	48.0 [4.00, 135]	
Missing	23129 (46.9%)	6129 (43.3%)	
LDL			
Mean (SD)	136 (39.5)	138 (40.0)	0.002
Median [Min, Max]	134 [10.5, 361]	135 [10.2, 371]	
Missing	23621 (47.9%)	6306 (44.6%)	
logTG			
Mean (SD)	4.70 (0.537)	4.74 (0.530)	<0.001
Median [Min, Max]	4.68 [2.08, 6.53]	4.74 [2.83, 6.52]	
Missing	23335 (47.3%)	6208 (43.9%)	
Total Cholesterol			
Mean (SD)	212 (44.0)	214 (44.7)	0.004
Median [Min, Max]	210 [38.2, 483]	210 [43.6, 464]	
Missing	23117 (46.9%)	6127 (43.3%)	
T2D status			
T2D cases	13018 (26.4%)	4370 (30.9%)	<0.001
Prediabetes	7295 (14.8%)	2465 (17.4%)	
T2D control	25603 (51.9%)	6350 (44.9%)	
Other controls	3422 (6.9%)	961 (6.8%)	
FBG			
Mean (SD)	5.39 (1.24)	5.56 (1.38)	<0.001
Median [Min, Max]	5.17 [0.160, 14.8]	5.28 [0.630, 14.7]	
Missing	16429 (33.3%)	4332 (30.6%)	
FBI			
Mean (SD)	9.86 (12.1)	12.0 (21.2)	<0.001
Median [Min, Max]	7.88 [0.288, 823]	9.36 [0.288, 1580]	
Missing	16656 (33.8%)	4342 (30.7%)	
HOMA-IR			
Mean (SD)	2.41 (2.18)	2.92 (2.54)	<0.001
Median [Min, Max]	1.80 [0.0169, 28.4]	2.20 [0.0498, 28.9]	
Missing	17230 (34.9%)	4544 (32.1%)	
HbA1c			
Mean (SD)	37.6 (8.82)	38.6 (9.92)	<0.001
Median [Min, Max]	36.6 [16.9, 121]	36.6 [14.7, 119]	
Missing	39322 (79.7%)	10975 (77.6%)	
Hypertension Status			
Normotensive	24380 (49.4%)	6679 (47.2%)	<0.001
Hypertenstive	24597 (49.9%)	7389 (52.2%)	
Missing	361 (0.7%)	78 (0.6%)	
SBP			
Mean (SD)	128 (20.8)	130 (21.2)	<0.001

Table 4.14. Distributions of variables by stratifying variables (BMI set) (6) Physical activity status

Cardiometabolic Profile	Physical Activity Non-sedentary	Sedentary	
Median [Min, Max]	126 [72.0, 247]	128 [61.0, 235]	
Missing	17762 (36.0%)	3848 (27.2%)	
DBP			
Mean (SD)	77.2 (11.7)	78.6 (12.0)	<0.001
Median [Min, Max]	77.0 [23.0, 135]	78.0 [37.0, 133]	
Missing	17767 (36.0%)	3849 (27.2%)	
MI status			
No	44856 (90.9%)	12686 (89.7%)	<0.001
Yes	4482 (9.1%)	1460 (10.3%)	
Stroke status			
No	44881 (91.0%)	12759 (90.2%)	0.005
Yes	4457 (9.0%)	1387 (9.8%)	
Physical status			
Non-sedentary	49338 (100%)	0 (0%)	<0.001
Sedantary	0 (0%)	14146 (100%)	
Missing			
Smoking status			
Never smokers	23726 (48.1%)	6832 (48.3%)	<0.001
Former smokers	18164 (36.8%)	4762 (33.7%)	
Current smokers	7055 (14.3%)	2395 (16.9%)	
Missing	393 (0.8%)	157 (1.1%)	

CHAPTER 5: MANUSCRIPT 2: GENETIC UNDERPINNINGS OF THE HETEROGENEOUS IMPACT OF OBESITY ON LIPID LEVELS AND CARDIOVASCULAR DISEASE

A. Overview

Obesity is thought to increase cardiovascular disease (CVD) risk through various CVD risk factors, including dyslipidemia. Yet, evidence suggests that obesity's effects on dyslipidemia are not uniform. One way to better understand the varied effects of obesity on dyslipidemia and downstream CVD is improved characterization of the underlying molecular mechanisms governing obesity and lipid traits, in particular, their shared genetic underpinnings. In this regard, we aimed to investigate the shared genetic underpinnings of obesity-related traits and dyslipidemia-related traits.

In this study, we examined three continuous proxies of dyslipidemia (high-density lipoprotein cholesterol (HDL), low-density lipoprotein cholesterol (LDL), and triglycerides (TG)) and one continuous proxy of obesity (BMI). To identify shared genetic underpinnings of BMI and lipid traits, we estimated local genetic correlations using genome-wide association study (GWAS) summary statistics of those traits from European ancestry UK Biobank (UKB) participants to identify shared genetic underpinnings of BMI and lipid traits. Based on the signs of local genetic correlation coefficients, we identified obesity genomic loci associated with lower risk of dyslipidemia (defined as "Ob/DysL(–)" loci; counter-intuitive to the phenotypic correlation between BMI and lipid traits) as compared to those associated with higher levels of dyslipidemia (defined as "Ob/DysL(+)" loci; as expected from the

phenotypic correlation). To identify causal genes at each locus, we integrated GWAS results with publicly available gene-expression data and performed gene-based association analyses. Lastly, we examined the potential differential effects of counter-intuitive Ob/DysL(–) loci on BMI, lipid levels, and subsequent CVD and its risk factors among diverse participants from the *Population Architecture using Genomics and Epidemiology (PAGE)* study using obesity polygenic risk scores (PRS; weights estimated using PRS-CS and UKB GWAS) constructed only with variants in loci that increase BMI and decrease dyslipidemia.

Out of 2,495 partitioned genomic regions, we identified 789 HDL, 26 LDL, and 494 TG significant local genetic correlations with BMI (including overlapping loci), of which three for HDL (0.4%), 10 for LDL (38.4%) and 8 for TG (1.6%) were Ob/DysL(–) loci. The gene-based analysis results prioritized plausible genes underlying the counter-intuitive local genetic correlations [Ob/DysL(–) loci], including a novel interesting candidate, *NEIL2*, predicted to be associated with muscular atrophy (characterized as reduced body weight with increased CVD risk) in mouse models. In PAGE, the PRS constructed using the BMI-HDL Ob/DysL(–) loci was associated with increased levels of obesity (and increased BMI) but decreased levels of dyslipidemia, CVD, and its risk factors, at least to some extent.

The identification and validation of genomic loci with shared genetic signals between obesity-related traits and dyslipidemia-related traits further support the importance of using genetics to define the heterogeneous impact of obesity on dyslipidemia and downstream CVD.

B. Introduction

Obesity has an enormous global public health burden[1,2] and increases the burden of many downstream sequelae, for example, cardiovascular diseases (CVD)[3], through its impact on CVD risk factors (e.g., dyslipidemia, hypertension, and type 2 diabetes).[201-204] However, there are significant gaps in research, and only a limited number of studies have explored the frequently referenced yet not well-understood variations in CVD risk within populations affected by obesity[31], which might be partly due to heterogeneous relationships between obesity and CVD risk factors, especially for obesity and lipid levels.[137,205] One plausible but largely unexplored source of heterogeneity in the obesity-lipid level relationship is the shared genetic architecture across obesity and lipid traits. Better characterization of the shared genetic underpinnings can help us better understand the heterogeneous impact of obesity on lipid traits and CVD.

Recent studies have suggested that pleiotropic obesity loci, especially those with counter-intuitive associations with CVD traits, could help explain the observed heterogeneous impact of obesity on CVD.[12-18] For example, two recent studies identified 36[16] and 62[12] obesity-increasing variants that also were associated with favorable or protective metabolic profiles, respectively. Several different variant-level approaches have been implemented to identify pleiotropic obesity variants[12]. However, no previous studies have used locus-level approaches and local genetic correlation analysis, an emerging genomic analysis tool to explore pleiotropy. In addition, previously identified pleiotropic loci have not been validated in populations with diverse ancestries. As with other genomic research, these loci were discovered in European ancestry populations, and it is unknown whether the identified bivariate loci show comparable influences on obesity and cardiometabolic traits in other ancestries.

In this regard, we aimed to identify genomic regions with significant shared genetic signals between continuous BMI and lipid traits (BMI-lipid bivariate loci) in opposing directions, to investigate the potential causal genes underlying counter-intuitive pleiotropy between BMI and lipid levels, and to examine the potential influence of BMI-lipid bivariate loci on BMI, lipid levels, and downstream CVD and its risk factors.

C. Methods

C.1 Study Population

C.1.1 UK Biobank (UKB)

The UKB is a large-scale prospective study of more than 500,000 individuals living in the United Kingdom. The stated goals of the UKB were to improve prevention, diagnosis, and treatment of various diseases onset later in life.[162] Participants aged 40 – 69 were recruited between 2006 – 2010,[162] and their phenotypic and genotypic information, including questionnaires, physical and blood measures, genome-wide genotyping data, imaging data, and health outcomes, has been collected.[162] The UKB is available to researchers for paid access.

C.1.2 Population Architecture using Genetics and Epidemiology: The PAGE study

The PAGE consortium was launched in 2008 along with NHGRI's effort to expand the ancestral diversity in genomic studies.[91,152] In this study, all participants with relevant genetic and phenotypic data from PAGE participating cohort studies were included. The PAGE cohort studies include the Atherosclerosis Risk in Communities (ARIC) study, Coronary Artery Risk Development in Young Adults Study (CARDIA), Hispanic Community Health Study / Study of Latinos (HCHS/SOL), Women's Health Initiative (WHI), Multiethnic Cohort Study (MEC), and Icahn School of Medicine at Mount Sinai BioMe biobank. Participants were self-classified as Asian American, Hispanic/Latino, Native Hawaiian, non-Hispanic White, and non-Hispanic

Black. A total of 83,376 participants with genetic data and BMI measure were included in the analysis. We used the PAGE study populations as the validation population for identified bivariate loci to assess generalizability. Brief descriptions of the PAGE-participating cohort studies were summarized in the ***Supplementary Information***.

C.2 Measurement

The present study utilized publicly available GWAS summary statistics from UKB as a discovery source for the genetic correlation between BMI and lipid traits. Individual-level genetic and phenotypic data were only utilized for the PAGE population.

C.2.1 Genetic Information

UKB. A total of 488,377 participants from the UKB were genotyped on the Applied Biosystems UKB Lung Exome Variant Evaluation (UK BiLEVE) Axiom Array (N=49,950) and the UKB Axiom Array (N=807,411).[206] Imputation was performed using IMPUTE4 with Haplotype Reference Consortium, UK10K, and the 1000 Genome Phase 3. Detailed methods were described in the previous literature.[206]

PAGE. In the original PAGE study, a total of 54,844 participants of diverse ancestry (African ancestry, Hispanic/Latino, East Asian, Native Hawaiian, and American Indian participants) who provided samples from different participating studies were genotyped on the MEGA array at the Center for Inherited Disease Research.[91,164] In addition to the MEGA genotyping platform, some participants from ARIC, BioMe, CARDIA, MEC, and WHI were genotyped separately on Illumina or Affymetrix arrays by each study or ancillary study. The number of samples included in this study by study, self-reported race/ethnicity, and the genotyping platform are shown in **Table 5.2**. A total of 34,373 samples that were included in the

current analysis were genotyped on MEGA, and the remaining 49,003 samples were genotyped on the non-MEGA array. **Table 5.3** summarizes the genotyping platform, QC criteria, imputation methods, and reference panel that each study and ancillary study implemented.

C.2.2 Phenotype Information

UKB. Anthropometric traits, including standing height and weight, were measured at the baseline visit for approximately 500,000 participants between 2006 and 2010.[162] BMI was calculated from measured height and weight (i.e., weight (kg) / [height (m)]^2). Participants' blood samples were also collected, and various biomarkers, including three lipid traits [high-density lipoprotein cholesterol (HDL), low-density lipoprotein cholesterol (LDL), and triglycerides (TG)] for the current study, were measured. Specifically, HDL levels were measured by enzyme immuno-inhibition method, LDL by enzymatic protective selection method, and TG by enzymatic colorimeter method (more information available at: https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/serum_biochemistry.pdf)

PAGE. BMI was used as a continuous proxy measure of overall body mass, and obesity status was determined by BMI > 30 kg/m^2 for non-Asians and BMI > 25 kg/m^2 for Asians. Three lipid measures (HDL, LDL, and TG) were used as continuous proxy measures of dyslipidemia. HDL and TG levels were quantified from fasting blood samples. LDL levels were computed using the Friedewald Equation, excluding individuals whose TG levels were > 400 mg/dL. Blood glucose and insulin levels were measured after an 8-hour fast. We determined participants' diabetes status according to the ADA criteria. Blood pressure was measured with a standardized protocol. Participants were classified as hypertensive if they met any of the following criteria using the following criteria: 1) SBP ≥ 140 mmHg, 2) DBP ≥ 90 mmHg, 3) use of any antihypertensive medication (self-report), or 4) ICD-9 codes 401. x or ICD-10 codes I10.x -

I15.x.[91] A subset of PAGE participating cohorts collected data on CVD, including their prevalence, incidence, or related deaths. Detailed descriptions of phenotypic measures were provided in the ***Supplementary Information***.

C.3 Statistical Analyses

C.3.1 Bivariate loci identification

Discovery GWAS. In 2018, the Pan-UKB team (https://pan.UKB.broadinstitute.org/docs/study-design) made results available from multi-ancestry GWAS of 7,221 phenotypes, including anthropometric and obesity-related measures. GWAS analysis was conducted using SAIGE[163] and a linear mixed model – with a kinship matrix considered as a random effect and covariates treated as fixed effects. Continuous traits were rank-based inverse normalized within each ancestry group, and covariates included in GWAS were age, sex, age*sex, age^2, age^2*sex, and the first 10 PCs (https://github.com/atgu/UKB_pan_ancestry/wiki/QC). In the current study, we utilized Pan-UKB GWAS summary statistics resulting from this project for BMI and lipid traits (HDL, LDL, and TG) for European ancestry as a discovery sample for the BMI-lipid bivariate loci identification (N: 419,163 for BMI, 367,021 for HDL, 400,223 for LDL, and 400,639 for TG).

Global SNP-based heritability and genetic correlations Prior to performing local genetic correlation analyses, we estimated global SNP-based heritability of BMI and three lipid traits and genetic correlation between three BMI-lipid pairs (BMI-HDL, BMI-LDL, and BMI-TG) in the UKB by performing LD score regression[207] based on GWAS summary statistics from UKB.

Bivariate loci identification We identified BMI-lipid bivariate loci (genomic loci with shared genetic signals between BMI and lipid levels) based on UKB GWAS summary statistics

for BMI and lipid traits (HDL, LDL, and TG) using local genetic correlation analyses (**Figure 1A**). As is standard in genetic epidemiological studies, we assessed obesity and dyslipidemia using continuous measures of these phenotypes. By looking at alleles that increase BMI, we are, in essence, identifying alleles that increase obesity risk. Also, by looking at alleles that increase LDL and TG or decrease HDL, we are identifying alleles that increase dyslipidemia risk. Three pairs of GWAS summary statistics from UKB (BMI-HDL, BMI-LDL, and BMI-TG) were used as input files, and local genetic correlation analyses were performed for the 2,495 pre-partitioned genomic regions (~1.12 Mb per locus on average; provided by the developers[208]) using the *LAVA* (Local Analysis of [co]Variant Association)[173]. Detailed descriptions of performing the analyses using *LAVA* are provided in the ***Supplementary Information.*** Significant local genetic correlation estimates [p for local genetic correlation coefficient estimates < (0.05 / the number of significant univariate loci for both traits)] were classified into two different groups based on their directions of effects with an obesity-related trait and dyslipidemia-related traits (**Table 5.4**). That is, if a given bivariate locus showed a positive local genetic correlation coefficient between obesity (using BMI as a continuous proxy of obesity) and dyslipidemia (using three lipid measures as proxies of dyslipidemia) (i.e., rg < 0 for BMI-HDL, rg > 0 for BMI-LDL and BMI-TG pairs), the locus was then classified as an Ob/DysL(+) locus whereas if the bivariate locus showed a negative local genetic correlation coefficient (i.e., rg > 0 for BMI-HDL, rg < 0 for BMI-LDL and BMI-TG), the locus was classified as an Ob/DysL(–) locus. We considered Ob/DysL(–) loci as counter-intuitive since the phenotypic correlations were in opposite directions (i.e., phenotypic correlation coefficient (r) < 0 for BMI-HDL, r > 0 for BMI-LDL and BMI-TG). Each BMI-lipid pair is tested separately, so there could be overlapping loci for

multiple BMI-lipid pairs, even with different signs (e.g., a locus can be Ob/DysL(+) for one BMI-lipid pair and Ob/DysL(–) for another BMI-lipid pair).

C.3.2 Biological interrogation for the bivariate loci

To investigate the biological implications of the identified BMI-lipid bivariate loci, we conducted TWAS-FUSION[174] and identified potential genes whose genetically predicted expression levels were associated with the BMI or lipid traits (**Figure 1B**). We integrated each GWAS summary result (BMI, HDL, LDL, and TG) with reference gene expression levels in Whole Blood samples from the Cardiovascular Risk in Young Finns Study (YFS)[175] and adipose tissue from Metabolic Syndrome in Men Study (METSIM).[176] Then, we filtered the genes located within the bivariate loci (based on the start and the end position of the genes) and identified the overlapping genes from the BMI and corresponding lipid trait. We also examined directional consistency by comparing TWAS Z scores for BMI and the corresponding lipid trait. For example, we verified if a gene within BMI-HDL Ob/DysL(–) loci had the same direction of TWAS Z-score for both BMI and HDL. Based on the known roles of the overlapped genes (reported in public databases (e.g.,) PubMed or Genecards), we inferred potential pathways simultaneously influencing BMI and lipid traits.

C.3.3 Potential Influence of the BMI-lipid bivariate loci among ancestrally diverse populations

We examined the generalizability of the identified loci by investigating the association of the Ob/DysL(–) and Ob/DysL(+) bivariate loci in ancestrally diverse PAGE participants (**Figure 1C**). We hypothesized that BMI-lipid Ob/DysL(–) loci and Ob/DysL(+) loci were involved in distinct biological pathways, linking adiposity with protective roles and detrimental roles in lipid metabolism, respectively, and that obesity polygenic risk scores (PRS) constructed with variants restricted to the identified bivariate loci would capture the genetic predisposition to the distinct

subtypes of adiposity. Based on these assumptions, we constructed Ob/DysL(–)-based and Ob/DysL(+)-based obesity PRS to capture the genetic predisposition to Ob/DysL(–) adiposity and Ob/DysL(+) adiposity. We utilized publicly available PRS weights for BMI prepared and provided by ExPRSweb[177] (https://exprsweb.sph.umich.edu/). The PRS weights for BMI were estimated using PRS-CS method ($N_{variants}$ = 1,113,832; Pearson correlation between PRS and BMI in testing sample = 0.321[177]) based on UKB GWAS summary statistics for BMI. We restricted genetic variants to those located in the Ob/DysL(–) bivariate loci and Ob/DysL(+) bivariate loci for the PRS-Ob/DysL(–) and PRS-Ob/DysL(+), respectively, and applied the weights to our target population, the PAGE study. The remaining variants that were located outside of the bivariate loci were not included in the PRS calculation.

Then, the associations of the bivariate loci-based obesity PRS with BMI (and obesity status), lipid traits (HDL, LDL, TG, total cholesterol, and dyslipidemia), CVD risk factors (fasting glucose, fasting insulin, HOMA-IR, HbA1c, T2D status, systolic blood pressure, diastolic blood pressures, and hypertension), and CVD outcomes (MI and stroke) were assessed. We hypothesized that higher PRS-Ob/DysL(–) would be associated with increased BMI or obesity but, counter-intuitively, with protective cardiometabolic profiles. Conversely, higher PRS-Ob/DysL(+) would be associated with increased BMI and increased probability of having dyslipidemia (decreased HDL and increased LDL and TG levels). We applied logistic regression models for binary outcomes and linear regression models for continuous outcomes. Covariates were age, sex, ten genetic PCs (for ancestry), study, genotype panel, and self-reported race/ethnicity as a social construct associated with the social determinant of health, racism, discrimination, and environmental factors.

D. Results

D.1. Global SNP-based heritability and genetic correlation

In line with our previous understanding, global SNP-based heritability estimates (SE) were 0.25 (0.01), 0.20 (0.02), 0.09 (0.01), and 0.18 (0.02) for BMI, HDL, LDL, and TG, respectively. Likewise, global genetic correlation coefficient estimates (SE) were -0.43 (0.03), -0.07 (0.03), and 0.31 (0.04) for BMI-HDL, BMI-LDL, and BMI-TG pair, respectively, as expected (**Table 5.6**).

D.2. BMI-lipid bivariate loci identification in UKB

Out of 2,495 genomic regions, 2,268 (BMI-HDL), 1,018 (BMI-LDL), and 2,017 (BMI-TG) loci demonstrated significant local heritability ($p < 2.00 \times 10^{-5}$) for both BMI and the respective lipid trait and were further tested for the local genetic correlation (**Table 5.7**). As such, we identified 789 HDL, 26 LDL, and 494 TG loci with significant local genetic correlation with BMI. The median and inter-quartile range of local genetic correlation coefficients among the tested regions were -0.44 (-0.56, -0.31) for BMI-HDL, 0.02 (-0.12, 0.16) for BMI-LDL, and 0.39 (0.25, 0.52) for BMI-TG. Of these, three for HDL (0.4%), 10 for LDL (38.4%), and 8 for TG (1.6%) were [Ob/DysL(–)] loci (**Table 5.1**). Also, as expected from the strong correlation between HDL and TG, many of the BMI-HDL bivariate loci overlapped the BMI-HDL and BMI-TG loci– i.e., a total of 400 Ob/DysL(+) loci and 2 Ob/DysL(–) were identified for both BMI-HDL pair and BMI-TG pair.

A total of four Ob/DysL(–) loci were identified across multiple BMI-lipid pairs. Loc1351 (Chr8:125,453,323-126,766,827) was identified as an Ob/DysL(–) locus for all three BMI-lipid pairs. Loc2351 (Chr19:45,040,933-45,893,307) was identified for BMI-LDL and BMI-HDL,

Loc 965 (Chr6: 32,586,785-32,629,239) for BMI-LDL and BMI-TG, and Loc 1851 (Chr12: 123,396,635-124,843,768) for BMI-HDL and BMI-TG. Of these four loci, three loci included previously reported Ob/DysL(–)-related variants, rs2980888[12] and rs7005992[14] in Loc1351, rs7133378[12,16,144], rs7973683[14], and rs863750[12] in Loc1851, and rs2075650[12] in Loc2351 (**Table 5.10**).

We compared the Ob/DysL(–) results with findings from five previous studies of counter-intuitive BMI-CVD risk factor pleiotropy.[12,14,16,144,145] All of these five studies were variant-based approaches (e.g., multivariate adiposity and cardiovascular traits GWAS). A total of 149 distinct variants have been identified as obesity variants associated with protective cardiometabolic profile, and they were located within 104 loci (out of the 2,495 genomic regions used for our local genetic correlation analyses). Although our analyses were locus-based and it is difficult to compare loci and variants directly, we identified 11 novel Ob/DysL(–) loci (7 from BMI-LDL results and 5 from BMI-TG loci; 1 overlapping locus) (**Table 5.10**). In addition, 3 BMI-HDL Ob/DysL(–) loci, 3 of 10 BMI-LDL Ob/DysL(–) loci, and 3 of 8 BMI-TG Ob/DysL(–) included at least one of the previously identified counter-intuitive variants. Differences across studies may be due to different discovery populations (though some of the studies utilized UKB) and different identification strategies and methods.

D.3 Identification of the genes within the BMI-lipid bivariate loci influencing both BMI and lipid traits

For BMI-HDL, we identified 3 Ob/DysL(–) genes (loc1851-*CCDC92, DDX55, DNAH10*). For BMI-LDL, we identified 3 Ob/DysL(–) genes (loc837-*ANKDD1B, POC5, POLK,* and loc970-*C6orf106*). For BMI-TG, we identified 5 Ob/DysL(–) genes (loc1247-*ERI1*, loc1251-*NEIL2*, and loc1851-*CCDC92, DDX55, DNAH10*) (**Table 5.8-5.9**).

D.4 Evaluating the associations of the identified BMI-lipid bivariate loci with BMI, lipid, and CVD and its risk factors among PAGE study participants

A total of 83,376 PAGE participants across four different self-identified race/ethnicity groups [Non-Hispanic White (EUR; N = 25,418), non-Hispanic Balck (AFR; N =25,255), Hispanic/Latino (HIS; N = 25,814), and East Asian (EAS; N = 6,889)] were included in the current analysis (**Table 5.5**). The participant mean age was 55.0 (SD: 11.5) years, and the proportion of male participants was 31.2% (N=26,017). Mean BMI, HDL, LDL, and TG levels were 28.6 (SD: 6.11) kg/m^2, 51.0 (SD: 15.9) mg/dL, 136 (SD: 40.7) mg/dL, and 132 (SD: 80.2) mg/dL, respectively.

From the BMI-HDL loci-based PRS, as expected based on the discovery, we observed clear differences in the direction of associations between PRS-Ob/DysL(+) (i.e., associated with adverse CVD risk profile) and PRS-Ob/DysL(–) (i.e., associated with protective CVD risk profile) for dyslipidemia, lipid levels (HDL, LDL, logTG, and total cholesterol), and glycemic traits (fasting glucose) in independent PAGE populations. Positive associations with the obesity traits (obesity status and BMI) were observed for both PRS-Ob/DysL(+) and PRS-Ob/DysL(–) despite the small number of loci included in the BMI-HDL PRS-Ob/DysL(–). Due to the substantial overlap with PRS-BMI (reference), the PRS-Ob/DysL(+) demonstrated similar patterns of associations to the reference PRS-BMI (**Figure 5.2A** and **Table 5.11-5.14**).

From the BMI-LDL loci-based PRS, we observed no clear distinction between PRS-Ob/DysL(–) and PRS-Ob/DysL(+) in the associations with outcome traits (**Figure 5.2B** and **Table 5.11-5.14**).

From the BMI-TG loci-based PRS, we observed positive associations with obesity-related traits for both PRS. However, we did not find evidence of the protective associations

between PRS-Ob/DysL(–) and lipid and other CVD-related traits. PRS-Ob/DysL(+) demonstrated a similar pattern of associations with the reference PRS-BMI (**Figure 5.2C** and **Table 5.11-5.14**).

E. Discussion

In this study, using large-scale GWAS summary statistics derived from the UKB, we identified 16 genomic regions with shared genetic underpinnings between BMI and lipid levels, which increased obesity risk yet were protective from dyslipidemia. We further explored the potential causal genes underlying the counter-intuitive Ob/DysL(–) loci using gene-based TWAS results and identified 8 genes. Using the BMI-lipid bivariate loci-based PRS, we were able to generalize our findings to the multi-ancestry PAGE study populations and explore the clinical significance of these bivariate loci on downstream CVD and its risk factors.

The smaller global genetic correlations between BMI and LDL in comparison to BMI-HDL and BMI-TG have been consistently reported in the literature.[209,210] By performing local-level genetic correlation analysis for BMI and LDL, we intended to investigate whether there is a true lack of genetic correlation (both locally and globally) between BMI and LDL or whether the lack of global genetic correlation is due to the presence of the comparable numbers of local-level correlations in opposite directions, resulting in nullifying each other's effects globally. The current study supported both possibilities – i) a much smaller number of correlated loci was identified, implying a lack of genetic correlation compared to the BMI-HDL or BMI-TG pairs, and ii) the comparable numbers of Ob/DysL(+) and Ob/DysL(–). Indeed, many more Ob/DysL(+) loci were discovered compared to Ob/DysL(–) loci for BMI-HDL and BMI-TG, as expected from the high phenotypic positive correlation between obesity and dyslipidemia[211]. It is

also true that, unlike BMI-HDL or BMI-TG results, a similar number of Ob/DysL(+) loci and Ob/DysL(–) were identified, and they might have nullified each others' effects, resulting in a small magnitude of global genetic correlation between BMI and LDL. These differences in BMI-lipid pairs (BMI-TG, BMI-HDL vs. BMI-LDL) may suggest the presence of distinct adiposity-lipid inter-relationships for HDL and TG vs. LDL.[212,213] Lastly, it is also possible that insufficient adjustment for lipid-lowering medications (i.e., residual confounding by lipid-lowering medication) could have contributed to our results.

By integrating TWAS results with the current local genetic correlation analysis, we prioritized potential causal genes, both novel and known genes, underlying the counter-intuitive genetic correlations. As an example of the novel genes, we identified the *NEIL2* gene for BMI-TG Ob/DysL(–) Loc1251. *NEIL2* is Nei-like DNA Glycosylase 2 and has been predicted to be involved in Autosomal Dominant Adult-Onset Proximal Spinal Muscular Atrophy by mice models[214], which is relevant for both reduced body weight and an adverse CVD risk profile. Moreover, *NEIL2* has been associated with reduced expression of adipose tissue TG lipase, causing TG accumulation in immobilized muscles by atrophy compared to that in control muscles.[215] According to the GWAS catalog, the *NEIL2* gene has been associated with both TG levels and BMI-adjusted WHR, along with other CVD traits, further supporting the *NEIL2* gene as a potential causal gene influencing decreased obesity yet increased CVD.

In some instances, we observed both Ob/DysL(–) and Ob/DysL(+) effects in the same loci, when considering different lipid traits. For example, we identified three novel genes (*POLK*, *ANKDD1B*, and *POC5*) for Loc837. However, unlike the BMI-LDL results, Loc837 was an Ob/DysL(+) locus for the BMI-HDL pair, and previous studies identified rs2112347 in this locus as Ob/DysL(+) SNP. Nevertheless, the follow-up partial local genetic correlation analyses

in the current study suggested that the strongest partial correlation was found from BMI-LDL in the Ob/DysL(–) direction. Such conflicting evidence regarding the direction of effects may be driven by allelic heterogeneity and/or population-specific variants. Indeed, rs2112347 in Loc837 was associated with both BMI and LDL in opposing directions[216], and SNPs in the *HMGCR-POC5* region were associated with LDL cholesterol and T2D in opposing directions.[217] Therefore, *POLK, ANKDD1B,* or *POC5* may harbor Ob/DysL(-) variants influencing BMI and LDL rather than Ob/DysL(+) variants influencing BMI-TG or BMI-HDL.

Several of our findings are consistent with previous studies. For example, in Loc1851, among three genes – *DDX55*, *DNAH10*, and *CCDC92* – associated with BMI, TG, and HDL in the current TWAS, *DNAH10,* and *CCDC92* were previously prioritized genes for the Ob/DysL(–) variants[12]. We also detected the *ERI1* gene related to the BMI-TG Ob/DysL(–) Loc1247. Although a different gene (*PPP1R3B-TNKS*) was prioritized, two previous SNPs (rs9987289 and rs17149279) were reported from the adjacent locus, Loc1247.[12]

The current study provides evidence of a generalizable heterogeneous genomic relationship between obesity and dyslipidemia, especially for the BMI-HDL loci. Despite the small number of loci (three loci) included in the PRS-Ob/DysL(–) for BMI-HDL, its effects on dyslipidemia were larger than that of the overall PRS-BMI or PRS-Ob/DysL(+). In terms of continuous CVD risk factors, the higher PRS-Ob/DysL(–) for BMI-HDL was associated with a protective cardiometabolic profile (except for HbA1c) and increased BMI. These results suggest that shared genetic underpinnings between obesity traits and lipid traits may partly explain the heterogeneous impact of BMI on CVD risk. We did not find clear evidence when applying the PRS from the counterintuitive BMI-LDL or BMI-TG pairs, possibly due to limited power.

Further studies with larger sample sizes (in discovery GWAS and in target populations) are needed to discover and generalize more of the heterogeneous BMI-lipid genomic loci.

The present study has some limitations. First, the genomic partitioning was based on a European LD structure (1000 Genome European population), so the partitioned genomic regions may not represent the independent LD blocks for non-European populations well. Thus, we may have missed ancestry-specific genetic correlations. Related to this, as BMI is a crude proxy for obesity, we may have missed important pleiotropic loci between adiposity and lipid traits. However, the advantages of leveraging much larger sample sizes may have counterbalanced this limitation. The current study has notable strengths as well. First, the total sample size of the PAGE study was large, and we were able to evaluate the relationships between BMI-lipid bivariate loci and various CVD profiles. In addition, the distribution of self-identified race/ethnicity in the PAGE study – especially across self-identified non-Hispanic White, non-Hispanic Black, and Hispanic/Latino populations– was well-balanced, thus equally contributing to the population-pooled results. Furthermore, this study implemented a novel locus-based approach to identify BMI-lipid bivariate loci and proposed a novel application of locus-restricted PRS to evaluate the influence of certain genomic loci on phenotypes.

In summary, we identified two distinct types of genomic loci with shared genetic underpinnings of BMI and lipid levels in opposing directions [Ob/DysL(–) and Ob/DysL(+)] and suggested potential causal genes (*NEIL2, POLK, ANKDD1B,* and *POC5)* underlying counter-intuitive Ob/DysL(–) loci. Notably, from the association test using PRS-Ob/DysL(–), the BMI-HDL Ob/DysL(–) loci demonstrated protective associations with dyslipidemia and downstream CVD risk profiles in an independent population. Indeed, as even larger GWAS of various CVD traits become available, this approach may be expanded to other CVD complications of obesity –

e.g., Obesity-T2D bivariate loci or Obesity-Hypertension bivariate loci, enabling the identification of possible subtypes of obesity.

F. Main Tables and Figures

Table 5.1. A list of genomic loci with shared genetic signals between BMI and a lipid trait in a counter-intuitive direction (opposite to the phenotypic correlation)

Locus	CHR	Start Position	Stop Position	Discovery pair	Prioritized genes†	$rho_{BMI-HDL}$	BMI_{-HDL}	$rho_{BMI-LDL}$	BMI_{-LDL}	rho_{BMI-TG}	BMI_{-TG}
158	1	205009624	205917548	BMI-LDL		-0.57	2.01E-09	-0.57	2.69E-05	0.32	8.46E-04
498	3	87411259	88375763	BMI-LDL		-0.61	6.34E-06	-0.67	3.12E-06	0.48	2.72E-03
692	4	102544804	104384534	BMI-LDL		-0.75	9.16E-37	-0.47	3.85E-05	0.68	3.08E-12
836	5	73314062	74245354	BMI-LDL		-0.52	5.78E-06	-0.56	9.74E-09	0.46	4.58E-04
837	5	74245355	75239302	BMI-LDL	*POLK*, *ANKDD1B*, *POC5*	-0.72	1.59E-10	-0.69	5.84E-27	0.11	4.26E-01
965*	6	32586785	32629239	BMI-LDL, BMI-TG		-0.33	8.97E-03	-0.70	5.83E-08	-0.54	1.80E-06
1185	7	98173565	99465540	BMI-LDL		-0.46	5.68E-06	-0.38	8.06E-06	0.35	1.62E-03
1246*	8	8064601	8589770	BMI-TG		-0.39	7.23E-04	0.15	2.12E-01	-0.48	8.74E-09
1247*	8	8589771	9167795	BMI-TG	*ERI1*	0.04	4.60E-01	0.26	1.82E-03	-0.35	2.28E-06
1248	8	9167796	9835863	BMI-TG		0.19	1.01E-03	0.36	2.45E-06	-0.62	1.95E-11
1249	8	9835864	10478851	BMI-TG		-0.02	7.06E-01	0.51	3.99E-05	-0.41	2.33E-08
1251*	8	11466762	12296849	BMI-TG	*NEIL2*	-0.10	2.55E-01	N/A	N/A	-0.36	2.27E-06
1351*	8	125453323	126766827	BMI-HDL, BMI-LDL, BMI-TG		0.33	1.33E-05	-0.54	1.51E-14	-0.50	7.14E-14
1851*	12	123396635	124843768	BMI-HDL, BMI-TG	*DDX55*, *DNAH10*, *CCDC92*	0.37	1.59E-08	-0.43	8.63E-05	-0.53	5.92E-12
2135	16	53393883	54866095	BMI-LDL		-0.56	4.69E-30	-0.73	5.36E-21	0.18	9.98E-03
2351	19	45040933	45893307	BMI-HDL, BMI-LDL		0.34	9.97E-13	-0.46	1.21E-27	0.21	2.23E-06

* No discordance across different BMI-lipid pairs (i.e., all three BMI-lipid pairs demonstrated Ob/DysL(–) direction or non-significant correlation)

† Genes identified from TWAS-FUSION analysis.

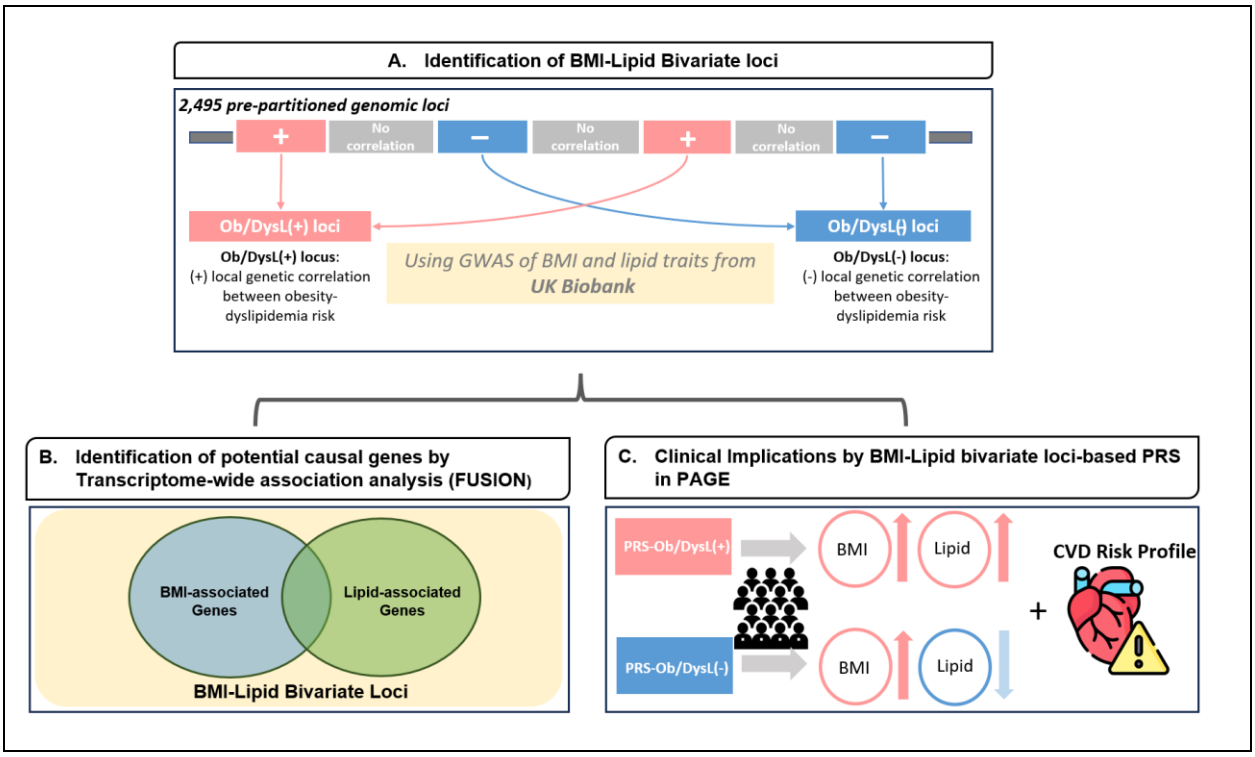


Figure 5.1. Summary of Statistical Analyses. In this study, we first identified pleiotropic genomic loci between BMI and lipid levels in opposing directions using local genetic correlation analyses (implemented through LAVA) [A]. Using TWAS-FUSION, potential causal genes for the pleiotropic loci [B]. Lastly, clinical implications of the distinct pleiotropic loci were assessed by using polygenic risk scores [C].

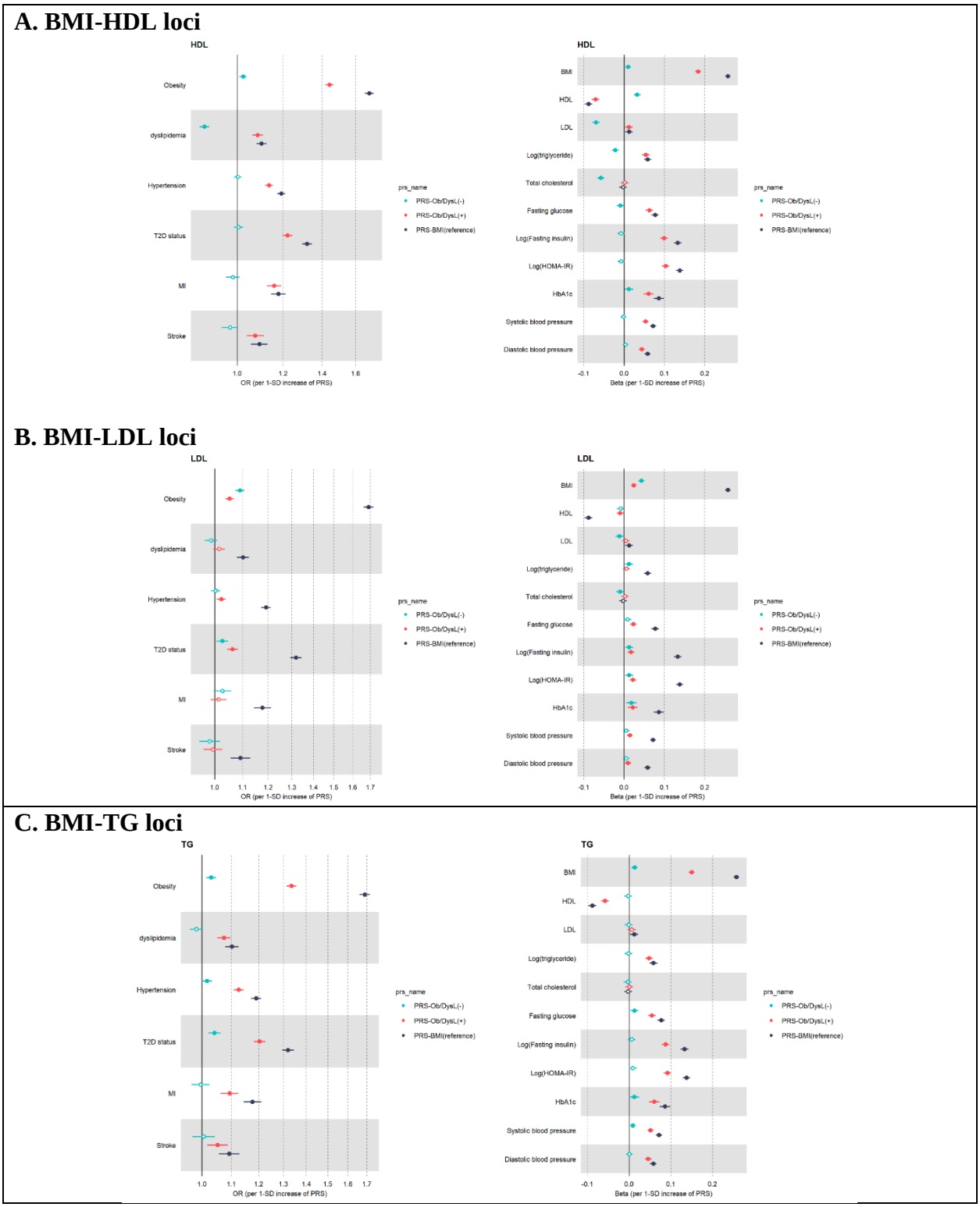


Figure 5.2. The associations of BMI-lipid PRS-Ob/DysL(–), PRS-Ob/DysL(+), or PRS-BMI with obesity-, lipid-, and CVD-related traits. The results showed the estimated associations of (95% CI) of PRS-Ob/DysL(–), PRS-Ob/DysL(+), and PRS-BMI (reference) with obesity-related traits, lipid-related traits, and other CVD-related factors in PAGE participants. Covariates were age, sex, study, genotype panel, self-reported race/ethnicity, and ten genetic PCs. PRS and outcome variables were standardized with a mean of 0 and a standard deviation of 1. Filled-in circles represent P<0.05, while empty circles represent P>0.05.

G. Supplement

G.1 Supplemental Methods

G.1.1 PAGE-participating cohort studies

ARIC, funded by the National Heart, Lung, and Blood Institute (NHLBI), is an ongoing community-based prospective cohort study primarily aiming to investigate the etiology of atherosclerosis and its clinical outcomes.[153] A random sample of 15,792 adults aged 45 – 64 years at baseline was initially recruited between 1987 and 1989 (approximately 4,000 participants for each of four communities in the U.S. – Forsyth County, NC; Jackson, MS; Washington County, MD; Minneapolis, MN).[153] Participants have received standardized examinations on their demographic, social, and health status approximately every five years.

BioMe, funded by the Charles Bronfman Institute for Personalized Medicine, is an electronic medical record-linked biobank whose participants were based on consented and volunteered patients in the Mount Sinai Medical Center (MSMC) (among over 70,000 inpatients and 800,000 outpatients annually).[154] The MSMC serves racially/ethnically diverse communities of the upper Manhattan area, which includes Central Harlem (predominantly African American), East Harlem (predominantly Hispanic/Latino), and Upper East Side (predominantly European American). There have been more than 57,843 participants (21% African American, 34% Hispanic/Latino, 31% European American, and 14% of other ancestry groups) enrolled in BioMe since 2007 (as of Feb 2021). Among them, a total of 32,344 participants have been genotyped (as of Feb 2021) so that they can be investigated in genomic studies (https://icahn.mssm.edu/research/ipm/programs/biome-biobank/facts).

CARDIA, funded by NHLBI, is a community-based prospective cohort study aiming to investigate the influencing factors for the development of coronary heart disease and its risk

factors during young adulthood.[155] Initial recruitment was done in 1985 – 1986, and a total of 5,116 African American (52%) and European American (48%), aged 18 – 30 years, participated from four urban communities – 1,179 from Birmingham, AL; 1,109 from Chicago, IL; 1,402 from Minneapolis, MN; and 1,426 from Oakland, CA.[155] In the recruiting step, participants were selected for the cohort to be balanced in age (> or ≤ 24 years), educational level (> or ≤ 12 years), sex, and race/ethnicity.[155] After the initial examination, participants were asked to respond to the follow-up examinations during 1987 – 1988 (Year 2), 1990 – 1991 (Year 5), 1992 – 1993 (Year 7), 1995 – 1996 (Year 10), 2000 – 2001 (Year 15), 2005 – 2006 (Year 20), 2010 – 2011 (Year 25), and 2015 – 2016 (Year 35) (and currently Year 40 exam is ongoing as of Dec 2022). Data collection included the potential influencing factors for coronary heart disease – e.g., blood pressure, glucose levels, blood cholesterol levels, anthropometric traits, lifestyle factors, and family history.

HCHS/SOL, funded by NHLBI and other institutes, is a community-based prospective cohort study of Hispanic/Latino populations in the U.S. aiming to determine the role of acculturation in the prevalence and incidence of diseases and to identify influencing factors for the health of Hispanic/Latino populations. A total of more than 16,000 participants who were self-identified as Hispanic/Latinos and aged 18 – 74 years were recruited between 2008 and 2011 from four study sites – Bronx, NY; Chicago, IL; Miami, FL; and San Diego, CA. The study was designed to enroll 4,000 participants (2,500 aged 45 – 74 years and 1,500 aged 18 – 44 years) in each study site and to have at least 2,000 participants in each of the four groups of origin – Cuban, Puerto Rican, Mexican, or Central/South American.[156] The participants received extensive baseline examinations on psych-social and clinical factors during 2008 – 2011. A

follow-up assessment for the cohort was done during 2015 – 2017, the third exam is in progress now and annual follow-up interviews via phone calls are ongoing.

MEC, funded by the National Cancer Institute, is a prospective cohort study to investigate lifestyle and genetic risk factors for cancer in the U.S. [157] A total of 215,251 adults aged 45 – 75 years at baseline were recruited between 1993 and 1996 from Hawaii and L.A. County, CA.[157] Ethnic distributions of the participants were 16.3% of African American, 22.0% of Hispanic/Latino, 26.4% of Japanese American, 6.5% of Native Hawaiian, 22.9% of European American, and 5.8% of other ethnic groups.[157] During 2001 – 2006, a prospective biospecimen collection (i.e., biospecimen collected before the onset of disease; blood, urine, mouthwash, saliva, or viable lymphocytes) was done for a subset of participants (75,928 as of April 2019) (https://www.uhcancercenter.org/for-researchers/mec-cohort-composition). In this study, eight ancillary studies were included – the Slim Initiative in Genomic Medicine for the Americas (MEC-Sigma) (a type 2 diabetes study in Hispanic/Latino adults); MEC-AAPC, MEC-JAPC, and MEC-LAPC (studies of prostate cancer in African American, Japanese American, and Hispanic/Latino men, respectively); MEC-AABC, MEC-JABC, MEC-LABC, and MEC-HIBC (studies of breast cancer in African American, Japanese American, Hispanic/Latino women, and Native Hawaiian women, respectively).

WHI, funded by NHLBI, is a prospective cohort study to investigate the health of postmenopausal women in the U.S., especially for preventing CVD, breast cancer, colon cancer, and osteoporotic fractures in women aged 50 – 79 years.[158] A total of 161,808 participants were recruited between 1993 and 1998 at 40 clinical centers across the U.S. There are two different parts in WHI – one is the WHI Clinical Trial (~64,500), a randomized clinical trial of hormone therapy, dietary intervention, and calcium/vitamin D supplements, and the other is WHI

Observational Study (~100,000), investigating incidence, risk factors, and potential interventions for CVD, cancer, and osteoporotic fractures.[158] Followings are ancillary studies that were included in our analyses – the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO); the Modification of PM-Mediate Arrhythmogenesis in Population stud(MOPMAP); the Genomics and Randomized Trials Networks (GARNET); the Hip Fracture GWAS (HIPFX); the Long Life Study (LLS); the Women's Health Initiative Memory Study (WHIMS); and the Women's Health Initiative-SNP Health Association Resource (WHI-SHARe).

G.1.2 Phenotype measurement

BMI. We used BMI as a continuous proxy measure of obesity risk. BMI was derived from weight and height measured at the baseline visit (at the time of enrollment) for ARIC, BioMe Biobank, CARDIA, HCHS/SOL, and WHI. For 140 WHI participants who were missing in height or weight at baseline, height and/or weight measures at 1-year or 3-year follow-up substituted the missing baseline measures.[165] In MEC, height and weight measures were self-reported, and these self-reported baseline height and weight measures were used to generate BMI at baseline.

Lipid traits. We used three lipid measures as continuous proxies of dyslipidemia risk**.** HDL-C and TG levels were measured from fasting blood, and the Friedewald Equation was used to calculate LDL-C levels from other lipid measures. If measured TG levels were greater than 400mg/dL, LDL-C levels were not calculated. In addition, following previous studies, medication status was adjusted by adding a constant (**Table S13**).[167,168] The largest constant was applied if more than one medication was reported. Those who had not fasted for 8 hours or were pregnant at measure were excluded from the harmonized phenotype database. Natural-log transformation was applied to TG levels after adjusting for medication.

Glycemic traits. Fasting blood glucose levels and insulin levels were measured at baseline visits using standard assays after 8 hours of fasting. HbA1c levels were measured during follow-up visits for all cohort studies except for HCHS/SOL. Participants without diabetes (normoglycemia) were defined as having fasting glucose < 5.6 mmol/L or HbA1c < 38 mmol/mol and aged over 40. We excluded those under 40 years old were glucose < 5.6 mmol/L or HbA1c < 38 mmol/L from the analysis. Participants with pre-diabetes were defined as having glucose ≥ 5.6 mmol/L or HbA1c ≥ 38 mmol/mol. Lastly, participants with diabetes were defined based on ADA criteria (by medication, report diagnosis, fasting glucose ≥ 7 mmol/L or HbA1c ≥ 48 mmol), or random glucose > 11.11 mmol/L, and aged ≥ 21 years at the time of diagnosis (to avoid potential misclassification between T1D and T2D).

Blood pressure was measured using a standardized protocol. Participants were considered hypertensive when they met at least one of the following criteria: 1) SBP ≥ 140 mmHg, 2) DBP ≥ 90 mmHg, 3) any antihypertensive medication reported, or 4) ICD-9 codes 401. x or ICD-10 codes I10.x - I15.x. [91]

Cardiovascular diseases. Some of the PAGE participating cohorts have information (prevalence, incidence, or death) on cardiovascular diseases. ARIC, MEC, and WHI ascertained the prevalence or incidence of myocardial infarction (MI) and stroke. Detailed descriptions of CVD ascertainment by studies were reported in the **following section**.

G.1.3 CVD Ascertainment by PAGE-participating studies

In ARIC, information on the CHD events including hospitalization and deaths were collected through annual follow-up interviews and community surveillance.[169] Definitions of

CHD events included acute hospitalized MI, definite fatal CHD, MI diagnosed by ECG, and revascularization.[169]

In MEC, As described in previous studies[170], CHD cases and controls from several nested case-control substudies in MEC were used in the current analysis. CHD cases were ascertained through the participants' medical record from the California Hospital Discharge Data (1990 - 2012) and the Centers for Medicare and Medicaid Services claim files (outpatients) (1999 - 2011), which were linked to MEC study - c.f., some participants from Hawaii (76.6% of Japanese American) were not available for hospital discharge data. Case definitions for CHD were ICD-9 codes (DX 410 - 414) for ischemic heart disease as the principal or first diagnosis code and the principal or first procedure code. Also, if a primary cause of death is MI (ICD-9 DX410, ICD-10 I21) or other CHD (ICD-9 DX411-414, ICD-10 I20, I22-25), these individuals were included as cases. Both prevalent (~20%; ascertained at baseline) and incident (~80%; ascertained during follow-up) CHD cases were ascertained.[169] Controls were selected among those without history of heart attack or angina from the questionnaire at baseline or all follow-up questions.

In WHI, CHD events were identified through self-reported questionnaire and adjudicated by physicians after reviewing the chart within 3 months.[171] CHD cases were defined as individuals who had a history of MI (self-reported) or a revascularization procedure at baseline, and/or manifested a definitive MI, went through a revascularization procedure, or died from CHD during follow-up. [171]

G.1.4 Performing local genetic correlation analyses using LAVA [173]

Ob/DysL(+) and Ob/DysL(–) loci were identified by local genetic correlation analysis using a pair of UKB GWAS summary statistics for obesity (BMI) and lipid traits (HDL, LDL, and TG). Local genetic correlation analyses were conducted using the *LAVA* R package. A total of 3 obesity-lipid trait pairs were analyzed.

Here is a brief description for the local genetic correlation approach implemented to this study. LAVA, like other local genetic correlation estimation tools, was developed to estimate the locus-level genetic correlation between two phenotypes. LAVA first estimated the local genetic signal (measured by local heritability (h^2)) as follows. [173]

$$Y_p = X\alpha_p + \epsilon_p$$

Y_p : Standardized phenotype vector

X: genotype matrix with K_{snp} $SNPs$ (standardized)

α_p: vector of joint SNP effects (accounting for LD)

ϵ_p : vector of normally distributed residuals with variance η_p^2

$\hat{\alpha}_p = (X^T X)^{-1} X^T \boldsymbol{Y_p}$, if the local SNP LD matrix is denoted as $S = cor(X)$ and the vector of estimated marginal SNP effects are denoted as $\hat{\beta}_p$ (not accounting for LD), $\hat{\alpha}_p = S^{-1}\hat{\beta}_p$. That is, if marginal SNP effects are obtained from GWAS summary statistics we can estimate the joint SNP effects ($\hat{\alpha}_p$) using a reference population's LD structure. Using the estimated joint SNP effects, local residual phenotypic variance (η_p^2) and the proportion of phenotypic variance explained by the SNPs wihtin the locus (local h^2) can be estimated. Then, it estimates bivariate local genetic correlations. The local genetic effects (G) can be defined as $G = X\alpha$ (α is a K

(number of SNPs in the locus) by P (number of phenotypes) matrix of joint SNP effects). The realized covariance matrix of G is denoted as follows (Ω).

$$\Omega = \begin{pmatrix} \omega_p^2 & \omega_{qp} \\ \omega_{pq} & \omega_p^2 \end{pmatrix}$$

ω_p^2 : local genetic variance of G_p for phenotype p

ω_{pq} : local genetic covariance of G_p and G_q for phenotype p and q

Then, the local r_g can be calculated by $\rho_{pq} = \frac{\omega_{pq}}{\sqrt{\omega_p^2 \omega_q^2}}$, and ρ_{pq}^2 is considered as the proportion of variance in the local genetic effects G_p explained by G_q. Since G is not actually observed, Ω should be estimated using the Method of Moments, not computed directly. Significance of the correlation was determined using simulation-based p-values. This local genetic correlation analysis would be especially useful for the situations where some signals appear in opposing directions at different regions and nullify each other in a global level – i.e., the absence of global genetic correlation despite the presence of local genetic correlation in opposing directions, whereas global genetic correlation captures only the average genetic correlation across the whole genome and sometimes cannot differentiate the null genetic correlation.[173]

LAVA utilizes pre-partitioned genomic regions to get a local genetic correlation estimate for each locus. We used 2,495 pre-partitioned genome that has been provided by the developers of LAVA (https://github.com/cadeleeuw/lava-partitioning). These partitioned genomic blocks were generated based on the 1000 Genome European reference population on build hg19/GRCh37 to get approximately LD-independent genomic blocks across the whole genome.

As described earlier, LAVA first performed the univariate test to filter in the loci where a significant local genetic influence (measured by local heritability (h2)) on adiposity or lipid traits was estimated. It excluded the loci without any significant local heritability for either of the two traits from the following bivariate analysis (correlation analysis). Then, local genetic correlation coefficients between a pair of obesity traits and lipid traits were estimated among the significant univariate loci.

We defined the bivariate loci as follows. Bivariate loci were genomic regions showing significant local heritability estimates (Bonferroni-corrected $p < 0.00002$ (=0.05/2,495); call it as "univariate loci") and local genetic correlation coefficients (Bonferroni-corrected $p < 0.05$ / number of tested loci (univariate loci) for each obesity-lipid pair). We classified the bivariate loci into two different groups based on their directions of association with dyslipidemia risk. In other words, if a given bivariate locus shows positive local genetic correlation coefficients between obesity and dyslipidemia (i.e., $rg < 0$ for BMI-HDL, $rg > 0$ for BMI-LDL and BMI-TG pairs), the locus was classified as Ob/DysL(+) locus whereas if the bivariate locus shows negative local genetic correlation coefficients (i.e., $rg > 0$ for BMI-HDL, $rg < 0$ for BMI-LDL and BMI-TG), the locus was classified as Ob/DysL(–) locus.

G.2 Supplmental Tables and Figures

Table 5.2. Number of participants in PAGE genotyped on MEGA and non-MEGA array by study and by ancestry

Study	Race/ethnicity	MEGA	non-MEGA (Illumina or Affymetrix)
ARIC	European	0	9233
	African	0	2811
BioMe	European	0	1970
	African	4188	1744
	Hispanic/Latino	4293	3764
	East Asian	716	0
	American Indian	0	0
	Other	0	0
CARDIA	European	0	1652
	African	0	889
MEC	African	4465	2513
	Hispanic/Latino	24	6330
	East Asian	2972	2845
HCHS/SOL	Hispanic/Latino	7234	0
WHI	European	0	12563
	African	6092	2553
	Hispanic/Latino	4098	71
	East Asian	291	65

Table 5.3. Summary of the non-MEGA genotype and quality control information in the PAGE

Study	Ancillary Study	Genotyping Platform	Sample Call Rate	HWE threshold	Imputation	Reference Panel
ARIC		Affymetrix GeneChip SNP Array 6.0	90%	$p>10^{-6}$	IMPUTE version 2.3.2	1000 Genome phase 3 v 5
BioMe		Affymetrix GeneChip SNP Array 6.0 and Illumina OmniExpressExome Array	95%	$p>5x10^{-5}$	IMPUTE version 2.3.2	1000 Genome phase 3 v 5
CARDIA		Affymetrix GeneChip SNP Array 6.0	95%	$p>10^{-6}$	IMPUTE version 2.3.2	1000 Genome phase 3 v 5
MEC	JAPC	Illumina Human660W_Quad_v1 Array	95%	NA	MACH	HapMap Phase 2
	LAPC	Illumina Human660W_Quad_v1 Array	95%	NA	MACH	HapMap Phase 2
	AAPC	Illumina Human1M-Duo Array	95%	NA	MACH	HapMap Phase 2
	LA T2D 2.5M	Illumina HumanOmni2.5-4v1_B Array	95%	NA	IMPUTE version 2.2.0	1000 Genomes Phase I integrated variant set
	AABC	Illumina Human1M-Duo Array	95%	NA	MACH	HapMap Phase 2
	LABC	Illumina Human660W_Quad_v1 Array	95%	NA	MACH	HapMap Phase 2
	JABC	Illumina Human660W_Quad_v1 Array	95%	NA	MACH	HapMap Phase 2
	HIBC	Illumina Human660W_Quad_v1 Array	95%	NA	MACH	HapMap Phase 2
WHI	GARNET	Illumina Human Omni1-Quad v1-0 B	98%	p>10-6	IMPUTE version 2.3.2	1001 Genome phase 3 v 5
	GECCO	Illumina 610 and Cytochip 370K	98%	p>10-6	IMPUTE version 2.3.2	1002 Genome phase 3 v 5
	HIPFX	Illumina 50K and 610K	98%	p>10-6	IMPUTE version 2.3.2	1003 Genome phase 3 v 5

Table 5.3. Summary of the non-MEGA genotype and quality control information in the PAGE

Study	Ancillary Study	Genotyping Platform	Sample Call Rate	HWE threshold	Imputation	Reference Panel
	MOPMAP	Affymetrix Gene Titan, Axiom Genome-Wide, Human CEU I Array Plate	90%	p>10-6	IMPUTE version 2.3.2	1004 Genome phase 3 v 5
	WHIMS	Human OmniExpress Exome-8v1_B Genome-Wide Human	98%	p>10-6	IMPUTE version 2.3.2	1005 Genome phase 3 v 5
	LLS	Human OmniExpress Exome-8v1_A Genome-Wide Human	98%	p>10-6	IMPUTE version 2.3.2	1006 Genome phase 3 v 5
	WHI-SHARe	Affymetrix Gene Chip SNP Array 6.0	98%	p>10-6	IMPUTE version 2.3.2	1007 Genome phase 3 v 5
*MEGA		Infinium Expanded Multi-Ethnic Genotyping Array	98%	p>10-6	IMPUTE version 2.3.2	1000 Genome phase 3 v 5

Human genome build 37 and dbSNP version 150 were used for all cases.

Table 5.4. Classification of Ob/DysL(–) and Ob/DysL(+) loci based on local heritability analysis and local genetic correlation analysis

	Ob/DysL(–)	**Ob/DysL(+)**
Step 1. Local heritability (h2)	$p < 0.00002$ (= 0.05/2495)	$p < 0.00002$ (= 0.05/2495)
Step 2. Local genetic correlation (rg)	$p < 0.05$ / N tested loci	$p < 0.05$ / N tested loci
	$rg > 0$ for BMI-HDL	$rg < 0$ for BMI-HDL
	$rg < 0$ for BMI-LDL, BMI-TG	$rg > 0$ for BMI-LDL, BMI-TG

Table 5.5. Distribution of variables

	Total (N=83376)	European (N=25418)	African American (N=25255)	Hispanic (N=25814)	Asian (N=6889)
Age (years)	55.0 (11.5)	57.3 (11.0)	54.8 (11.2)	52.3 (12.2)	57.2 (9.51)
Sex					
Male	26017 (31.2%)	6200 (24.4%)	6994 (27.7%)	9528 (36.9%)	3295 (47.8%)
Female	57359 (68.8%)	19218 (75.6%)	18261 (72.3%)	16286 (63.1%)	3594 (52.2%)
Study					
ARIC	12044 (14.4%)	9233 (36.3%)	2811 (11.1%)	0 (0%)	0 (0%)
BioME	16675 (20.0%)	1970 (7.8%)	5932 (23.5%)	8057 (31.2%)	716 (10.4%)
CARDIA	2541 (3.0%)	1652 (6.5%)	889 (3.5%)	0 (0%)	0 (0%)
MEC	19149 (23.0%)	0 (0%)	6978 (27.6%)	6354 (24.6%)	5817 (84.4%)
SOL	7234 (8.7%)	0 (0%)	0 (0%)	7234 (28.0%)	0 (0%)
WHI	25733 (30.9%)	12563 (49.4%)	8645 (34.2%)	4169 (16.2%)	356 (5.2%)
Obesity measure					
BMI (kg/m^2)	28.6 (6.11)	27.7 (5.60)	30.0 (6.72)	29.2 (5.91)	25.1 (4.08)
Obesity					
No	52550 (63.0%)	18103 (71.2%)	14541 (57.6%)	16225 (62.9%)	3681 (53.4%)
Yes	30826 (37.0%)	7315 (28.8%)	10714 (42.4%)	9589 (37.1%)	3208 (46.6%)
Lipid profile					
HDL (mg/dL)	51.0 (15.9)	51.0 (16.5)	54.4 (16.2)	48.6 (14.6)	49.6 (17.2)
Missing in HDL	40211 (48.2%)	14090 (55.4%)	12113 (48.0%)	9108 (35.3%)	4900 (71.1%)
LDL (mg/dL)	136 (40.7)	134 (38.6)	140 (43.5)	133 (39.7)	141 (38.7)
Missing in LDL	41076 (49.3%)	14261 (56.1%)	12470 (49.4%)	9399 (36.4%)	4946 (71.8%)
TG (mg/dL)	132 (80.2)	127 (78.0)	111 (64.1)	150 (88.0)	139 (82.0)
Missing in TG	40178 (48.2%)	14078 (55.4%)	12222 (48.4%)	8979 (34.8%)	4899 (71.1%)
Dyslipidemia					
No	23532 (28.2%)	6250 (24.6%)	7485 (29.6%)	8857 (34.3%)	940 (13.6%)
Yes	20043 (24.0%)	5127 (20.2%)	5793 (22.9%)	8063 (31.2%)	1060 (15.4%)
Missing in Dyslipidemia	39801 (47.7%)	14041 (55.2%)	11977 (47.4%)	8894 (34.5%)	4889 (71.0%)
Blood Pressure					
DBP (mmHg)	78.4 (12.4)	76.2 (11.5)	82.5 (12.7)	77.6 (12.3)	77.9 (12.0)

Table 5.5. Distribution of variables

	Total (N=83376)	European (N=25418)	African American (N=25255)	Hispanic (N=25814)	Asian (N=6889)
Missing in DBP	21745 (26.1%)	527 (2.1%)	7659 (30.3%)	7629 (29.6%)	5930 (86.1%)
SBP (mmHg)	130 (21.6)	127 (20.4)	136 (22.4)	129 (21.2)	126 (21.4)
Missing in SBP	21761 (26.1%)	533 (2.1%)	7662 (30.3%)	7636 (29.6%)	5930 (86.1%)
Hypertension					
No	41635 (49.9%)	15394 (60.6%)	9729 (38.5%)	14344 (55.6%)	2168 (31.5%)
Yes	40611 (48.7%)	9301 (36.6%)	15359 (60.8%)	11266 (43.6%)	4685 (68.0%)
Missing in hypertension	1130 (1.4%)	723 (2.8%)	167 (0.7%)	204 (0.8%)	36 (0.5%)
Glycemic profile					
Fasting Glucose	5.45 (1.24)	5.46 (1.00)	5.47 (1.51)	5.48 (1.25)	5.12 (1.16)
Missing in Fasting Glucose	37580 (45.1%)	7514 (29.6%)	12312 (48.8%)	13153 (51.0%)	4601 (66.8%)
Fasting Insulin	10.5 (14.5)	9.70 (7.89)	12.0 (23.6)	11.0 (9.55)	6.68 (5.75)
Missing in fasting insulin	37811 (45.4%)	7394 (29.1%)	12460 (49.3%)	13221 (51.2%)	4736 (68.7%)
HOMA-IR	2.58 (2.30)	2.42 (2.04)	2.83 (2.58)	2.73 (2.37)	1.56 (1.48)
Missing in HOMA-IR	38612 (46.3%)	7817 (30.8%)	12709 (50.3%)	13314 (51.6%)	4772 (69.3%)
HbA1c	40.6 (13.2)	37.8 (9.99)	46.8 (17.9)	40.5 (12.2)	42.5 (12.9)
Missing in HbA1c	61516 (73.8%)	16304 (64.1%)	20898 (82.7%)	17565 (68.0%)	6749 (98.0%)
T2D Status					
T2D	21645 (26.0%)	3531 (13.9%)	7882 (31.2%)	7566 (29.3%)	2666 (38.7%)
Pre-diabetes	12454 (14.9%)	5863 (23.1%)	2613 (10.3%)	3756 (14.6%)	222 (3.2%)
T2D controls	40039 (48.0%)	14276 (56.2%)	12067 (47.8%)	10104 (39.1%)	3592 (52.1%)
Other controls	9238 (11.1%)	1748 (6.9%)	2693 (10.7%)	4388 (17.0%)	409 (5.9%)
CVD outcome					
Myocardial Infarction					
No	50419 (60.5%)	19117 (75.2%)	16249 (64.3%)	9539 (37.0%)	5514 (80.0%)
Yes	7589 (9.1%)	2650 (10.4%)	2433 (9.6%)	1890 (7.3%)	616 (8.9%)
Missing in MI	25368 (30.4%)	3651 (14.4%)	6573 (26.0%)	14385 (55.7%)	759 (11.0%)
Stroke					
No	39834 (47.8%)	11660 (45.9%)	13769 (54.5%)	9410 (36.5%)	4995 (72.5%)
Yes	7532 (9.0%)	1763 (6.9%)	2771 (11.0%)	1859 (7.2%)	1139 (16.5%)
Missing in Stroke	36010 (43.2%)	11995 (47.2%)	8715 (34.5%)	14545 (56.3%)	755 (11.0%)

Table 5.6. Global SNP heritability and genetic correlation estimated by LDSC

	Pan UKB EUR GWAS (N ~ 400,000)				
Trait/Trait pairs	Global SNP h2	SE	Global rg	SE	P
BMI	0.25	0.01			
HDL	0.20	0.02			
LDL	0.09	0.01			
TG	0.18	0.02			
BMI-HDL			-0.43	0.03	5.87E-56
BMI-LDL			-0.07	0.03	7.20E-03
BMI-TG			0.31	0.04	1.62E-18

Table 5.7. Summary of local genetic correlation results from LAVA

	Pan UKB EUR GWAS		
Trait pair	**BMI-HDL**	**BMI-LDL**	**BMI-TG**
Significant (p < 0.05/2495) local heritability (BMI + lipid) [a]	2268	1018	2017
Nominally significant (p < 0.05) local genetic correlation (Ob/DysL(-))	11	109	21
Nominally significant (p < 0.05) local genetic correlation (Ob/DysL(+))	1902	146	1494
Significant (p < 0.05 / number of test (a)) local genetic correlation (Ob/DysL(-))	3	10	8
Significant (p < 0.05 / number of test (a)) local genetic correlation (Ob/DysL(+))	786	16	486

Table 5.8. Summary of overlapping genes between TWAS of BMI and lipid traits within BMI-lipid loci

BMI-lipid pair	Overlapping genes within Ob/DysL(-) loci	Overlapping genes* within Ob/DysL(+) loci
BMI-HDL	loc1851 (DDX55, DNAH10, CCDC92)	loc36-PABPC4, loc155-LMOD1, loc158-PM20D1, loc178|179-GALNT2, loc201-SH3YL1, loc240-AC007401.1, FEZ2, loc388-CPS1, loc464-MST1R, RBM6, RNF123, UBA7, loc649-SNORA26, loc692-SLC39A8, loc837-POC5, POLK, loc888-SAR1B, loc902-FAM114A2, loc970-C6orf106, SNRPC, UHRF1BP1, loc1277-FUT10, loc1556-VDAC2, loc1655-HSD17B12, loc1658-ACP2, C1QTNF4, MYBPC3, PSMC3, loc1674-CTSW, FIBP, SNX32, loc1724-HMBS, RP11-110I1.14, loc2028-TRAF3, loc2050-NDUFAF1, NUSAP1, loc2127-EIF3C, SULT1A2, XPO6, loc2128-HSD3B7, MAPK3, loc2255-NPC1, loc2353-SAE1
BMI-LDL	loc837 (POLK, ANKDD1B, POC5)	-
BMI-TG	loc1247 (ERI1), loc1251 (NEIL2), loc1851 (DDX55, DNAH10, CCDC92)	loc36-PABPC4, loc155-LMOD1, loc179-GALNT2, loc201-SH3YL1, loc599-GRK4, MFSD10, loc649-SNORA26, loc1477-MED27, RP11-32B11.2, loc1625-ARNTL, loc1658-ACP2, C1QTNF4, MYBPC3, PSMC3, loc1804-LYZ, RP11-1143G9.4, loc2128-HSD3B7, KAT8, ZNF668, loc2148-CLEC18A, NOB1, RP11-296I10.6, WWP2, loc2255-NPC1, loc2327-CILP2, loc2454-HMGN1, PSMG1

Table 5.9. Genes associated with both BMI and a lipid trait within BMI-Lipid bivariate loci

Pair	Direction	Locus	eQTL.Tissue	Gene_ID	H2 (BMI)	Z-score (BMI)	P (BMI)	H2 (lipid)	Z-score (lipid)	P (lipid)
BMI-HDL	Ob/DysL(-)	1851	METSIM.ADIPOSE.RNASEQ	CCDC92	0.103	5.989	2.11E-09	0.103	21.582	2.64E-103
BMI-HDL	Ob/DysL(-)	1851	YFS.Whole_Blood	DDX55	0.130	4.577	4.72E-06	0.130	6.924	4.40E-12
BMI-HDL	Ob/DysL(-)	1851	METSIM.ADIPOSE.RNASEQ	DNAH10	0.033	4.570	4.87E-06	0.033	18.999	1.73E-80
BMI-LDL	Ob/DysL(-)	837	METSIM.ADIPOSE.RNASEQ	ANKDD1B	0.077	4.498	6.86E-06	0.077	-5.414	6.16E-08
BMI-LDL	Ob/DysL(-)	837	METSIM.ADIPOSE.RNASEQ	POC5	0.107	-7.493	6.75E-14	0.107	9.578	9.93E-22
BMI-LDL	Ob/DysL(-)	837	METSIM.ADIPOSE.RNASEQ	POLK	0.048	-12.249	1.70E-34	0.048	20.794	4.91E-96
BMI-TG	Ob/DysL(-)	1247	YFS.Whole_Blood	ERI1	0.148	-5.822	5.81E-09	0.148	4.497	6.89E-06
BMI-TG	Ob/DysL(-)	1251	YFS.Whole_Blood	NEIL2	0.094	5.455	4.91E-08	0.094	-14.038	9.15E-45
BMI-TG	Ob/DysL(-)	1851	METSIM.ADIPOSE.RNASEQ	CCDC92	0.103	5.989	2.11E-09	0.103	-17.728	2.57E-70
BMI-TG	Ob/DysL(-)	1851	YFS.Whole_Blood	DDX55	0.130	4.577	4.72E-06	0.130	-4.497	6.89E-06
BMI-TG	Ob/DysL(-)	1851	METSIM.ADIPOSE.RNASEQ	DNAH10	0.033	4.570	4.87E-06	0.033	-14.330	1.42E-46
BMI-HDL	Ob/DysL(+)	36	METSIM.ADIPOSE.RNASEQ	PABPC4	0.048	-5.715	1.10E-08	0.048	18.145	1.42E-73
BMI-HDL	Ob/DysL(+)	155	METSIM.ADIPOSE.RNASEQ	LMOD1	0.086	-11.124	9.59E-29	0.086	4.652	3.29E-06
BMI-HDL	Ob/DysL(+)	158	METSIM.ADIPOSE.RNASEQ	PM20D1	0.255	4.723	2.33E-06	0.255	-4.779	1.76E-06
BMI-HDL	Ob/DysL(+)	178	METSIM.ADIPOSE.RNASEQ	GALNT2	0.098	-5.562	2.67E-08	0.098	33.327	1.57E-243
BMI-HDL	Ob/DysL(+)	179	METSIM.ADIPOSE.RNASEQ	GALNT2	0.098	-5.562	2.67E-08	0.098	33.327	1.57E-243
BMI-HDL	Ob/DysL(+)	201	YFS.Whole_Blood	SH3YL1	0.214	5.803	6.52E-09	0.214	-7.304	2.79E-13
BMI-HDL	Ob/DysL(+)	240	YFS.Whole_Blood	AC007401.1	0.057	4.760	1.94E-06	0.057	-5.191	2.09E-07
BMI-HDL	Ob/DysL(+)	240	YFS.Whole_Blood	FEZ2	0.113	4.761	1.92E-06	0.113	-5.577	2.45E-08
BMI-HDL	Ob/DysL(+)	388	METSIM.ADIPOSE.RNASEQ	CPS1	0.123	4.744	2.10E-06	0.123	-4.690	2.73E-06
BMI-HDL	Ob/DysL(+)	464	METSIM.ADIPOSE.RNASEQ	MST1R	0.025	13.549	8.05E-42	0.025	-9.304	1.36E-20
BMI-HDL	Ob/DysL(+)	464	YFS.Whole_Blood	RBM6	0.135	-14.497	1.27E-47	0.135	12.646	1.18E-36
BMI-HDL	Ob/DysL(+)	464	METSIM.ADIPOSE.RNASEQ	RBM6	0.216	-14.497	1.27E-47	0.216	12.650	1.12E-36
BMI-HDL	Ob/DysL(+)	464	METSIM.ADIPOSE.RNASEQ	RNF123	0.034	12.875	6.22E-38	0.034	-9.378	6.73E-21
BMI-HDL	Ob/DysL(+)	464	YFS.Whole_Blood	UBA7	0.031	-12.909	4.00E-38	0.031	9.280	1.69E-20
BMI-HDL	Ob/DysL(+)	649	YFS.Whole_Blood	SNORA26	0.041	4.861	1.17E-06	0.041	-4.787	1.69E-06
BMI-HDL	Ob/DysL(+)	692	METSIM.ADIPOSE.RNASEQ	SLC39A8	0.175	5.461	4.73E-08	0.175	-8.760	1.95E-18
BMI-HDL	Ob/DysL(+)	837	METSIM.ADIPOSE.RNASEQ	POC5	0.107	-7.493	6.75E-14	0.107	4.684	2.82E-06
BMI-HDL	Ob/DysL(+)	837	METSIM.ADIPOSE.RNASEQ	POLK	0.048	-12.249	1.70E-34	0.048	6.225	4.81E-10

Table 5.9. Genes associated with both BMI and a lipid trait within BMI-Lipid bivariate loci

Pair	Direction	Locus	eQTL.Tissue	Gene_ID	H2 (BMI)	Z-score (BMI)	P (BMI)	H2 (lipid)	Z-score (lipid)	P (lipid)
BMI-HDL	Ob/DysL(+)	888	METSIM.ADIPOSE.RNASEQ	SAR1B	0.062	6.165	7.06E-10	0.062	-4.665	3.08E-06
BMI-HDL	Ob/DysL(+)	902	METSIM.ADIPOSE.RNASEQ	FAM114A2	0.039	-5.365	8.09E-08	0.039	5.335	9.55E-08
BMI-HDL	Ob/DysL(+)	970	METSIM.ADIPOSE.RNASEQ	C6orf106	0.040	7.624	2.46E-14	0.040	-9.001	2.24E-19
BMI-HDL	Ob/DysL(+)	970	METSIM.ADIPOSE.RNASEQ	SNRPC	0.046	-11.068	1.80E-28	0.046	8.280	1.23E-16
BMI-HDL	Ob/DysL(+)	970	YFS.Whole_Blood	UHRF1BP1	0.136	12.705	5.54E-37	0.136	-11.075	1.66E-28
BMI-HDL	Ob/DysL(+)	970	METSIM.ADIPOSE.RNASEQ	UHRF1BP1	0.161	12.326	6.53E-35	0.161	-10.212	1.75E-24
BMI-HDL	Ob/DysL(+)	1277	YFS.Whole_Blood	FUT10	0.254	-4.956	7.20E-07	0.254	4.515	6.33E-06
BMI-HDL	Ob/DysL(+)	1556	YFS.Whole_Blood	VDAC2	0.047	-5.317	1.05E-07	0.047	5.678	1.36E-08
BMI-HDL	Ob/DysL(+)	1556	METSIM.ADIPOSE.RNASEQ	VDAC2	0.099	-5.158	2.49E-07	0.099	5.323	1.02E-07
BMI-HDL	Ob/DysL(+)	1655	YFS.Whole_Blood	HSD17B12	0.292	-11.159	6.47E-29	0.292	5.712	1.12E-08
BMI-HDL	Ob/DysL(+)	1655	METSIM.ADIPOSE.RNASEQ	HSD17B12	0.195	-11.114	1.07E-28	0.195	5.936	2.92E-09
BMI-HDL	Ob/DysL(+)	1658	METSIM.ADIPOSE.RNASEQ	ACP2	0.052	6.831	8.44E-12	0.052	-21.886	3.51E-106
BMI-HDL	Ob/DysL(+)	1658	YFS.Whole_Blood	C1QTNF4	0.087	-10.488	9.81E-26	0.087	20.612	2.14E-94
BMI-HDL	Ob/DysL(+)	1658	METSIM.ADIPOSE.RNASEQ	C1QTNF4	0.242	-10.536	5.90E-26	0.242	20.608	2.33E-94
BMI-HDL	Ob/DysL(+)	1658	METSIM.ADIPOSE.RNASEQ	MYBPC3	0.107	-8.729	2.56E-18	0.107	19.559	3.48E-85
BMI-HDL	Ob/DysL(+)	1658	METSIM.ADIPOSE.RNASEQ	PSMC3	0.057	8.039	9.03E-16	0.057	-20.994	7.52E-98
BMI-HDL	Ob/DysL(+)	1674	METSIM.ADIPOSE.RNASEQ	CTSW	0.127	-7.461	8.58E-14	0.127	6.483	8.98E-11
BMI-HDL	Ob/DysL(+)	1674	METSIM.ADIPOSE.RNASEQ	FIBP	0.050	7.663	1.82E-14	0.050	-6.581	4.67E-11
BMI-HDL	Ob/DysL(+)	1674	METSIM.ADIPOSE.RNASEQ	SNX32	0.168	5.545	2.94E-08	0.168	-5.004	5.63E-07
BMI-HDL	Ob/DysL(+)	1724	METSIM.ADIPOSE.RNASEQ	HMBS	0.121	7.321	2.46E-13	0.121	-8.590	8.70E-18
BMI-HDL	Ob/DysL(+)	1724	YFS.Whole_Blood	RP11-110I1.14	0.111	-5.587	2.31E-08	0.111	7.507	6.06E-14
BMI-HDL	Ob/DysL(+)	2028	METSIM.ADIPOSE.RNASEQ	TRAF3	0.037	7.451	9.29E-14	0.037	-5.767	8.07E-09
BMI-HDL	Ob/DysL(+)	2050	YFS.Whole_Blood	NDUFAF1	0.252	4.717	2.39E-06	0.252	-4.664	3.10E-06
BMI-HDL	Ob/DysL(+)	2050	METSIM.ADIPOSE.RNASEQ	NUSAP1	0.034	4.528	5.95E-06	0.034	-4.928	8.31E-07
BMI-HDL	Ob/DysL(+)	2127	YFS.Whole_Blood	EIF3C	0.057	11.290	1.47E-29	0.057	-4.993	5.96E-07
BMI-HDL	Ob/DysL(+)	2127	YFS.Whole_Blood	SULT1A2	0.195	11.202	3.99E-29	0.195	-4.494	6.99E-06
BMI-HDL	Ob/DysL(+)	2127	METSIM.ADIPOSE.RNASEQ	XPO6	0.043	-5.242	1.59E-07	0.043	5.349	8.86E-08
BMI-HDL	Ob/DysL(+)	2128	YFS.Whole_Blood	HSD3B7	0.059	-9.782	1.35E-22	0.059	4.579	4.67E-06
BMI-HDL	Ob/DysL(+)	2128	YFS.Whole_Blood	MAPK3	0.095	-7.973	1.54E-15	0.095	5.170	2.34E-07
BMI-HDL	Ob/DysL(+)	2255	METSIM.ADIPOSE.RNASEQ	NPC1	0.062	-9.823	8.93E-23	0.062	7.123	1.06E-12

Table 5.9. Genes associated with both BMI and a lipid trait within BMI-Lipid bivariate loci

Pair	Direction	Locus	eQTL.Tissue	Gene_ID	H2 (BMI)	Z-score (BMI)	P (BMI)	H2 (lipid)	Z-score (lipid)	P (lipid)
BMI-HDL	Ob/DysL(+)	2353	METSIM.ADIPOSE.RNASEQ	SAE1	0.122	-7.717	1.19E-14	0.122	7.025	2.14E-12
BMI-TG	Ob/DysL(+)	36	METSIM.ADIPOSE.RNASEQ	PABPC4	0.048	-5.715	1.10E-08	0.048	-9.984	1.79E-23
BMI-TG	Ob/DysL(+)	155	METSIM.ADIPOSE.RNASEQ	LMOD1	0.086	-11.124	9.59E-29	0.086	-4.605	4.12E-06
BMI-TG	Ob/DysL(+)	179	METSIM.ADIPOSE.RNASEQ	GALNT2	0.098	-5.562	2.67E-08	0.098	-24.971	1.26E-137
BMI-TG	Ob/DysL(+)	201	YFS.Whole_Blood	SH3YL1	0.214	5.803	6.52E-09	0.214	4.567	4.95E-06
BMI-TG	Ob/DysL(+)	599	YFS.Whole_Blood	GRK4	0.076	5.368	7.98E-08	0.076	6.467	1.00E-10
BMI-TG	Ob/DysL(+)	599	METSIM.ADIPOSE.RNASEQ	GRK4	0.196	5.626	1.84E-08	0.196	5.172	2.32E-07
BMI-TG	Ob/DysL(+)	599	METSIM.ADIPOSE.RNASEQ	MFSD10	0.160	5.714	1.10E-08	0.160	6.792	1.11E-11
BMI-TG	Ob/DysL(+)	649	YFS.Whole_Blood	SNORA26	0.041	4.861	1.17E-06	0.041	5.077	3.84E-07
BMI-TG	Ob/DysL(+)	1477	YFS.Whole_Blood	MED27	0.121	5.114	3.15E-07	0.121	4.570	4.88E-06
BMI-TG	Ob/DysL(+)	1477	YFS.Whole_Blood	RP11-32B11.2	0.085	5.481	4.23E-08	0.085	5.130	2.90E-07
BMI-TG	Ob/DysL(+)	1625	YFS.Whole_Blood	ARNTL	0.125	-8.305	9.98E-17	0.125	-5.350	8.80E-08
BMI-TG	Ob/DysL(+)	1658	METSIM.ADIPOSE.RNASEQ	ACP2	0.052	6.831	8.44E-12	0.052	9.095	9.50E-20
BMI-TG	Ob/DysL(+)	1658	YFS.Whole_Blood	C1QTNF4	0.087	-10.488	9.81E-26	0.087	-5.954	2.62E-09
BMI-TG	Ob/DysL(+)	1658	METSIM.ADIPOSE.RNASEQ	C1QTNF4	0.242	-10.536	5.90E-26	0.242	-5.690	1.27E-08
BMI-TG	Ob/DysL(+)	1658	METSIM.ADIPOSE.RNASEQ	MYBPC3	0.107	-8.729	2.56E-18	0.107	-7.277	3.42E-13
BMI-TG	Ob/DysL(+)	1658	METSIM.ADIPOSE.RNASEQ	PSMC3	0.057	8.039	9.03E-16	0.057	6.139	8.31E-10
BMI-TG	Ob/DysL(+)	1804	METSIM.ADIPOSE.RNASEQ	LYZ	0.061	4.930	8.22E-07	0.061	6.597	4.21E-11
BMI-TG	Ob/DysL(+)	1804	YFS.Whole_Blood	RP11-1143G9.4	0.149	4.744	2.10E-06	0.149	7.194	6.29E-13
BMI-TG	Ob/DysL(+)	2128	YFS.Whole_Blood	HSD3B7	0.059	-9.782	1.35E-22	0.059	-6.918	4.58E-12
BMI-TG	Ob/DysL(+)	2128	METSIM.ADIPOSE.RNASEQ	KAT8	0.167	7.491	6.84E-14	0.167	5.963	2.48E-09
BMI-TG	Ob/DysL(+)	2128	YFS.Whole_Blood	ZNF668	0.050	-9.208	3.32E-20	0.050	-6.986	2.83E-12
BMI-TG	Ob/DysL(+)	2148	METSIM.ADIPOSE.RNASEQ	CLEC18A	0.260	-6.694	2.17E-11	0.260	-5.479	4.27E-08
BMI-TG	Ob/DysL(+)	2148	METSIM.ADIPOSE.RNASEQ	NOB1	0.025	-5.463	4.68E-08	0.025	-5.160	2.47E-07
BMI-TG	Ob/DysL(+)	2148	YFS.Whole_Blood	RP11-296I10.6	0.077	-6.184	6.26E-10	0.077	-4.619	3.86E-06
BMI-TG	Ob/DysL(+)	2148	METSIM.ADIPOSE.RNASEQ	WWP2	0.078	-5.206	1.93E-07	0.078	-5.255	1.48E-07
BMI-TG	Ob/DysL(+)	2255	METSIM.ADIPOSE.RNASEQ	NPC1	0.062	-9.823	8.93E-23	0.062	-6.926	4.32E-12
BMI-TG	Ob/DysL(+)	2327	METSIM.ADIPOSE.RNASEQ	CILP2	0.043	-5.584	2.36E-08	0.043	-9.028	1.75E-19

Table 5.9. Genes associated with both BMI and a lipid trait within BMI-Lipid bivariate loci

Pair	Direction	Locus	eQTL.Tissue	Gene_ID	H2 (BMI)	Z-score (BMI)	P (BMI)	H2 (lipid)	Z-score (lipid)	P (lipid)
BMI-TG	Ob/DysL(+)	2454	METSIM.ADIPOSE.RNASEQ	HMGN1	0.108	4.912	9.02E-07	0.108	6.304	2.90E-10
BMI-TG	Ob/DysL(+)	2454	METSIM.ADIPOSE.RNASEQ	PSMG1	0.047	4.760	1.93E-06	0.047	6.290	3.17E-10

Table 5.10. Comparison with previously reported adiposity variants with protective cardiometabolic profile

Known variants	locus	CHR	START	STOP	UKB:BMI-HDL	UKB:BMI-LDL	UKB:BMI-TG
rs1010447 (PMID 33619380)	13	1	10753428	11709173	neutral	neutral	neutral
rs683135 (PMID 27841877)	36	1	38474037	40200950	Ob/DysL(+)	neutral	Ob/DysL(+)
rs17386142 (PMID 27841877);rs3789588 (PMID 33619380)	44	1	49185415	51777577	neutral	N/A	neutral
rs6603981 (PMID 33619380)	81	1	92904466	94168576	neutral	neutral	neutral
rs11577194 (PMID 27841877)	97	1	110224231	111134062	neutral	neutral	neutral
rs9425291 (PMID 27841877)	129	1	171044256	172465153	Ob/DysL(+)	neutral	neutral
rs2802774 (PMID 33980691)	156	1	202583885	204092537	neutral	N/A	neutral
rs11118306 (PMID 30352878);rs12130231 (PMID 33980691);rs2820446 (PMID 33619380);rs4846565 (PMID 25048195;27841877)	169	1	218563961	220073132	neutral	neutral	neutral
rs1260326 (PMID 33619380)	230	2	26894103	28819510	neutral	neutral	neutral
rs2249105 (PMID 27841877)	271	2	64696090	65938002	neutral	neutral	neutral
rs4988235 (PMID 33619380)	327	2	135160198	137061003	Ob/DysL(+)	neutral	neutral
rs10195252 (PMID 25048195;27841877);rs1128249 (PMID 33619380);rs13389219 (PMID 30352878;33980691)	349	2	164702313	165917788	neutral	neutral	neutral
rs1427445 (PMID 33619380);rs492400 (PMID 27841877)	395	2	218396259	219678783	neutral	neutral	neutral
rs2943645 (PMID 25048195;27841877);rs2943652 (PMID 33619380);rs2943653 (PMID 30352878;33980691)	403	2	226242843	227557841	neutral	neutral	neutral
rs11563251 (PMID 33619380)	410	2	234115093	234945577	neutral	neutral	Ob/DysL(+)

Table 5.10. Comparison with previously reported adiposity variants with protective cardiometabolic profile

Known variants	locus	CHR	START	STOP	UKB:BMI-HDL	UKB:BMI-LDL	UKB:BMI-TG
rs17036328 (PMID 25048195);rs1801282 (PMID 30352878);rs2881654 (PMID 33619380);rs308971 (PMID 27841877);rs4684847 (PMID 33980691)	434	3	11997659	12859209	neutral	neutral	neutral
rs3864041 (PMID 27841877)	437	3	15150877	16084005	Ob/DysL(+)	neutral	Ob/DysL(+)
rs295449 (PMID 27841877)	463	3	45844192	47588461	neutral	N/A	Ob/DysL(+)
rs4392441 (PMID 33619380)	464	3	47588462	50387742	Ob/DysL(+)	N/A	neutral
rs11130329 (PMID 27841877)	466	3	51953969	54074844	neutral	neutral	neutral
rs4616635 (PMID 33619380)	477	3	64662374	65326751	Ob/DysL(+)	N/A	Ob/DysL(+)
rs11708067 (PMID 33619380);rs9881942 (PMID 27841877)	528	3	122340833	123518507	Ob/DysL(+)	neutral	neutral
rs645040 (PMID 27841877)	540	3	135288395	137372141	Ob/DysL(+)	N/A	Ob/DysL(+)
rs9851766 (PMID 33980691)	541	3	137372142	138693846	neutral	N/A	neutral
rs62271373 (PMID 33980691)	552	3	149998412	151131307	neutral	neutral	Ob/DysL(+)
rs4481184 (PMID 33619380)	582	3	184524269	185709996	neutral	neutral	neutral
rs4234589 (PMID 33619380)	583	3	185709997	186602045	neutral	neutral	neutral
rs2699429 (PMID 27841877)	599	4	2468936	3549229	Ob/DysL(+)	neutral	Ob/DysL(+)
rs4450871 (PMID 33980691)	601	4	4266179	5051834	neutral	neutral	neutral
rs13132853 (PMID 33980691)	638	4	37880861	38984838	Ob/DysL(+)	N/A	neutral
rs2276936 (PMID 30352878);rs3822072 (PMID 33619380;25048195;27841877);rs987469 (PMID 33980691)	680	4	89244555	90236971	neutral	neutral	neutral
rs13107325 (PMID 33619380)	692	4	102544804	104384534	Ob/DysL(+)	Ob/DysL(-)	Ob/DysL(+)
rs974801 (PMID 25048195)	694	4	105319196	106479155	Ob/DysL(+)	N/A	neutral
rs6822892 (PMID 25048195;27841877)	740	4	157597310	159176073	neutral	N/A	neutral

Table 5.10. Comparison with previously reported adiposity variants with protective cardiometabolic profile

Known variants	locus	CHR	START	STOP	UKB:BMI-HDL	UKB:BMI-LDL	UKB:BMI-TG
rs3776717 (PMID 33619380);rs4865796 (PMID 25048195;27841877)	820	5	52837226	53747856	neutral	neutral	neutral
rs30351 (PMID 33980691);rs40271 (PMID 30352878);rs459193 (PMID 33619380;25048195;27841877);rs9686661 (PMID 33619380)	822	5	55221399	55968966	neutral	neutral	neutral
rs4976033 (PMID 33619380;27841877;33980691)	833	5	67096192	68006993	Ob/DysL(+)	N/A	neutral
rs7713317 (PMID 33619380)	854	5	95117260	96467377	Ob/DysL(+)	N/A	Ob/DysL(+)
rs6887914 (PMID 27841877)	869	5	111983116	113121555	neutral	N/A	Ob/DysL(+)
rs1045241 (PMID 27841877);rs9764678 (PMID 33980691)	875	5	118605252	119664173	neutral	N/A	neutral
rs11135038 (PMID 33980691);rs2434612 (PMID 33619380;27841877)	907	5	157191082	158484775	neutral	neutral	neutral
rs6861681 (PMID 33619380);rs966544 (PMID 27841877)	919	5	172285683	173606995	neutral	N/A	neutral
rs3094222 (PMID 33619380)	957	6	30715007	31106493	Ob/DysL(+)	neutral	neutral
rs12525532 (PMID 27841877)	971	6	34979271	36346353	Ob/DysL(+)	neutral	Ob/DysL(+)
rs998584 (PMID 33619380;30352878;33980691)	977	6	42103739	43770626	neutral	neutral	neutral
rs6937438 (PMID 27841877)	978	6	43770627	44596897	neutral	neutral	neutral
rs2745353 (PMID 25048195;27841877);rs72959041 (PMID 33980691);rs9385400 (PMID 33619380)	1054	6	125365055	127545459	neutral	neutral	neutral
rs9492443 (PMID 27841877)	1057	6	129850179	130550137	neutral	neutral	neutral

Table 5.10. Comparison with previously reported adiposity variants with protective cardiometabolic profile

Known variants	locus	CHR	START	STOP	UKB:BMI-HDL	UKB:BMI-LDL	UKB:BMI-TG
rs3861397 (PMID 27841877);rs573454216 (PMID 33980691);rs632057 (PMID 30352878)	1065	6	139716714	141449453	Ob/DysL(+)	neutral	neutral
rs17080091 (PMID 33619380)	1073	6	150635316	151629954	neutral	neutral	Ob/DysL(+)
rs539958 (PMID 33619380)	1084	6	160583919	161371014	neutral	neutral	neutral
rs702485 (PMID 33619380)	1101	7	6020654	6716905	neutral	Ob/DysL(+)	Ob/DysL(+)
rs17169104 (PMID 27841877)	1115	7	15877565	16739013	neutral	N/A	neutral
rs864745 (PMID 33619380)	1126	7	27351287	28890886	neutral	N/A	neutral
rs4731702 (PMID 33619380);rs972283 (PMID 27841877;30352878;33980691)	1209	7	130418705	131856481	neutral	N/A	neutral
rs6977416 (PMID 33980691)	1223	7	149843000	150897818	Ob/DysL(+)	neutral	neutral
rs17149279 (PMID 33619380);rs2126259 (PMID 27841877);rs9987289 (PMID 33619380)	1248	8	9167796	9835863	neutral	Ob/DysL(+)	Ob/DysL(-)
rs1011685 (PMID 27841877)	1264	8	19488889	20135628	neutral	neutral	neutral
rs10090367 (PMID 33619380);rs12681990 (PMID 33980691)	1280	8	36641175	38803980	Ob/DysL(+)	N/A	Ob/DysL(+)
rs4738141 (PMID 27841877)	1307	8	72013185	72917489	neutral	N/A	neutral
rs2980888 (PMID 33980691;30352878);rs7005992 (PMID 27841877)	1351	8	125453323	126766827	Ob/DysL(-)	Ob/DysL(-)	Ob/DysL(-)
rs498313 (PMID 27841877)	1423	9	77862309	78630915	neutral	neutral	neutral
rs7896600 (PMID 33619380)	1499	10	11856925	12581571	neutral	neutral	N/A
rs10995441 (PMID 27841877)	1545	10	64069688	65400431	Ob/DysL(+)	neutral	Ob/DysL(+)
rs10883832 (PMID 33619380)	1581	10	104206838	106142283	neutral	N/A	neutral
rs7903146 (PMID 33619380)	1589	10	114255955	115588903	neutral	neutral	neutral
rs740746 (PMID 33619380)	1590	10	115588904	116845213	neutral	neutral	neutral
rs7928810 (PMID 33619380)	1628	11	16383387	17583948	neutral	N/A	N/A

Table 5.10. Comparison with previously reported adiposity variants with protective cardiometabolic profile

Known variants	locus	CHR	START	STOP	UKB:BMI-HDL	UKB:BMI-LDL	UKB:BMI-TG
rs113222038 (PMID 33980691)	1672	11	61717118	62800368	neutral	neutral	neutral
rs11231693 (PMID 27841877);rs2845885 (PMID 33619380)	1673	11	62800369	64594822	Ob/DysL(+)	neutral	neutral
rs11603334 (PMID 33619380)	1679	11	71242835	72875068	neutral	N/A	neutral
rs17402950 (PMID 27841877)	1754	12	13559528	14656849	Ob/DysL(+)	N/A	neutral
rs11045172 (PMID 30352878;33980691);rs7134375 (PMID 33619380)	1760	12	20074931	20866285	neutral	neutral	neutral
rs718314 (PMID 33619380;27841877)	1766	12	25990814	26958056	neutral	neutral	neutral
rs10876529 (PMID 33980691);rs754133 (PMID 33619380)	1792	12	54371449	55416802	Ob/DysL(+)	neutral	neutral
rs3741414 (PMID 33619380)	1794	12	56987106	58748139	Ob/DysL(+)	neutral	neutral
rs10774625 (PMID 33619380)	1841	12	111592382	113947983	neutral	neutral	neutral
rs11057405 (PMID 33619380);rs12369179 (PMID 33980691)	1850	12	121817510	123396634	neutral	N/A	neutral
rs7133378 (PMID 33619380;30352878;33980691);rs7973683 (PMID 27841877);rs863750 (PMID 33619380)	1851	12	123396635	124843768	Ob/DysL(-)	neutral	Ob/DysL(-)
rs7323406 (PMID 27841877)	1950	13	111621245	112319064	neutral	N/A	neutral
rs17522122 (PMID 33619380)	1965	14	32382246	33591113	Ob/DysL(+)	N/A	neutral
rs72697297 (PMID 33980691)	2019	14	92101229	93386328	neutral	neutral	neutral
rs12441543 (PMID 33980691)	2042	15	30604120	32177320	neutral	N/A	neutral
rs7176058 (PMID 27841877)	2049	15	39238841	40604780	neutral	neutral	neutral
rs8032586 (PMID 27841877)	2074	15	72058130	73375718	neutral	N/A	neutral
rs1378940 (PMID 33619380)	2076	15	74458114	76401952	neutral	neutral	neutral
rs879620 (PMID 33619380)	2105	16	3379997	4816145	Ob/DysL(+)	N/A	Ob/DysL(+)
rs4985155 (PMID 33619380)	2118	16	13893408	15921108	neutral	neutral	neutral

Table 5.10. Comparison with previously reported adiposity variants with protective cardiometabolic profile

Known variants	locus	CHR	START	STOP	UKB:BMI-HDL	UKB:BMI-LDL	UKB:BMI-TG
rs754814 (PMID 27841877)	2178	17	4463972	5573784	Ob/DysL(+)	neutral	neutral
rs12940684 (PMID 33980691);rs2955617 (PMID 33619380)	2181	17	7264459	8554763	Ob/DysL(+)	neutral	neutral
rs6504872 (PMID 33619380)	2208	17	44865833	45883901	neutral	neutral	neutral
rs142186653 (PMID 33980691)	2230	17	73741322	74908266	Ob/DysL(+)	neutral	neutral
rs11664106 (PMID 33980691)	2241	18	2839843	3722828	neutral	neutral	neutral
rs7233512 (PMID 33980691)	2271	18	41425455	42974165	neutral	N/A	neutral
rs7227237 (PMID 27841877)	2275	18	46558307	47455925	neutral	neutral	neutral
rs12454712 (PMID 33619380)	2289	18	60780195	62074993	neutral	neutral	neutral
rs4804833 (PMID 27841877);rs8101064 (PMID 27841877)	2315	19	7249360	8199016	Ob/DysL(+)	neutral	Ob/DysL(+)
rs4804311 (PMID 27841877)	2316	19	8199017	9105577	neutral	neutral	neutral
rs7258937 (PMID 30352878;33980691);rs731839 (PMID 33619380;25048195;27841877)	2340	19	33785836	34633274	neutral	neutral	neutral
rs2075650 (PMID 33619380)	2351	19	45040933	45893307	Ob/DysL(-)	Ob/DysL(-)	Ob/DysL(+)
rs555162510 (PMID 33980691)	2352	19	45893308	46765060	neutral	neutral	neutral
rs6029180 (PMID 33980691)	2403	20	38427595	40272390	neutral	neutral	neutral
rs1211644 (PMID 33619380);rs6066149 (PMID 27841877)	2408	20	44072211	45673603	neutral	N/A	neutral
rs132985 (PMID 27841877);rs2267373 (PMID 30352878);rs3761445 (PMID 33619380);rs4821764 (PMID 33980691)	2482	22	37364005	38718589	neutral	neutral	neutral

Table 5.11. The associations of PRS-Ob/DysL(-) or PRS-Ob/DysL(+) with CMD

		PRS-Ob/DysL(-)					PRS-Ob/DysL(+)				
BMI-Lipid loci	**Outcome**	**OR**	**SE**	**95% LCL**	**95% UCL**	**p-value**	**OR**	**SE**	**95% LCL**	**95% UCL**	**p-value**
BMI-HDL	Obesity	1.024	0.007	1.009	1.039	1.82E-03	1.442	0.008	1.419	1.466	<1E-08
	dyslipidemia	0.878	0.010	0.861	0.895	<1E-08	1.085	0.011	1.062	1.107	<1E-08
	T2D status	1.005	0.009	0.988	1.023	5.57E-01	1.221	0.010	1.198	1.245	<1E-08
	Hypertension	1.002	0.008	0.987	1.017	8.09E-01	1.134	0.008	1.116	1.153	<1E-08
	Myocardial Infarction	0.983	0.014	0.957	1.009	2.01E-01	1.158	0.015	1.125	1.191	<1E-08
	Stroke	0.972	0.016	0.941	1.003	8.07E-02	1.074	0.017	1.039	1.111	3.04E-05
BMI-LDL	Obesity	1.090	0.008	1.073	1.108	<1E-08	1.052	0.008	1.037	1.068	<1E-08
	dyslipidemia	0.988	0.011	0.967	1.009	2.57E-01	1.015	0.010	0.995	1.036	1.30E-01
	T2D status	1.026	0.010	1.006	1.046	1.01E-02	1.063	0.009	1.044	1.083	<1E-08
	Hypertension	1.004	0.008	0.987	1.020	6.71E-01	1.023	0.008	1.007	1.039	3.90E-03
	Myocardial Infarction	1.027	0.015	0.997	1.058	7.90E-02	1.013	0.014	0.986	1.041	3.59E-01
	Stroke	0.983	0.018	0.949	1.018	3.44E-01	0.995	0.016	0.963	1.027	7.41E-01
BMI-TG	Obesity	1.030	0.008	1.014	1.047	2.38E-04	1.334	0.008	1.313	1.355	<1E-08
	dyslipidemia	0.982	0.011	0.962	1.003	9.31E-02	1.073	0.010	1.052	1.096	<1E-08
	T2D status	1.041	0.010	1.021	1.062	5.73E-05	1.204	0.010	1.181	1.227	<1E-08
	Hypertension	1.017	0.008	1.000	1.033	4.52E-02	1.126	0.008	1.108	1.144	<1E-08
	Myocardial Infarction	0.995	0.015	0.967	1.025	7.52E-01	1.093	0.015	1.062	1.124	<1E-08
	Stroke	1.005	0.018	0.970	1.042	7.64E-01	1.052	0.017	1.017	1.087	2.94E-03

Table 5.12. The associations of PRS-BMI (reference) with CMD

			PRS-BMI(reference)		
Outcome	**OR**	**SE**	**95% LCL**	**95% UCL**	**p-value**
Obesity	1.689	0.009	1.661	1.717	<1E-08
dyslipidemia	1.101	0.011	1.079	1.124	<1E-08
T2D status	1.319	0.010	1.294	1.345	<1E-08
Hypertension	1.191	0.008	1.172	1.210	<1E-08
Myocardial Infarction	1.177	0.015	1.144	1.211	<1E-08
Stroke	1.092	0.017	1.056	1.129	2.60E-07

Table 5.13. The associations of PRS-Ob/DysL(-) or PRS-Ob/DysL(+) with CMD risk factors

		PRS-Ob/DysL(-)					PRS-Ob/DysL(+)				
BMI-Lipid loci	**Outcome**	**Beta**	**SE**	**95% LCL**	**95% UCL**	**p-value**	**Beta**	**SE**	**95% LCL**	**95% UCL**	**p-value**
BMI-HDL	BMI	0.010	0.003	0.004	0.016	2.54E-03	0.183	0.003	0.177	0.190	<1E-08
	HDL	0.033	0.005	0.024	0.041	<1E-08	-0.071	0.005	-0.080	-0.061	<1E-08
	LDL	-0.069	0.005	-0.079	-0.060	<1E-08	0.012	0.005	0.002	0.022	1.51E-02
	Log(triglyceride)	-0.022	0.004	-0.030	-0.013	1.67E-06	0.054	0.005	0.044	0.063	<1E-08
	Total cholesterol	-0.057	0.005	-0.066	-0.049	<1E-08	0.001	0.005	-0.008	0.011	7.68E-01
	Fasting glucose	-0.009	0.004	-0.018	0.000	3.89E-02	0.063	0.005	0.053	0.072	<1E-08
	Log(Fasting insulin)	-0.008	0.004	-0.016	0.001	8.76E-02	0.099	0.005	0.090	0.108	<1E-08
	Log(HOMA-IR)	-0.007	0.005	-0.016	0.002	1.12E-01	0.103	0.005	0.094	0.112	<1E-08
	HbA1c	0.012	0.006	0.000	0.024	4.24E-02	0.061	0.006	0.048	0.073	<1E-08
	Diastolic blood pressure	0.003	0.004	-0.004	0.011	3.62E-01	0.044	0.004	0.036	0.052	<1E-08
	Systolic blood pressure	-0.002	0.004	-0.009	0.005	5.98E-01	0.053	0.004	0.046	0.061	<1E-08
BMI-LDL	BMI	0.043	0.004	0.036	0.050	<1E-08	0.024	0.003	0.018	0.031	<1E-08
	HDL	-0.008	0.005	-0.018	0.001	9.68E-02	-0.010	0.005	-0.019	-0.001	2.77E-02
	LDL	-0.011	0.005	-0.021	-0.001	2.88E-02	0.004	0.005	-0.005	0.014	3.53E-01
	Log(triglyceride)	0.012	0.005	0.003	0.022	1.31E-02	0.006	0.005	-0.003	0.015	1.66E-01
	Total cholesterol	-0.010	0.005	-0.020	0.000	4.96E-02	0.003	0.005	-0.007	0.012	5.88E-01
	Fasting glucose	0.008	0.005	-0.001	0.018	8.42E-02	0.023	0.004	0.014	0.032	2.70E-07
	Log(Fasting insulin)	0.013	0.005	0.003	0.022	1.02E-02	0.017	0.005	0.008	0.026	1.66E-04
	Log(HOMA-IR)	0.013	0.005	0.003	0.023	1.02E-02	0.022	0.005	0.013	0.031	1.55E-06
	HbA1c	0.018	0.007	0.005	0.031	5.23E-03	0.021	0.006	0.009	0.033	4.59E-04
	Diastolic blood pressure	0.005	0.004	-0.003	0.013	2.32E-01	0.010	0.004	0.002	0.017	1.05E-02
	Systolic blood pressure	0.006	0.004	-0.002	0.013	1.37E-01	0.015	0.004	0.007	0.022	6.31E-05
BMI-TG	BMI	0.014	0.004	0.007	0.021	1.50E-04	0.150	0.003	0.143	0.157	<1E-08
	HDL	-0.002	0.005	-0.012	0.007	6.40E-01	-0.058	0.005	-0.067	-0.048	<1E-08
	LDL	-0.001	0.005	-0.011	0.009	8.31E-01	0.006	0.005	-0.003	0.016	1.91E-01
	Log(triglyceride)	-0.001	0.005	-0.011	0.008	7.79E-01	0.048	0.005	0.039	0.057	<1E-08
	Total cholesterol	-0.003	0.005	-0.013	0.007	5.36E-01	0.000	0.005	-0.009	0.010	9.35E-01

Table 5.13. The associations of PRS-Ob/DysL(-) or PRS-Ob/DysL(+) with CMD risk factors

BMI-Lipid loci	Outcome	PRS-Ob/DysL(-) Beta	SE	95% LCL	95% UCL	p-value	PRS-Ob/DysL(+) Beta	SE	95% LCL	95% UCL	p-value
	Fasting glucose	0.013	0.005	0.004	0.022	5.31E-03	0.055	0.005	0.046	0.064	<1E-08
	Log(Fasting insulin)	0.006	0.005	-0.003	0.016	1.68E-01	0.087	0.005	0.078	0.097	<1E-08
	Log(HOMA-IR)	0.009	0.005	0.000	0.019	5.05E-02	0.092	0.005	0.083	0.101	<1E-08
	HbA1c	0.012	0.006	0.000	0.024	4.72E-02	0.060	0.006	0.048	0.073	<1E-08
	Diastolic blood pressure	0.001	0.004	-0.007	0.008	8.49E-01	0.046	0.004	0.038	0.054	<1E-08
	Systolic blood pressure	0.009	0.004	0.002	0.017	1.29E-02	0.051	0.004	0.044	0.059	<1E-08

Table 5.14. The associations of PRS-BMI (reference) with CMD risk factors

	PRS-BMI (reference)				
Outcome	**Beta**	**SE**	**95% LCL**	**95% UCL**	**p-value**
BMI	0.257	0.003	0.251	0.264	<1E-08
HDL	-0.088	0.005	-0.097	-0.079	<1E-08
LDL	0.012	0.005	0.003	0.022	1.27E-02
Log(triglyceride)	0.058	0.005	0.049	0.068	<1E-08
Total cholesterol	-0.002	0.005	-0.012	0.007	6.07E-01
Fasting glucose	0.077	0.005	0.068	0.086	<1E-08
Log(Fasting insulin)	0.133	0.005	0.124	0.142	<1E-08
Log(HOMA-IR)	0.138	0.005	0.128	0.147	<1E-08
HbA1c	0.086	0.006	0.074	0.099	<1E-08
Diastolic blood pressure	0.058	0.004	0.050	0.066	<1E-08
Systolic blood pressure	0.072	0.004	0.064	0.079	<1E-08

CHAPTER 6: CONCLUSIONS

A. Recapitulation of Specific Aims

Obesity is an enormous global public health burden[1,2] and is a major risk factor for numerous health outcomes, including cardiovascular diseases[3] through its impact on CVD risk factors (e.g., dyslipidemia, hypertension, and type 2 diabetes).[201-204] Although the current obesogenic environment has been a critical component of the secular trends of increasing obesity, inter-individual variability in response to external environmental factors for obesity is largely driven by genetics (heritability estimates ranged from 40% to 70%).[4] Indeed, thousands of obesity-associated genetic loci have been identified, enabling improved risk prediction of obesity by using polygenic risk scores (PRS).[70] The application of obesity PRS has strong clinical utilities and may elucidate novel biological underpinnings of obesity pathogenicity (biological utilities).

Nevertheless, heterogeneities in the performance of PRS across populations need to be documented and explored prior to the clinical application of these PRS. Thus, the overarching goal of this dissertation is to better understand the heterogeneous impact of obesity genetic susceptibility and to provide the clinical and biological implications of these heterogeneities. Specifically, Aim 1 addressed the lack of understanding of the heterogeneities in the prediction performance of obesity PRS in various settings – e.g., PRS estimation modeling, race/ethnicity, demographic, lifestyle, and comorbidities. Though novel PRS estimation methods and new large-scale discovery GWAS are now available, the potential differences in prediction performance by

PRS estimation methods and by self-reported race/ethnicity and population-specific contexts have not been investigated. Failing to account for heterogeneity across contexts could dramatically limit the clinical utility of obesity PRS. Aim 2 addressed the lack of understanding of the heterogeneous impact of obesity genetic factors on downstream disease, in particular for CVD. Even though there has been extensive research linking obesity to CVD and CVD risk factors, considerable gaps in research persist. Only limited studies have examined the often-mentioned but poorly comprehended heterogeneity in CVD risk observed among individuals with obesity[31]. An understudied factor that could potentially contribute to the heterogeneities in obesity-related CVD risk is the impact of shared genetic underpinnings, in particular for the underlying genetic correlation between obesity and dyslipidemia.[137,205] A better understanding of the shared genetic effects of obesity and dyslipidemia could improve our understanding of obesity's heterogeneous effect on CVD.

In this dissertation, to address these two research gaps, we investigated the performance of obesity PRS across different PRS estimation methods and diverse contexts, including race/ethnicity, demographic background, lifestyle factors, and various obesity-related comorbidities. We also explored the shared genetic underpinnings of obesity-related traits and dyslipidemia-related traits to better understand if this shared genetic architecture could improve our understanding of obesity's heterogeneous impact on downstream disease. We conducted our analyses using data from PAGE study participants (Aim 1 and Aim 2), GWAS of BMI and WHRadjBMI from the GIANT consortium (Aim 1), and GWAS of BMI from UKBB (Aim 2).

B. Summary of Main Findings

For the first aim, we explored the prediction performance of PRS for obesity-related traits (PRS-BMI and PRS-WHRadjBMI) in the diverse PAGE study populations. Specifically, we evaluated two different PRS estimation methods [P+T and PRS-CS (PRS-CSx)] in population-pooled and population-specific sets (based on self-reported race/ethnicity). We observed substantial differences in the performance of PRS across statistical methods and across self-reported race/ethnic groups. We also characterized the performance of PRS-BMI and PRS-WHRadjBMI in various demographic, lifestyle, and comorbidity contexts. We observed differences in the prediction performance of PRS by age group, sex, smoking status, T2D status, and hypertension status. These findings suggested that 1) there is room for an improved prediction performance of obesity PRS by applying better PRS estimation methods, 2) more discovery GWAS are needed for under-represented populations (e.g., non-Hispanic Black), and 3) individuals' contextual variables should be considered as important modifying factors of PRS performance.

For the second aim, we identified two distinct types of genomic loci with shared genetic underpinnings of BMI and lipid levels in opposing directions [Ob/DysL(–) Ob/DysL(+)] using local genetic correlation analysis in the UKBB GWAS results. We further identified potential causal genes (*NEIL2, POLK, ANKDD1B,* and *POC5)* underlying the counter-intuitive Ob/DysL(–) loci through integration with gene expression data and a gene-based TWAS approach. To generalize our findings to other populations with distinct ancestries, we constructed the BMI-lipid bivariate loci-based PRS [PRS-Ob/DysL(–) and PRS-Ob/DysL(+)] in the diverse PAGE study populations and evaluated the associations of the PRS-Ob/DysL(–) and PRS-Ob/DysL(+) with BMI, lipid levels, and downstream CVD and its risk factors. As a result, from

the analysis of the BMI-HDL pair, PRS constructed using the counter-intuitive Ob/DysL(–) loci demonstrated protective associations with dyslipidemia and some downstream CVD risk factors in this independent population. Thus, the results suggested that distinct types of correlated genomic loci (shared genetic underpinnings) between obesity-related traits and lipid traits partly explained heterogeneities in CVD risk among people with obesity. Also, the findings illustrated that obesity-associated variants can be involved in distinct biological mechanisms and demonstrate heterogeneous impacts on downstream diseases. Furthermore, it can be inferred that each obesity variant should be weighted and characterized differently by their influence on downstream diseases.

B.1. Strengths

This study has notable strengths. First, the total sample size of the PAGE study was large, enabling comprehensive characterization of the PRS-BMI or PRS-WHRadjBMI (Aim 1) and evaluation of the relationships between BMI-lipid bivariate loci and various CVD profiles (Aim 2). The distribution of self-identified race/ethnicity (across non-Hispanic White, non-Hispanic Black, and Hispanic/Latino populations) in the PAGE study was also relatively balanced, ensuring that the race/ethnicity-pooled results were not biased toward one particular population. In addition, PAGE participants were extensively phenotyped, facilitating the thorough characterization of the PRS-BMI and PRS-WHRadjBMI in various contexts (Aim 1) and association testing with various CVD and CVD risk factors (Aim 2).

B.2. Limitations

The current study also has limitations. First, since the discovery GWAS was predominantly from populations that self-identified as non-Hispanic White, the PRS predictive performance across diverse populations was likely limited. Therefore, this could have also

impacted the performance of PRS across different contexts as well (Aim 1). Related to this point, the genomic partitioning conducted in Aim 2 used a European ancestry LD structure (1000 Genome European population), thus genomic regions may not well represent LD blocks from populations of diverse ancestries. Thus, the identified BMI-lipid bivariate loci may not generalize across all populations (Aim 2). In addition, we used broad self-identified race/ethnicity categories, which likely introduced heterogeneity into our populations and possibly limited our consideration of prediction performance across populations (Aim 1). Furthermore, as all of our analyses were cross-sectional, our inference on the differences in prediction performance over time and across the life course is limited (Aim 2).

C. Overall Conclusion

We observed substantial heterogeneities in the prediction performance of PRS for obesity-related traits across different PRS estimation methods, diverse self-reported race/ethnicity, demographic factors, lifestyle factors, and comorbidities. We also identified distinct genomic regions with heterogeneous shared genetic signals between obesity and dyslipidemia and observed the heterogeneous downstream CVD risk profile by the shared genetic loci. All in all, these heterogeneities in obesity genomics should be considered before being utilized in clinical and public health settings.

APPENDIX 1: GWAS OF OBESITY-RELATED TRAITS REPORTING GENOME-WIDE SIGNIFICANT SIGNALS

FIRST.AUTHOR	YEAR	STUDY	PUBMED
Frayling TM	2007	A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity.	17434869
Scuteri A	2007	Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits.	17658951
Chambers JC	2008	Common genetic variation near MC4R is associated with waist circumference and insulin resistance.	18454146
Loos RJ	2008	Common variants near MC4R are associated with fat mass, weight and risk of obesity.	18454148
Thorleifsson G	2008	Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity.	19079260
Willer CJ	2008	Six new loci associated with body mass index highlight a neuronal influence on body weight regulation.	19079261
Meyre D	2009	Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations.	19151714
Liu XG	2009	Genome-wide association and replication studies identified TRHR as an important gene for lean body mass.	19268274
Cotsapas C	2009	Common body mass index-associated variants confer risk of extreme obesity.	19553259
Heard-Costa NL	2009	NRXN3 is a novel locus for waist circumference: a genome-wide association study from the CHARGE Consortium.	19557197
Scherag A	2010	Two new Loci for body-weight regulation identified in a joint analysis of genome-wide association studies for early-onset extreme obesity in French and german study groups.	20421936
Heid IM	2010	Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution.	20935629
Speliotes EK	2010	Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index.	20935630
Wan ES	2010	Genome-wide association analysis of body mass in chronic obstructive pulmonary disease.	21037115
Kraja AT	2011	A bivariate genome-wide approach to metabolic syndrome: STAMPEED consortium.	21386085
Wang K	2011	A genome-wide association study on obesity and obesity-related traits.	21552555
Jiao H	2011	Genome wide association study identifies KCNMA1 contributing to human obesity.	21708048
Paternoster L	2011	Genome-wide population-based association study of extremely overweight young adults--the GOYA study.	21935397
Melka MG	2011	Genome-wide scan for loci of adolescent obesity and their relationship with blood pressure.	22013104
Wen W	2012	Meta-analysis identifies common variants associated with body mass index in east Asians.	22344219
Okada Y	2012	Common variants at CDKAL1 and KLF9 are associated with body mass index in east Asian populations.	22344221
Bradfield JP	2012	A genome-wide association meta-analysis identifies new childhood obesity loci.	22484627

FIRST.AUTHOR	YEAR	STUDY	PUBMED
Fox CS	2012	Genome-wide association for abdominal subcutaneous and visceral adipose reveals a novel locus for visceral fat in women.	22589738
Yang J	2012	FTO genotype is associated with phenotypic variability of body mass index.	22982992
Guo YF	2012	Suggestion of GLYAT gene underlying variation of bone size and body lean mass as revealed by a bivariate genome-wide association study.	23108985
Comuzzie AG	2012	Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population.	23251661
Melen E	2013	Genome-wide association study of body mass index in 23 000 individuals with and without asthma.	23517042
Berndt SI	2013	Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture.	23563607
Wheeler E	2013	Genome-wide SNP and CNV analysis identifies common and low-frequency variants associated with severe early-onset obesity.	23563609
Monda KL	2013	A meta-analysis identifies new loci associated with body mass index in individuals of African ancestry.	23583978
Graff M	2013	Genome-wide analysis of BMI in adolescents and young adults reveals additional insight into the effects of genetic loci over the life course.	23669352
Liu CT	2013	Genome-wide association of body fat distribution in African ancestry populations suggests new loci.	23966867
Pei YF	2013	Meta-analysis of genome-wide association data identifies novel susceptibility loci for obesity.	24064335
Namjou B	2013	EMR-linked GWAS study: investigation of variation landscape of loci for body mass index in children.	24348519
Wen W	2014	Meta-analysis of genome-wide association studies in East Asian-ancestry populations identifies four new loci for body mass index.	24861553
Scannell Bryan M	2014	Genome-wide association studies and heritability estimates of body mass index related phenotypes in Bangladeshi adults.	25133637
Shungin D	2015	New genetic loci link adipose and insulin biology to body fat distribution.	25673412
Locke AE	2015	Genetic studies of body mass index yield new insights for obesity biology.	25673413
Warrington NM	2015	A genome-wide association study of body mass index across early life and childhood.	25953783
Wilson CL	2015	Genetic and clinical factors associated with obesity among adult survivors of childhood cancer: A report from the St. Jude Lifetime Cohort.	25963547
Ahmad S	2015	A novel interaction between the FLJ33534 locus and smoking in obesity: a genome-wide study of 14 131 Pakistani adults.	26278006
Winkler TW	2015	The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study.	26426971
Sung YJ	2015	Genome-wide association studies suggest sex-specific loci associated with abdominal and visceral fat.	26480920
Wen W	2016	Genome-wide association studies in East Asians identify new loci for waist-hip ratio and waist circumference.	26785701

FIRST.AUTHOR	YEAR	STUDY	PUBMED
Lu Y	2016	New loci for body fat percentage reveal link between adiposity and cardiometabolic disease risk.	26833246
Wood AR	2016	Variants in the FTO and CDKAL1 loci have recessive effects on risk of obesity and type 2 diabetes, respectively.	26961502
Minster RL	2016	A thrifty variant in CREBRF strongly influences body mass index in Samoans.	27455349
Chu AY	2016	Multiethnic genome-wide meta-analysis of ectopic fat depots identifies loci associated with adipocyte development and differentiation.	27918534
McDonald MN	2017	Body mass index change in gastrointestinal cancer and chronic obstructive pulmonary disease is associated with Dedicator of Cytokinesis 1.	28044437
Pei YF	2017	Genomic variants at 20p11 associated with body fat mass in the European population.	28224759
Nagy R	2017	Exploration of haplotype research consortium imputation for genome-wide association studies in 20,032 Generation Scotland participants.	28270201
Chen G	2017	Genome-wide analysis identifies an african-specific variant in SEMA4D associated with body mass index.	28296344
Ng MCY	2017	Discovery and fine-mapping of adiposity loci using high density imputation of genome-wide association studies in individuals of African ancestry: African Ancestry Anthropometry Genetics Consortium.	28430825
Justice AE	2017	Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits.	28443625
Graff M	2017	Genome-wide physical activity interactions in adiposity - A meta-analysis of 200,452 adults.	28448500
Southam L	2017	Whole genome sequencing and imputation in isolated populations identify genetic associations with medically-relevant complex traits.	28548082
Tachmazidou I	2017	Whole-Genome Sequencing Coupled to Imputation Discovers Genetic Signals for Anthropometric Traits.	28552196
Akiyama M	2017	Genome-wide association study identifies 112 new loci for body mass index in the Japanese population.	28892062
Turcot V	2017	Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity.	29273807
Gong J	2017	Trans-ethnic analysis of metabochip data identifies two new loci associated with BMI.	29381148
Lee MR	2018	Causal association of body mass index with hypertension using a Mendelian randomization design.	30045251
Hoffmann TJ	2018	A Large Multi-ethnic Genome-Wide Association Study of Adult Body Mass Index Identifies Novel Loci.	30108127
Granot-Hershkovitz E	2018	A study of Kibbutzim in Israel reveals risk factors for cardiometabolic traits and subtle population structure.	30108283
Pulit SL	2018	Meta-analysis of genome-wide association studies for body fat distribution in 694,649 individuals of European ancestry.	30239722
Clifton EAD	2018	Genome-wide association study for risk taking propensity indicates shared pathways with body mass index.	30271922
Cha EDK	2018	Using Adipose Measures from Health Care Provider-Based Imaging Data for Discovery.	30363675

FIRST.AUTHOR	YEAR	STUDY	PUBMED
Emdin CA	2018	DNA Sequence Variation in ACVR1C Encoding the Activin Receptor-Like Kinase 7 Influences Body Fat Distribution and Protects Against Type 2 Diabetes.	30389748
Lotta LA	2018	Association of Genetic Variants Related to Gluteofemoral vs Abdominal Fat Distribution With Type 2 Diabetes, Coronary Disease, and Cardiovascular Risk Factors.	30575882
Hubel C	2018	Genomics of body fat percentage may contribute to sex bias in anorexia nervosa.	30593698
Kichaev G	2018	Leveraging Polygenic Functional Enrichment to Improve GWAS Power.	30595370
Rask-Andersen M	2019	Genome-wide association study of body fat distribution identifies adiposity loci and sex-specific genetic effects.	30664634
Riveros-McKay F	2019	Genetic architecture of human thinness compared to severe obesity.	30677029
Justice AE	2019	Protein-coding variants implicate novel genes related to lipid homeostasis contributing to body-fat distribution.	30778226
Jiao H	2019	Genome-Wide Interaction and Pathway Association Studies for Body Mass Index.	31118946
Wojcik GL	2019	Genetic analyses of diverse populations improves discovery for complex traits.	31217584
Wang H	2019	Genotype-by-environment interactions inferred from genetic effects on phenotypic variability in the UK Biobank.	31453325
Kusic DM	2019	rs11670527 Upstream of ZNF264 Associated with Body Mass Index in the Coriell Personalized Medicine Collaborative.	31498392
Jeon S	2019	Structural equation modeling for hypertension and type 2 diabetes based on multiple SNPs and multiple phenotypes.	31513605
Helgeland O	2019	Genome-wide association study reveals dynamic role of genetic variation in infant and early childhood growth.	31575865
Zhu Z	2019	Shared Genetic and Experimental Links between Obesity-Related Traits and Asthma Subtypes in UK Biobank.	31669095
Gurdasani D	2019	Uganda Genome Resource Enables Insights into Population History and Genomic Discovery in Africa.	31675503
Costa-Urrutia P	2019	Genome-Wide Association Study of Body Mass Index and Body Fat in Mexican-Mestizo Children.	31752434
Couto Alves A	2019	GWAS on longitudinal growth traits reveals different genetic factors influencing infant, child, and adult BMI.	31840077
Chiang KM	2019	Genome-wide association study of morbid obesity in Han Chinese.	31852448
Schlauch KA	2019	A Comprehensive Genome-Wide and Phenome-Wide Examination of BMI and Obesity in a Northern Nevadan Cohort.	31888951
Pei YF	2020	Bivariate genome-wide association analysis identified three pleiotropic loci underlying osteoporosis and obesity.	31903547
Lind L	2020	Genetic Determinants of Clustering of Cardiometabolic Risk Factors in U.K. Biobank.	31928498
Andersen MK	2020	The derived allele of a novel intergenic variant at chromosome 11 associates with lower body mass index and a favorable metabolic phenotype in Greenlanders.	31978080
Giri AK	2020	Multifaceted genome-wide study identifies novel regulatory loci in SLC22A11 and ZNF45 for body mass index in Indians.	32363570

FIRST.AUTHOR	YEAR	STUDY	PUBMED
Salinas YD	2020	Discovery and Mediation Analysis of Cross-Phenotype Associations With Asthma and Body Mass Index in 12q13.2.	32700739
Wei XT	2020	Pleiotropic genomic variants at 17q21.31 associated with bone mineral density and body fat mass: a bivariate genome-wide association analysis.	32963334
Richard MA	2020	Genetic variation in the body mass index of adult survivors of childhood acute lymphoblastic leukemia: A report from the Childhood Cancer Survivor Study and the St. Jude Lifetime Cohort.	33048379
Ahn Y	2020	Identification of Genetic Variants for Female Obesity and Evaluation of the Causal Role of Genetically Defined Obesity in Polycystic Ovarian Syndrome.	33209044
Li Z	2021	Bivariate genome-wide association study (GWAS) of body mass index and blood pressure phenotypes in northern Chinese twins.	33539483
Huang LO	2021	Genome-wide discovery of genetic loci that uncouple excess adiposity from its comorbidities.	33619380
Jung HU	2021	Identification of genetic loci affecting body mass index through interaction with multiple environmental factors using structured linear mixed model.	33654129
Fjukstad KK	2021	Genetic variants associated with cardiometabolic abnormalities during treatment with selective serotonin reuptake inhibitors: a genome-wide association study.	33824429
Lee S	2021	Novel recessive locus for body mass index in childhood asthma.	33888571
Zhuang Z	2021	Shared genetic etiology and causality between body fat percentage and cardiovascular diseases: a large-scale genome-wide cross-trait analysis.	33910581
Martin S	2021	Genetic evidence for different adiposity phenotypes and their opposing influence on ectopic fat and risk of cardiometabolic disease.	33980691
Christakoudi S	2021	GWAS of allometric body-shape indices in UK Biobank identifies loci suggesting associations with morphogenesis, organogenesis, adrenal cell renewal and cancer.	34021172
Cho HW	2021	A Genome-Wide Association Study of Novel Genetic Variants Associated With Anthropometric Traits in Koreans.	34054925
Wan JY	2021	Genome-wide association analysis of metabolic syndrome quantitative traits in the GENNID multiethnic family study.	34074324
Liu Y	2021	Genetic architecture of 11 organ traits derived from abdominal MRI using deep learning.	34128465
Akbari P	2021	Sequencing of 640,000 exomes identifies GPR75 variants associated with protection from obesity.	34210852
Barton AR	2021	Whole-exome imputation within UK Biobank powers rare coding variant association and fine-mapping analyses.	34226706
Livingstone KM	2021	Discovery Genome-Wide Association Study of Body Composition in 4,386 Adults From the UK Biobank's Pilot Imaging Enhancement Study.	34239500
Sakaue S	2021	A cross-population atlas of genetic associations for 220 human phenotypes.	34594039
Park S	2021	Interactions between Polygenic Risk Scores, Dietary Pattern, and Menarche Age with the Obesity Risk in a Large Hospital-Based Cohort.	34836030

FIRST.AUTHOR	YEAR	STUDY	PUBMED
Wong HS	2022	Genome-wide association study identifies genetic risk loci for adiposity in a Taiwanese population.	35051171
Wood AC	2022	Identification of genetic loci simultaneously associated with multiple cardiometabolic traits.	35168826
Helgeland O	2022	Characterization of the genetic architecture of infant and early childhood body mass index.	35315439
Fernandez-Rhodes L	2022	Ancestral diversity improves discovery and fine-mapping of genetic loci for anthropometric traits-The Hispanic/Latino Anthropometry Consortium.	35399580
Wang SH	2022	Causality of abdominal obesity on cognition: a trans-ethnic Mendelian randomization study.	35538205
Chung W	2022	Bayesian analysis of longitudinal traits in the Korea Association Resource (KARE) cohort.	35794696
Huang QQ	2022	Transferability of genetic loci and polygenic scores for cardiometabolic traits in British Pakistani and Bangladeshi individuals.	35945198
Akbari P	2022	Multiancestry exome sequencing reveals INHBE mutations associated with favorable fat distribution and protection from diabetes.	35999217
Lee CJ	2022	Phenome-wide analysis of Taiwan Biobank reveals novel glycemia-related loci and genetic risks for diabetes.	36329257

APPENDIX 2: PREVIOUSLY REPORTED SNPS ASSOCIATED WITH BOTH ADIPOSITY AND CARDIOMETABOLIC PROFILE

Study (PMID)	Category [1)]	SNP	CHR	BP (GRCh37)	EA	OA	Nearest genes	Bivariate (lipid + adiposity) loci [2)]
25048195	Protective	rs4846565	1	219722104	A	G	*LYPLAL1*	
25048195	Protective	rs10195252	2	165513091	C	T	*GRB14*	
25048195	Protective	rs2943645	2	227099180	C	T	*IRS1*	
25048195	Protective	rs17036328	3	12390484	C	T	*PPARG*	
25048195	Protective	rs3822072	4	89741269	G	A	*FAM13A*	
25048195	Protective	rs974801	4	106071064	A	G	*TET2*	
25048195	Protective	rs6822892	4	157734675	G	A	*PDGFC*	
25048195	Protective	rs4865796	5	53272664	G	A	*ARL15*	
25048195	Protective	rs459193	5	55806751	A	G	*ANKRD55*	
25048195	Protective	rs2745353	6	127452935	C	T	*RSPO3*	
25048195	Protective	rs731839	19	33899065	A	G	*PEPD*	
27841877	Protective	rs683135	1	39895460	A	G	*MACF1*	
27841877	Protective	rs17386142	1	50815783	C	T	*DMRTA2*	
27841877	Protective	rs11577194	1	110500175	T	C	*CSF1*	
27841877	Protective	rs9425291	1	172312769	A	G	*DNM3*	
27841877	Protective	rs4846565	1	219722104	G	A	*RNU5F-1/LYPLAL1*	
27841877	Protective	rs2249105	2	65287896	A	G	*CEP68*	
27841877	Protective	rs10195252	2	165513091	T	C	*COBLL1/GRB14*	
27841877	Protective	rs492400	2	219349752	T	C	*USP37*	
27841877	Protective	rs2943645	2	227099180	T	C	*IRS1*	
27841877	Protective	rs308971	3	12116620	G	A	*SYN2/PPARG*	
27841877	Protective	rs3864041	3	15185634	T	C	*COL6A4P1*	
27841877	Protective	rs295449	3	47375955	A	G	*KLHL18*	
27841877	Protective	rs11130329	3	52896855	A	C	*TMEM110-MUSTN1*	
27841877	Protective	rs9881942	3	123082416	A	G	*ADCY5*	
27841877	Protective	rs645040	3	135926622	T	G	*MSL2*	
27841877	Protective	rs2699429	4	3480136	C	T	*DOK7*	

Study (PMID)	Category [1)]	SNP	CHR	BP (GRCh37)	EA	OA	Nearest genes	Bivariate (lipid + adiposity) loci [2)]
27841877	Protective	rs3822072	4	89741269	A	G	*FAM13A*	
27841877	Protective	rs6822892	4	157734675	A	G	*PDGFC*	
27841877	Protective	rs4865796	5	53272664	A	G	*ARL15/FST*	
27841877	Protective	rs459193	5	55806751	G	A	*ANKRD55*	
27841877	Protective	rs4976033	5	67714246	G	A	*PIK3R1*	
27841877	Protective	rs6887914	5	112711486	C	T	*MCC*	
27841877	Protective	rs1045241	5	118729286	C	T	*TNFAIP8*	
27841877	Protective	rs2434612	5	158022041	G	A	*EBF1*	
27841877	Protective	rs966544	5	173350405	G	A	*CPEB4*	
27841877	Protective	rs12525532	6	35004819	T	C	*ANKS1A*	
27841877	Protective	rs6937438	6	43815364	A	G	*LOC100132354*	
27841877	Protective	rs2745353	6	127452935	T	C	*RSPO3*	
27841877	Protective	rs9492443	6	130398731	C	T	*L3MBTL3*	
27841877	Protective	rs3861397	6	139828916	G	A	*LOC645434*	
27841877	Protective	rs17169104	7	15883727	G	C	*MEOX2*	
27841877	Protective	rs972283	7	130466854	G	A	*KLF14*	
27841877	Protective	rs2126259	8	9185146	T	C	*PPP1R3B*	
27841877	Protective	rs1011685	8	19830769	C	T	*LPL*	
27841877	Protective	rs4738141	8	72469742	G	A	*EYA1*	
27841877	Protective	rs7005992	8	126528955	C	G	*TRIB1*	
27841877	Protective	rs498313	9	78034169	A	G	*MIR548H3*	
27841877	Protective	rs10995441	10	64869239	G	T	*NRBF2*	
27841877	Protective	rs11231693	11	63862612	A	G	*MACROD1*	
27841877	Protective	rs17402950	12	14571671	G	A	*ATF7IP*	
27841877	Protective	rs718314	12	26453283	G	A	*ITPR2*	
27841877	Protective	rs7973683	12	124449223	C	A	*CCDC92/DNAH10*	
27841877	Protective	rs7323406	13	111628195	A	G	*ANKRD10*	
27841877	Protective	rs7176058	15	39464167	A	G	*C15orf54*	
27841877	Protective	rs8032586	15	73081067	C	T	*LOC100287559*	

Study (PMID)	Category [1)]	SNP	CHR	BP (GRCh37)	EA	OA	Nearest genes	Bivariate (lipid + adiposity) loci [2)]
27841877	Protective	rs754814	17	4657034	T	C	*ZMYND15*	
27841877	Protective	rs7227237	18	47174679	C	T	*LIPG*	
27841877	Protective	rs8101064	19	7293119	T	C	*INSR*	
27841877	Protective	rs4804833	19	7970635	A	G	*MAP2K7*	
27841877	Protective	rs4804311	19	8615589	A	G	*MYO1F*	
27841877	Protective	rs731839	19	33899065	G	A	*PEPD*	
27841877	Protective	rs6066149	20	45602638	G	A	*EYA2*	
27841877	Protective	rs132985	22	38563471	C	T	*PLA2G6*	
30352878	Protective	rs11118306	1	219627486	A	G	*LYPLAL1, SLC30A10*	
30352878	Protective	rs13389219	2	165528876	T	C	*GRB14, COBLL1*	
30352878	Protective	rs2943653	2	227047771	C	T	*NYAP2, IRS1*	
30352878	Protective	rs1801282	3	12393125	G	C	*PPARG*	
30352878	Protective	rs2276936	4	89726283	A	C	*FAM13A*	
30352878	Protective	rs40271	5	55796319	C	T	*ANKRD55, MAP3K1*	
30352878	Protective	rs998584	6	43757896	C	A	*VEGFA, C6orf223*	
30352878	Protective	rs632057	6	139834012	G	T	*CITED2*	
30352878	Protective	rs972283	7	130466854	A	G	*KLF14, MKLN1*	
30352878	Protective	rs2980888	8	126507308	C	T	*TRIB1*	
30352878	Protective	rs11045172	12	20470221	C	A	*AEBP2, PDE3A*	
30352878	Protective	rs7133378	12	124409502	A	G	*DNAH10*	
30352878	Protective	rs7258937	19	33938800	T	C	*PEPD*	
30352878	Protective	rs2267373	22	38600542	C	T	*MAFF*	
33619380	Protective	rs1010447	1	11269795	T	C	*MTOR*	
33619380	Protective	rs3789588	1	51266522	A	G	*FAF1*	
33619380	Protective	rs6603981	1	92993806	C	T	*EVI5*	LDL
33619380	Protective	rs2820446	1	219748817	G	C	*ZC3H11B-LYPLAL1*	

Study (PMID)	**Category** [1)]	**SNP**	**CHR**	**BP (GRCh37)**	**EA**	**OA**	**Nearest genes**	**Bivariate (lipid + adiposity) loci** [2)]
33619380	Protective	rs1260326	2	27730939	C	T	*GCKR*	LDL;TG
33619380	Protective	rs4988235	2	136608645	A	G	*MCM6*	LDL
33619380	Protective	rs1128249	2	165528623	T	G	*GRB14-COBLL1*	HDL;LDL;TG
33619380	Protective	rs1427445	2	219555572	A	C	*STK36*	TG
33619380	Protective	rs2943652	2	227108445	C	T	*NYAP2-IRS1*	HDL;TG
33619380	Protective	rs11563251	2	234679383	C	T	*UGT1A8:UGT1A10:UGT1A9:UGT1A7:UGT1A6:UGT1A5:UGT1A4:UGT1A3:UGT1A1*	LDL
33619380	Protective	rs2881654	3	12396954	A	G	*PPARG*	
33619380	Protective	rs4392441	3	48077700	T	C	*MAP4*	
33619380	Protective	rs4616635	3	64702274	G	C	*ADAMTS9-AS2*	
33619380	Protective	rs11708067	3	123065777	G	A	*ADCY5*	
33619380	Protective	rs4481184	3	185505786	C	T	*IGF2BP2*	
33619380	Protective	rs4234589	3	185818881	A	G	*ETV5*	HDL
33619380	Protective	rs3822072	4	89741268	G	A	*FAM13A*	HDL;TG
33619380	Protective	rs13107325	4	103188708	T	C	*SLC39A8*	
33619380	Protective	rs3776717	5	53298761	G	A	*ARL15*	HDL
33619380	Protective	rs459193	5	55806750	A	G	*ANKRD55-MAP3K1*	HDL;TG
33619380	Protective	rs9686661	5	55861785	C	T	*ANKRD55-MAP3K1*	HDL;TG
33619380	Protective	rs4976033	5	67714245	A	G	*PIK3R1*	HDL
33619380	Protective	rs7713317	5	95716721	G	A	*PCSK1*	
33619380	Protective	rs2434612	5	158022040	A	G	*EBF1*	
33619380	Protective	rs6861681	5	173362457	G	A	*CPEB4*	
33619380	Protective	rs3094222	6	31081433	A	G	*C6orf15-PSORS1C1*	
33619380	Protective	rs998584	6	43757895	C	A	*VEGFA*	HDL;TG
33619380	Protective	rs9385400	6	126764189	T	G	*CENPW*	
33619380	Protective	rs17080091	6	150997400	T	C	*PLEKHG1*	

Study (PMID)	Category [1)]	SNP	CHR	BP (GRCh37)	EA	OA	Nearest genes	Bivariate (lipid + adiposity) loci [2)]
33619380	Protective	rs539958	6	160772841	T	C	*SLC22A3*	LDL;TG
33619380	Protective	rs702485	7	6449271	G	A	*DAGLB*	
33619380	Protective	rs864745	7	28180555	T	C	*JAZF1*	
33619380	Protective	rs4731702	7	130433383	T	C	*KLF14*	HDL;TG
33619380	Protective	rs9987289	8	9183357	G	A	*PPP1R3B-TNKS*	
33619380	Protective	rs17149279	8	9195637	T	C	*PPP1R3B-TNKS*	HDL;TG
33619380	Protective	rs10090367	8	36825079	A	G	*KCNU1*	TG
33619380	Protective	rs7896600	10	12255174	G	C	*CDC123*	
33619380	Protective	rs10883832	10	104871278	G	T	*NT5C2*	
33619380	Protective	rs7903146	10	114758348	C	T	*TCF7L2*	
33619380	Protective	rs740746	10	115792786	G	A	*NHLRC2-ADRB1*	
33619380	Protective	rs7928810	11	17372442	A	C	*NCR3LG1*	
33619380	Protective	rs2845885	11	63869061	T	C	*MACROD1*	HDL;TG
33619380	Protective	rs11603334	11	72432984	A	G	*ARAP1*	
33619380	Protective	rs7134375	12	20473757	A	C	*PDE3A*	HDL
33619380	Protective	rs718314	12	26453282	A	G	*SSPN-ITPR2*	HDL
33619380	Protective	rs754133	12	54418919	A	G	*HOXC4-HOXC6*	
33619380	Protective	rs3741414	12	57844048	T	C	*INHBC*	HDL;TG
33619380	Protective	rs10774625	12	111910218	G	A	*ATXN2*	HDL
33619380	Protective	rs11057405	12	122781896	G	A	*CLIP1*	HDL
33619380	Protective	rs7133378	12	124409501	A	G	*CCDC92-DNAH10*	HDL;LDL;TG
33619380	Protective	rs863750	12	124505443	C	T	*FAM101A*	HDL;TG
33619380	Protective	rs17522122	14	33302881	T	G	*AKAP6*	
33619380	Protective	rs1378940	15	75083493	A	C	*CSK*	
33619380	Protective	rs879620	16	4015728	T	C	*ADCY9*	
33619380	Protective	rs4985155	16	15129458	A	G	*PDXDC1*	HDL;TG
33619380	Protective	rs2955617	17	7538784	A	C	*SHBG-ATP1B2*	HDL

Study (PMID)	Category [1)]	SNP	CHR	BP (GRCh37)	EA	OA	Nearest genes	Bivariate (lipid + adiposity) loci [2)]
33619380	Protective	rs6504872	17	45438951	C	T	*EFCAB13*	LDL
33619380	Protective	rs12454712	18	60845883	C	T	*BCL2*	HDL;TG
33619380	Protective	rs731839	19	33899064	A	G	*PEPD*	HDL;TG
33619380	Protective	rs2075650	19	45395618	A	G	*TOMM40*	HDL;LDL;TG
33619380	Protective	rs1211644	20	45592841	C	T	*EYA2*	TG
33619380	Protective	rs3761445	22	38595410	G	A	*PLA2G6-MAFF*	HDL;TG
33980691	Protective	rs2802774	1	203527812	A		*OPTC, ATP2B4*	
33980691	Protective	rs12130231	1	219631304	A		*LYPLAL1, SLC30A10*	
33980691	Protective	rs13389219	2	165528876	T		*GRB14, COBLL1*	
33980691	Protective	rs2943653	2	227047771	C		*NYAP2, IRS1*	
33980691	Protective	rs4684847	3	12386337	T		*SYN2, PPARG*	
33980691	Protective	rs9851766	3	138121509	A		*MRAS*	
33980691	Protective	rs62271373	3	150066540	T		*PFN2, TSC22D2*	
33980691	Protective	rs4450871	4	4990298	G		*MSX1, CYTL1*	
33980691	Protective	rs13132853	4	38680015	A		*KLF3*	
33980691	Protective	rs987469	4	89706643	C		*FAM13A*	
33980691	Protective	rs30351	5	55794632	G		*ANKRD55*	
33980691	Protective	rs4976033	5	67714246	A		*PIK3R1, SLC30A5*	
33980691	Protective	rs9764678	5	118726662	C		*TNFAIP8*	
33980691	Protective	rs11135038	5	157930133	G		*CLINT1, EBF1*	
33980691	Protective	rs998584	6	43757896	C		*VEGFA, C6orf223*	
33980691	Protective	rs72959041	6	127454893	G		*RSPO3*	
33980691	Protective	rs573454216	6	139837429	A		*CITED2*	
33980691	Protective	rs972283	7	130466854	A		*KLF14, MKLN1*	
33980691	Protective	rs6977416	7	150542711	G		*TMEM176A, ABP1*	

Study (PMID)	Category [1)]	SNP	CHR	BP (GRCh37)	EA	OA	Nearest genes	Bivariate (lipid + adiposity) loci [2)]
33980691	Protective	rs12681990	8	36859186	T		*KCNU1, ZNF703*	
33980691	Protective	rs2980888	8	126504383	C		*TRIB1*	
33980691	Protective	rs113222038	11	62380027	C		*EML3*	
33980691	Protective	rs11045172	12	20470221	C		*AEBP2, PDE3A*	
33980691	Protective	rs10876529	12	54421810	C		*HOXC8, HOXC6*	
33980691	Protective	rs12369179	12	122963550	C		*ZCCHC8*	
33980691	Protective	rs7133378	12	124409502	A		*DNAH10*	
33980691	Protective	rs72697297	14	93069989	T		*RIN3*	
33980691	Protective	rs12441543	15	31689543	A		*KLF13, OTUD7A*	
33980691	Protective	rs12940684	17	7453919	C		*TNFSF12, TNFSF13*	
33980691	Protective	rs142186653	17	73879851	C		*TRIM47, TRIM65*	
33980691	Protective	rs11664106	18	2846812	T		*SMCHD1, EMILIN2*	
33980691	Protective	rs7233512	18	42595076	G		*SETBP1*	
33980691	Protective	rs7258937	19	33938800	T		*PEPD*	
33980691	Protective	rs555162510	19	46183031	A		*NA*	
33980691	Protective	rs6029180	20	39178923	G		*MAFB*	
33980691	Protective	rs4821764	22	38599364	G		*MAFF*	
33980691	Unfavorable	1:72767554_CA_C	1	72767554	CA		*NA*	
33980691	Unfavorable	rs71658797	1	77967507	A		*AK5*	
33980691	Unfavorable	1:113202203_TCTCTC_T	1	113202203	TC TC TC		*NA*	
33980691	Unfavorable	rs539515	1	177889025	C		*FAM5B, SEC16B*	
33980691	Unfavorable	rs11122450	1	230301811	T		*GALNT2*	
33980691	Unfavorable	rs143684747	2	633053	AC		*NA*	

Study (PMID)	Category [1)]	SNP	CHR	BP (GRCh37)	EA	OA	Nearest genes	Bivariate (lipid + adiposity) loci [2)]
33980691	Unfavorable	rs6752378	2	25150116	A		*ADCY3, DNAJC27*	
33980691	Unfavorable	rs1471740	3	136328270	C		*STAG1*	
33980691	Unfavorable	rs10938397	4	45182527	G		*GNPDA2, GABRG1*	
33980691	Unfavorable	rs13107325	4	103188709	T		*SLC39A8*	
33980691	Unfavorable	rs2112347	5	75015242	T		*POC5, SV2C*	
33980691	Unfavorable	5:87969925_CGG_C	5	87969925	C		*TMEM161B, MEF2C*	
33980691	Unfavorable	rs10623997	5	107478679	T		*NA*	
33980691	Unfavorable	rs17764730	5	127357526	C		*CTXN3, SLC12A2*	
33980691	Unfavorable	rs9358912	6	26211146	G		*HIST1H4E, HIST1H2BG*	
33980691	Unfavorable	6:34650934_CGT_C	6	34650934	C		*C6orf106*	
33980691	Unfavorable	rs72892910	6	50816887	T		*TFAP2B, PKHD1*	
33980691	Unfavorable	rs236660	7	75050086	C		*NA*	
33980691	Unfavorable	rs4876611	8	116671848	G		*TRPS1*	
33980691	Unfavorable	rs10756713	9	15880555	A		*CCDC171*	
33980691	Unfavorable	rs2274224	10	96039597	G		*PLCE1*	
33980691	Unfavorable	rs61888762	11	27709630	G		*BDNF*	
33980691	Unfavorable	rs4755725	11	43637975	C		*NA*	
33980691	Unfavorable	rs7124681	11	47529947	A		*CELF1, PTPMT1*	
33980691	Unfavorable	rs7132908	12	50263148	A		*FAIM2*	
33980691	Unfavorable	rs3764002	12	108618630	C		*WSCD2*	
33980691	Unfavorable	14:79940130_TAGGAGTTTTTCCAGATCATTAGCCACTTATACGGAG_T	14	79940130	T		*NA*	
33980691	Unfavorable	rs4776985	15	68123021	T		*SKOR1*	

Study (PMID)	Category [1)]	SNP	CHR	BP (GRCh37)	EA	OA	Nearest genes	Bivariate (lipid + adiposity) loci [2)]
33980691	Unfavorable	15:73322940_AT_A	15	73322940	A		*NA*	
33980691	Unfavorable	rs6602997	15	84521398	T		*ADAMTSL3*	
33980691	Unfavorable	rs56186137	16	28825953	G		*NPIPL1*	
33980691	Unfavorable	rs11642015	16	53802494	T		*FTO*	
33980691	Unfavorable	rs8049669	16	69551467	A		*CYB5B, NFAT5*	
33980691	Unfavorable	rs4790292	17	1824305	C		*RPA1, RTN4RL1*	
33980691	Unfavorable	rs55931203	17	65854602	T		*BPTF*	
33980691	Unfavorable	rs771025058	18	21122207	AA	G	*NPC1*	
33980691	Unfavorable	rs6567160	18	57829135	C		*PMAIP1, MC4R*	
33980691	Unfavorable	rs11666808	19	18383506	T		*KIAA1683*	

1) Four of five studies focused on the adiposity-related genetic variants simultaneously associated with protective cardiometabolic profile, wheareas one study (PMID: 33980691) investigated both protective and unfavorable adiposity variants.

2) One study (PMID: 33619380) explicitly identified bivariate (a pair of one adiposity trait and one cardiometabolic trait) genetic variants, and this column indicated the genetic variants identified by bivaraite analyses for pairs of adiposity and lipid traits.

REFERENCES

1. Collaboration NCDRF. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128.9 million children, adolescents, and adults. *Lancet.* 2017;390(10113):2627-2642.

2. World Health Organization. Obesity and overweight. https://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight. Accessed July 19, 2022.

3. Prospective Studies C, Whitlock G, Lewington S, et al. Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies. *Lancet.* 2009;373(9669):1083-1096.

4. Loos RJF, Yeo GSH. The genetics of obesity: from discovery to biology. *Nat Rev Genet.* 2022;23(2):120-133.

5. Maes HH, Neale MC, Eaves LJ. Genetic and environmental factors in relative body weight and human adiposity. *Behav Genet.* 1997;27(4):325-351.

6. Lanigan J. Prevention of overweight and obesity in early life. *Proc Nutr Soc.* 2018;77(3):247-256.

7. Dashti HS, Hivert MF, Levy DE, McCurley JL, Saxena R, Thorndike AN. Polygenic risk score for obesity and the quality, quantity, and timing of workplace food purchases: A secondary analysis from the ChooseWell 365 randomized trial. *PLoS Med.* 2020;17(7):e1003219.

8. Ding M, Ellervik C, Huang T, et al. Diet quality and genetic association with body mass index: results from 3 observational studies. *Am J Clin Nutr.* 2018;108(6):1291-1300.

9. Day FR, Loos RJ. Developments in obesity genetics in the era of genome-wide association studies. *J Nutrigenet Nutrigenomics.* 2011;4(4):222-238.

10. Goodarzi MO. Genetics of obesity: what genetic association studies have taught us about the biology of obesity and its complications. *Lancet Diabetes Endocrinol.* 2018;6(3):223-236.

11. Gutierrez-Cuevas J, Santos A, Armendariz-Borunda J. Pathophysiological Molecular Mechanisms of Obesity: A Link between MAFLD and NASH with Cardiovascular Diseases. *Int J Mol Sci.* 2021;22(21).

12. Huang LO, Rauch A, Mazzaferro E, et al. Genome-wide discovery of genetic loci that uncouple excess adiposity from its comorbidities. *Nat Metab.* 2021;3(2):228-243.

13. Kilpelainen TO, Zillikens MC, Stancakova A, et al. Genetic variation near IRS1 associates with reduced adiposity and an impaired metabolic profile. *Nat Genet.* 2011;43(8):753-760.

14. Lotta LA, Gulati P, Day FR, et al. Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. *Nat Genet.* 2017;49(1):17-26.

15. Lu Y, Day FR, Gustafsson S, et al. New loci for body fat percentage reveal link between adiposity and cardiometabolic disease risk. *Nat Commun.* 2016;7:10495.

16. Martin S, Cule M, Basty N, et al. Genetic Evidence for Different Adiposity Phenotypes and Their Opposing Influences on Ectopic Fat and Risk of Cardiometabolic Disease. *Diabetes.* 2021.

17. Scott RA, Fall T, Pasko D, et al. Common genetic variants highlight the role of insulin resistance and body fat distribution in type 2 diabetes, independent of obesity. *Diabetes.* 2014;63(12):4378-4387.

18. Yaghootkar H, Lotta LA, Tyrrell J, et al. Genetic Evidence for a Link Between Favorable Adiposity and Lower Risk of Type 2 Diabetes, Hypertension, and Heart Disease. *Diabetes.* 2016;65(8):2448-2460.

19. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience.* 2019;8(7).

20. Ge T, Chen CY, Ni Y, Feng YA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun.* 2019;10(1):1776.

21. Ruan Y, Lin YF, Feng YA, et al. Improving polygenic prediction in ancestrally diverse populations. *Nat Genet.* 2022;54(5):573-580.

22. Freedman DS, Horlick M, Berenson GS. A comparison of the Slaughter skinfold-thickness equations and BMI in predicting body fatness and cardiovascular disease risk factor levels in children. *Am J Clin Nutr.* 2013;98(6):1417-1424.

23. Wohlfahrt-Veje C, Tinggaard J, Winther K, et al. Body fat throughout childhood in 2647 healthy Danish children: agreement of BMI, waist circumference, skinfolds with dual X-ray absorptiometry. *Eur J Clin Nutr.* 2014;68(6):664-670.

24. Centers for Disease Control and Prevention. Defining Adult Overweight and Obesity. https://www.cdc.gov/obesity/basics/adult-defining.html. Accessed August 3, 2022.

25. Collaborators GBDO, Afshin A, Forouzanfar MH, et al. Health Effects of Overweight and Obesity in 195 Countries over 25 Years. *N Engl J Med.* 2017;377(1):13-27.

26. Ward ZJ, Bleich SN, Long MW, Gortmaker SL. Association of body mass index with health care expenditures in the United States by age and sex. *PLoS One.* 2021;16(3):e0247307.

27. Centers for Disease Control and Prevention. Adult Obesity Facts https://www.cdc.gov/obesity/data/adult.html. Accessed August 3, 2022.

28. Stierman BA, Joseph;Carroll, Margaret D.;Chen, Te-Ching;Davy, Orlando;Fink, Steven;Fryar, Cheryl D.;Gu, Qiuping;Hales, Craig M.;Hughes, Jeffery P.;Ostchega, Yechiam;Storandt, Renee J.;Akinbami, Lara J.;. National Health and Nutrition Examination Survey 2017–March 2020 Prepandemic Data Files Development of Files and Prevalence Estimates for Selected Health Outcomes. *National Health Statistics Reports.* 2021(NHSR No. 158).

29. Carbone S, Lavie CJ, Arena R. Obesity and Heart Failure: Focus on the Obesity Paradox. *Mayo Clin Proc.* 2017;92(2):266-279.

30. Fox CS, Massaro JM, Hoffmann U, et al. Abdominal visceral and subcutaneous adipose tissue compartments: association with metabolic risk factors in the Framingham Heart Study. *Circulation.* 2007;116(1):39-48.

31. Wildman RP, Muntner P, Reynolds K, et al. The obese without cardiometabolic risk factor clustering and the normal weight with cardiometabolic risk factor clustering: prevalence and correlates of 2 phenotypes among the US population (NHANES 1999-2004). *Arch Intern Med.* 2008;168(15):1617-1624.

32. Hu FB. Measurements of adiposity and body composition. *Obesity epidemiology.* 2008;416:53-83.

33. Chooi YC, Ding C, Magkos F. The epidemiology of obesity. *Metabolism.* 2019;92:6-10.

34. Malik VS, Willett WC, Hu FB. Global obesity: trends, risk factors and policy implications. *Nat Rev Endocrinol.* 2013;9(1):13-27.

35. Hruby A, Manson JE, Qi L, et al. Determinants and Consequences of Obesity. *Am J Public Health.* 2016;106(9):1656-1662.

36. Swinburn B, Egger G, Raza F. Dissecting obesogenic environments: the development and application of a framework for identifying and prioritizing environmental interventions for obesity. *Prev Med.* 1999;29(6 Pt 1):563-570.

37. French SA, Story M, Jeffery RW. Environmental influences on eating and physical activity. *Annu Rev Public Health.* 2001;22:309-335.

38. Swinburn B, Egger G. Preventive strategies against weight gain and obesity. *Obes Rev.* 2002;3(4):289-301.

39. Swinburn B, Egger G. The runaway weight gain train: too many accelerators, not enough brakes. *BMJ.* 2004;329(7468):736-739.

40. Gortmaker SL, Must A, Sobol AM, Peterson K, Colditz GA, Dietz WH. Television viewing as a cause of increasing obesity among children in the United States, 1986-1990. *Arch Pediatr Adolesc Med.* 1996;150(4):356-362.

41. Hebebrand J, Hinney A. Environmental and genetic risk factors in obesity. *Child Adolesc Psychiatr Clin N Am.* 2009;18(1):83-94.

42. Lamerz A, Kuepper-Nybelen J, Wehle C, et al. Social class, parental education, and obesity prevalence in a study of six-year-old children in Germany. *Int J Obes (Lond).* 2005;29(4):373-380.

43. Toschke AM, Koletzko B, Slikker W, Jr., Hermann M, von Kries R. Childhood obesity is associated with maternal smoking in pregnancy. *Eur J Pediatr.* 2002;161(8):445-448.

44. von Kries R, Toschke AM, Wurmser H, Sauerwald T, Koletzko B. Reduced risk for overweight and obesity in 5- and 6-y-old children by duration of sleep--a cross-sectional study. *Int J Obes Relat Metab Disord.* 2002;26(5):710-716.

45. Garvey WT. Is Obesity or Adiposity-Based Chronic Disease Curable: The Set Point Theory, the Environment, and Second-Generation Medications. *Endocr Pract.* 2022;28(2):214-222.

46. Berthoud HR, Morrison CD, Munzberg H. The obesity epidemic in the face of homeostatic body weight regulation: What went wrong and how can it be fixed? *Physiol Behav.* 2020;222:112959.

47. Berthoud HR, Munzberg H, Morrison CD. Blaming the Brain for Obesity: Integration of Hedonic and Homeostatic Mechanisms. *Gastroenterology.* 2017;152(7):1728-1738.

48. Elks CE, den Hoed M, Zhao JH, et al. Variability in the heritability of body mass index: a systematic review and meta-regression. *Front Endocrinol (Lausanne).* 2012;3:29.

49. Bouchard C. Genetics of Obesity: What We Have Learned Over Decades of Research. *Obesity (Silver Spring).* 2021;29(5):802-820.

50. Rokholm B, Silventoinen K, Angquist L, Skytthe A, Kyvik KO, Sorensen TI. Increased genetic variance of BMI with a higher prevalence of obesity. *PLoS One.* 2011;6(6):e20816.

51. Bouchard C, Perusse L, Leblanc C, Tremblay A, Theriault G. Inheritance of the amount and distribution of human body fat. *Int J Obes.* 1988;12(3):205-215.

52. Kansra AR, Lakkunarajah S, Jay MS. Childhood and Adolescent Obesity: A Review. *Front Pediatr.* 2020;8:581461.

53. Fairbrother U, Kidd E, Malagamuwa T, Walley A. Genetics of Severe Obesity. *Curr Diab Rep.* 2018;18(10):85.

54. Ayers KL, Glicksberg BS, Garfield AS, et al. Melanocortin 4 Receptor Pathway Dysfunction in Obesity: Patient Stratification Aimed at MC4R Agonist Treatment. *J Clin Endocrinol Metab.* 2018;103(7):2601-2612.

55. Clement K, Mosbah H, Poitou C. Rare genetic forms of obesity: From gene to therapy. *Physiol Behav.* 2020;227:113134.

56. Farooqi IS. Genetic, molecular and physiological insights into human obesity. *Eur J Clin Invest.* 2011;41(4):451-455.

57. Jequier E. Leptin signaling, adiposity, and energy balance. *Ann N Y Acad Sci.* 2002;967:379-388.

58. Zhang Y, Proenca R, Maffei M, Barone M, Leopold L, Friedman JM. Positional cloning of the mouse obese gene and its human homologue. *Nature.* 1994;372(6505):425-432.

59. Wasim M, Awan FR, Najam SS, Khan AR, Khan HN. Role of Leptin Deficiency, Inefficiency, and Leptin Receptors in Obesity. *Biochem Genet.* 2016;54(5):565-572.

60. Gibson WT, Farooqi IS, Moreau M, et al. Congenital leptin deficiency due to homozygosity for the Delta133G mutation: report of another case and evaluation of response to four years of leptin therapy. *J Clin Endocrinol Metab.* 2004;89(10):4821-4826.

61. Baxter J, Armijo PR, Flores L, Krause C, Samreen S, Tanner T. Updates on Monogenic Obesity in a Multifactorial Disease. *Obes Surg.* 2019;29(12):4077-4083.

62. Micioni Di Bonaventura E, Botticelli L, Tomassoni D, Tayebati SK, Micioni Di Bonaventura MV, Cifani C. The Melanocortin System behind the Dysfunctional Eating Behaviors. *Nutrients.* 2020;12(11).

63. Tao YX. The melanocortin-4 receptor: physiology, pharmacology, and pathophysiology. *Endocr Rev.* 2010;31(4):506-543.

64. Huszar D, Lynch CA, Fairchild-Huntress V, et al. Targeted disruption of the melanocortin-4 receptor results in obesity in mice. *Cell.* 1997;88(1):131-141.

65. Blevins JE, Morton GJ, Williams DL, et al. Forebrain melanocortin signaling enhances the hindbrain satiety response to CCK-8. *Am J Physiol Regul Integr Comp Physiol.* 2009;296(3):R476-484.

66. Hinney A, Korner A, Fischer-Posovszky P. The promise of new anti-obesity therapies arising from knowledge of genetic obesity traits. *Nat Rev Endocrinol.* 2022.

67. Frayling TM, Timpson NJ, Weedon MN, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science.* 2007;316(5826):889-894.

68. Albuquerque D, Nobrega C, Manco L, Padez C. The contribution of genetics and environment to obesity. *Br Med Bull.* 2017;123(1):159-173.

69. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature.* 2015;518(7538):197-206.

70. Khera AV, Chaffin M, Wade KH, et al. Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. *Cell.* 2019;177(3):587-596 e589.

71. Yang J, Bakshi A, Zhu Z, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat Genet.* 2015;47(10):1114-1120.

72. Speakman JR, Loos RJF, O'Rahilly S, Hirschhorn JN, Allison DB. GWAS for BMI: a treasure trove of fundamental insights into the genetic basis of obesity. *Int J Obes (Lond).* 2018;42(8):1524-1531.

73. Muller MJ, Geisler C, Blundell J, et al. The case of GWAS of obesity: does body weight control play by the rules? *Int J Obes (Lond).* 2018;42(8):1395-1405.

74. Yengo L, Sidorenko J, Kemper KE, et al. Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. *Hum Mol Genet.* 2018;27(20):3641-3649.

75. Loos RJ. The genetics of adiposity. *Curr Opin Genet Dev.* 2018;50:86-95.

76. Ghosh S, Bouchard C. Convergence between biological, behavioural and genetic determinants of obesity. *Nat Rev Genet.* 2017;18(12):731-748.

77. Turcot V, Lu Y, Highland HM, et al. Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. *Nat Genet.* 2018;50(1):26-41.

78. Akbari P, Gilani A, Sosina O, et al. Sequencing of 640,000 exomes identifies GPR75 variants associated with protection from obesity. *Science.* 2021;373(6550).

79. Shungin D, Winkler TW, Croteau-Chonka DC, et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature.* 2015;518(7538):187-196.

80. Pulit SL, Stoneman C, Morris AP, et al. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum Mol Genet.* 2019;28(1):166-174.

81. Justice AE, Karaderi T, Highland HM, et al. Protein-coding variants implicate novel genes related to lipid homeostasis contributing to body-fat distribution. *Nat Genet.* 2019;51(3):452-469.

82. Chu AY, Deng X, Fisher VA, et al. Multiethnic genome-wide meta-analysis of ectopic fat depots identifies loci associated with adipocyte development and differentiation. *Nat Genet.* 2017;49(1):125-130.

83. Livingstone KM, Tan MH, Abbott G, et al. Discovery Genome-Wide Association Study of Body Composition in 4,386 Adults From the UK Biobank's Pilot Imaging Enhancement Study. *Front Endocrinol (Lausanne).* 2021;12:692677.

84. Sung YJ, Perusse L, Sarzynski MA, et al. Genome-wide association studies suggest sex-specific loci associated with abdominal and visceral fat. *Int J Obes (Lond).* 2016;40(4):662-674.

85. Rask-Andersen M, Karlsson T, Ek WE, Johansson A. Genome-wide association study of body fat distribution identifies adiposity loci and sex-specific genetic effects. *Nat Commun.* 2019;10(1):339.

86. Young KL, Graff M, Fernandez-Rhodes L, North KE. Genetics of Obesity in Diverse Populations. *Curr Diab Rep.* 2018;18(12):145.

87. Akiyama M, Okada Y, Kanai M, et al. Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nat Genet.* 2017;49(10):1458-1467.

88. Wen W, Zheng W, Okada Y, et al. Meta-analysis of genome-wide association studies in East Asian-ancestry populations identifies four new loci for body mass index. *Hum Mol Genet.* 2014;23(20):5492-5504.

89. Ng MCY, Graff M, Lu Y, et al. Discovery and fine-mapping of adiposity loci using high density imputation of genome-wide association studies in individuals of African ancestry: African Ancestry Anthropometry Genetics Consortium. *PLoS Genet.* 2017;13(4):e1006719.

90. Gong J, Nishimura KK, Fernandez-Rhodes L, et al. Trans-ethnic analysis of metabochip data identifies two new loci associated with BMI. *Int J Obes (Lond).* 2018;42(3):384-390.

91. Wojcik GL, Graff M, Nishimura KK, et al. Genetic analyses of diverse populations improves discovery for complex traits. *Nature.* 2019;570(7762):514-518.

92. Fernandez-Rhodes L, Graff M, Buchanan VL, et al. Ancestral diversity improves discovery and fine-mapping of genetic loci for anthropometric traits-The Hispanic/Latino Anthropometry Consortium. *HGG Adv.* 2022;3(2):100099.

93. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature.* 2009;461(7265):747-753.

94. Genin E. Missing heritability of complex diseases: case solved? *Hum Genet.* 2020;139(1):103-113.

95. Wray NR, Lin T, Austin J, et al. From Basic Science to Clinical Application of Polygenic Risk Scores: A Primer. *JAMA Psychiatry.* 2021;78(1):101-109.

96. Lambert SA, Abraham G, Inouye M. Towards clinical utility of polygenic risk scores. *Hum Mol Genet.* 2019;28(R2):R133-R142.

97. Ma Y, Zhou X. Genetic prediction of complex traits with polygenic scores: a statistical review. *Trends Genet.* 2021;37(11):995-1011.

98. Kullo IJ, Lewis CM, Inouye M, Martin AR, Ripatti S, Chatterjee N. Polygenic scores in biomedical research. *Nat Rev Genet.* 2022.

99. Abraham G, Havulinna AS, Bhalala OG, et al. Genomic prediction of coronary heart disease. *Eur Heart J.* 2016;37(43):3267-3278.

100. Sharp SA, Rich SS, Wood AR, et al. Development and Standardization of an Improved Type 1 Diabetes Genetic Risk Score for Use in Newborn Screening and Incident Diagnosis. *Diabetes Care.* 2019;42(2):200-207.

101. Mavaddat N, Michailidou K, Dennis J, et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am J Hum Genet.* 2019;104(1):21-34.

102. Censin JC, Peters SAE, Bovijn J, et al. Causal relationships between obesity and the leading causes of death in women and men. *PLoS Genet.* 2019;15(10):e1008405.

103. Ni G, Zeng J, Revez JA, et al. A Comparison of Ten Polygenic Score Methods for Psychiatric Disorders Applied Across Multiple Cohorts. *Biol Psychiatry.* 2021;90(9):611-620.

104. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res.* 2007;17(10):1520-1528.

105. Prive F, Vilhjalmsson BJ, Aschard H, Blum MGB. Making the Most of Clumping and Thresholding for Polygenic Scores. *Am J Hum Genet.* 2019;105(6):1213-1221.

106. Prive F, Arbel J, Vilhjalmsson BJ. LDpred2: better, faster, stronger. *Bioinformatics.* 2020.

107. Lloyd-Jones LR, Zeng J, Sidorenko J, et al. Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nat Commun.* 2019;10(1):5086.

108. International Schizophrenia C, Purcell SM, Wray NR, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature.* 2009;460(7256):748-752.

109. Wang Y, Tsuo K, Kanai M, Neale BM, Martin AR. Challenges and Opportunities for Developing More Generalizable Polygenic Risk Scores. *Annu Rev Biomed Data Sci.* 2022;5:293-320.

110. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. *Bioinformatics.* 2015;31(9):1466-1468.

111. Zhang C, Ye Y, Zhao H. Comparison of Methods Utilizing Sex-Specific PRSs Derived From GWAS Summary Statistics. *Front Genet.* 2022;13:892950.

112. Choi SW, Mak TS, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc.* 2020;15(9):2759-2772.

113. Baker JL, Olsen LW, Sorensen TI. Childhood body-mass index and the risk of coronary heart disease in adulthood. *N Engl J Med.* 2007;357(23):2329-2337.

114. Loos RJF, Janssens A. Predicting Polygenic Obesity Using Genetic Information. *Cell Metab.* 2017;25(3):535-543.

115. Murthy VL, Xia R, Baldridge AS, et al. Polygenic Risk, Fitness, and Obesity in the Coronary Artery Risk Development in Young Adults (CARDIA) Study. *Jama Cardiol.* 2020;5(3):40-48.

116. Peterson RE, Maes HH, Holmans P, et al. Genetic risk sum score comprised of common polygenic variation is associated with body mass index. *Hum Genet.* 2011;129(2):221-230.

117. Shi M, Chen W, Sun X, et al. Association of Genome-Wide Polygenic Risk Score for Body Mass Index With Cardiometabolic Health From Childhood Through Midlife. *Circ Genom Precis Med.* 2022;15(4):e003375.

118. Hollands GJ, French DP, Griffin SJ, et al. The impact of communicating genetic risks of disease on risk-reducing health behaviour: systematic review with meta-analysis. *BMJ.* 2016;352:i1102.

119. Meisel SF, Beeken RJ, van Jaarsveld CH, Wardle J. Genetic susceptibility testing and readiness to control weight: Results from a randomized controlled trial. *Obesity (Silver Spring).* 2015;23(2):305-312.

120. Rimm EB, Stampfer MJ, Giovannucci E, et al. Body size and fat distribution as predictors of coronary heart disease among middle-aged and older US men. *Am J Epidemiol.* 1995;141(12):1117-1127.

121. Havlik RJ, Hubert HB, Fabsitz RR, Feinleib M. Weight and hypertension. *Ann Intern Med.* 1983;98(5 Pt 2):855-859.

122. West KM, Kalbfleisch JM. Influence of nutritional factors on prevalence of diabetes. *Diabetes.* 1971;20(2):99-108.

123. Denke MA, Sempos CT, Grundy SM. Excess body weight. An underrecognized contributor to high blood cholesterol levels in white American men. *Arch Intern Med.* 1993;153(9):1093-1103.

124. Kim MS, Kim WJ, Khera AV, et al. Association between adiposity and cardiovascular outcomes: an umbrella review and meta-analysis of observational and Mendelian randomization studies. *Eur Heart J.* 2021;42(34):3388-3403.

125. Diaz-Melean CM, Somers VK, Rodriguez-Escudero JP, et al. Mechanisms of adverse cardiometabolic consequences of obesity. *Curr Atheroscler Rep.* 2013;15(11):364.

126. Trayhurn P, Wang B, Wood IS. Hypoxia in adipose tissue: a basis for the dysregulation of tissue function in obesity? *Br J Nutr.* 2008;100(2):227-235.

127. Hotamisligil GS. Endoplasmic reticulum stress and the inflammatory basis of metabolic disease. *Cell.* 2010;140(6):900-917.

128. Nguyen MT, Favelyukis S, Nguyen AK, et al. A subpopulation of macrophages infiltrates hypertrophic adipose tissue and is activated by free fatty acids via Toll-like receptors 2 and 4 and JNK-dependent pathways. *J Biol Chem.* 2007;282(48):35279-35292.

129. Veljic I, Polovina M, Seferovic JP, Seferovic PM. Adipokine profile as a novel screening method for cardiometabolic disease: Help or hindrance? *Eur J Prev Cardiol.* 2018;25(14):1543-1547.

130. Perez-Hernandez AI, Catalan V, Gomez-Ambrosi J, Rodriguez A, Fruhbeck G. Mechanisms linking excess adiposity and carcinogenesis promotion. *Front Endocrinol (Lausanne).* 2014;5:65.

131. Singh P, Hoffmann M, Wolk R, Shamsuzzaman AS, Somers VK. Leptin induces C-reactive protein expression in vascular endothelial cells. *Arterioscler Thromb Vasc Biol.* 2007;27(9):e302-307.

132. Bravo PE, Morse S, Borne DM, Aguilar EA, Reisin E. Leptin and hypertension in obesity. *Vasc Health Risk Manag.* 2006;2(2):163-169.

133. Westerbacka J, Vehkavaara S, Bergholm R, Wilkinson I, Cockcroft J, Yki-Jarvinen H. Marked resistance of the ability of insulin to decrease arterial stiffness characterizes human obesity. *Diabetes.* 1999;48(4):821-827.

134. Arcaro G, Zamboni M, Rossi L, et al. Body fat distribution predicts the degree of endothelial dysfunction in uncomplicated obesity. *Int J Obes Relat Metab Disord.* 1999;23(9):936-942.

135. Ghosh A, Gao L, Thakur A, Siu PM, Lai CWK. Role of free fatty acids in endothelial dysfunction. *J Biomed Sci.* 2017;24(1):50.

136. Blokhin IO, Lentz SR. Mechanisms of thrombosis in obesity. *Curr Opin Hematol.* 2013;20(5):437-444.

137. Neeland IJ, Poirier P, Despres JP. Cardiovascular and Metabolic Heterogeneity of Obesity: Clinical Challenges and Implications for Management. *Circulation.* 2018;137(13):1391-1406.

138. Li Q, Wang Q, Xu W, et al. C-Reactive Protein Causes Adult-Onset Obesity Through Chronic Inflammatory Mechanism. *Front Cell Dev Biol.* 2020;8:18.

139. Rey-Lopez JP, de Rezende LF, Pastor-Valero M, Tess BH. The prevalence of metabolically healthy obesity: a systematic review and critical evaluation of the definitions used. *Obes Rev.* 2014;15(10):781-790.

140. National Cholesterol Education Program Expert Panel on Detection E, Treatment of High Blood Cholesterol in A. Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. *Circulation.* 2002;106(25):3143-3421.

141. Smith GI, Mittendorfer B, Klein S. Metabolically healthy obesity: facts and fantasies. *J Clin Invest.* 2019;129(10):3978-3989.

142. Loos RJF, Kilpelainen TO. Genes that make you fat, but keep you healthy. *J Intern Med.* 2018;284(5):450-463.

143. Scott RA, Lagou V, Welch RP, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet.* 2012;44(9):991-1005.

144. Ji Y, Yiorkas AM, Frau F, et al. Genome-Wide and Abdominal MRI Data Provide Evidence That a Genetically Determined Favorable Adiposity Phenotype Is Characterized by Lower Ectopic Liver Fat and Lower Risk of Type 2 Diabetes, Heart Disease, and Hypertension. *Diabetes.* 2019;68(1):207-219.

145. Yaghootkar H, Scott RA, White CC, et al. Genetic evidence for a normal-weight "metabolically obese" phenotype linking insulin resistance, hypertension, coronary artery disease, and type 2 diabetes. *Diabetes.* 2014;63(12):4369-4377.

146. Manning AK, Hivert MF, Scott RA, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet.* 2012;44(6):659-669.

147. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet.* 2019;51(4):584-591.

148. Kaisinger LR, Kentistou KA, Stankovic S, et al. Large-scale exome sequence analysis identifies sex- and age-specific determinants of obesity. *Cell Genom.* 2023;3(8):100362.

149. Rask-Andersen M, Karlsson T, Ek WE, Johansson A. Gene-environment interaction study for BMI reveals interactions between genetic factors and physical activity, alcohol consumption and socioeconomic status. *PLoS Genet.* 2017;13(9):e1006977.

150. Gong W, Li H, Song C, et al. Effects of Gene-Environment Interaction on Obesity among Chinese Adults Born in the Early 1960s. *Genes (Basel).* 2021;12(2).

151. Justice AE, Winkler TW, Feitosa MF, et al. Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits. *Nat Commun.* 2017;8:14977.

152. Matise TC, Ambite JL, Buyske S, et al. The Next PAGE in understanding complex traits: design for the analysis of Population Architecture Using Genetics and Epidemiology (PAGE) Study. *Am J Epidemiol.* 2011;174(7):849-859.

153. The ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol.* 1989;129(4):687-702.

154. Gottesman O, Kuivaniemi H, Tromp G, et al. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. *Genet Med.* 2013;15(10):761-771.

155. Friedman GD, Cutter GR, Donahue RP, et al. CARDIA: study design, recruitment, and some characteristics of the examined subjects. *J Clin Epidemiol.* 1988;41(11):1105-1116.

156. Lavange LM, Kalsbeek WD, Sorlie PD, et al. Sample design and cohort selection in the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol.* 2010;20(8):642-649.

157. Kolonel LN, Henderson BE, Hankin JH, et al. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. *Am J Epidemiol.* 2000;151(4):346-357.

158. The Women's Health Initiative Study Group. Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. *Control Clin Trials.* 1998;19(1):61-109.

159. Yengo L, Vedantam S, Marouli E, et al. A saturated map of common genetic variants associated with human height. *Nature.* 2022;610(7933):704-712.

160. Winkler TW, Day FR, Croteau-Chonka DC, et al. Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc.* 2014;9(5):1192-1212.

161. Feng S, Liu D, Zhan X, Wing MK, Abecasis GR. RAREMETAL: fast and powerful meta-analysis for rare variants. *Bioinformatics.* 2014;30(19):2828-2829.

162. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* 2015;12(3):e1001779.

163. Zhou W, Nielsen JB, Fritsche LG, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet.* 2018;50(9):1335-1341.

164. Bien SA, Wojcik GL, Zubair N, et al. Strategies for Enriching Variant Coverage in Candidate Disease Loci on a Multiethnic Genotyping Array. *PLoS One.* 2016;11(12):e0167758.

165. Fesinmeyer MD, North KE, Ritchie MD, et al. Genetic risk factors for BMI and obesity in an ethnically diverse population: results from the population architecture using genomics and epidemiology (PAGE) study. *Obesity (Silver Spring).* 2013;21(4):835-846.

166. Carty CL, Bhattacharjee S, Haessler J, et al. Analysis of metabolic syndrome components in >15 000 african americans identifies pleiotropic variants: results from the population architecture using genomics and epidemiology study. *Circ Cardiovasc Genet.* 2014;7(4):505-513.

167. Zubair N, Graff M, Luis Ambite J, et al. Fine-mapping of lipid regions in global populations discovers ethnic-specific signals and refines previously identified lipid loci. *Hum Mol Genet.* 2016;25(24):5500-5512.

168. Hu Y, Graff M, Haessler J, et al. Minority-centric meta-analyses of blood lipid levels identify novel loci in the Population Architecture using Genomics and Epidemiology (PAGE) study. *PLoS Genet.* 2020;16(3):e1008684.

169. Tcheandjieu C, Zhu X, Hilliard AT, et al. Large-scale genome-wide association study of coronary artery disease in genetically diverse populations. *Nat Med.* 2022;28(8):1679-1692.

170. Ke W, Rand KA, Conti DV, et al. Evaluation of 71 Coronary Artery Disease Risk Variants in a Multiethnic Cohort. *Front Cardiovasc Med.* 2018;5:19.

171. Curb JD, McTiernan A, Heckbert SR, et al. Outcomes ascertainment and adjudication methods in the Women's Health Initiative. *Ann Epidemiol.* 2003;13(9 Suppl):S122-128.

172. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell.* 2017;169(7):1177-1186.

173. Werme J, van der Sluis S, Posthuma D, de Leeuw CA. An integrated framework for local genetic correlation analysis. *Nat Genet.* 2022;54(3):274-282.

174. Gusev A, Ko A, Shi H, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet.* 2016;48(3):245-252.

175. Laaksonen J, Taipale T, Seppala I, et al. Blood pathway analyses reveal differences between prediabetic subjects with or without dyslipidaemia. The Cardiovascular Risk in Young Finns Study. *Diabetes Metab Res Rev.* 2017;33(7).

176. Laakso M, Kuusisto J, Stancakova A, et al. The Metabolic Syndrome in Men study: a resource for studies of metabolic and cardiovascular diseases. *J Lipid Res.* 2017;58(3):481-493.

177. Ma Y, Patil S, Zhou X, Mukherjee B, Fritsche LG. ExPRSweb: An online repository with polygenic risk scores for common health-related exposures. *Am J Hum Genet.* 2022;109(10):1742-1760.

178. Petersen R, Pan L, Blanck HM. Racial and Ethnic Disparities in Adult Obesity in the United States: CDC's Tracking to Inform State and Local Action. *Prev Chronic Dis.* 2019;16:E46.

179. Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM. Pitfalls of predicting complex traits from SNPs. *Nat Rev Genet.* 2013;14(7):507-515.

180. So HC, Sham PC. Improving polygenic risk prediction from summary statistics by an empirical Bayes approach. *Sci Rep.* 2017;7:41262.

181. Duncan L, Shen H, Gelaye B, et al. Analysis of polygenic risk score usage and performance in diverse human populations. *Nat Commun.* 2019;10(1):3328.

182. Dikilitas O, Schaid DJ, Kosel ML, et al. Predictive Utility of Polygenic Risk Scores for Coronary Heart Disease in Three Major Racial and Ethnic Groups. *Am J Hum Genet.* 2020;106(5):707-716.

183. Bitarello BD, Mathieson I. Polygenic Scores for Height in Admixed Populations. *G3 (Bethesda).* 2020;10(11):4027-4036.

184. Mostafavi H, Harpak A, Agarwal I, Conley D, Pritchard JK, Przeworski M. Variable prediction accuracy of polygenic scores within an ancestry group. *Elife.* 2020;9.

185. Lambert SA, Gil L, Jupp S, et al. The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. *Nat Genet.* 2021;53(4):420-425.

186. Cavazos TB, Witte JS. Inclusion of variants discovered from diverse populations improves polygenic risk score transferability. *HGG Adv.* 2021;2(1).

187. Verweij LM, Terwee CB, Proper KI, Hulshof CT, van Mechelen W. Measurement error of waist circumference: gaps in knowledge. *Public Health Nutr.* 2013;16(2):281-288.

188. Man T, Nolte IM, Jaju D, et al. Heritability and genetic correlations of obesity indices with ambulatory and office beat-to-beat blood pressure in the Oman Family Study. *J Hypertens.* 2020;38(8):1474-1480.

189. Dashti HS, Miranda N, Cade BE, et al. Interaction of obesity polygenic score with lifestyle risk factors in an electronic health record biobank. *Bmc Med.* 2022;20(1):5.

190. Shi M, O'Brien KM, Weinberg CR. Interactions between a Polygenic Risk Score and Non-genetic Risk Factors in Young-Onset Breast Cancer. *Sci Rep.* 2020;10(1):3242.

191. Chiolero A, Faeh D, Paccaud F, Cornuz J. Consequences of smoking for body weight, body fat distribution, and insulin resistance. *Am J Clin Nutr.* 2008;87(4):801-809.

192. Hu J, Hu Y, Hertzmark E, et al. Weight Change, Lifestyle, and Mortality in Patients With Type 2 Diabetes. *J Clin Endocrinol Metab.* 2022;107(3):627-637.

193. Tucci SA. Phytochemicals in the Control of Human Appetite and Body Weight. *Pharmaceuticals (Basel).* 2010;3(3):748-763.

194. Filozof C, Fernandez Pinilla MC, Fernandez-Cruz A. Smoking cessation and weight gain. *Obes Rev.* 2004;5(2):95-103.

195. Pistelli F, Aquilini F, Carrozzi L. Weight gain after smoking cessation. *Monaldi Arch Chest Dis.* 2009;71(2):81-87.

196. Strelitz J, Ahern AL, Long GH, et al. Moderate weight change following diabetes diagnosis and 10 year incidence of cardiovascular disease and mortality. *Diabetologia.* 2019;62(8):1391-1402.

197. Ferreira JP, Rossignol P, Bakris G, Mehta C, White WB, Zannad F. Body weight changes in patients with type 2 diabetes and a recent acute coronary syndrome: an analysis from the EXAMINE trial. *Cardiovasc Diabetol.* 2021;20(1):187.

198. Trujillo JM, Nuffer W, Ellis SL. GLP-1 receptor agonists: a review of head-to-head clinical studies. *Ther Adv Endocrinol Metab.* 2015;6(1):19-28.

199. Silventoinen K, Jelenkovic A, Sund R, et al. Differences in genetic and environmental variation in adult BMI by sex, age, time period, and region: an individual-based pooled analysis of 40 twin cohorts. *Am J Clin Nutr.* 2017;106(2):457-466.

200. Clarke SL, Huang RDL, Hilliard AT, et al. Race and Ethnicity Stratification for Polygenic Risk Score Analyses May Mask Disparities in Hispanics. *Circulation.* 2022;146(3):265-267.

201. Bamba V, Rader DJ. Obesity and atherogenic dyslipidemia. *Gastroenterology.* 2007;132(6):2181-2190.

202. Marinou K, Tousoulis D, Antonopoulos AS, Stefanadi E, Stefanadis C. Obesity and cardiovascular disease: from pathophysiology to risk stratification. *Int J Cardiol.* 2010;138(1):3-8.

203. Mathew B, Francis L, Kayalar A, Cone J. Obesity: effects on cardiovascular disease and its diagnosis. *J Am Board Fam Med.* 2008;21(6):562-568.

204. Van Gaal LF, Mertens IL, De Block CE. Mechanisms linking obesity with cardiovascular disease. *Nature.* 2006;444(7121):875-880.

205. Vekic J, Zeljkovic A, Stefanovic A, Jelic-Ivanovic Z, Spasojevic-Kalimanovska V. Obesity and dyslipidemia. *Metabolism.* 2019;92:71-81.

206. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 2018;562(7726):203-209.

207. Bulik-Sullivan BK, Loh PR, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet.* 2015;47(3):291-295.

208. de Leeuw CA. LAVA partitioning algorithm (v1.0.0). In. Zenodo2021.

209. Wielscher M, Amaral AFS, van der Plaat D, et al. Genetic correlation and causal relationships between cardio-metabolic traits and lung function impairment. *Genome Med.* 2021;13(1):104.

210. Gao B, Yang C, Liu J, Zhou X. Accurate genetic and environmental covariance estimation with composite likelihood in genome-wide association studies. *PLoS Genet.* 2021;17(1):e1009293.

211. Nguyen NT, Magno CP, Lane KT, Hinojosa MW, Lane JS. Association of hypertension, diabetes, dyslipidemia, and metabolic syndrome with obesity: findings from the National Health and Nutrition Examination Survey, 1999 to 2004. *J Am Coll Surg.* 2008;207(6):928-934.

212. Laclaustra M, Lopez-Garcia E, Civeira F, et al. LDL Cholesterol Rises With BMI Only in Lean Individuals: Cross-sectional U.S. and Spanish Representative Data. *Diabetes Care.* 2018;41(10):2195-2201.

213. Shamai L, Lurix E, Shen M, et al. Association of body mass index and lipid profiles: evaluation of a broad spectrum of body mass index patients including the morbidly obese. *Obes Surg.* 2011;21(1):42-47.

214. NEIL2 GeneCard. 2023. https://www.genecards.org/cgi-bin/carddisp.pl?gene=NEIL2&keywords=NEIL2. Accessed 2023. 9. 3.

215. Kimura K, Morisasa M, Mizushige T, et al. Lipid Dynamics due to Muscle Atrophy Induced by Immobilization. *J Oleo Sci.* 2021;70(7):937-946.

216. van Vliet-Ostaptchouk JV, den Hoed M, Luan J, et al. Pleiotropic effects of obesity-susceptibility loci on metabolic traits: a meta-analysis of up to 37,874 individuals. *Diabetologia.* 2013;56(10):2134-2146.

217. Klimentidis YC, Arora A, Newell M, et al. Phenotypic and Genetic Characterization of Lower LDL Cholesterol and Increased Type 2 Diabetes Risk in the UK Biobank. *Diabetes.* 2020;69(10):2194-2205.